Rational billiards and Teichmüller flow

Anton Zorich

June 27, 2007

Acknowledgements

Substantial part of my results in this subject were obtained during numerous visits to MPI. I would like to use this opportunity to thank members, visitors and staff of the institute for their hospitality and for a stimulating aura.

I also want to thank my collaborators, J. Athreya, A. Eskin, M. Kontsevich and H. Masur for the pleasure to work with them.

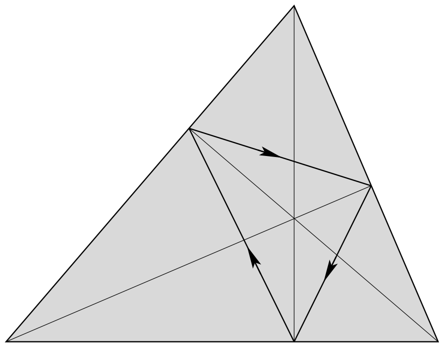
2 / 49

From Billiards to Flat Surfaces

3 / 49

Closed trajectories

It is easy to find a closed billiard trajectory in an acute triangle.



Exercise. Prove that the broken line joining the bases of hights of an acute triangle is a billiard trajectory (it is called *Fagnano trajectory*). Show that it realizes an inscribed triangle of minimal perimeter.

4 / 49

Challenge

It is difficult to believe, but a similar problem for an obtuse triangle is open.

Open problem. *Is there at least one closed trajectory for (almost) any obtuse triangle?*

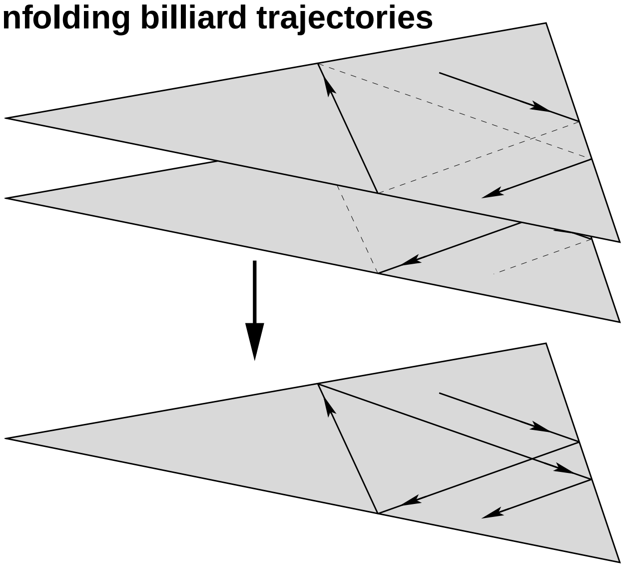
It seems like the answer is affirmative (see an extensive computer search performed by R. Schwartz and P. Hooper: `www.math.brown.edu/∼res/Billiards/index.html`)

Then, one can ask further questions:

- *Estimate the number $N(L)$ of periodic trajectories of length bounded by $L \gg 1$ when $L \to +\infty$.*
- *Is the billiard flow ergodic for almost any triangle?*

5 / 49

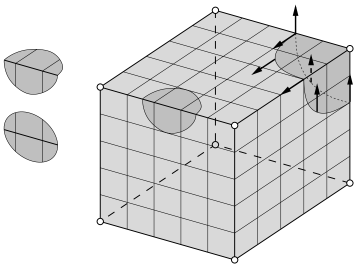
Unfolding billiard trajectories

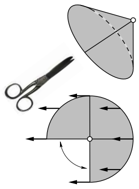


Identifying the boundary of two triangles we get a flat sphere. A billiard trajectory unfolds to a geodesic on this flat sphere.

6 / 49

Flat surfaces

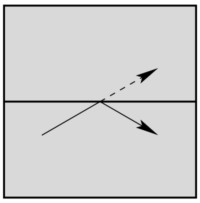




The surface of the cube represents a flat sphere with eight conical singularities. The metric *does not* have singularities on the edges. After parallel transport around a conical singularity a vector comes back pointing to a direction different from the initial one, so this flat metric has *nontrivial holonomy*.

Rational polygons

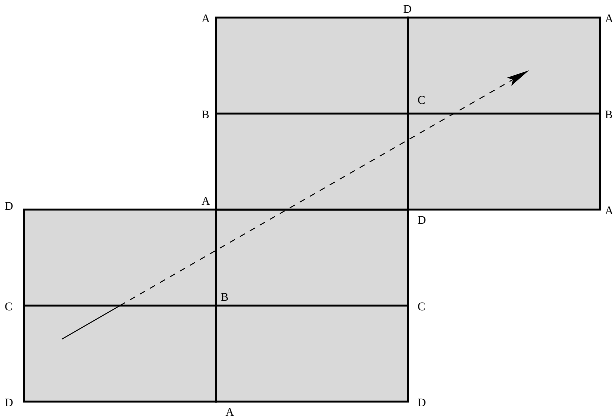
A polygon Π is called *rational* if all the angles of Π are rational multiples $\frac{p_i}{q_i}\pi$ of π. The properties of rational polygons are known much better. Consider a model case of a rectangular billiard.



As before instead of reflecting the trajectory we can reflect the billiard table. The trajectory unfolds to a straight line. Folding back the copies of the billiard table we project this line to the actual trajectory.

Billiard in a rectangle

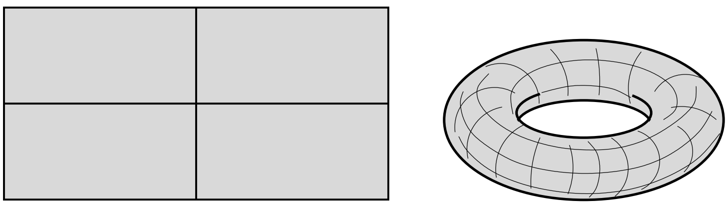
Fix a (generic) trajectory. At any moment the ball moves in one of four directions. They correspond to four copies of the billiard table; other copies can be obtained from these four by a parallel translation.



9 / 49

Billiard in a rectangle

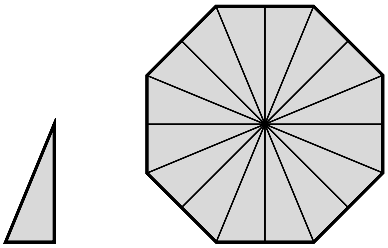
Identifying the equivalent patterns by a parallel translation we obtain a torus; the billiard trajectory unfolds to a “straight line” on the corresponding torus.



10 / 49

Unfolding rational billiards

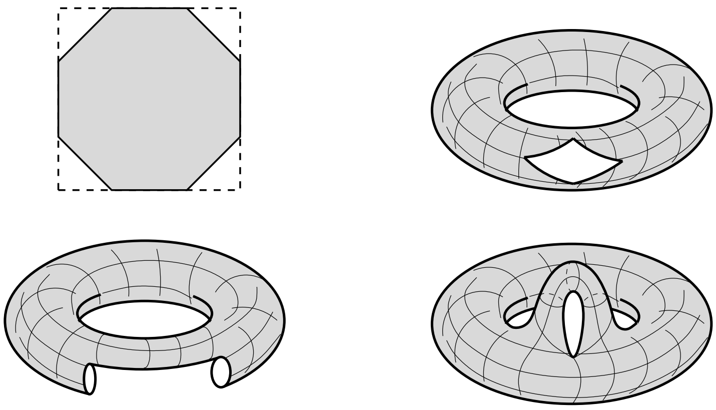
We can apply an analogous procedure to any rational billiard.



Consider, for example, a triangle with angles $\pi/8, 3\pi/8, \pi/2$. It is easy to check that a generic trajectory in such billiard table has 16 directions (instead of 4 for a rectangle). Using 16 copies of the triangle we unfold the billiard into a regular octagon with opposite sides identified by parallel translations.

11 / 49

Flat surface of genus 2



Identifying the pair of horizontal sides and then the pair of vertical sides of a regular octagon we get a torus with a single square hole. Identifying two opposite sides of the hole we get a torus with two distinct holes. Identifying the holes we get a surface of genus two.

12 / 49

Very flat surfaces

Note that the flat metric on the resulting surface has *trivial holonomy*, since the identifications of the sides were performed by parallel translations. As before, a billiard trajectory is unfolded to a geodesic on the surface. But now geodesics resemble geodesics on the torus: they do not have self-intersections!

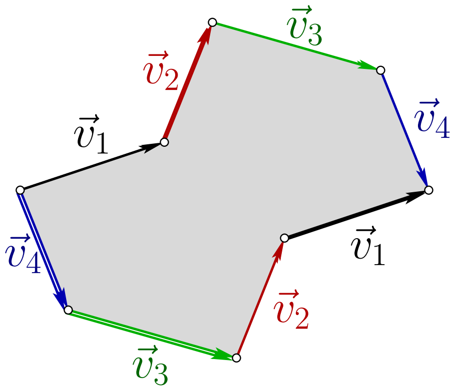
We abandon rational billiards for a while and pass to a more systematic study of *"very flat"* surfaces.

13 / 49

Very flat surfaces

Very flat surfaces

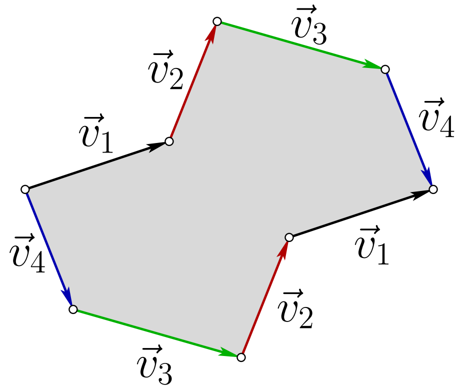
Consider a broken line constructed from vectors $\vec{v}_1, \ldots, \vec{v}_k$.



and another one constructed from the same vectors taken in another order. If we are lucky enough the two broken lines do not intersect and form a polygon.

15 / 49

Very flat surfaces



Identifying the corresponding pairs of sides by parallel translations we get a closed surface endowed with a flat metric.

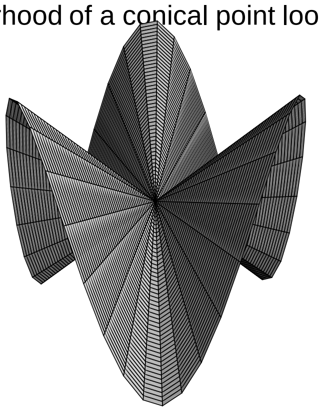
Properties of very flat surfaces

- The flat metric is nonsingular outside of a finite number of conical singularities (inherited from the vertices of the polygon).
- The flat metric has trivial holonomy, i.e. parallel transport along any closed path brings a tangent vector to itself.
- In particular, all cone angles are integer multiples of 2π.
- By convention, the choice of the vertical direction ("direction to the North") will be considered as a part of the "very flat structure".

  For example, a surface obtained from a rotated polygon is considered as a different very flat surface.
- A conical singularity with the cone angle $2\pi \cdot N$ has N outgoing directions to the North.

Example: conical singularity with cone angle 6π

Locally a neighborhood of a conical point looks like a “*monkey saddle*”.



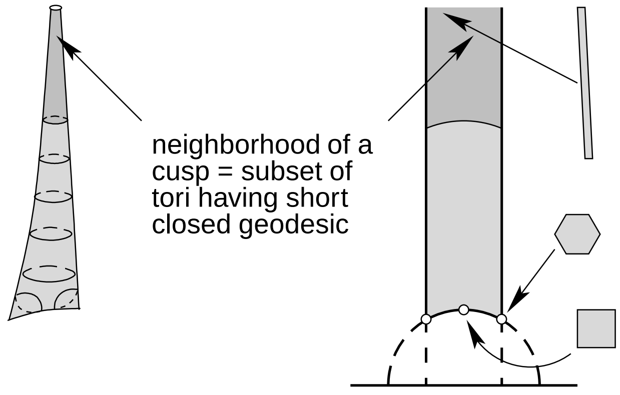
A neighborhood of a conical point with a cone angle 6π can be glued from six metric half discs. At this conical point we have 3 distinct directions to the North.

Families of flat surfaces

The polygon in our construction depends continuously on the vectors $\vec{v}_j$. This means that the combinatorial geometry of the resulting flat surface (its genus g, the number m and types $2\pi(d_1+1), \dots, 2\pi(d_m+1)$ of the resulting conical singularities) does not change under small deformations of the vectors $\vec{v}_j$. This allows to consider a flat surface as an element of a **family** of flat surfaces sharing common combinatorial geometry.
As an example of such family one can consider a family of flat tori of area one, which can be identified with the space of lattices of area one:

$$\mathrm{SO}(2,\mathbb{R})\backslash\, \mathrm{SL}(2,\mathbb{R})\,/\,\mathrm{SL}(2,\mathbb{Z}) \quad = \quad \mathbb{H}^2/\,\mathrm{SL}(2,\mathbb{Z})$$

Family of flat tori



The corresponding “modular surface” is not compact: flat tori representing points, which are close to the cusp, are almost degenerate: they have a very short closed geodesic.

20 / 49

Holomorphic 1-forms and quadratic differentials versus very flat surfaces

21 / 49

Holomorphic 1-form associated to a flat structure

Consider the natural coordinate z in the complex plane, where lives the polygon. In this coordinate the parallel translations which we use to identify the sides of the polygon are represented as $z' = z + const$.

Since this correspondence is holomorphic, our flat surface S with punctured conical points inherits the complex structure. This complex structure extends to the punctured points.

Consider now a holomorphic 1-form dz in the complex plane. The coordinate z is not globally defined on the surface S. However, since the changes of local coordinates are defined as $z' = z + const$, we see that $dz = dz'$. Thus, the holomorphic 1-form dz on $\mathbb{C}$ defines a holomorphic 1-form ω on S which in local coordinates has the form $\omega = dz$.

The form ω has zeroes exactly at those points of S where the flat structure has conical singularities.

22 / 49

Flat structure defined by a holomorphic 1-form

- Reciprocally a pair (Riemann surface, holomorphic 1-form) uniquely defines a flat structure.
- In a neighborhood of zero a holomorphic 1-form can be represented as $w^d\,dw$, where d is the **degree** of zero. The form ω has a zero of degree d at a conical point with cone angle $2\pi(d+1)$. Moreover, $d_1 + \cdots + d_m = 2g-2$.
- The moduli space $\mathcal{H}_g$ of pairs (complex structure, holomorphic 1-form) is a $\mathbb{C}^g$-vector bundle over the moduli space $\mathcal{M}_g$ of complex structures.
- The space $\mathcal{H}_g$ is naturally stratified by the strata $\mathcal{H}(d_1,\dots,d_m)$ enumerated by unordered partitions $d_1 + \cdots + d_m = 2g-2$.
- Any holomorphic 1-forms corresponding to a fixed stratum $\mathcal{H}(d_1,\dots,d_m)$ has exactly m zeroes of degrees $d_1,\dots,d_m$.

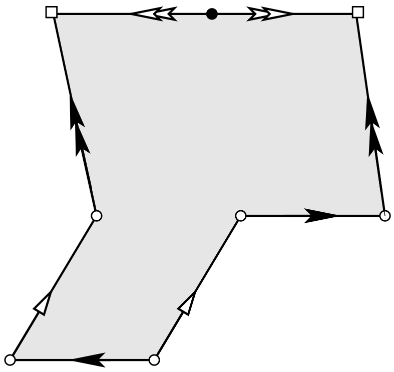
23 / 49

Dictionary

flat structure (including a choice of the vertical direction)	complex structure and a choice of a holomorphic 1-form ω
conical point with a cone angle $2\pi(d+1)$	zero of degree d of the holomorphic 1-form ω (in local coordinates $\omega = w^d\,dw$)
side $\vec{v}_j$ of a polygon	relative period $\int_{P_j}^{P_{j+1}} \omega = \int_{\vec{v}_j} dz$ of the 1-form ω
family of flat surfaces sharing the same cone angles $2\pi(d_1+1),\dots,2\pi(d_m+1)$	stratum $\mathcal{H}(d_1,\dots,d_m)$ in the moduli space of holomorphic 1-forms
coordinates in the family: vectors $\vec{v}_i$ defining the polygon	coordinates in $\mathcal{H}(d_1,\dots,d_m)$: relative periods of ω in $H^1(S,\{P_1,\dots,P_m\};\mathbb{C})$

24 / 49

Flat surfaces and quadratic differentials



Identifying pairs of sides of this polygon by isometries we obtain a surface of genus $g = 1$. Now the flat metric has holonomy group $\mathbb{Z}/2\mathbb{Z}$. The cone angles are multiples of π.
Flat surfaces of this type correspond to quadratic differentials.
For example, the quadratic differential representing the surface from the picture belongs to the stratum $\mathcal{Q}(2, -1, -1)$.

25 / 49

Volume element

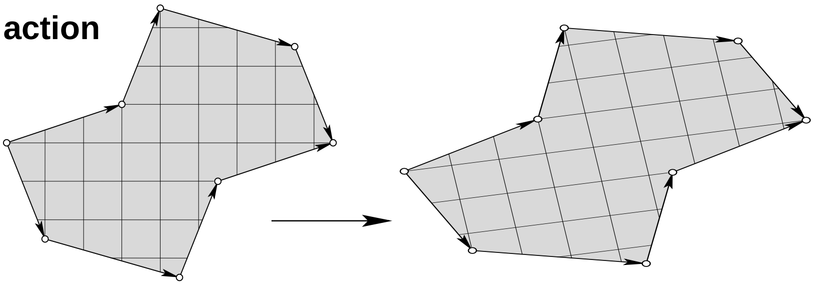
Note that the vector space $H^1(S, \{P_1, \dots, P_m\} ; \mathbb{C})$ contains a natural integer lattice $H^1(S, \{P_1, \dots, P_m\} ; \mathbb{Z} \oplus \sqrt{-1}\,\mathbb{Z})$. Consider a linear volume element $d\nu$ normalized in such a way that the volume of the fundamental domain in this lattice equals one. Consider now the real hypersurface $\mathcal{H}_1(d_1, \dots, d_m) \subset \mathcal{H}(d_1, \dots, d_m)$ defined by the equation $area(S) = 1$. The volume element $d\nu$ can be naturally restricted to the hypersurface defining the volume element $d\nu_1$ on $\mathcal{H}_1(d_1, \dots, d_m)$.

Theorem (H. Masur; W. A. Veech) *The total volume* $\mathrm{Vol}(\mathcal{H}_1(d_1, \dots, d_m))$ *of every stratum is finite.*

The values of these volumes were computed by A. Eskin and A. Okounkov.

26 / 49

Group action



The subgroup $SL(2,\mathbb{R})$ of area preserving linear transformations acts on the "unit hyperboloid" $\mathcal{H}_1(d_1,\dots,d_m)$. The diagonal subgroup $\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \subset SL(2,\mathbb{R})$ induces a natural flow on the stratum, which is called the *Teichmüller geodesic flow*.

Key Theorem (H. Masur; W. A. Veech) *The action of the groups $SL(2,\mathbb{R})$ and $\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$ preserves the measure $d\nu_1$. Both actions are ergodic with respect to this measure on each connected component of every stratum $\mathcal{H}_1(d_1,\dots,d_m)$.*

27 / 49

Hope for a magic wand

Suppose that we need some information about geometry or dynamics of an individual flat surface S. Consider the element S in the corresponding family of flat surfaces $\mathcal{H}(d_1,\dots,d_m)$. Denote by $\mathcal{C}(S) = \overline{GL^+(2,\mathbb{R})\, S} \subset \mathcal{H}(d_1,\dots,d_m)$ the closure of the $GL^+(2,\mathbb{R})$-orbit of S in $\mathcal{H}(d_1,\dots,d_m)$. In numerous cases knowledge about the structure of $\mathcal{C}(S)$ gives a comprehensive information about geometry and dynamics of the initial flat surface S. Moreover, some delicate numerical characteristics of S can be expressed as averages of simpler characteristics over $\mathcal{C}(S)$.

Examples:

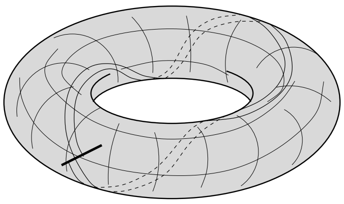
- Veech surfaces;
- Nonergodic directions;

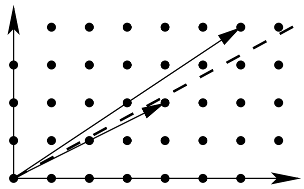
28 / 49

Asymptotic cycle

Theorem (S. Kerckhoff, H. Masur, J. Smillie) *For any flat surface directional flow in almost any direction is ergodic.*

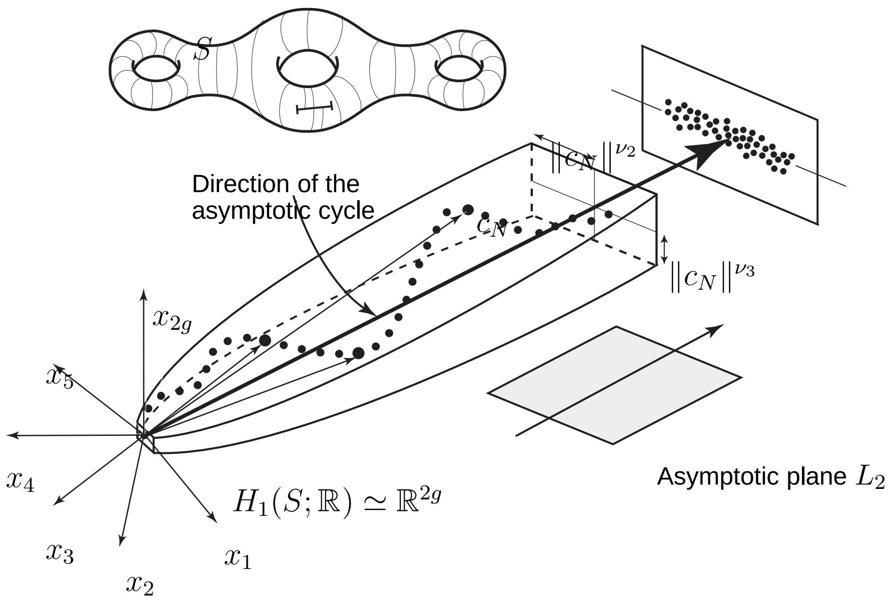
Consider an "irrational" geodesic on a torus. Choose a short transversal segment X. Each time when the geodesic crosses X we join the crossing point with the point x_0 along X obtaining a closed loop. Consecutive return points $x_1, x_2, \dots$ define a sequence of cycles $c_1, c_2, \dots$.





The *asymptotic cycle* is defined as $\lim_{n\to\infty} \frac{c_n}{\|c_n\|} = c \in H_1(\mathbb{T}^2;\mathbb{R})$

Asymptotic flag



Asymptotic flag

Theorem *For almost any surface S in any stratum $\mathcal{H}_1(d_1, \dots, d_m)$ there exists a flag of subspaces $L_1 \subset L_2 \subset \cdots \subset L_g \subset H_1(S;\mathbb{R})$ such that for any $j = 1, \dots, g-1$ one has*

$$\limsup_{N\to\infty} \frac{\log \operatorname{dist}(c_N, L_j)}{\log N} = \nu_{j+1}$$

and

$$\operatorname{dist}(c_N, L_g) \le const,$$

where the constant depends only on S and on the choice of the Euclidean structure in the homology space.
The numbers $2, 1+\nu_2, \dots, 1+\nu_g$ are the top g Lyapunov exponents of the Teichmüller geodesic flow on the corresponding connected component of the stratum $\mathcal{H}(d_1, \dots, d_m)$; in particular, they do not depend on the individual generic flat surface S in the connected component.

32 / 49

Lyapunov exponents

Consider a map $f : X \to X$ ergodic with respect to a finite measure $d\mu$ on X. Consider a matrix-valued function $F : X \to \mathrm{GL}(m,\mathbb{R})$ such that $\int_X \max(\log \|F\|, 0) d\mu < \infty$. Consider a product of matrices along a trajectory $P_0, P_1 = f(P_0), \dots$ of the map f:

$$F^{(n)}(P_0) = F(P_{n-1}) \cdot \cdots \cdot F(P_1) \cdot F(P_0)$$

Lyapunov exponents are defined as mean values of logarithms of eigenvalues of these product matrices : $\nu_i(P_0) := \lim_{n\to\infty} \frac{\log|\lambda_i(n)|}{n}$.

Theorem (Oseledets) *Under conditions above Lyapunov exponents are well defined and constant almost everywhere.*

When $f : X^m \to X^m$ is smooth one can consider the differential $F(P) := Df_P$ as a matrix-valued function. In this case the Lyapunov exponents are responsible for the rate of divergence (convergence) between two trajectories starting nearby.

33 / 49

The theorem above was initially formulated in as a conditional statement under the conjecture that all the exponents ν_j, for $j = 2, \ldots, g$, are distinct. This conjectures was partly proved by G. Forni and in full generality by A. Avila and M. Viana.

Currently there are no methods of calculation of individual Lyapunov exponents ν_j (though there is some experimental knowledge of their approximate values). Nevertheless, for any connected component of any stratum (and, more generally, for any $\mathrm{GL}^+(2;\mathbb{R})$-invariant suborbifold) it is possible to evaluate the *sum* of the Lyapunov exponents $\nu_1 + \cdots + \nu_g$, where g is the genus. The formula for this sum was discovered by M. Kontsevich; morally, it is given in terms of characteristic numbers of some natural vector bundles over the strata $\mathcal{H}(d_1, \ldots, d_m)$. In some particular cases this formula was developed by I. Bouw and M. Möller and by M. Bainbridge. This is also a subject of our unfinished project with M. Kontsevich.

34 / 49

Saddle connections and closed geodesics

35 / 49

Saddle connections

A *saddle connection* is a geodesic segment joining a pair of conical singularities or a conical singularity to itself without any singularities in its interior.

Similar to the torus case regular closed geodesics on flat surface always appear in families; any such family fills a maximal cylinder bounded on each side by a closed saddle connection or by a chain of parallel saddle connections.

Let $N_{sc}(S, L)$ be the number of saddle connections of length at most L on a flat surface S. Let $N_{cg}(S, L)$ be the number of maximal cylinders filled with closed regular geodesics of length at most L on S. It was proved by H. Masur that for any flat surface S both counting functions $N(S, L)$ grow quadratically in L:

$$const_1(S) \leq N(S, L)/L^2 \leq const_2(S)$$

36 / 49

Exact quadratic asymptotics

Theorem (A. Eskin and H. Masur) *For almost all flat surfaces S in any stratum $\mathcal{H}(d_1,\dots,d_m)$ the counting functions $N_{sc}(S,L)$ and $N_{cg}(S,L)$ have exact quadratic asymptotics*

$$\lim_{L\to\infty}\frac{N_{sc}(S,L)}{\pi L^2}=c_{sc}(S) \qquad \lim_{L\to\infty}\frac{N_{cg}(S,L)}{\pi L^2}=c_{cg}(S)$$

where the Siegel–Veech constants $c_{sc}(S)$ and $c_{cg}(S)$ depend only on the connected component of the stratum.

Consider some saddle connection $\gamma_1=[P_1P_2]$ with an endpoint at P_1. Memorize its direction, say, let it be the North-West direction. Let us launch a geodesic from the same starting point P_1 in one of the remaining remaining $k-1$ North-West directions. Let us study how big is the chance to hit P_2 ones again, and how big is the chance to hit it after passing the same distance as before.

Phenomenon of multiple saddle connections

Theorem (A. Eskin, H. Masur, A. Zorich) *For almost any flat surface S in any stratum and for any pair P_1,P_2 of conical singularities on S the function $N_2(S,L)$ counting the number of pairs of parallel saddle connections of the same length joining P_1 to P_2 also has exact quadratic asymptotics*

$$\lim_{L\to\infty}\frac{N_2(S,L)}{\pi L^2}=c_2>0.$$

For almost all flat surfaces S in any stratum one cannot find neither a single pair of parallel saddle connections on S of different length, nor a single pair of parallel saddle connections joining different pairs of singularities.

Rigid collections of saddle connections

- Any saddle connection on a flat surface persists under small deformations of S inside the ambient stratum.
- It might happen that any deformation of a given flat surface which shortens some specific saddle connection necessarily shortens some other saddle connections.
- We say that a collection $\{\gamma_1, \ldots, \gamma_n\}$ of saddle connections is *rigid* if any sufficiently small deformation of the flat surface inside the stratum preserves the proportions $|\gamma_1| : |\gamma_2| : \cdots : |\gamma_n|$ of the lengths of all saddle connections in the collection.

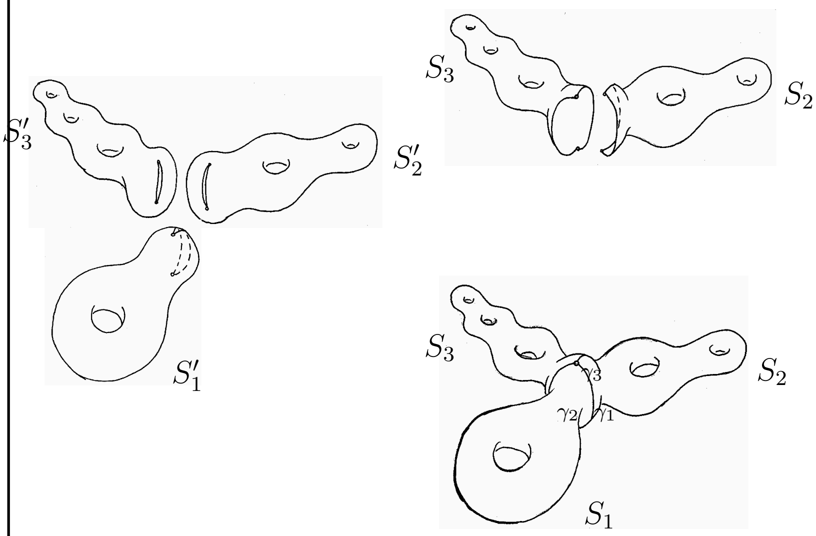
Theorem (Eskin, Masur, Zorich) *Let S be a flat surface corresponding to a holomorphic 1-form ω. A collection $\gamma_1, \ldots, \gamma_n$ of saddle connections on S is rigid if and only if all saddle connections $\gamma_1, \ldots, \gamma_n$ are homologous in $H^1(S, \{\text{zeroes of } \omega\}; \mathbb{C})$.*

Homologous saddle connections

The directions and lengths of saddle connections can be expressed in terms of integrals of the holomorphic 1-form ω along corresponding paths. Hence

- Homologous saddle connections $\gamma_1, \ldots, \gamma_n$ are parallel and have equal length and
  - either all of them join the same pair of distinct singular points,
  - or all γ_i are closed loops.

Saddle connections joining distinct zeroes



Multiple homologous saddle connections, topological picture.

41 / 49

Siegel–Veech formula

To every closed regular geodesic γ on a flat surface S we associate a vector $\vec{v}(\gamma)$ in $\mathbb{R}^2$ having the length and the direction of γ. In other words, $\vec{v} = \int_\gamma \omega$, where we consider a complex number as a vector in $\mathbb{R}^2 \simeq \mathbb{C}$. Applying this construction to all closed regular geodesic on S we construct a discrete set $V(S) \subset \mathbb{R}^2$. Consider the following operator $f \mapsto \hat{f}$ from functions with compact support on $\mathbb{R}^2$ to functions on the stratum $\mathcal{H}_1(\beta) = \mathcal{H}_1(d_1, \ldots, d_m)$:

$$\hat{f}(S) := \sum_{\vec{v} \in V(S)} f(\vec{v})$$

Function $\hat{f}(S)$ generalizes the counting function $N_{cg}(S, L)$ introduced in the beginning of this section. Namely, when $f = \chi_L$ is the characteristic function χ_L of the disc of radius L with the center at the origin of $\mathbb{R}^2$, the function $\hat{\chi}_L(S)$ counts the number of regular closed geodesics of length at most L on a flat surface S.

42 / 49

Siegel–Veech formula

Theorem (W. Veech *For any function* $f : \mathbb{R}^2 \to \mathbb{R}$ *with compact support the following equality is valid:*

$$\frac{1}{\operatorname{Vol}\mathcal{H}_1^{comp}(\beta)} \int_{\mathcal{H}_1^{comp}(\beta)} \hat{f}(S)\, d\nu_1 = C \int_{\mathbb{R}^2} f(x,y)\, dx\, dy\ ,$$

where the constant C *does not depend on the function* f.
The same theorem is valid for configurations of homologous saddle connections (but the constant will change).
Theorem (A. Eskin, H. Masur, A. Z.)

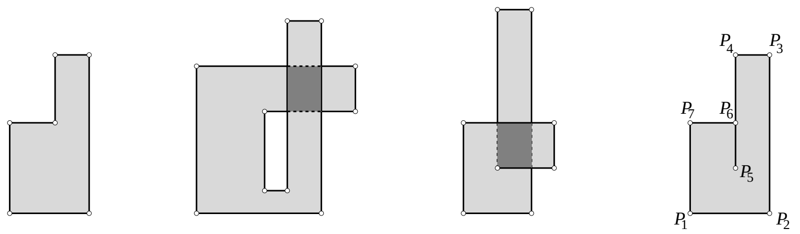
$$c(\mathcal{C}) = \lim_{\varepsilon\to 0} \frac{1}{\pi\varepsilon^2} \frac{\operatorname{Vol}(\text{“}\varepsilon\text{-neighborhood of the cusp } \mathcal{C}\text{ ”})}{\operatorname{Vol}\mathcal{H}_1^{comp}(\beta)} =$$

$$= (\text{explicit combinatorial factor}) \cdot \frac{\prod_{j=1}^{k} \operatorname{Vol}\mathcal{H}_1(\beta'_k)}{\operatorname{Vol}\mathcal{H}_1^{comp}(\beta)}$$

43 / 49

Applications: billiards in rectangular polygons. (With J. Athreya and A. Eskin) 44 / 49

Rectangular polygons



By a **rectangular polygon** Π we call a topological disc endowed with a flat metric, such that the boundary $\partial\Pi$ is presented by a finite broken line of geodesic segments and such that the angle between any two consecutive sides equals $k\pi/2$, where $k \in \mathbb{N}$.

45 / 49

Billiards in rectangular polygons versus quadratic differentials on $\mathbb{C}\mathrm{P}^1$

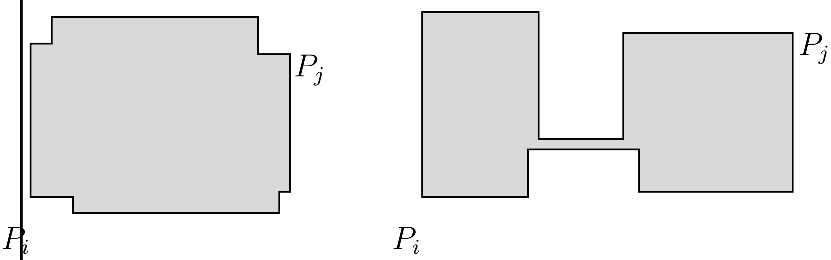
We want to count trajectories emitted from a corner of such billiard and getting to some other corner. We also want to count closed billiard trajectories. In both cases we count trajectories of length bounded by $L \gg 1$ and study the asymptotics as L tends to infinity.

Note that the topological sphere obtained by gluing two copies of the billiard table by the boundary is a flat surface, or, in other words, a meromorphic quadratic differential with simple poles on $\mathbb{C}\mathrm{P}^1$.

Moreover, geodesics on this flat sphere project to billiard trajectories! Thus, to count billiard trajectories we may count geodesics on flat spheres!

46 / 49

Number of generalized diagonals



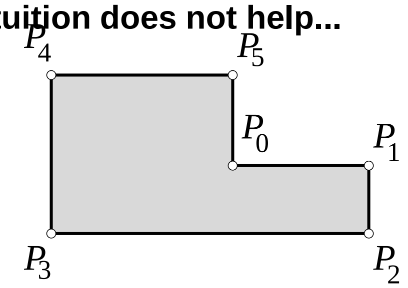
We prove that for a generic rectangular polygon with angles $\pi/2$ and $3\pi/2$ the number of trajectories joining any two fixed corners with right angles is “approximately” the same as for a rectangle:

$$\frac{1}{2\pi} \cdot \frac{(\text{bound for the length})^2}{\text{area of the table}}$$

and does not depend on the shape of the polygon.

47 / 49

Naive intuition does not help...



However, say, for a typical L-shaped polygon the number of trajectories joining the corner with the angle $3\pi/2$ to some other corner is "approximately"

$$\frac{2}{\pi} \cdot \frac{(\text{bound for the length})^2}{\text{area of the table}}$$

which is 4 times (and not 3) times bigger than the number of trajectories joining a fixed pair of right corners...

48 / 49

Open problems and bibliography

● Compactification.
● Classification of orbit closures of $GL(2;\mathbb{R})$.
– done for genus two (C. McMullen);
– "invariant algebraic" implies "affine" (M. Kontsevich) and sometimes even more (M. Möller);
– even extremely particular flat surfaces might produce the entire stratum as their orbit closure
(P. Hubert, E. Lanneau, M. Möller).
● "Ratner-type" theorem: classification of closures of the unipotent subgroup?
...

For a short survey and essential bibliography see:

A. Zorich, *Geodesics on flat surfaces*, Proceedings of the ICM 2006, Madrid, Vol. III. EMS Publishing House, 2006, 121–146.

For a more detailed survey and extended bibliography see:

A. Zorich, *Flat surfaces*, in collect. "Frontiers in Number Theory, Physics and Geometry. Vol. 1: On random matrices, zeta functions and dynamical systems"; Ecole de physique des Houches, France, March 9-21 2003, P. Cartier; B. Julia; P. Moussa; P. Vanhove (Editors), Springer-Verlag, Berlin, 2006, 439–586.

49 / 49