

MATH 421 DYNAMICS

Week 9 Lecture 1 Notes

1. BILLIARDS

We return to billiard maps now and present a more general situation. The 2-particle billiards is really a part of an entire field of study called convex billiards. To start, let D be a bounded, closed domain in the plane, where B is the boundary of D , so $B = \partial D$. Orbits of motion are line segments in D with endpoints in B , and adjacent line segments meet in B . When B , as a curve in the plane, is C^1 , the angle which a line segment makes with the tangent to B at the end point is the same as the angle the adjacent line segment makes. This is what is meant by “angle of incidence equals angle of reflection”. Should B also contain corners (points where B is not C^1), declare that an orbit entering the corner end there (this is sometimes referred to as “pocket” billiards. Motion is always considered with constant velocity on line segments, and collisions with B are specular (elastic).

Some dynamical criteria:

- Every orbit is completely determined by its starting point and direction.
- Recall for polygonal billiards, a billiard flow is continuous flow per unit time. It is certainly not a differentiable flow, as it fails at the collisions with B (Note: One can certainly define a smooth flow whose trajectory has corners. All that is necessary is for the flow to slow up and momentarily stop at the corner, to allow it to change direction smoothly. This is quite common for parameterized curves. Here, though, the flow does not slow up.)
- In the billiard flow on the triangle, we cured the non differentiable flow points by “unfolding” the table. Here, instead, we will analyze this situation by creating a completely different state space which collects only the relevant information from the actual billiard.

First, ignore the time between collisions of line segments with B , and consider orbits as simply a sequence of points on B , along with their angle of incidence. For each collision of an orbit with B , the point and the angle completely determine the next point and angle of collision. In the “space” of points of B and possible angles of collision, we get an assignment of the next point of collision and angle for each previous one. It turns out that this assignment is quite well defined. Call this assignment Φ , where $(x_1, \theta_1) \mapsto (x_2, \theta_2) \mapsto \dots \mapsto (x_n, \theta_n) \mapsto \dots$. For now, let B be C^1 . Collect up all of the points of B , and you get a copy of S^1 . Collect up all possible angles of incidence and you get the interval $[0, \pi]$ (really one gets the open interval, but one can limit to the orbits that simply run along the length of B . This is not such an important factor here. The state space is all of the points in B along with all of the incidence angles is a copy of $C = S^1 \times [0, \pi]$, the cylinder. The assignment takes

$(x_1, \theta_1) \mapsto (x_2, \theta_2) \mapsto \cdots \mapsto (x_n, \theta_n)$. See the figure. The resulting cylinder, along with the evolution map Φ is called the billiard map.

Example 1. Let

$$D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

be the unit disk in the plane. Here $B = \partial D = S^1$ is the unit circle, parameterized by the standard angular coordinate θ from polar coordinates in the plane (note that this parameter takes values in $[0, 2\pi)$ and is quite different from the parameterization we have been using for S^1 given by the exponential map $x \mapsto e^{2\pi i x}$). The state space is then $C = S^1 \times I$, where $I = [0, \pi]$. What are the dynamics? Go back to the light-ray in a circular mirrored room exercise from before. You will find that the initial angle of incidence never changes, and the evolution map is constant on the second coordinate.

Exercise 1. Show for S^1 the unit circle, that $\Phi(s, \theta) = (s + 2\theta, \theta)$.

Exercise 2. Show that this is not quite true for a billiard table whose radius is not 1.

Now do you recognize the evolution map on the state space in this dynamical system? This is basically the twist map on the cylinder, a map that you already showed was area preserving. And you already know the dynamics of this map. To continue our study, we can say more about the orbit structure within each invariant cross-section (constant θ section) of the cylinder: To each $\theta = \theta_0$ is associated a *caustic*:

- In optics, a caustic is the envelope of light rays reflected or refracted by a curved surface or object, or the projection of that envelope of rays on another surface.
- Or the caustic is a curve or surface to which each of the light rays is tangent, defining a boundary of an envelope of rays as a curve of concentrated light.
- In differential geometry and geometric optics (mathematics, in general), a caustic is the envelope of rays (directed line segments) either reflected or refracted by a manifold.

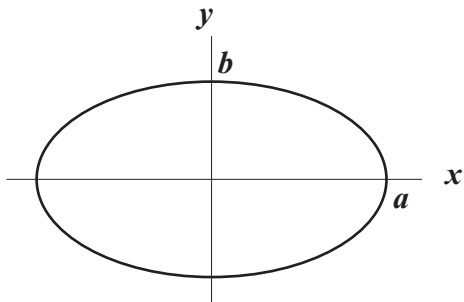
Exercise 3. For the circle billiard, let $\theta \notin \mathbb{Q}$. Then the caustic is the edge of the region basically filled with light. What shape is this caustic, and can you write the equation for this caustic as a function of the angle θ_0 .

Experiment 1. Shine a light from a small hole horizontally into a circular mirrored room. Try to pass the light beam directly through the center of the room (force $\theta_0 = \frac{\pi}{2}$). What happens as you “focus” the light? How does the light fill the room as you approach $\frac{\pi}{2}$, and when you reach $\frac{\pi}{2}$?

Example 2. Let

$$D = \left\{ (x, y) \in \mathbb{R}^2 \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \right\}$$

be an ellipse, where the *diameter* is $2a$ and the *width* is $2b$.



Recall that a definition of an ellipse is the set of points in the plane whose combined distance from two reference points is a constant. The two reference points are the two foci of the ellipse, and in our case, the combined distance constant is the diameter $2a$. Written as above, the ellipse is centered at the origin in the plane, and the diameter and the width are along the horizontal and vertical axes, respectively.

Upon inspection, one can see that it will not be the case, as with the circle, that there is a period-2 orbit passing through each point. There are four such points, though, and all four of these lie on one of the axes. Why is this true? We will see.

This billiard table has notable differences from the circular one, beyond the relative lack of period-2 orbits. To understand these difference better, we introduce a technique of study common in billiards: Generating Functions.

Parameterize the boundary by arc-length s and let p and p' be 2 points on $B = \partial D$. Now define a real-valued function on $B \times B$ by

$$H((s, s')) = -d(p, p'),$$

where d is the standard Euclidean metric in the plane. This function H is called the *generating Function* for the billiard:

Some notes:

- This function helps to identify points on the same orbit.
- Critical points of H determine period-2 orbits (think about what this means for the ellipse.)
- rarely can we find a good working expression for H in terms of s and s' . But we can discuss its properties and use them effectively.

Example 3. Let $a = b = 1$, and we are back at the circular billiard. Here $H(s, s') = -2 \sin \frac{1}{2}(s - s')$.

Exercise 4. Derive this function using the geometry of the unit circle.

Exercise 5. For $a > b$, we do not have a good expression for H . However, we can surmise that the diameter boundary points are at a minimum for H (remember the minus sign), and the width boundary points are a saddle point for H . Why is this? Can you see it?

1.1. Dynamics of elliptic billiards. As in circular billiards, one way to discuss the orbit structure for an elliptic billiard is to try to describe any possible caustics (curves tangent to orbits, which help to define edges of envelopes of orbit regions. We have two results here:

Proposition 4. *Every smaller confocal (having the same foci) ellipse is a caustic.*

The proof here is constructive and can be found in the book. This family of ellipses works as a caustic for any orbit segment that does not pass between or meet the foci. Convince yourself that if an orbit segment does not meet or pass between the foci, then the entire orbit will not intersect the closed line segment connecting the foci. And once an orbit segment crosses that line, it will continue to cross that line both forward and backward for each line segment in the orbit. And if an orbit segment passes through a focus, where will it go next? Where will it go over the long term?

Proposition 5. *There exists a caustic for every ray between the foci. The caustic of the orbit corresponding to this ray is both pieces of a hyperbola confocal to the ellipse.*

Note: Ellipses and hyperbolas are both conic sections, and related via their *eccentricity*, a nonnegative number that parameterizes conic sections via a ratio of their data. Indeed, along the major axis (the diameter) of a conic section, one can measure the distance from the curve to the origin (let's keep all conic section centered at the origin for now). Call this the radius a . One can also measure the distance from the center to one of the foci. call this c . Then eccentricity e is the ration of these two numbers:

- For $e = \frac{c}{a} = 0$ (implying that $c = 0$), the section is a circle.
- For $0 < e = \frac{c}{a} < 1$, the section is an ellipse.
- For $e = \frac{c}{a} = 1$, the section is a parabola.
- For $e = \frac{c}{a} > 1$, the section is a hyperbola.

For the circle case, the equation is elliptical, with $a = b$, and we have $\frac{x^2}{a^2} + \frac{y^2}{a^2} = 1$, or $x^2 + y^2 = a^2$. For the hyperbolic case, we have $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$.

And as a couple of final dynamic notes;

- An orbit that passes through one focus must pass through the other . What are the implications of this for the resulting orbit?
- There are tons of periodic orbits in elliptic billiards, of all periods. Can you draw some? Period-4 should be easy to see, as is period-2. How about period-3?

Exercise 6. Construct a period-4 orbit for an elliptic billiard and show analytically that it exists.

Exercise 7. Describe the long term behavior of ANY orbit that has a orbit segment that pass through one of the foci.

Going back to the generating functions, we can say more about orbits in general. Here are some properties. Recall for any convex billiard table, $H(s, s') = -d(p, p')$, where d is the standard Euclidean metric in $\mathbb{R}^2$.

Lemma 6. $\frac{\partial}{\partial s'} H(s, s') = -\cos \theta'$, and $\frac{\partial}{\partial s} H(s, s') = \cos \theta$.

Proof. This really is simply calculus. For the first result, fix and parameterize a small arc in the ellipse centered at s' , $c(t)$, where $c(t_0) = p'$. Choose a parameterization such that the tangent vector is unit length. Then, noting that $d(p, p') = \|p - p'\|$, we have

$$\begin{aligned} \frac{\partial}{\partial s'} H(s, s') &= \frac{d}{dt} \bigg|_{t=t_0} -d(p, c(t)) = \frac{-1}{2d(p, c(t))} (2(c'(t) \cdot (p - c(t)))) \bigg|_{t=t_0} \\ &= \frac{-\|c'(t_0)\| \|p - c(t_0)\| \cos \theta'}{\|p - c(t_0)\|} = -\cos \theta' \end{aligned}$$

by the cosine formula for the dot product of two vectors and since $\|c'(t_0)\| = 1$ by the parameterization. Hence $\frac{\partial H}{\partial s'} = -\cos \theta'$. The other result is similar. $\square$

Now apply this idea to any three points s_{-1} , s_0 , and s_1 on the ellipse. can these three points lie successively on an orbit? The answer is , yes, if

$$\frac{\partial}{\partial s'} H(s, s') + \frac{\partial}{\partial s} H(s, s') = 0.$$

That is, if s_0 is a critical point of the assignment

$$s \longmapsto H(s_{-1}, s) + H(s, s_1).$$

This is a variational approach to the construction of orbits, and techniques like this form the content of our course 110.427 Introduction to the Calculus of Variations.

Experiment 2. Consider a convex billiard with one pocket (corner) p . Find all possible bank shots to sink a ball at p .

Recall the notion of a strictly convex domain, where $B = \partial D$ has non-zero curvature (where B is C^2 and where the second derivative is non-zero). Visually, this means that there are no straight-line segments on B , and certainly no inflection points (changes in concavity). It also means that we can effectively take the angle of incidence to be from the open interval $(0, \pi)$ instead of the closed interval. Thus the state space is the open cone.

Here are some quick results: First, switch from the angular coordinate θ to the rectilinear coordinate $r = -\cos \theta$, so that for $\theta \in (0, \pi)$, we have $r \in (-1, 1)$.

Proposition 7. For a convex billiard, the billiard map

$$\Phi(s, r) = (\mathcal{S}(s, r), \mathcal{R}(s, r)) : C \rightarrow C$$

is area and orientation preserving.

Proof. The proof is constructive and based on simply calculus. □

Proposition 8. *If $B = \partial D$ is C^k (which means that the Euclidean coordinates are C^k functions of the length parameter), then both $\mathcal{S}$ and $\mathcal{R}$ are C^{k-1} .*

Proof. This is the Implicit Function Theorem. □

Proposition 9. *For D strictly convex, the billiard map has at least two period-2 orbits; at the diameter and at the width.*

One can describe the width in the following way. Take two distinct vertical parallel lines tangent to the billiard table (necessarily on “opposite sides” of the table). As one rotates the table, the distance between these lines will change. When one reaches the diameter of the table (the largest possible Euclidean distance between 2 points on the boundary), the two points will lie along the line perpendicular to the vertical lines. This perpendicular line segment represents one of the period-2 orbits. The other comes at the point when the two vertical lines reach a local minimum distance (which is the minimum distance for a strictly convex table). At this point again, the line segment joining the two tangencies will be perpendicular to the vertical lines and represent another period-2 orbit. This is the width of the table.

One final note, finding these period-2 orbits using this method involves finding where the vertical lines reach a minimum and maximum distance from each other. But this is what the generating function H is also doing, and why the generating function is particular good at finding period-2 orbits. It is actually good at finding period- n orbits also, but this goes a bit beyond this course.