Diana Davis

Billiards, Surfaces and Geometry: a problem-centered approach

March 22, 2023

Contents

Preface

The study of mathematical billiards is a beautiful field connecting many classical objects of study – rational numbers, regular polygons, and paper folding, to name a few – and its ideas are accessible to students at any level. I've written this book so that more instructors can create, and more students can take, a one-term introductory billiards course. I'm glad you're here.

This book has no content prerequisites beyond things like angles, slopes, and rational vs. irrational numbers. A few useful ideas from algebra and analysis are explained as they are needed. Some background in proof-writing is helpful. I first taught this course to undergraduate math majors at Williams College in 2016, and then to high school juniors and seniors at Phillips Exeter Academy in 2021 and 2023. These problems are also a great introduction for people who wish to contribute to the flat surfaces literature and community.

How to teach a course with this text

The intended format of a course associated to this book is as follows:

- First day of class: do Day 1 problems in class. For homework, students do Day 2 problems.
- Second day of class: students discuss their solutions to Day 2 problems. For homework, students do Day 3 problems.
- Third day of class: students discuss their solutions to Day 3 problems. For homework...

The problems teach the material on their own; no lectures are necessary. The problems are hard: I do not expect every student to solve every problem before class. I pitch them at this level on purpose, so that students have something to discuss with each other when they get to class. As a consequence, this book is not ideal for an independent study.

Materials needed

- pencil (for drawing right in this book)
- ruler (for creating accurate billiards trajectories)
- graph paper (also for creating accurate billiards trajectories)
- a set of colored pens or pencils (for color coding in many colors)
- scissors (for cutting up surfaces)
- tape (for reassembling the cut-up surfaces)
- string (for making an ellipse)
- access to a printer (for printing out extra copies of these pages for cutting and folding)

The problems in this text

This style of curriculum is integrated: rather than a single problem set with many problems about continued fractions (for example), problems on each topic are sprinkled across many days, gradually increasing in sophistication. This way, students have a chance to discuss each problem on a given topic before moving on to a harder one. For this reason, it's essential to leave class understanding each of the previous night's homework problems, as the next set of problems usually builds directly on them.

The vast majority of the problems in this book were written by its author, some based on her previous book *Lines in positive genus: An introduction to flat surfaces* and some specifically for this course; these are labeled in the margin as **DD**. Some problems are taken from *Geometry and Billiards* by Serge Tabachnikov, labeled in the margin as **ST**. The dissection problems are taken from *Mostly Surfaces* by Rich Schwartz, labeled in the margin as **RS**. Problems from Phillips Exeter Academy's materials are labeled **PEA**.

Deer Isle, Maine *Diana Davis*
June 2021

1

Introduction to billiards in many forms

In life, *billiards* is a game where a ball bounces around inside a rectangular table. In mathematics, we'll extend the notion of billiards considerably. In this first chapter, we'll meet billiards inside polygonal tables, billiards inside smooth tables, and billiards *outside* of a table. The idea of the first chapter is to introduce all of the big ideas of the course, in their simplest forms. We will understand the simple case very well, and then later when we study more complicated things, we will have a solid background on the simple cases to build on.

In this chapter, you'll learn to draw beautiful, accurate pictures of periodic billiard paths in a square billiard table. Drawing accurate pictures is an excellent tool that you can use to understand what's going on. I recommend that you draw a picture for every problem, or draw right on the picture that's in the book.

The most powerful tool in the study of billiards in polygons is *unfolding* the billiard table. In its unfolded form, the table becomes a surface, and the path of the ball becomes an infinite line. This opens up the study of linear paths on *flat surfaces*, which is a big area of current research and a main object of our study. We'll start with the square torus surface, and later study more complicated surfaces.

Another powerful tool is turning the *geometric* problem of a billiard path into a *combinatorial* problem about the list of edges that the ball hits. A list of symbols (edge names) is much simpler than a picture of a path, and these lists (called *bounce sequences* or *cutting sequences*) have a lot of beautiful structure.

Let's get started!

1.1 What are periodic billiard paths and where can we find them?

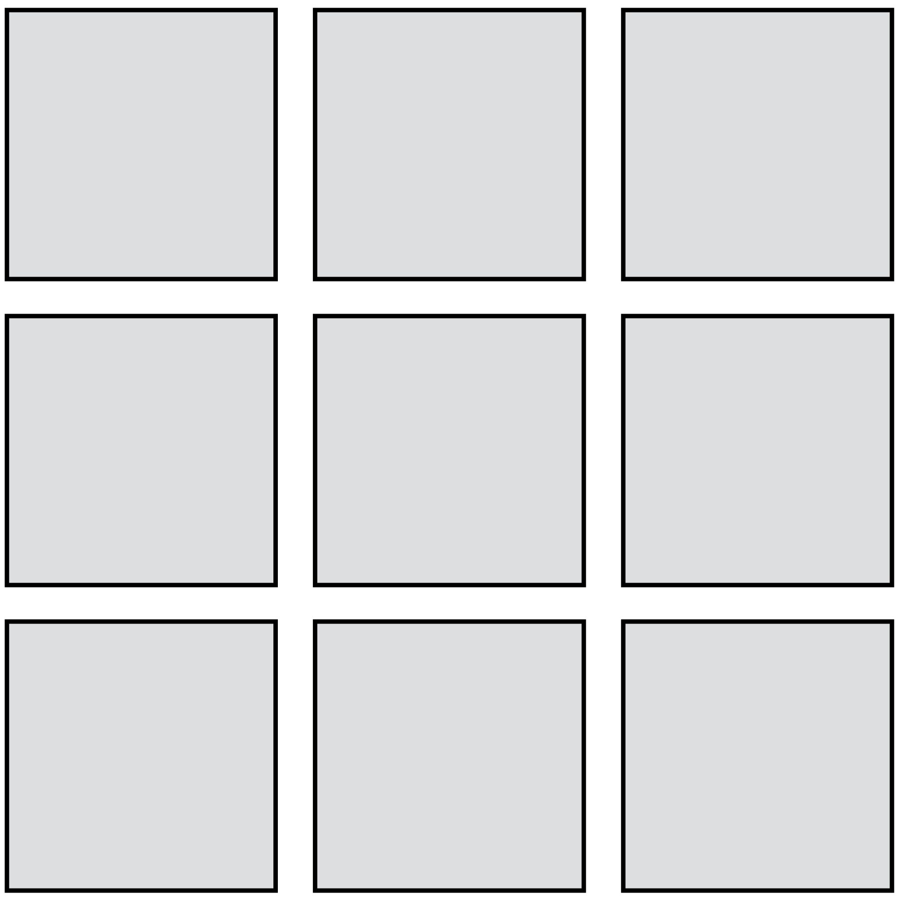
First day of class: in-class problems

DD

1.1. Consider a ball bouncing around inside a square billiard table. We'll assume that the table has no "pockets" (it's a billiard table, not a pool table!), that the ball is just a point, and that when it hits a wall, it reflects off and the angle of incidence equals the angle of reflection, as in real life.
(a) A billiard path is called *periodic* if it repeats, and the *period* is the number of bounces before repeating. Construct a periodic billiard path of period 2.

Note: Please consider drawing right on the page! This book belongs to you.

(b) For which other periods can you construct periodic paths?

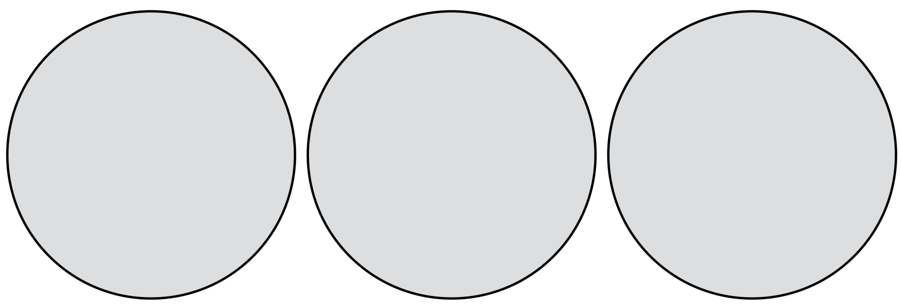


DD

1.2. Now consider a *circular* billiard table. Again assume that the ball is just a point, and that when it bounces off, the angle of incidence equals the angle of reflection. Note that in a billiard table with curved edges, the ball reflects off of the *tangent line* to the point of impact.
(a) Draw several accurate of billiard trajectories in a circular billiard table.

Write right on the pictures! That's why they're here.

(b) Consider paths that close up (*periodic* paths), and also paths that don't (*aperiodic* paths). What is the probability that a billiard path in the circular table is periodic?
(c) Describe the behavior in general.



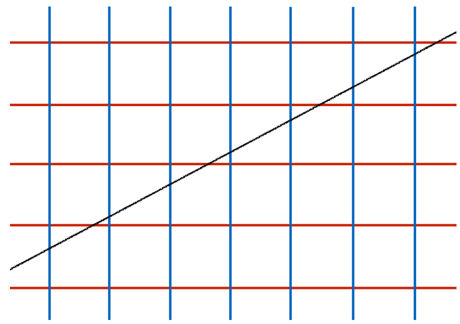
ST

1.3. Suppose 100 ants are on a log 1 meter long, each moving either to the left or right with speed 1 meter per minute. Assume the ants collide elastically (when they hit each other, each ant immediately turns around and goes the other way), and that when they reach the end of the log, they fall off. What is the longest possible waiting time until all the ants are off the log?

1.2 Billiards further afield

DD

1.4. Draw a line on an infinite square grid, and record each time the line crosses a horizontal or vertical edge. We will assume that the direction of travel along a line is always left to right. We could record the line to the right with the sequence . . . ●●●●●●●●●●●●●● . . ., or we could assign A to horizontal and B to vertical edges, and record it as $\ldots BABBABBABBA \ldots$.



(a) What is the slope of the line in the picture?
(b) Record this *cutting sequence* of colors, or of As and Bs, for several different lines. Describe any patterns you notice. What can you predict about the cutting sequence, from the line?
(c) What should you do if the line hits a vertex?

Here are the ways that people typically deal with lines that hit vertices, or billiard trajectories that hit corners of the table:

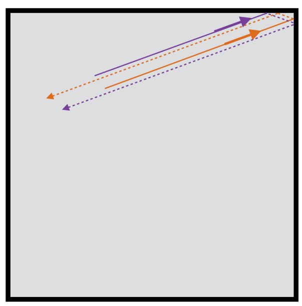
- *Authoritarian*: Trajectories are not allowed to hit vertices.
- *Minimalist*: If a trajectory hits a vertex, it stops.
- *Uncertain*: The vertex belongs to both sides, so it's ambiguous.
- *Positive attitude*: If the ball hits the pocket in the corner, you win!

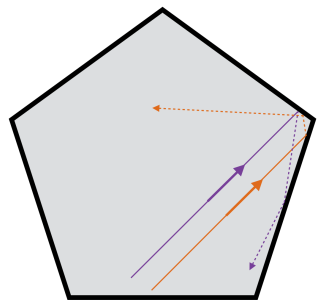
In any case, we generally consider trajectories that do not hit vertices.

Contextual note. The sequences of symbols in Problem 1.4 are called *cutting sequences*. Caroline Series (pictured to the left), a British mathematician, wrote a series of papers exploring cutting sequences on the square grid and linking them to other areas of mathematics. We will see that cutting sequences are related to group theory and continued fractions; Caroline also explained their relationship with hyperbolic geometry.

True story: A few years ago when I was teaching this course, I told my students that we don't let trajectories hit vertices, and they were dissatisfied with my explanation. Then two of them went and played squash together (for real) which is essentially billiards in a cube. The next day, they said: "now we agree, the ball should not be allowed to hit the vertex – when the squash ball hits the corner of the room, it bounces in a totally unpredictable direction!"

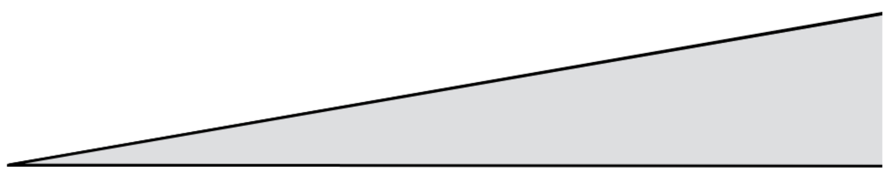
In fact, while the *cutting sequence* corresponding to a trajectory that hits a vertex is ambiguous, the forward *trajectory* itself is not necessarily ambiguous. For example, on the square billiard table, nearby parallel trajectories continue to be nearby and parallel after two reflections (left picture below). But on the regular pentagon, two nearby parallel trajectories have very different futures if they hit different sides of a vertex (right picture below). It turns out that if the vertex angle is a divisor of 2π, the behavior is like the square, and otherwise, the behavior is like the pentagon. (Thanks to Barak Weiss for pointing this out.) Since in the squash court situation described above, the vertex angle between the wall and the floor evenly divides 2π, perhaps the issue there is that the squash ball has a positive radius, and the problem arises when the ball hits both walls simultaneously.





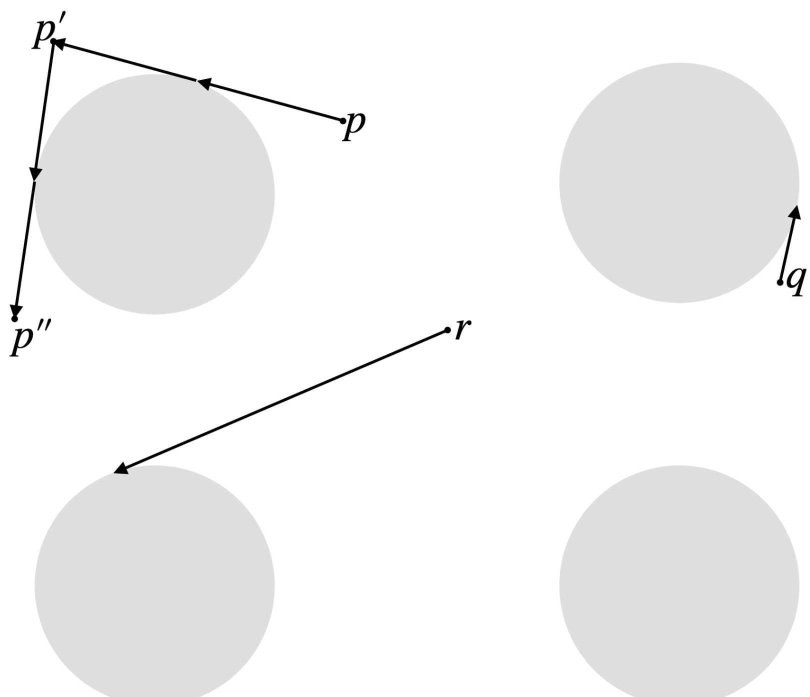
DD

1.5. Consider a billiard "table" in the shape of an infinite sector with a small vertex angle, say $10°$. Draw several examples of billiard trajectories in this sector (calculate the angles at each bounce so that your sketch is accurate). Is it possible for a trajectory (that does not hit the vertex) to go in toward the vertex and get "stuck"? Find an example of a trajectory that does this, or explain why it cannot happen.

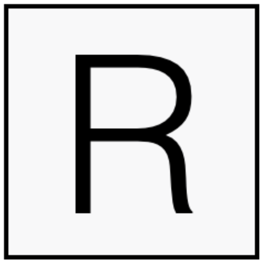


DD

1.6. *Outer billiards.* Though it may seem strange to call it "billiards," we can also define a billiard map on the *outside* of a billiard table. First, choose a starting point p, and a direction, either clockwise or counter-clockwise. Then draw the tangent line from p to the table in that direction to find the point of tangency. Double the vector from p to the point of tangency, and add this to p to get p', as in the picture. Repeat to find p'', and so on.
(a) Work out the first five or six iterations for the starting points given below, and then describe the behavior in general.
(b) What is the probability that p eventually returns to its starting point?
(c) What does the set of *all* the images of p look like? Consider the case when p returns to its starting point, and also the case when it doesn't.
(d) Can you make a periodic path of period 5?



DD



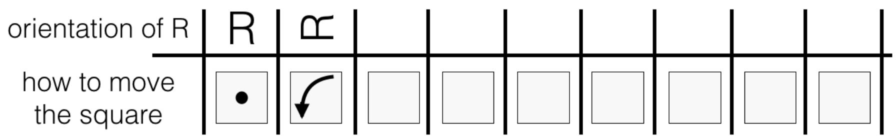
1.7. *Symmetries of the square.* If you turn a square 90° counter-clockwise, it looks the same as before. We call a 90° counter-clockwise rotation a *symmetry* of the square, because after you do it, you have a square just like the original.

In this problem, we'll find all the symmetries of the square. Of course, if you rotate a square by 90°, it looks identical to the original, so to keep track of the square's orientation, we'll draw an **R** on it.

(a) Cut out a square and draw an **R** on one side, as shown, and also hold it up to the light and trace through a backwards **R** on the back.

(b) How many different symmetries of the square can you find? Record in the first line of the table below the appearance of the **R** for each one.

(c) In the second line of the table, indicate how to move the square to achieve that position.

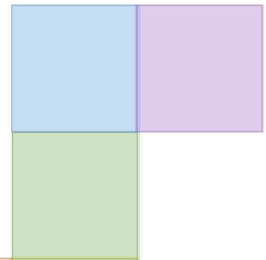


(d) Do you have all of them? If so, explain how you know.

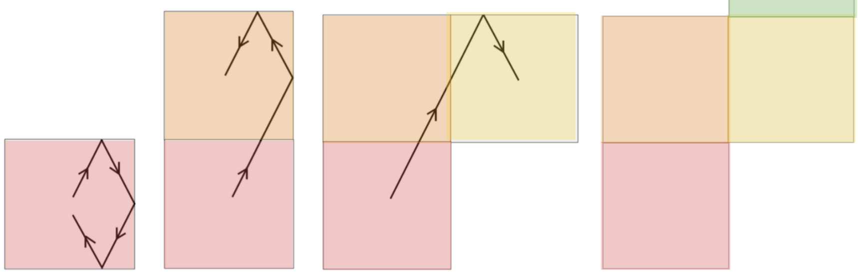
DD

1.3 We get more experience and build some tools

1.8. A powerful tool for understanding inner billiards is *unfolding* a trajectory into an infinite line, by creating a new copy of the billiard table each time the ball hits an edge. Two steps of the unfolding process are shown for a small piece of trajectory of slope ± 2 in the square.



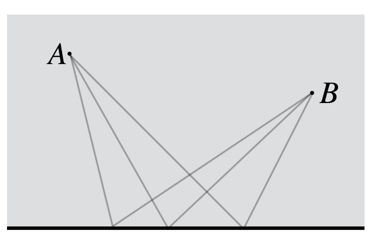
(a) Draw some more steps of the unfolding.
(b) Draw the complete billiard path in the original square: keep going until it closes up.



(c) Use the unfolding to explain why a trajectory with slope 2 yields a *periodic* (repeating) billiard trajectory on the square.
(d) Which other slopes yield a periodic billiard trajectory?

When we say “a trajectory with slope 2,” we are assuming that one edge of the square table is horizontal. If our billiard table is tilted, we just rotate it until it does have a horizontal edge. This is one way of reducing our problem (to polygons with a horizontal edge) and making it easier to talk to each other (“slope 2” instead of “with the edge, the trajectory makes an angle whose tangent is 2”). Another way to reduce our work is to only consider trajectories in a small sector of directions; this is what our work in Problem 1.7 will do for us in the future (Problem 2.10).

1.9. *The billiard reflection law, polygonal case.* We wish to show that, when a billiard trajectory hits the edge of the table, the angle of incidence equals the angle of reflection. We will use the *Fermat principle*: when the ball travels from point A, to the table's edge, to B, it follows the (locally) shortest path. We will consider the case when the ball hits a linear edge of the table. Use reflection in the edge to show that the shortest path from A to the edge to B satisfies the billiard reflection law.



It turns out that billiards on the square are related to number theory, via *continued fractions*. Continued fractions are generally not in the standard curriculum, but they are an efficient (and honestly quite fun) way of expressing real numbers as nested fractions. We'll play with continued fractions for a while to develop our skills, and then see how everything fits together in a few weeks.

1.10. The *continued fraction expansion* gives an expanded expression of a given number. To obtain the continued fraction expansion for a number, say 15/11, we do the following:

$$\frac{15}{11} = 1+\frac{4}{11} = 1+\frac{1}{11/4} = 1+\frac{1}{1+7/4} = 1+\frac{1}{2+3/4} = 1+\cfrac{1}{2+\cfrac{1}{4/3}} = \mathbf{1}+\cfrac{1}{\mathbf{2}+\cfrac{1}{\mathbf{1}+\cfrac{1}{\mathbf{3}}}}.$$

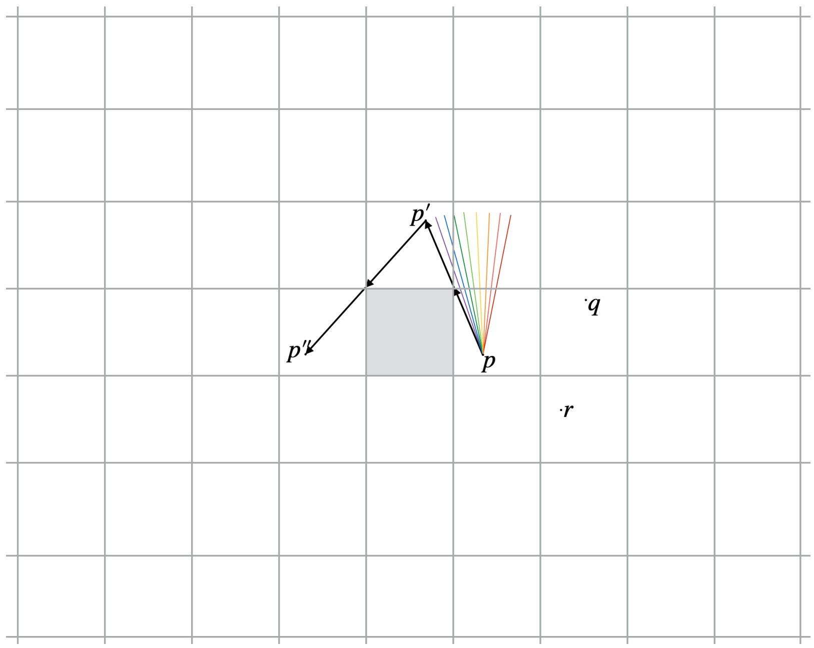
The idea is to pull off 1s until the number is less than 1, take the reciprocal of what is left, and repeat until the reciprocal is a whole number. Since all the numerators are 1, we can denote the continued fraction expansion compactly by recording only the bolded numbers: $15/11 = [1; 2, 1, 3]$. The semicolon indicates that the initial 1 is outside the fraction.
(a) Find the first few steps of the continued fraction expansion of π, and explain why the common approximation 22/7 is a good choice. What is the best fraction to use, if you want a ratio of integers that have 3 or fewer digits?
(b) Find the continued fraction expansion of $3.14 = 157/50$.
(c) Find a rational approximation of the number whose continued fraction expansion is $[1; 1, 1, 1, \ldots]$. This number, known as the golden ratio ϕ, is sometimes called the "most irrational number." Explain.

Contextual note. Above, you found that the continued fraction expansion of 3.14 is $[3; 7, 7]$. Is this the best approximation for π that we can get with a ratio of integers with three digits or fewer? No, Problem 1.10 **(a)** and **(b)** show that we can find a better rational approximation by using the continued fraction expansion, and truncating it at a convenient point.

DD

1.11. We can also play outer billiards on polygonal tables. Here, the "tangent line" is always through a vertex − you can think of sweeping a line counterclockwise until it hits a vertex, as shown.

Find the forward orbits of the points p, q and r shown below. Can you find any periodic trajectories? Can you find any aperiodic trajectories? *Hint:* be accurate. Consider using a ruler.



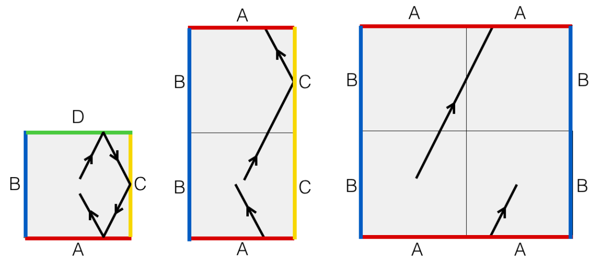
Contextual note. The outer billiards system was proposed as a toy model for planetary motion: the table is the sun, and the point is the planet bouncing around it. It is easier to analyze a *discrete* dynamical system, where a planet jumps from place to place, than a *continuous* dynamical system where planets move smoothly.

It is important to know whether our solar system is stable or whether Earth will spin out away from the sun, or something else. Related to this, it was for a long time an open problem whether there exists a shape of table, and a point outside the table, such that under the outer billiard map the point eventually bounces off to infinity. The answer is yes: Rich Schwartz (left) showed that the *Penrose kite* has this property, and Dmitry Dolgopyat and Bassam Fayad showed the same for the half disk, both in 2009.

1.4 We learn to draw accurate pictures

DD

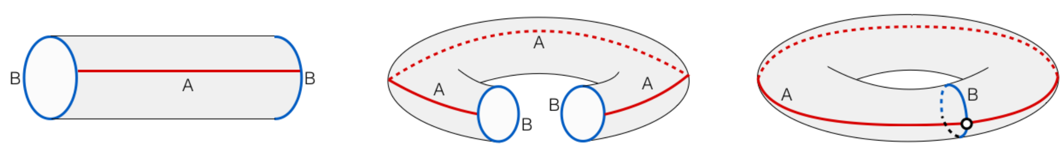
1.12. Here's another way that we can unfold the square billiard table. First, unfold across the top edge of the table, creating another copy in which the ball keeps going straight. The new top edge is just a copy of the bottom edge, so we now label them both A to remember that they are the same. Similarly, we can unfold across the right edge of the table, creating another copy of the unfolded table. The new right edge is a copy of the left edge, so we now label them both B. When the trajectory hits the top edge A, it reappears in the same place on the bottom edge A and keeps going. Similarly, when the trajectory hits the right edge B, it reappears on the left edge B.



(a) The partial billiard trajectory shown on the left part of the top figure repeats after 6 bounces. Sketch in the rest of the trajectory in each of the three pictures above. What is its corresponding *cutting sequence* for trajectory on the surface on the right part of the figure?

(b) When we unfolded the trajectory to a line in Problem 1.8, we created a new copy of the table every time the trajectory crossed an edge. Explain why, in the picture above, just 4 copies is enough.

(c) Suppose that you have a rectangular sheet of very stretchy rubber. You tape together the top and bottom edges (edge A) to create a tube, and then you curl around the tube and attach the open ends to each other along their edges (edge B). These steps are shown in the figure below. Explain. The result is called a *torus*, the surface of a donut.



DD

1.13. Show that the cutting sequence corresponding to a line of slope $1/2$ on the square grid is periodic. Which other slopes yield periodic cutting sequences? What can you say about the period, from the slope? Write proofs of your claims.

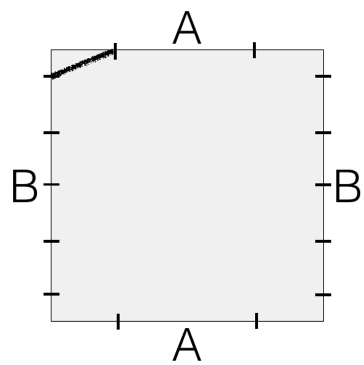
DD

1.14. Prove that every billiard trajectory on the square with irrational slope is aperiodic.

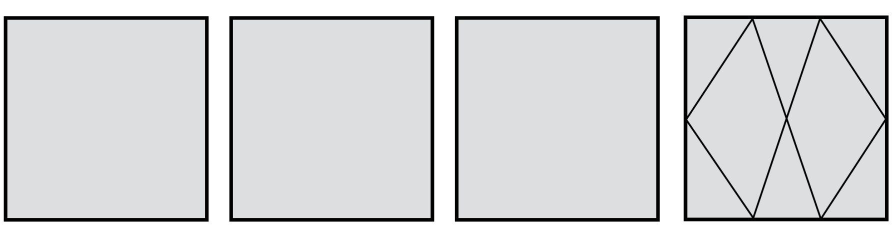
Contextual note. The field of mathematics devoted to the study of objects like the square torus that we just constructed is called *flat surfaces*. There are hundreds of mathematicians (including those featured in this book) working on flat surfaces, spread across the globe and particularly concentrated in France. It is currently a "hot" field, with many papers posted every week with new results. Two of the 2014 Fields Medals were awarded to mathematicians working in this area.

Amie Wilkinson (left) created a phenomenal animation showing how, as we did with the square in Problem 1.12, we can make an octagon into a flat surface. It is at 26:00 of her Fields Symposium public lecture from 2018, available here:
`https://www.youtube.com/watch?v=zjccKzHIniw&t=1560s`

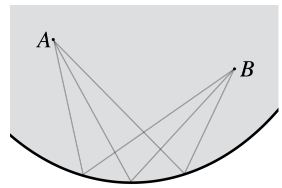
DD



1.15. In problem 1.12, we ended up with a trajectory of slope 2 on the *square torus* surface. The picture to the right shows some scratchwork for drawing a trajectory of slope 2/5 on the square torus. Starting at the top-left corner, connect the top mark on the left edge to the left-most mark on the top edge with a line segment, as shown. Then connect the other six pairs with parallel segments, down to the bottom-right corner.
(a) Explain why, on the torus surface, these line segments connect up to form a continuous trajectory. Follow the trajectory along, and write down the corresponding cutting sequence of As and Bs.
(b) Exactly where should you place the tick marks so that all of the segments have the same slope? Prove your claim.
(c) Create an accurate picture for a trajectory of slope 1/2 and then 3/2. *Hint*: make sure that all of your segments look parallel.
(d) Draw a picture of a *billiard* trajectory with slope $\pm 2/5$.
(e) Something is wrong with the "billiard trajectory" on the right. Explain.



ST



1.16. *The billiard reflection law, curved case.* We proved this law for linear boundaries in Problem 1.9; now we will prove it for curved boundaries. Again, we will use the principle that when the ball travels from point A, to the table's edge, to B, it follows the (locally) shortest path. Prove that, when the ball follows the shortest path, the angle of incidence equals the angle of reflection.

Hint: One way is use the multivariable calculus principle that the gradient vector of the distance function points in the direction of greatest increase of the function, and apply this to both A and B. Another way is to apply an equilibrium tension argument from physics, imagining the boundary of the table as a wire, and the billiard trajectory as an elastic string fixed at A and B that passes through a small ring threaded through the boundary wire.

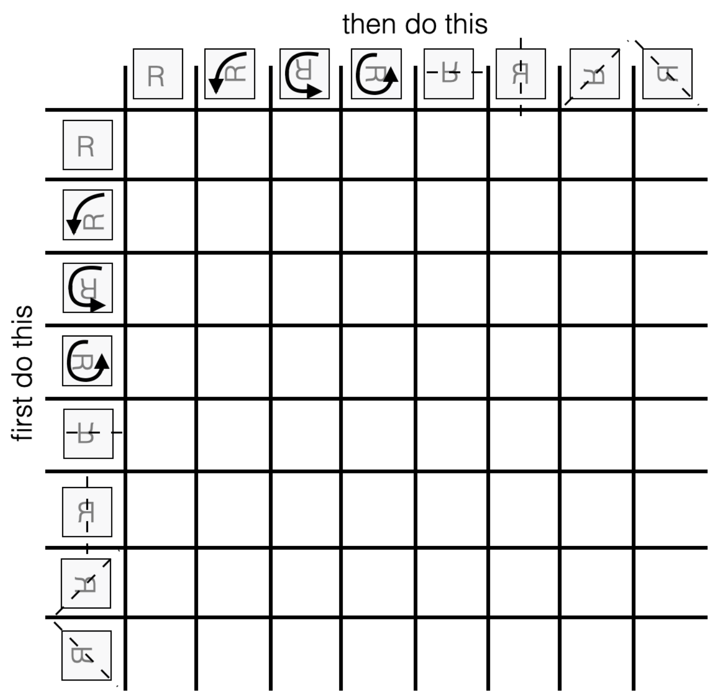
1.5 We do a little bit of group theory

DD

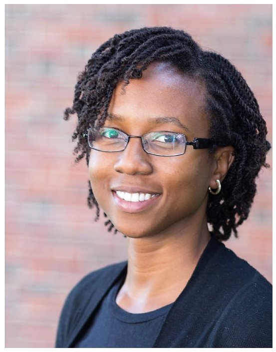
1.17. In Problem 1.7, you found the eight symmetries of the square. It turns out that these eight symmetries form a *group*, called the *dihedral group* of the square. For a set of symmetries to be a group, it must have the following properties:

1. It contains an *identity element*, a symmetry that does nothing;
2. Each symmetry has an *inverse*, a symmetry that “undoes” its action;
3. It is *closed*: composing two symmetries (doing one and then the other) yields a symmetry in the group.
4. Composing symmetries is *associative*, i.e. $a(bc) = (ab)c$ for symmetries a, b, c.

(a) Explain why parts (1), (2) and (3) hold for the symmetries of the square.
(b) Fill in the following table (known as a *Cayley table*). Do you see any patterns? Prove that they exist. *Note:* it is much easier to see patterns if you denote a symmetry by its arrow or dashed line; it is much more difficult to see patterns if you use the oriented R.



(c) Does this group of symmetries commute, i.e. is $ab = ba$ true for every pair of symmetries a, b? If not, is there *any* pair of symmetries that commutes?

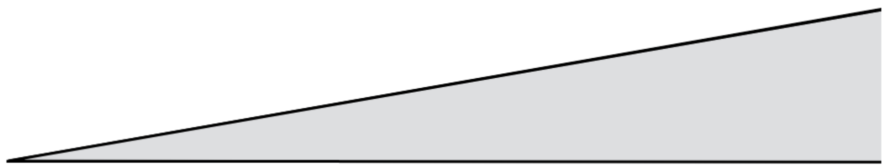


Contextual note. Some groups, such as the group of integers under multiplication, are commutative. Others, like the one above, are not. For some sets, such as the set of integers, you can actually define *two* operations (e.g. addition and multiplication) on them, and this makes the set into a *ring*. Professor Haydee Lindo of Harvey Mudd College (left) studies commutative rings.

DD

1.18. Consider again (following Problem 1.5) a billiard table in the shape of an infinite sector, with vertex angle α. Use unfolding to show that any billiard on such a table makes **(a)** finitely many bounces, and in fact **(b)** at most $\lceil \pi/\alpha \rceil$ bounces. *Hint*: Unfold the sector as many times as you can.

Here the notation $\lceil \cdot \rceil$ is the "ceiling" and means "round up," e.g. $\lceil \pi \rceil = 4$.



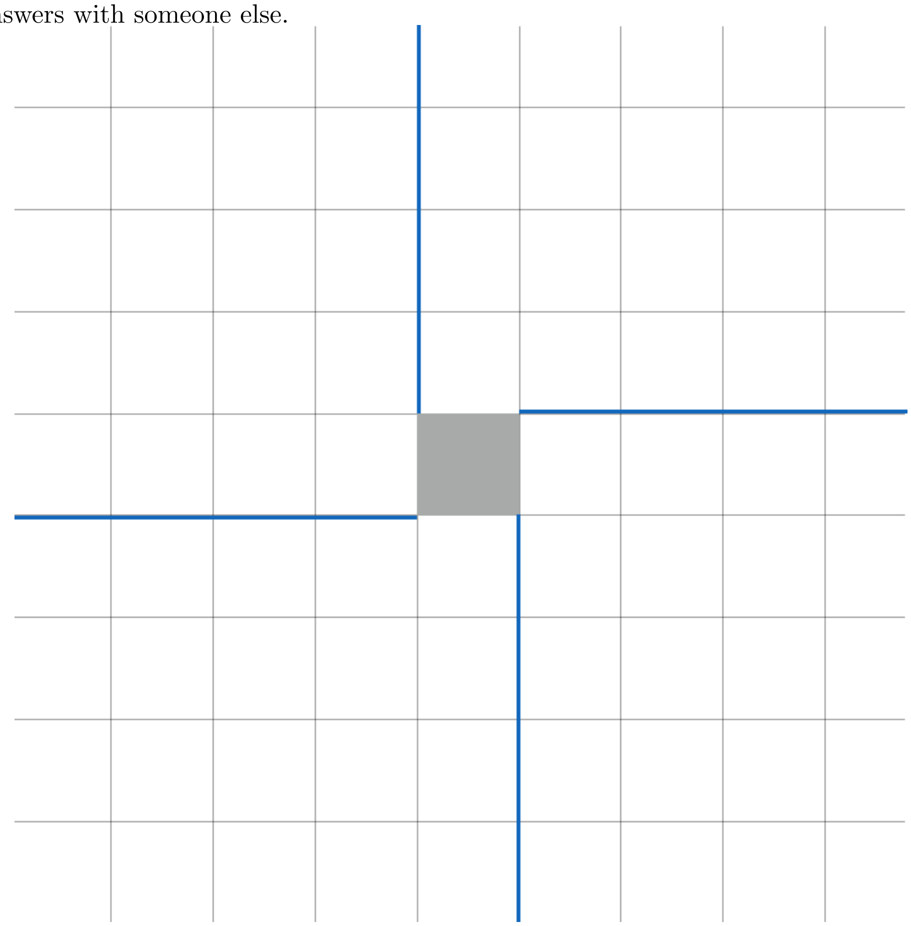
DD

1.19. Let's gather some data and make some conjectures.
(a) Construct an *accurate* (see Problem 1.15) picture of a trajectory on the square torus with slope 3/4. Repeat for two more slopes of your choice.
(b) For each of your trajectories, find the corresponding cutting sequence.
(c) Note down any observations. What is the relationship between the slope and the cutting sequence?

DD

1.20. You will need: colored pens or pencils. Consider again outer billiards on the square table, in the counter-clockwise direction.
(a) Points p on the blue lines are not allowed, because their images p' are ambiguously defined. Explain.
(b) Points p whose image p' is on a blue line are also not allowed. Explain. These are the *inverse images* of the blue points. Color these points red.
(c) The inverse images of the red lines are also not allowed. Explain. Color these points green. *Hint:* each one has two pieces.
(d) Color the inverse images of the green points black. Keep going, with different colors at each step. Describe the full set of disallowed points.

The purpose of specifying the colors above is so that you can check your answers with someone else.



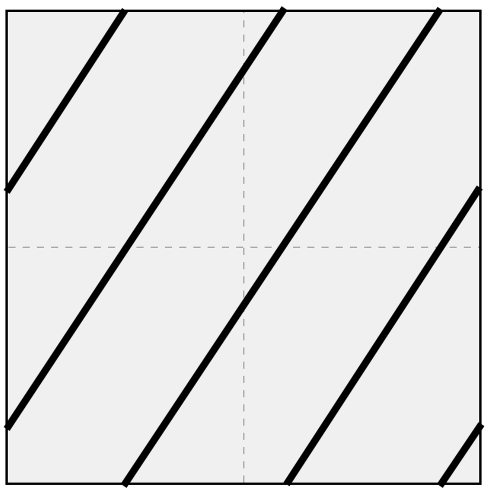
Note: The resemblance of the (incomplete) diagram to a swastika is unfortunately impossible to avoid. This symbol was first used 12,000 years ago; it is a natural construction that has become synonymous with an odious regime.

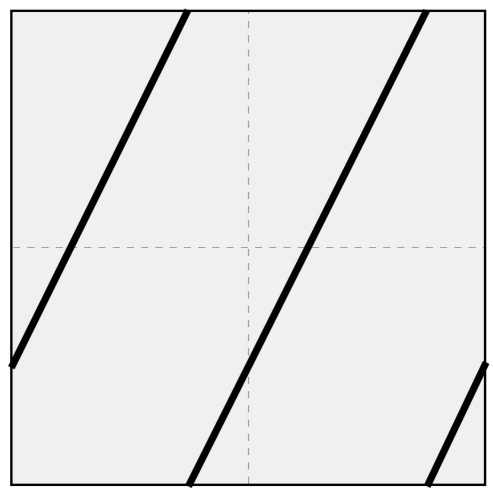
1.6 We fold up our billiard trajectories

DD

1.21. You will need: tissue paper, or other thin paper. We saw that a billiard trajectory on the square table can be *unfolded* to a line on the square torus. Going the other way, a trajectory on the square torus can be *folded* to a billiard trajectory on the square table.
(a) Confirm that each trajectory below is a closed path on the square torus.
(b) Carefully trace the first figure onto a piece of thin paper. Fold it in quarters as indicated by dashed lines, and then hold it up to the light: behold, a billiard trajectory!

Repeat for the second figure.
(c) For each picture, find the corresponding cutting sequence on the square torus, and also on the square table. Note any observations.





As previously explained, the study of flat surfaces is a very hot field these days, and many people are proving results about them. Sometimes, people are perfectly satisfied with results about flat surfaces, and they don't fold up their surfaces to get a billiard table back. You will not be one of these people.

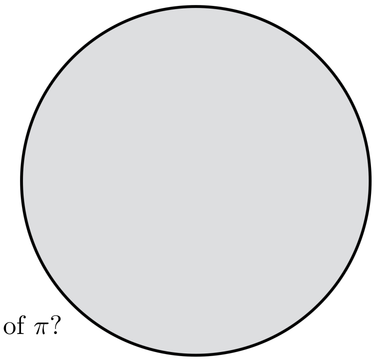
DD

1.22. In Problem 1.20 you showed that for outer billiards on the square, all of the points on the square grid lines are not allowed. Choose a point p that is *not* on one of the grid lines. Under the outer billiard map, this point reflects through a sequence of vertices $v_1, v_2, \ldots$ where each v_i is one of the four vertices of the square table. Explain why *every* point that is in the same (open) square as p reflects through that *same* sequence of vertices.

SPACE FOR A MATHEMATICIAN SPOTLIGHT HERE.

ST

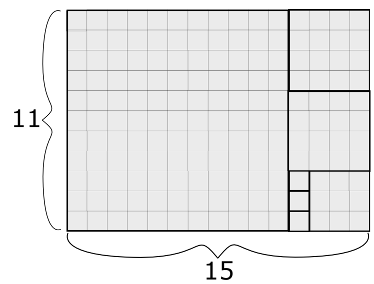
1.23. Consider a billiard trajectory in the unit circle, where at each impact the trajectory makes angle α with the (tangent line to the) circle.
(a) Find the central angle θ from the circle's center, between each impact point and the next one, as a function of α.
(b) Prove that if $\theta = 2\pi p/q$ for integers p and q, then every billiard orbit is q-periodic and makes p turns around the circle before repeating.
(c) What happens if θ is *not* a rational multiple of π?



DD

1.24. Geometrically, the continued fraction algorithm for a number x is:

1. Begin with a $1 \times x$ rectangle (or $p \times q$ if $x = p/q$).
2. Cut off the largest possible square, as many times as possible. Count how many squares you cut off; this is a_1.
3. With the remaining rectangle, cut off the largest possible squares; the number of these is a_2.
4. Continue until there is no remaining rectangle. The continued fraction expansion of x is then $[a_1, a_2, \ldots]$ or possibly $[a_1; a_2, \ldots]$.



(a) Draw the rectangle picture for 5/7 to geometrically compute its continued fraction expansion, and
(b) compute the continued fraction expansion for 5/7 in the way explained in Problem 1.10, and check that your results agree. Explain why this geometric method is equivalent to the fraction method previously explained, for determining the continued fraction expansion.

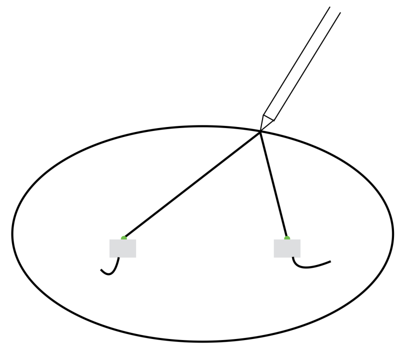
DD

1.25. In Problem 1.15, we put 2 marks on edge A and 5 marks on edge B and connected the marks to create a trajectory with slope 2/5. Do the same with 4 marks on edge A and 10 marks on edge B. Explain what you get.

1.7 Automorphisms come for the torus

DD

1.26. Prove that a trajectory on the square torus is periodic if and only if its slope is rational.



1.27. You will need: string, tape.
(a) Mark two dots on a piece of paper, and tape down your piece of string on each dot, leaving a lot of slack in the string. With your pencil, pull out the string until it is taut and trace out all the points the pencil can reach, as shown.
(b) Each of the two endpoints of the string is called a *focus* of the ellipse. Show that a billiard trajectory through one focus reflects through the other focus. In other words, the string is a billiard path in the ellipse.

Contextual note. The reflection property of ellipses is well known, and appears in architecture as the *whispering gallery.* Several U.S. state house rotundas, and the Statuary Hall at the U.S. Capitol building, have ellipsoidal ceilings, so if you stand at one focus, you can hear someone whisper at the other. Since legislative chambers are often arranged with members of the two political parties on opposite sides, people can actually sit at one focus and listen to what members of the other party are saying at the other focus!

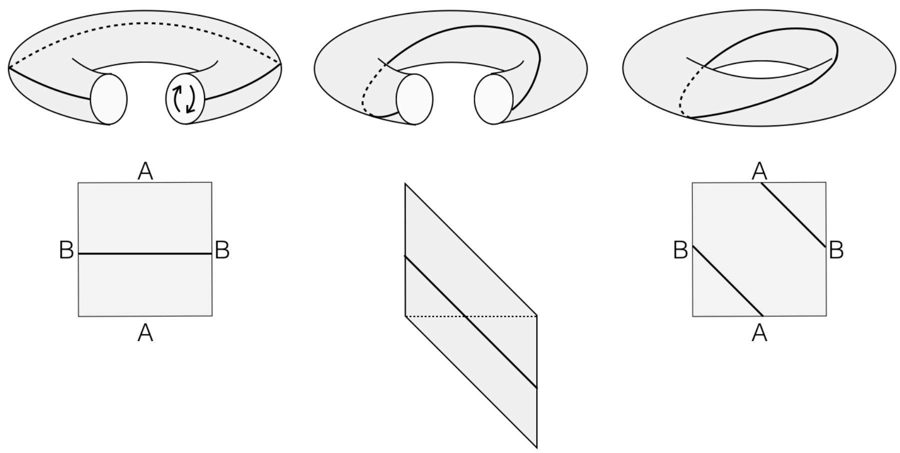
An accessible and impressive example of this is in Grand Central Station in New York City (above), where although the background noise is very loud, if you speak into one column, someone on the opposite column can hear you.

DD

1.28. An *automorphism* of a surface is a bijective action that takes the surface to itself. In other words, it modifies the surface but creates neither holes nor overlaps, and preserves the surface's structure. Two types of automorphisms of the square *torus* come from symmetries of the *square* itself: reflections and rotations, as we found in Problems 1.7 and 1.17.
(a) Explain what a vertical reflection of the square torus looks like on the torus surface. You might think about what it does to the surface, or to a closed path drawn on the surface.
(b) Do the same for a horizontal reflection. What about diagonal reflections, or rotations?

DD

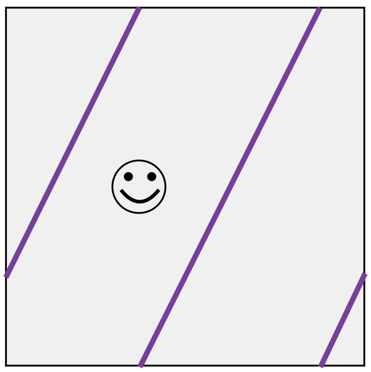
1.29. It turns out that there is a third type of automorphism of the square torus, that is *not* a symmetry of the square: a *shear*. The shear is shown below on the square (bottom) and on the 3D surface (top), where its effect is to twist the torus.
(a) Explain the effect of this shear on the surface, and on a trajectory drawn on that surface.
(b) What 2×2 matrix, applied to the "unit square" $[0,1] \times [0,1]$ shown in the bottom-left picture, gives the parallelogram shown in the bottom-middle picture?

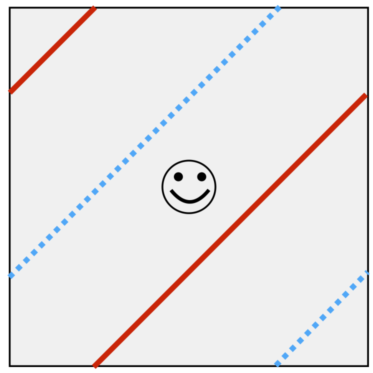


SPACE FOR A MATHEMATICIAN SPOTLIGHT HERE.

1.8 Hands-on activities for Chapter 1

1.30. The pictures below show linear trajectories on the square torus, as usual.
(a) Explain why the purple trajectory (left) is a single trajectory, while the red and blue trajectories (solid and dashed, right) are two different trajectories.
(b) The red and blue trajectories partition the square torus into two pieces. In other words, if the trajectories were walls, the smiley person could only explore half of the torus. Justify this statement.
(c) Also explain why the purple trajectory does *not* partition the torus into two pieces – the smiley person can explore the whole thing.





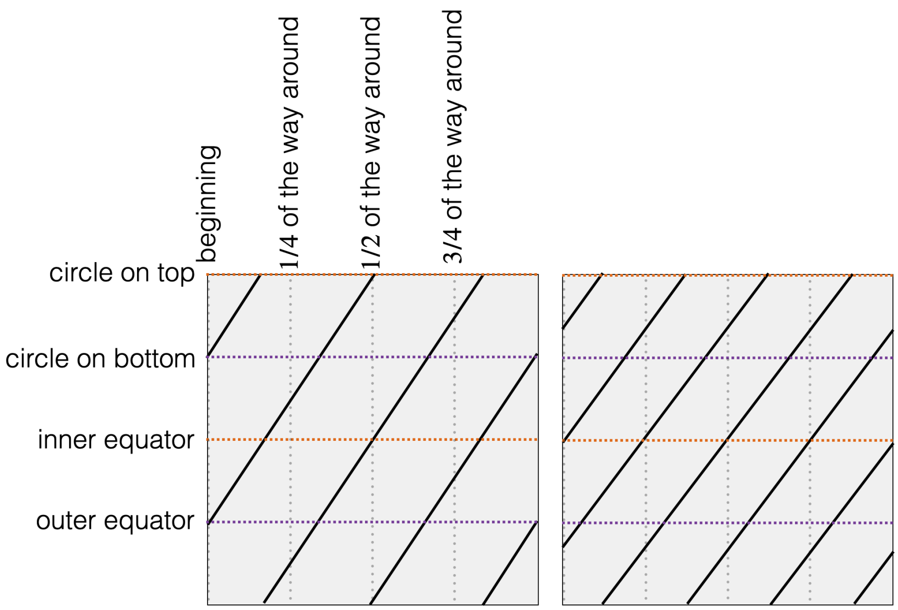
1.31. *Cutting a bagel into two linked rings.* **You will need: bagel with a large hold in the middle; serrated knife; tray to catch the crumbs.**
1. Draw the red and blue trajectories on your bagel.
2. Cut the bagel: The pointy end of the knife should follow the red trajectory, while the handle follows the blue trajectory. Flip the associated colors halfway through, to keep the handle on the outside.
3. Separate your bagel into linked rings!
(a) Explain why the procedure above leads to linked rings.
(b) Explain what would have happened if you had cut along the purple trajectory instead.

This activity originally came from George Hart's website.

Below are pictures of bagels with trajectories that correspond to slopes $1/2$, 2 and $3/2$, respectively, on the square torus.

DD

1.32. You will need: bagel with a large hole in it, marker. Choose a periodic trajectory (such as one of those below), and find a way to mark your bagel to indicate where to draw the trajectory. One method is suggested below. Then connect up your marks with smooth curves!



Connection to knot theory: Imagine that the bagel disappears, and all that is left is the trajectory corresponding to slope p/q, now made out of a piece of string. It turns out that if p and q are relatively prime, then you get a *knot* – a knotted-up loop that you can't untangle into a circle. The trajectory with slope p/q corresponds to the (p, q) *torus knot*, meaning that it goes through the center p times and around the outside q times.

2

Billiards, automorphisms and continued fractions

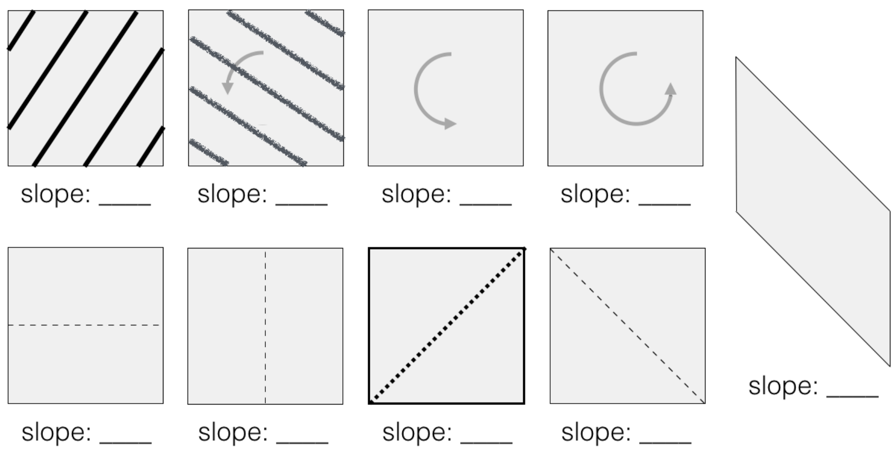
In Chapter 1, we became acquainted with our main protagonists: billiards, automorphisms, and continued fractions. In Chapter 2, we will build a grand unifying theory of how they all relate. It's a beautiful theory, and you'll see how it all comes together as you work through the problems in this chapter.

Our goal is to understand the ideas of billiards, automorphisms and continued fractions really well for the square torus and square billiard table. In subsequent chapters, we will generalize these ideas to more complicated systems, where things will be analogous to our work on the square.

2.1 We apply symmetries to trajectories

DD

2.1. Given a trajectory on the square torus, we want to know what happens to that trajectory if we apply a symmetry of the surface. To do this, we can sketch the trajectory before and after applying the symmetry. Do so below for each of the eight symmetries of the square, as indicated by the curved arrow or the reflection line, and for the shear. I've done one for you.



The flip across the positive diagonal is in bold because we will use it later.

DD

2.2. (Continuation) For each symmetry above, make a guess about what it does to a starting slope of the form p/q. Then prove your answers correct!

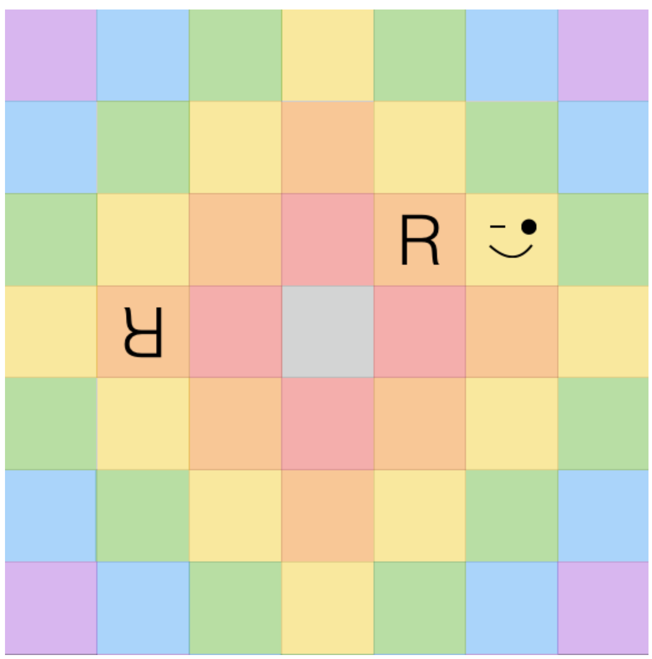
DD

2.3. An active area of research is to describe all possible cutting sequences on a given surface. On the square torus, that question is: "Which infinite sequences of As and Bs are cutting sequences corresponding to a trajectory?" Let's answer an easier question: How can you tell that a given infinite sequence of As and Bs is *not* a cutting sequence? You have computed many examples of cutting sequences that *do* correspond to a line on the square grid or square torus. Now make up an example of an infinite sequence of As and Bs that *cannot* be a cutting sequence on the square grid or square torus, and justify your answer.

Contextual note. As described above, an active area of research is to describe all possible cutting sequences on a given surface. John Smillie and Corinna Ulcigrai (left) classified all cutting sequences on the *regular octagon* surface, which is created similarly to the square torus. Because cutting sequences are infinite, and most are not periodic, it turns out that there is no finite criterion for deciding whether a given cutting sequence is valid: the algorithm necessarily requires a possibly unbounded number of steps. We will see that the same is true for cutting sequences on the square.

DD

2.4. In Problem 1.22, we showed that under the outer billiard map on the square, points in a given square move together. Let's explore *how* they move.
(a) Plot the complete orbit (meaning, until you get back to where you started) of the R and of the winky face under the counter-clockwise outer billiard map. One step is shown for the R. *Hint:* to determine the orientation of the image square, you can consider the image of each corner of the square. *Another hint:* the Rs are on orange squares, and the winky faces are on yellow squares.
(b) Prove that the square of the outer billiard map (this means that you apply it twice) is a *translation*.



2.2 We dream of an action on cutting sequences

ST

2.5. Theorem (billiards in an ellipse). If one segment of a billiard trajectory doesn't pass through the focal segment, then no segments of that trajectory pass through the focal segment, and furthermore all the segments of the trajectory are tangent to the same confocal ellipse.

More precisely: *Consider an ellipse E with foci F_1, F_2. If some segment of a billiard trajectory does not intersect the focal segment F_1F_2 of E, then no segment of this trajectory intersects F_1F_2, and all segments are tangent to the same ellipse E' with foci F_1 and F_2.*

Let's prove it! Steps of the proof below are color-coded to match the corresponding part of the picture.

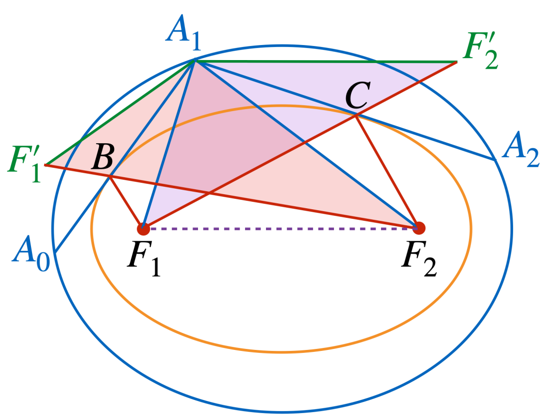
(a) Consider the billiard trajectory $A_0A_1A_2$ in the larger ellipse E shown in the figure. Explain why $\angle A_0A_1F_1 = \angle A_2A_1F_2$.

(b) Reflect F_1 across $\overline{A_0A_1}$ to create F_1', and reflect F_2 across $\overline{A_1A_2}$ to create F_2'. Explain why $\angle A_0A_1F_1' = \angle A_0A_1F_1$ and $\angle A_2A_1F_2' = \angle A_2A_1F_2$.

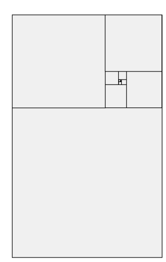
(c) Show that $\Delta F_1'A_1F_2$ and $\Delta F_1A_1F_2'$ are congruent.

(d) Mark the intersection of $\overline{F_1'F_2}$ with $\overline{A_0A_1}$ as B, and the intersection of $\overline{F_1F_2'}$ with $\overline{A_1A_2}$ as C. Show that the *string length* $|\overline{F_1B}| + |\overline{BF_2}|$ is the same as the string length $|\overline{F_1C}| + |\overline{CF_2}|$.

(e) Prove the theorem as stated above.



DD



2.6. Explain why a cutting sequence on the square torus can have blocks of multiple As separated by single Bs, or blocks of multiple Bs separated by single As, but not both.

DD

2.7. Find the continued fraction expansions of $3/2, 5/3, 8/5$, and $13/8$. Describe any patterns you notice, and explain why they occur.

Contextual note. So far, to determine the effect of a surface automorphism (symmetry) on a trajectory lying on that surface, we have drawn a picture of original trajectory and of the transformed trajectory (Problem 2.1). It's a great way to understand what's going on, but it's not super efficient.

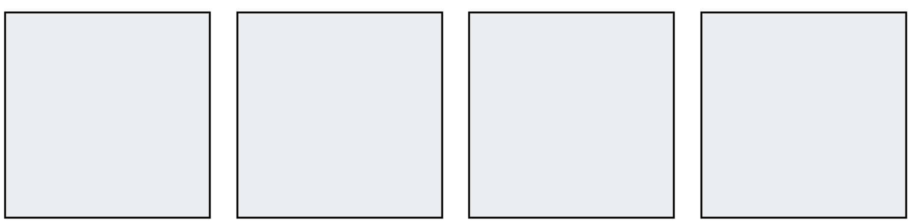
A much more efficient way to write down the effect of the automorphism would be to record how it affects the *cutting sequence* corresponding to a trajectory. Then we could just act on the cutting sequence – an operation on symbols, not on pictures! – and get the cutting sequence corresponding to the transformed trajectory. Irene Pasquinelli (left) figured out how to do this for a large class of surfaces. We'll work on this problem for the square torus now.

DD

2.8. Given a trajectory τ on the square torus, we want to know what happens to that trajectory under an automorphism of the surface.[1] We'll do this by comparing their cutting sequences: the cutting sequence $c(\tau)$ corresponding to the original trajectory τ, and the cutting sequence $c(\tau')$ corresponding to the transformed trajectory τ'. The goal is to figure out how to get $c(\tau')$ directly from $c(\tau)$.
(a) Let τ_2 be the trajectory of slope 2. Sketch τ_2, and find $c(\tau_2)$.
(b) For each automorphism (1)-(5) below, apply it to τ_2 to get a transformed trajectory τ_2', sketch τ_2', and compute $c(\tau_2')$.

1. reflection across a horizontal line;
2. reflection across a vertical line;
3. reflection across the positive diagonal;
4. reflection across the negative diagonal;
5. rotation by 90° counter-clockwise.

(c) Explain how to obtain $c(\tau')$ from $c(\tau)$ for a general trajectory τ, for each of the five automorphisms. Prove your answer correct.



DD

2.9. For each of the five automorphisms in the previous question:
(a) Find the 2×2 matrix that performs this automorphism. For the purpose of this question, assume that the square torus is centered at the origin.
(b) Find the determinant of each matrix and give a geometric explanation for why they all turn out to be ± 1.

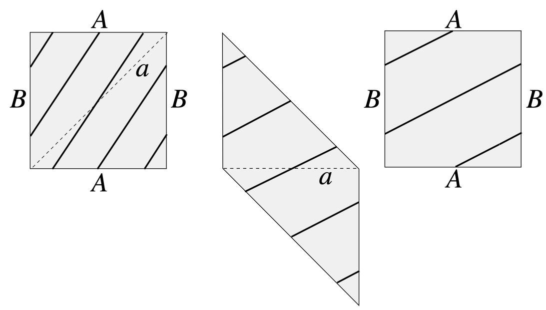
[1] τ is spelled *tau* and rhymes with "cow."

2.3 The dream comes true

In these problems, we will determine the effect of the shearing automorphism from Problem 1.29 on a trajectory τ and its cutting sequence $c(\tau)$.

DD

2.10. First, we will apply symmetry to reduce our work to just one set of trajectories. Show that, given a linear trajectory in *any* direction on the square torus, we can apply rotations and reflections so that it is going left to right with slope ≥ 1.



Since we have reduced to the case of slopes that are ≥ 1, we will analyze the effect of the vertical shear $\left[\begin{smallmatrix}1&0\\-1&1\end{smallmatrix}\right]$, because these slopes work nicely with this shear. Later we will show that every shear can be reduced to this case.

As an example, we'll use the trajectory τ with slope $3/2$, with corresponding cutting sequence $c(\tau) = \overline{BAABA}$ (left picture). We shear it via $\left[\begin{smallmatrix}1&0\\-1&1\end{smallmatrix}\right]$, which transforms the square into a parallelogram (middle picture), and then we reassemble the two triangles back into a square torus, while respecting the edge identifications (right picture). The new cutting sequence is $c(\tau') = \overline{BAB}$.

2.11. Notice that the horizontal edge A in the right picture corresponds to dashed edge a in the left and middle pictures. We can use this *auxiliary edge*, and its corresponding edge crossings, to form an *augmented cutting sequence* $\overline{BAaABA}$, which leads us to the *derived* cutting sequence $\overline{BAA}$:

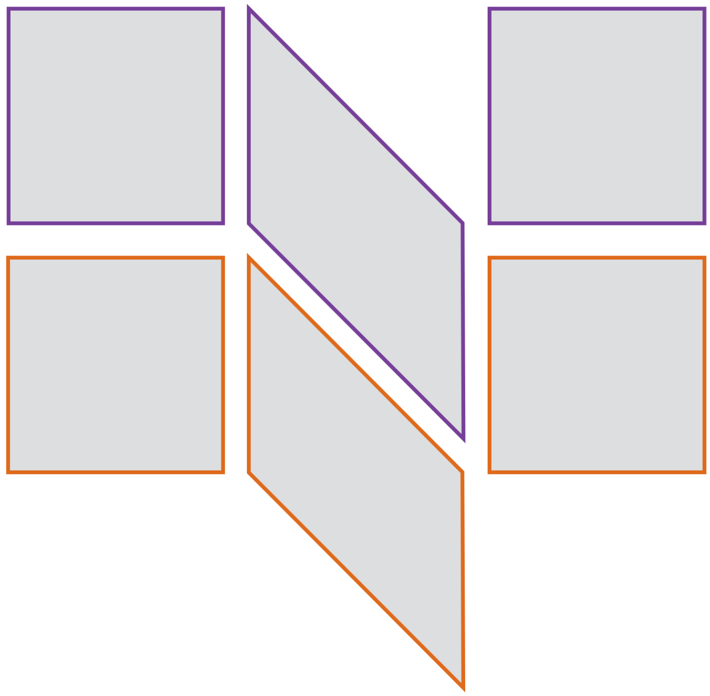
$$\overline{BAABA} \longrightarrow \overline{BAaABA} \longrightarrow \overline{BaB} \longrightarrow \overline{BAB}.$$

Explain.

DD

2.12. Perform the geometric process described above for two different trajectories τ of your choice with slope ≥ 1: Sketch a trajectory τ, sketch its image as a parallelogram after shearing by $\left[\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right]$, and then sketch the reassembled square with the new trajectory τ'. For each, record $c(\tau)$ and $c(\tau')$. Try to find the pattern: a rule to get $c(\tau')$ from $c(\tau)$. Then prove your conjecture.

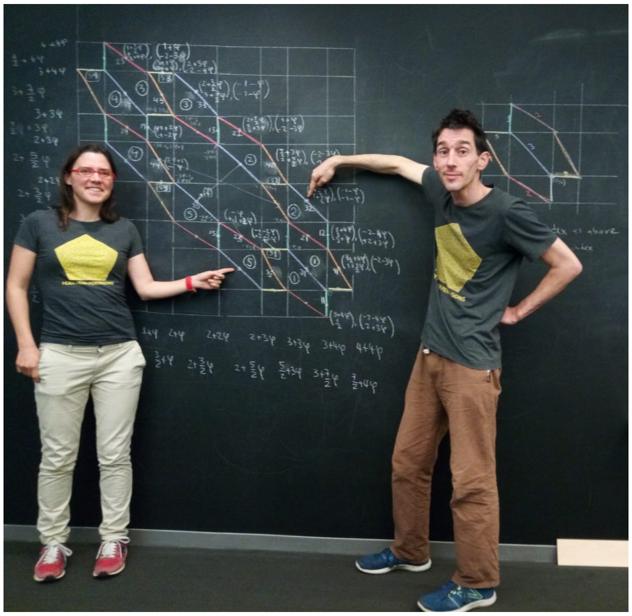
Hint: Apply the "edge marks" technique from Problem 1.15 on the parallelogram edges to make accurate pictures.



2.13. Find the continued fraction expansion of $\sqrt{2}-1$. Then solve the equation $x = \frac{1}{2+x}$ and explain how these are related.

2.14. How many billiard paths of period 10 are there on the square billiard table? Of period 12? Construct an accurate sketch of each of them.

Contextual note. Above, we asked how many different billiard paths of period 10 and 12 there are on the *square* billiard table. You could ask the same question about any other shape of billiard table – pentagon, hexagon, heptagon, isosceles triangle, reflex quadrilateral, half disk – any shape you want to do billiards in. Samuel Lelièvre collaborated with the author (below, in the midst of said collaboration, at ICERM) to answer this and many related questions for the regular pentagon billiard table.



Do you think that *every* possible billiard table has a periodic billiard path of length 12?

DD

2.15. We have identified the top and bottom edges, and the left and right edges, of a square to obtain a surface: the square torus. If we identify opposite parallel edges of a parallelogram, what surface do we get?

2.4 We consolidate our gains

We are about to formulate a grand unifying theory relating a trajectory on the square torus, its corresponding cutting sequence, and the continued fraction expansion of its slope. We need these two results:

DD

2.16. Show that if we apply the flip $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ to the square torus:
(a) The effect on the slope of a trajectory is to take its reciprocal.
(b) The induced effect on the cutting sequence corresponding to a trajectory is to switch As and Bs.

DD

2.17. Show that if we apply the shear $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ to the square torus:
(a) The effect on the slope of a trajectory is to decrease it by 1.
(b) The induced effect on the cutting sequence corresponding to a trajectory *whose slope is greater than 1* is to remove one A between each pair of Bs.

Let's nail down these results, which we have previously conjectured:

DD

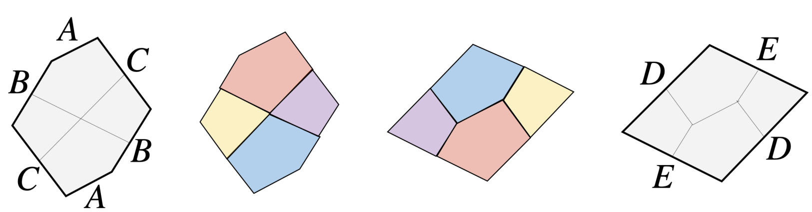
2.18. Show that a trajectory with slope p/q (in lowest terms) on the square billiard table has period $2(p+q)$.

DD

2.19. Show that the continued fraction expansion of a number terminates (stops) if and only if the number is rational.

We will soon study surfaces other than the square torus. Here we go: DD

2.20. If we identify opposite parallel edges of a hexagon, what surface do we get? Let's explore this question:
(a) The picture below shows one way to figure it out: a hexagon surface is *scissors equivalent* to a parallelogram surface. This means that you can cut up the pieces of a hexagon surface and reassemble them, respecting the edge identifications, into a parallelogram whose opposite parallel edges are also identified. Explain, and check that the steps in the picture respect the edge identifications.



(b) An alternative approach is to sketch a "movie" of what it looks like to glue identified edges together, assuming that the hexagon is made out of stretchy material. Try this, too.

Contextual note. In the problem above, we created a surface from an arbitrary hexagon that has three pairs of opposite parallel sides. We could consider the *space* of all possible hexagons, or the space of all of the *surfaces* created by identifying the opposite parallel sides of such hexagons. You might expect that the surface created from a regular hexagon, or other special cases of hexagons, would appear in an identifiable place in the surface, and indeed the symmetries of the surfaces help us to understand the symmetries of the space of surfaces.

Maryam Mirzakhani (left) studied spaces of surfaces, and their symmetries. She received the Fields Medal in 2014.

DD

2.21. In the picture with Problem 2.20, we tiled the plane with a “random” hexagon that has three pairs of parallel edges.
(a) Does a hexagon with three pairs of parallel edges always tile the plane?
(b) A polygon is *convex* if all of its angles are less than 180°, or equivalently if every line segment connecting two points of the polygon lies completely within the polygon. Does a non-convex hexagon with three pairs of parallel edges always tile the plane?

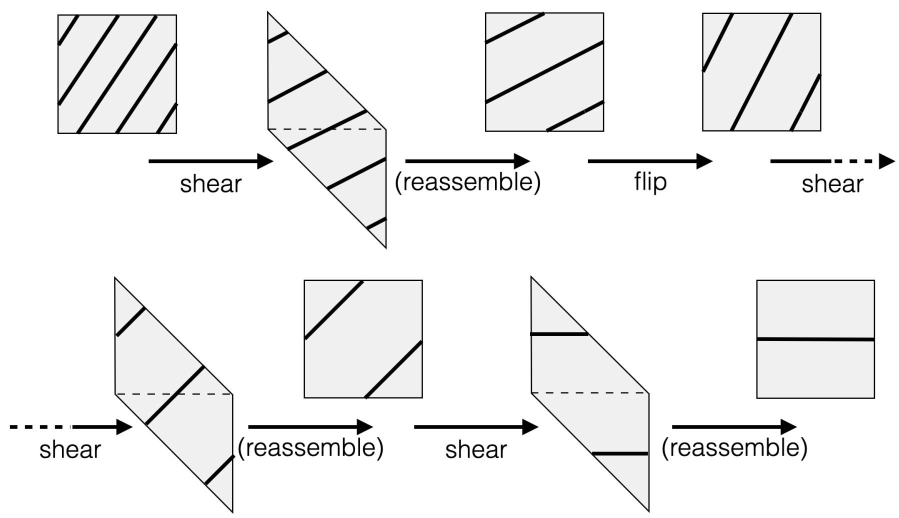
2.5 A grand unifying theory emerges

DD

2.22. Starting with a trajectory on the square torus with positive slope, apply the following algorithm:

1. If the slope is ≥ 1, apply the shear $\left[\begin{smallmatrix}1 & 0\\ -1 & 1\end{smallmatrix}\right]$.
2. If the slope is between 0 and 1, apply the flip $\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]$.
3. If the slope is 0, stop.

An example is shown below. (The second line is a continuation of the first.)



We can note down the steps we took: shear, flip, shear, shear. We ended with a slope of 0. Work backwards, using this information and your work in Problems 2.16 and 2.17, to determine the slope of the initial trajectory. Keep track of each step.

DD

2.23. (Continuation) Write down the continued fraction expansion for the slope at each step.

DD

2.24. (Continuation) Write down the cutting sequence for the trajectory at each step.

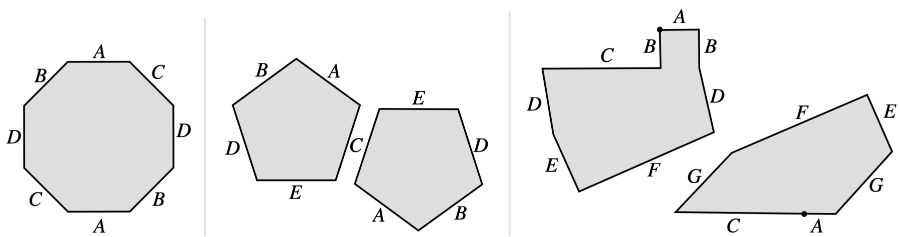
DD

2.25. (Continuation) Formulate a Grand Unifying Theory relating a trajectory on the square torus, its corresponding cutting sequence, and the continued fraction expansion of its slope.

DD

2.26. We can create a surface by identifying opposite parallel edges of a single polygon, as we have done with the square and hexagon. We'll call such a surface a *translation surface*, since parallel edges are translates of each other, and you can translate the polygon to identify the edges. *Parallel edges* must be parallel and also the same length. *Opposite edges* means that the polygon is on the left side of one of the edges, and on the right side of the other.

In a similar way, we can create a surface from two polygons, or from any number of polygons. Some examples are below. Edges with the same letter are identified, as with A and B on the square torus. For the surfaces in the middle and on the right, *two* polygons glued together form a single surface.



(a) Review the part of Amie Wilkinson's talk[2] from 26 to 29 minutes, which shows how to wrap the flat octagon surface (far left) into a curved surface in 3-space. What is its *genus* – how many holes does it have?
(b) Do your best to repeat her stretching methods for the double pentagon surface (center) to make it into a curved surface in 3-space.
(c) The flat octagon surface has 4 edges. How many edges do the other two surfaces have?

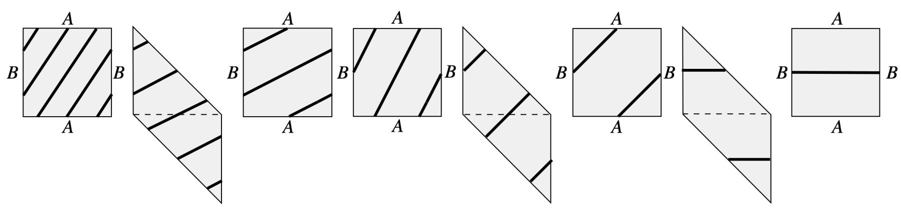
Contextual note. Some people love translation surfaces, and other people *really* love translation surfaces. Jayadev Athreya (left, with the author in Marseille in 2015) has made many contributions to the field, but his most unique contribution just might be having a double pentagon tattooed on his forearm.

[2] YouTube: "Dr. Amie Wilkinson - Public Opening of the Fields Symposium 2018," available at `https://www.youtube.com/watch?v=zjccKzHIniw&t=1560s`

2.6 We expand from familiar friends to new examples

DD

2.27. In Problem 2.22, we gave an algorithm that gradually simplified a trajectory on the square torus with slope ≥ 1, by un-twisting it step by step, until it was a horizontal trajectory. Transform that algorithm into an equivalent algorithm for the *cutting sequence* corresponding to a trajectory. You should translate each of the four sentences ("Starting with...," $1, 2$ and 3) to act purely on sequences of As and Bs. Then apply your algorithm to the cutting sequence $\overline{ABAAB}$ and check that your result at each step is consistent with the pictures in the figure.



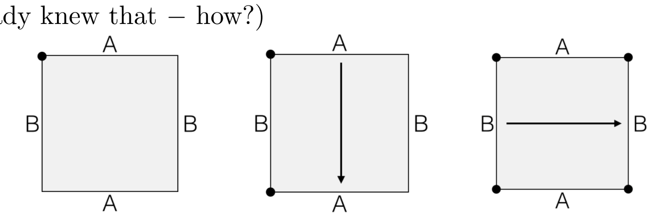
DD

2.28. Make up an example of a translation surface (recall Problem 2.26), made from *three* polygons. Try to choose an example that no one else will think of. How many edges does your surface have?

To count the *faces* of a surface, we count how many polygons it's made of. To count how many *edges* it has, it might be easiest to count the edge *labels*, remembering that pairs of opposite parallel edges are identified in order to create a surface. Finally, we need to know how to count its *vertices*, which again requires understanding the edge identifications:

DD

2.29. *Vertex chasing.* To explain how to count the vertices of a surface, we will use the square torus. First, mark any vertex (say, the top left). We want to see which other vertices are the same as this one. The marked vertex is at the left end of edge A, so we also mark the left end of the bottom edge A. We can see that the top and bottom ends of edge B on the left are now both marked, so we mark the top and bottom ends of edge B on the right, as well. Now all of the vertices are marked, so the square torus has just one vertex. (We already knew that $-$ how?)



Determine the number of vertices for

(a) a hexagon with opposite parallel edges identified (Problem 2.20);

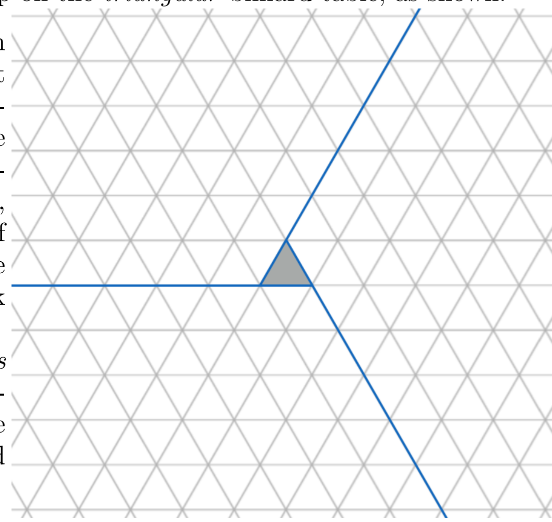
(b) each surface in Problem 2.26; and

(c) your surface created in Problem 2.28.

DD

2.30. You will need: colored pencils or pens. Consider the counter-clockwise outer billiard map on the *triangular* billiard table, as shown.

(a) Explain why points on the thick blue lines are not allowed. Then color the inverse images (red) of the blue lines, the inverse images (green) of the red lines, the inverse images (black) of the green lines, the inverse images (purple) of the black lines, etc.
(b) Identify some *necklaces* of iterated images of triangles, and color each necklace a different color, as we did in Problem 2.4.



SPACE FOR A MATHEMATICIAN SPOTLIGHT HERE.

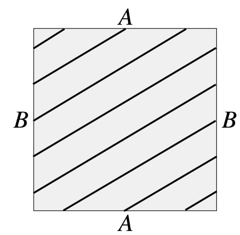
ST

2.31. SPACE FOR A NEW PROBLEM HERE!

2.7 We figure out how to ignore trajectories completely

DD

2.32. Apply the geometric algorithm from Problems 2.22 and 2.27 to the trajectory shown to the right, to reduce it to slope 0. Note down the steps you take (shears and flips). Then use this information to work backwards from an ending slope of 0 to determine the slope of the initial trajectory. Show all of your steps.



DD

2.33. (Continuation) Explain how shears and flips on the square torus are related to continued fraction expansions.

DD

2.34. (Continuation) Find the cutting sequence corresponding to the trajectory above. Apply your algorithm from Problem 2.27 to it, and check that your results at each step are consistent with each step of your work in Problem 2.32.

The following problem is, at long last, the payoff for all of our work with continued fractions, shears, flips and cutting sequences:

DD

2.35. Find the cutting sequence corresponding to a trajectory on the square torus whose slope has continued fraction expansion $[0; 1, 2, 2]$. *Hint*: you don't need to draw any pictures; just use your algorithm and the Grand Unifying Theory (Problem 2.25).

Contextual note. The above problem is an example of abstracting all the way away from trajectories, to working with only continued fractions and symbolic cutting sequences. Curtis McMullen (left, sailing with the author in Boston in 2018) is a master of plumbing the depths of abstraction in billiards and related areas. He received a Fields Medal in 1998. He was also Maryam Mirzakhani's Ph.D. advisor (§ 2.4).

Once we've made a surface, the *Euler characteristic* gives us a way of easily determining what kind of surface we obtain, without needing to come up with a clever trick like cutting up and reassembling hexagons into parallelograms (as we did in Problem 2.20):

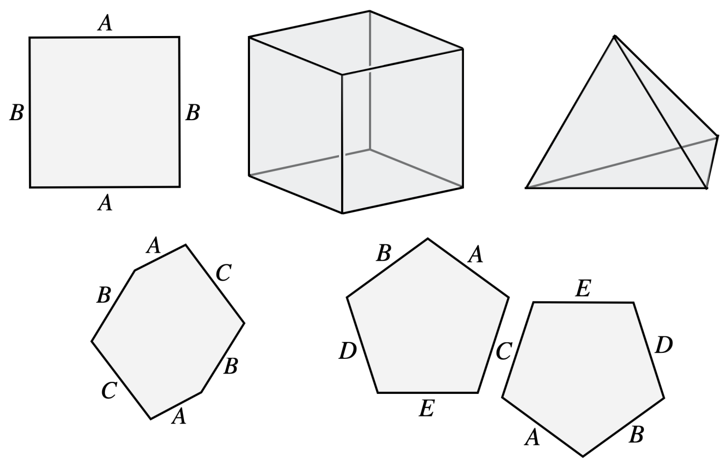
Given a surface S made by identifying edges of polygons, with V vertices, E edges, and F faces, its Euler characteristic $\chi(S)$ is[3]

$$\chi(S) = V - E + F.$$

Note that a "face" must be a *simply connected* polygon, without holes.

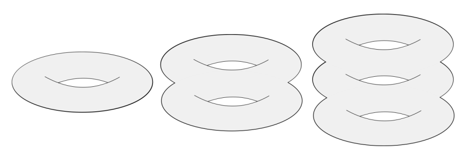
DD

2.36. Find the Euler characteristic of each of the surfaces below (the square torus, the cube, the tetrahedron, the hexagon and the double pentagon). Comment on any patterns you notice. Can you prove your conjectures?



DD

2.37. (Continuation) One of the main goals of the field of *topology* is to classify surfaces by their *genus*, which, informally speaking, is the number of "holes" it has. The surfaces shown have genus $1, 2$ and 3.



We can use the Euler characteristic to determine the genus of a surface: A surface S with genus g has Euler characteristic $\chi(S) = 2 - 2g$. Use this to compute the genus of each of your surfaces from the previous problem, and check that your answers agree with reality.

2.38. (Challenge) Prove the formula $\chi(S) = 2 - 2g$. One way is to proceed by induction: First, show that $\chi(S) = 2$ for the tetrahedron or some other simplest surface of your choice (base case). Then, show that subdividing by adding a vertex, edge or face maintains the same Euler characteristic. Finally, show that adding a hole decreases the Euler characteristic by 2.

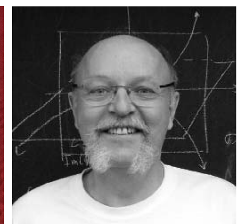
[3] χ is spelled *chi* and rhymes with "hi."

2.8 Hands-on activities for Chapter 2

As Chapter 2 comes to a close, we will bring the flat torus to life in three (!) dimensions.

One way to make a model of a torus from a piece of paper is as follows: Tape the left side to the right side, creating a tube. Then wrap it around to attach the bottom edge to the top edge. To do this, you'll have to flatten the tube. The resulting object looks like a paper wallet. It is not very satisfying; the volume inside the torus is zero.

Many people believe that the above description is the *only* way to create torus that is flat everywhere – that is, it has 360° of angle around every point – out of a piece of paper. But it turns out that we can do better! The picture below shows an example of a flat torus that encloses a positive volume.



The layout for this object was designed by Pierre Arnoux, (above, top) Samuel Lelièvre (see § 2.3) and Alba Málaga Sabogal (above, bottom). They call it *diplotorus*. The idea has been around for a while, and the path it took to get to them was a long and winding road:

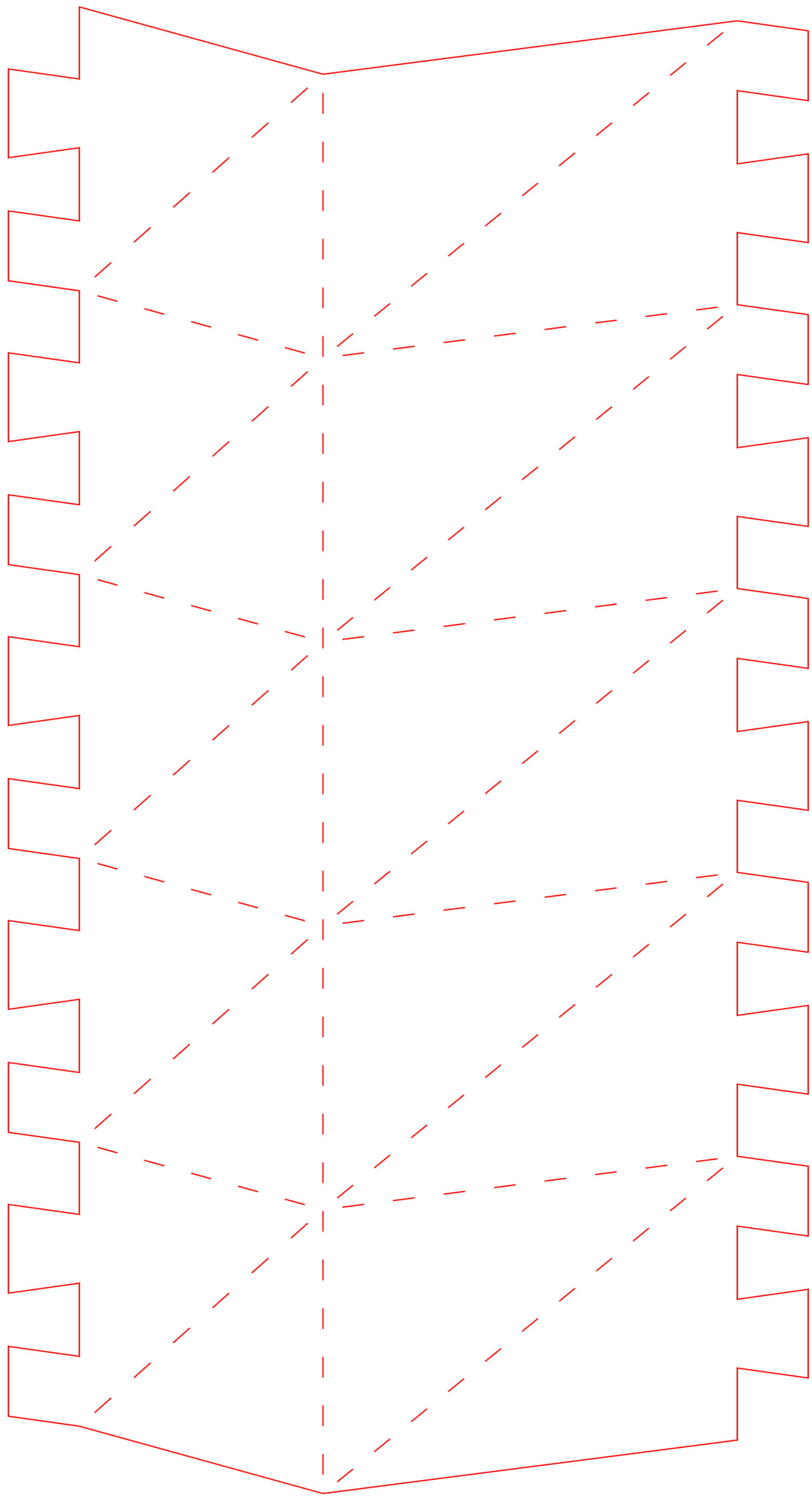
- Ulrich Brehm explained how to construct such an object during a talk at Oberwolfach in 1978.
- Then in 1984, Geoffrey Shephard gave another talk about it at Oberwolfach, and brought a model.
- Guy Valette was in the audience for that talk, and made a model of his own when he got home.
- Guy told Robert Ferréol about it, and Robert included it on his web site `http://mathcurve.com`, where Henry Segerman saw it.
- Henry made a 3D-printed version, which Pierre, Samuel and Alba saw at ICERM in 2019. Glen Whitney also brought a paper model of such a torus to ICERM.

- Then in 2020 they noticed that when you fire up the polyhedron tool in the software Grasshopper (a plugin for Rhino), the default polyhedron is one of these exact objects – which Grasshopper calls an *iris toroid*.
- Finally, the idea has made it to *you*!

2.39. You will need: scissors, perseverance. Cut out the layout on the next page, and crease it along the indicated lines. The dotted lines should be valley folds, and the rest should mountain folds. (*Hint*: The last part of the folding can be challenging; it is easier to succeed if the folds you make now are very strong creases.) Then bring the edges with the same colors together, and attach the flaps. Behold, a torus that encloses positive volume, and is *flat* (2π of angle around every point) everywhere!

The famous 4-*Color Theorem* tells us that if you want to color a geographic map in the plane so that regions meeting along an edge are always different colors, you only need at most 4 colors. This is because it is possible for 4 regions to all border each other (e.g. France, Belgium, Germany, Luxembourg), but not 5. For the torus, the minimum number of colors is 7, because it is possible for 7 regions to all border each other. The coloring on the torus on the previous page is such an example of 7 mutually adjacent regions. (Thanks to Moira Chas for suggesting this line of inquiry to Samuel.)

2.40. Now switch the mountain and valley folds, so that the other side of the paper is on the outside. Behold, a closed geodesic on a flat torus!



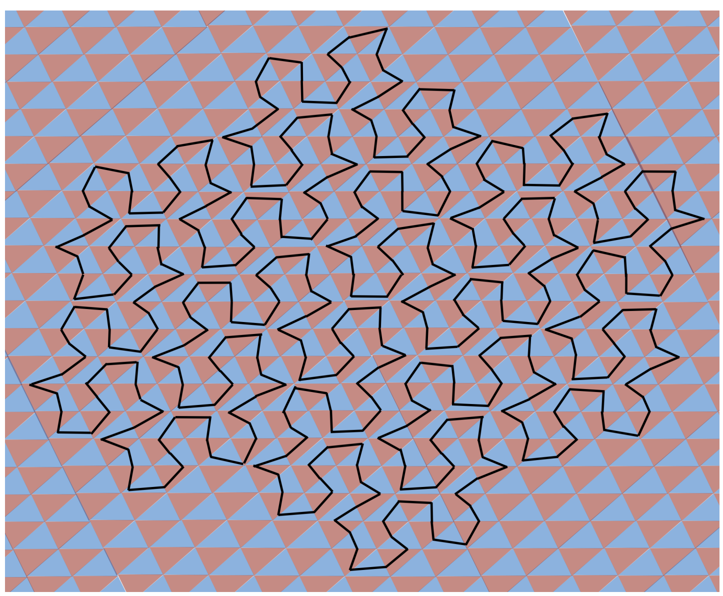
This page will have a few line segments across it.

3

Periodicity beyond the square

In Chapter 1, we met our main characters; in Chapter 2, we delved deeply into the structure of periodic directions on the square billiard table. We saw how periodic trajectories on the square billiard table are connected to continued fractions, and to automorphisms of the square torus surface. In so doing, we can see links between billiards, number theory, and group theory.

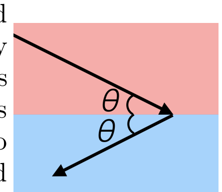
In Chapter 3, we will use the tools and insights we gained in our study of the square, as we more further afield to many different types of billiards. We will do billiards on non-square tables, and we will meet yet another kind of billiards. For billiards on non-square tables, the situation is often "the situation is analogous to the square, but not quite as elegant." For other types of billiards, the behavior is often not at all like the square. Exciting!



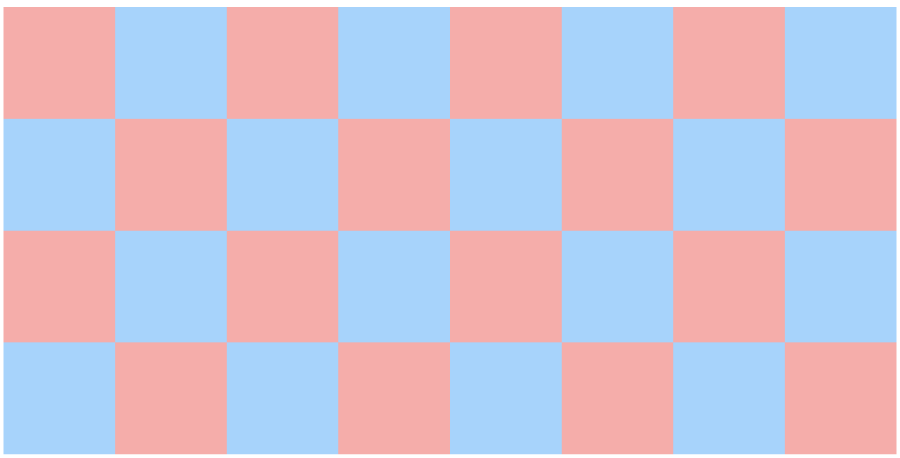
3.1 We meet tiling billiards

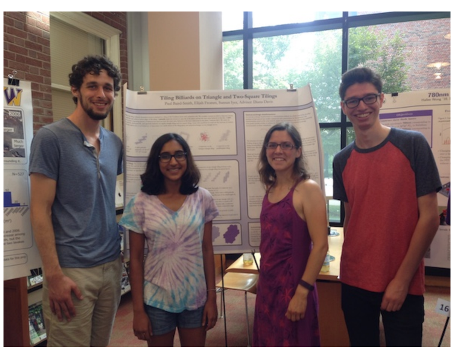
Now that we have explored the simplest case of classical billiards – inner billiards on the square – in detail, and understood it deeply, we will expand our view to other types of billiards.

Tiling billiards. Another type of billiards (besides inner and outer) that we will study is *tiling billiards*, where a trajectory refracts through a tiling of the plane. The *refraction rule* is that when the trajectory hits an edge of the tiling, it passes through in such a way that the angle of incidence is equal to the angle of reflection, and the trajectory has been reflected across the edge, as shown to the right.



3.1. Sketch some trajectories on the square grid tiling. What kinds of behaviors can you find? Prove that you have found them all.



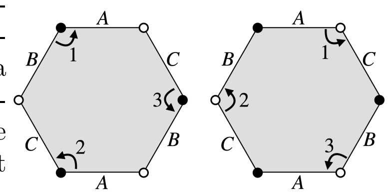


Contextual note. Tiling billiards is motivated by the existence of *metamaterials*, solids that have a negative index of refraction. Typical materials such as water and glass have a positive index of refraction; you have likely worked with these in physics, with *Snell's Law.* The idea here is to create a two-colorable tiling out of materials with opposite indices of refraction.

The results about tiling billiards that we will explore in this chapter come from three undergraduate students at the SMALL REU in 2016: Elijah

Fromm, Sumun Iyer, and Paul Baird-Smith (shown at their poster session, with the author).

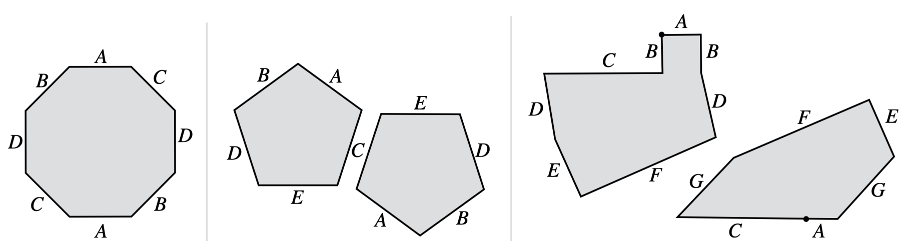
Walking around a vertex. We can determine the angle around a vertex by "walking around" it, as shown in the figure for a hexagon with opposite parallel edges identified. The left picture shows that the angle around the black vertex is $3 \cdot \frac{2\pi}{3}$, and the right picture shows the same for the white vertex.



To do this, first choose a vertex (say, the top left vertex of the hexagon, between edges A and B, marked as black) and walk counter-clockwise around the vertex. In our example, we go from the top end of edge B to the left end of edge A. Now, find where that we "come out" on the identified edge A at the bottom of the hexagon, and keep going counter-clockwise: we go from the left end of the bottom edge A to the bottom end of the left edge C. We keep going counter-clockwise from the bottom of the right edge C to the top end of the right edge B. We find the identified point on the top end of left edge B, and see that this is where we started! So the angle around the black vertex is $3 \cdot 2\pi/3 = 2\pi$. By the same method, or by symmetry, we can see that the angle around the white vertex is also 2π.

Since the black and white vertices each have 2π of angle around them, all the corners of the hexagon surface come together in a flat plane, as we have already seen in Problem 2.20.

3.2. For each of the surfaces below, count its vertices, and then determine the angle around each one.



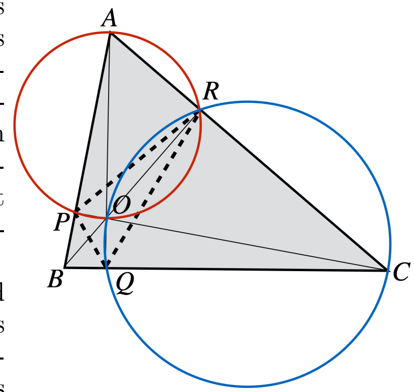
DD

3.3. (Continuation) A surface is called *flat* if it looks like the flat plane everywhere, meaning that there is 2π of angle around every point, *except* possibly at finitely many *cone points* (also known as *singularities*), where the angle around each cone point is a multiple of 2π. For example, the regular octagon surface is flat everywhere except at its single cone point, whose angle is 6π.

Prove that every translation surface (Problem 2.26) is flat.

3.4. (Challenge) Is the converse true? In other words, is it true that every flat surface can be represented by a collection of polygons, identified along opposite parallel edges? Prove it or find a counterexample.

3.5. *The Fagnano trajectory.* You have constructed several periodic billiard paths in the square billiard table; other polygons also have periodic paths. A classical theorem says that *Fagnano trajectory* connecting the feet of the three altitudes of an acute triangle is a 3-periodic billiard trajectory. We will prove this by showing that angles ARP and CRQ are equal; the argument is the same for the other bounces.



(a) Opposite angles of a quadrilateral add up to π if and only if the quadrilateral is *cyclic*. Use this result to show that quadrilaterals $APOR$ and $CROQ$ are cyclic, as suggested by the diagram.
(b) Another classic theorem of geometry says that two angles supporting the same circular arc are equal. Use this to show that $\measuredangle PAO = \measuredangle PRO$, and $\measuredangle ORQ = \measuredangle OCQ$.
(c) Use triangles BAQ and BCP to show that $\measuredangle PAO = \measuredangle OCQ$.
(d) Show that $\measuredangle ARP = \measuredangle CRQ$, as desired.

An active area of research is to *characterize* all possible cutting sequences on a given surface. Now we can do this for the square torus.

Theorem (cutting sequence characterization). Cutting sequences on the square torus are infinite sequences of As and Bs that do not fail under the following algorithm:

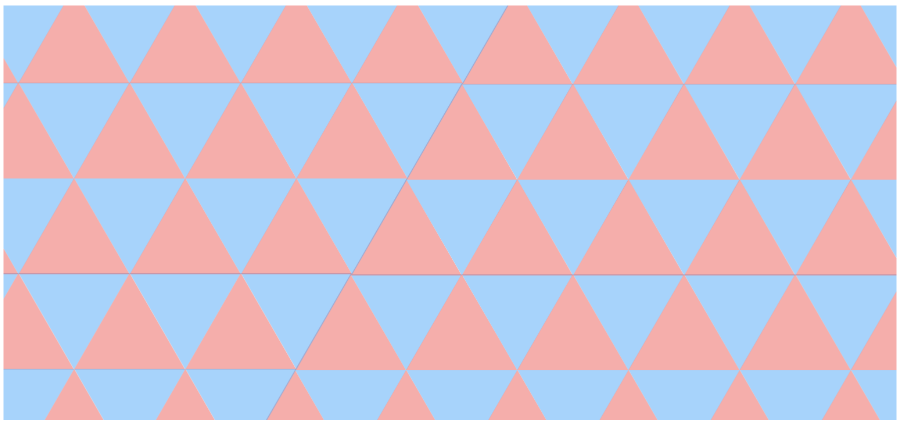
1. If there are multiple Bs separated by single As, switch As and Bs.
2. If there are multiple As separated by single Bs, remove an A between each pair of Bs.
3. If the sequence has AA somewhere and BB somewhere else, stop; it fails to be a valid cutting sequence.

DD

3.6. Earlier in the course, you probably conjectured that a cutting sequence could only have two consecutive numbers of As, such as 2 and 3, between each pair of Bs, e.g. $\overline{BABAAA}$ is not allowed. Use the theorem to prove this conjecture true.

3.2 Then, we unfolded; now, we fold

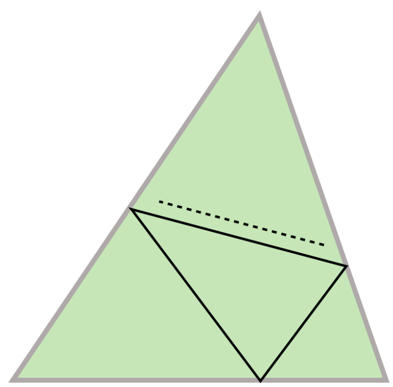
3.7. We saw that for tiling billiards on the square grid, there are only two types of trajectories: those that go to the opposite edge and zig-zag, and those that go to the adjacent edge and make a 4-periodic path. How many types of trajectories are there on the equilateral triangle grid? DD



In billiards on the square, we *unfolded* a billiard trajectory into a line on the square grid, and onto a linear trajectory on the square torus. In an analogous way, *folding* is a powerful technique for understanding tiling billiards trajectories:

3.8. Consider a tiling billiards trajectory that crosses an edge e of the tiling. Show that, if you fold the tiling along edge e, the two pieces of trajectory that intersect edge e lie on top of each other. DD

3.9. Consider again the 3-periodic Fagnano trajectory from Problem 3.5. The picture shows a piece of a trajectory that is parallel to the one in the construction and nearby. Continue the new trajectory until it closes up. What is its period? DD



Notice that as you follow the dashed trajectory around, initially it says "the solid trajectory is on my right!" and then after a bounce, "the solid trajectory is on my left!" and so on, switching sides at every bounce.

DD

3.10. Consider again the cutting sequence characterization theorem that precedes Problem 3.6.

(a) The vexing part of this characterization is that it doesn't have a step saying, "Stop! Congratulations; you have a valid cutting sequence." It only says, "Keep going; your cutting sequence hasn't proven to be invalid yet." But it turns out that it's the best we can do. Explain why this algorithm *does* stop for a *periodic* cutting sequence.

I left out one technical point of the theorem: It actually characterizes the *closure* of the space of all cutting sequences. Valid cutting sequences are in the interior of the space, and cutting sequences such as $\ldots AAABAAA\ldots$ are on the boundary of the space.

(b) Explain why the above cutting sequence does not fail in the algorithm, and also explain why it is nonetheless not a valid cutting sequence.

(c) Another cutting sequence on the boundary is $\ldots BBBABBB\ldots$. Find yet another example of a cutting sequence on the boundary of the space of cutting sequences.

Contextual note. The *space* of cutting sequences is a rather abstract notion, like the *space* of hexagon surfaces that we discussed in § 2.4. Typically, the first examples we would think of are on the interior of such a space, and degenerate cases are on the boundary of the space. Alex Wright has studied spaces of translation surfaces, and their orbit closures. This picture shows Rodrigo Treviño, the author, and Alex Wright on their way home from a translation surfaces conference at Oberwolfach in 2014.

3.11. In Chapter 2, our strategy for "untwisting" a periodic trajectory on the square torus (see Problem 2.22) was:

- If the slope is greater than 1, apply a vertical shear, and
- if the slope is less than 1, first flip so that the slope is greater than 1, and *then* apply a vertical shear.

Alternatively, we could say:

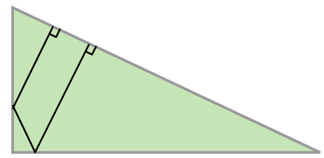
- If the slope is greater than 1, apply a vertical shear, and
- if the slope is less than 1, apply a *horizontal* shear.

Explain.

3.3 We meet the biggest open problem in billiards

DD

3.12. *Lots of triangles have periodic trajectories.*



(a) Explain why the Fagnano trajectory (Problem 3.5) gives a periodic trajectory in every acute triangle, and only in acute triangles.
(b) Rich Schwartz (§ 1.3) showed me the construction to the right. He calls it "shooting into the corner." Fill in the details, and show that it gives a periodic trajectory for every right triangle.
(c) Find an example of a periodic trajectory in an obtuse triangle.
(d) In fact, the Fagnano trajectory, the shooting into the corner trajectory, and the construction you probably used in **(c)** are all variations on the exact same idea. Explain. (Thanks to Alan Bu for pointing this out.)

Contextual note. The biggest open problem in billiards is: *does every triangular billiard table have a periodic trajectory?* The Fagnano trajectory shows that every *acute* triangle has a periodic billiard trajectory, and the construction above shows that every *right* triangle has one.

Howie Masur (left) showed that every polygon – including triangles, and also every other polygon – whose angles are *rational* numbers of degrees has a periodic path. Rich Schwartz (§ 1.3) used a computer-aided proof to show that every triangle whose largest angle is less than 100° has a periodic billiard trajectory, and in 2018 a team of four researchers extended that result to 112.3°. The problem is open in general for irrational-angled obtuse triangles with an angle larger than 112.3°. It seems that the methods of proof used for the 100° and 112.3° theorems do not work past about 112.5°, so a new idea is needed to move forward.

We have talked a little bit about the space of all possible translation surfaces. The space is divided into *strata* based on:

- how many cone points the surface has, and
- how many extra multiples of 2π are around each cone point.

(Recall that in Problem 3.3, we proved that the angle at a cone point of a translation surface is always a multiple of 2π.) We say that the double pentagon surface is in the stratum $\mathcal{H}(2)$ because it has one cone point, with two extra multiples of 2π around it: 6π total, so $2 \cdot 2\pi$ extra. A surface with two cone points, each with angle 4π, is in the stratum $\mathcal{H}(1,1)$. The "$\mathcal{H}$" stands for "hyperbolic," which means that the surface has extra angle around some of its points.

3.13. Identify which stratum each of the remaining surfaces in Problem 3.2 belongs to. Then come up with an example of a surface in $\mathcal{H}(1,1)$.

Note that a point with 2π of angle around it is not really a cone point; we can call it a *marked point* or a *removable singularity*. Depending on how much you care about such points, you can include 0s in your stratum, or not. For example, while we could say that the square torus is in $\mathcal{H}(0)$, we could also note that it is *flat*, not actually hyperbolic at all.

DD

3.14. Here is our dream: To understand the effect of *every* automorphism of the square torus, on the cutting sequence corresponding to a trajectory. Here is our progress so far (fill in the blanks):
(a) There are three types of automorphisms: rotations, reflections and shears. We understood the effects of rotations and reflections in Problems __________.
(b) Using rotations and reflections, we reduced our work, now only for shears, to the case of trajectories whose slope is greater than 1, in Problem ______.
(c) We understood the effect of the matrix $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ on a trajectory on the square torus in Problems __________.

By the way, we used $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ because it works nicely with trajectories whose slope is greater than 1: it makes them simpler, like taking a derivative in calculus, while $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ makes them more complicated, like taking an integral.
(d) Find the analogous effects on slopes of trajectories, of the matrices $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$.

DD

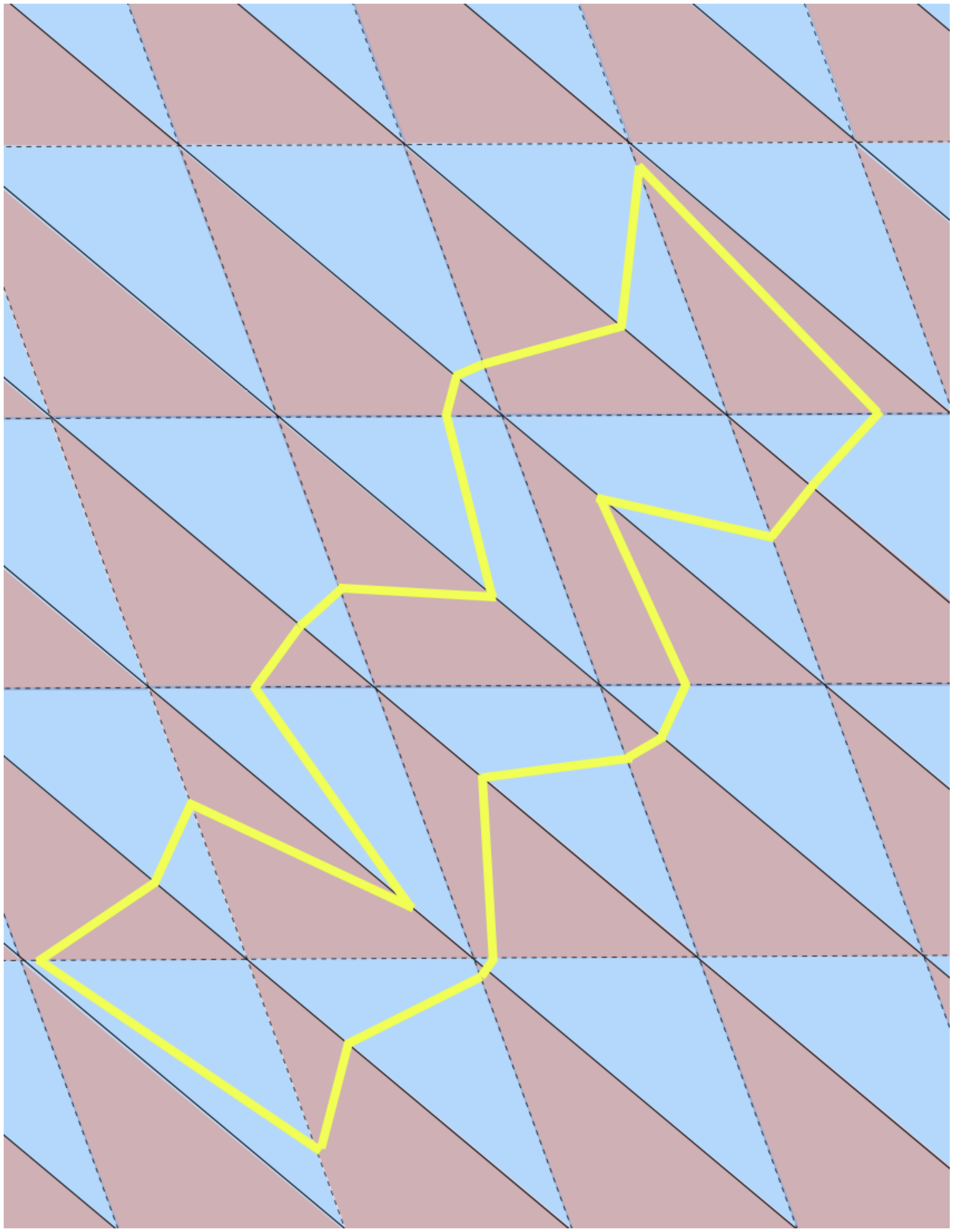
3.15. (Challenge) There is just one more step, to show that every shear can be reduced to the ones we understand. Prove the following:
(4) Every 2×2 matrix with nonnegative integer entries and determinant 1 is a product of powers of the shears $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. For example,

$$\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^2 \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^2 .$$

Thus we know the effect of every matrix with determinant 1 on slopes of trajectories, and we could work out the induced effects of $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ on cutting sequences just as we did for $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$.

DD

3.16. You will need: scissors, perseverance. The figure on the next page shows a periodic tiling billiards trajectory on a triangle tiling. Cut off the white part and then fold along all the edges of the tiling, in such a way that every part of the trajectory lies on a single line. The solid lines should be "valley folds" and the dashed lines should be "mountain folds." *Hint*: make sure all of your folds are well creased. Flat fold a little patch at first, and then gradually extend it to the whole paper. Save your folded paper, as we will use it in subsequent problems.



3.4 Families of parallel trajectories

DD

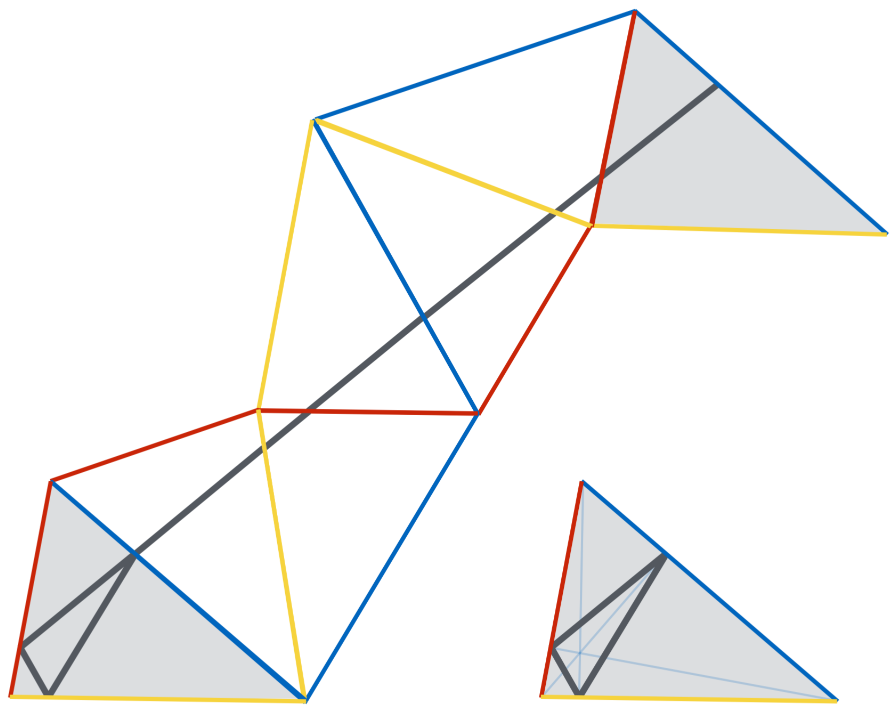
3.17. The figure below shows the Fagnano trajectory in the 40-60-80 triangle. In Problem 3.9 we showed that there are nearby parallel billiard trajectories of period 6.

(a) In the triangle in the lower right, sketch a period-6 trajectory that is parallel to the given Fagnano trajectory.

(b) How far can you push the period-6 trajectory until it disappears? Add to your picture a period-6 trajectory that is as far as you can make it from the given Fagnano trajectory.

(c) Sketch one of your period-6 trajectories in the copy of the triangle that is in the lower left of the picture. Then draw the "unfolding" of your trajectory. The unfolding of the Fagnano trajectory is given.

(d) Imagine the family of *all* possible period-6 trajectories that are parallel to the Fagnano trajectory. Can you sketch *all* of their unfoldings in the picture?

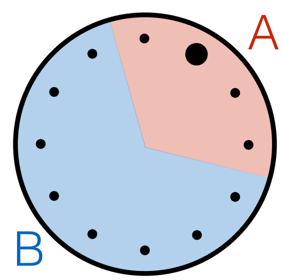


DD

3.18. You will need: your folded triangles from Problem 3.16. Consider a tiling by congruent triangles, created from a tiling by edge-to-edge parallelograms by splitting the parallelograms along parallel diagonals, such as the one you folded up in the previous problem.
(a) Given two adjacent triangles in the tiling, prove that, if you fold along their shared edge, the circumcenters of the triangles coincide, and the triangles share the same circumscribing circle.
(b) Prove that this result extends globally: if you fold along *all* of the edges of the tiling simultaneously, *all* the triangles, in the folded state, are circumscribed in the same circle.
(c) Use the above, and the result of Problem 3.8, to show that for a given tiling billiards trajectory on a triangle tiling, the (perpendicular) distance from the trajectory to the circumcenter of a triangle is the same for *every* triangle that the trajectory enters.

DD

3.19. Consider a circle broken into a red arc and a blue arc, taking up 1/3 and 2/3 of the circle respectively, as shown. The game is to start with any point on the circle, repeatedly rotate it by a 1/3 turn, and each time note down which part of the circle it lands in – say, an A if it lands in the red arc and a B if it lands in the blue arc. Try this for several different starting points, and rotate each of them until you see a pattern.



DD

3.20. One reason why people like cutting sequences on the square torus is that they have very low *complexity*: The *complexity function* $f(n)$ on a sequence is the number of different "words" of length n in the sequence. In other words, imagine a "window" n letters wide that you slide along the sequence, and you count how many different words appear in the window.
(a) Confirm that the sequence $\overline{ABABB}$ below has complexity $f(n) = n + 1$ for $n = 1, 2, 3, 4$ and complexity $f(n) = 5$ for $n \geq 5$.

$$\ldots ABABBABABBABABBABABBABABBABABBABABBAB \ldots$$

(b) Explain why a periodic cutting sequence on the square torus with period p has complexity $f(n) = n + 1$ for $n < p$ and complexity $f(n) = p$ for $n \geq p$.
(c) (Challenge) Aperiodic sequences on the square torus are called *Sturmian sequences*. Show that Sturmian sequences have complexity $f(n) = n + 1$.

DD

3.21. The *Gauss-Bonnet Theorem* says that the total (Gaussian) curvature K of a closed surface S is

$$\int_S \kappa \, dA = 2\pi \ \chi(S).$$

Here κ is the curvature at each point of the surface – a circle of radius r has curvature $\kappa = 1/r$ – and $\chi(S)$ is the Euler characteristic.

(a) Compute each side of this equation for a sphere S of radius r.

The *defect* of a cone point is 2π minus the cone angle at the cone point. The *total defect* of a surface (or of any polyhedron made from identifying edges of polygons) is the sum of the defects of all of its cone points.

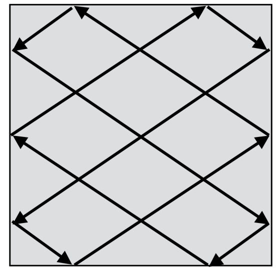
(b) Descartes' special case of the Gauss-Bonnet Theorem says that the total defect of a polyhedron is $2\pi \ \chi(S)$. Check this formula for the cube, the square torus, and the octagon surface, and check your results against your answers to Problem 2.36.

Bill Thurston (left, with the author at the Cornell Topology Festival in 2012) was hugely influential in 20$^{\text{th}}$-century mathematics, particularly in geometry. In addition to his own work, he was the Ph.D. advisor, and the advisor's advisor ("academic grandfather"), of many mathematicians currently working in billiards.

Bill received the Fields Medal in 1982. One of his later projects was working to smooth out the angle defect in polyhedra.

3.5 A cone point with angle 6π

ST

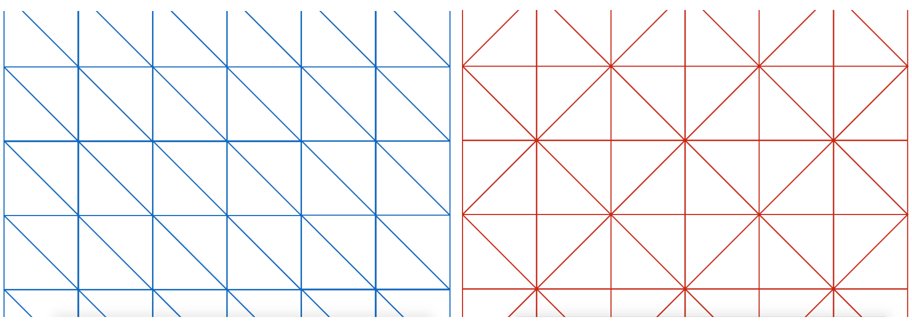


3.22. Show that a periodic trajectory on a polygonal billiard table is never isolated: an even-periodic trajectory belongs to a family of parallel periodic trajectories of the same period and length, and an odd-periodic trajectory is contained in a strip consisting of trajectories whose period and length is twice as great. One way to think about this is that there is a wide ribbon whose center line is the trajectory, wrapping around the table.

Based on the above, we say that there is a *1-parameter family* of billiard trajectories in a given direction on the square table. The idea is that, once you've chosen the direction, the only other thing left to choose is the starting point – say, along the bottom horizontal edge. There is only one dimension, or *parameter*, of such choices.

DD

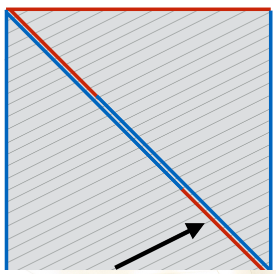
3.23. The pictures below show two different tilings of the plane by isosceles right triangles. Consider tiling billiards on each of them.



(a) For each tiling, consider: are there periodic trajectories on the tiling? If so, explain how to construct one and sketch it; if not, prove that periodic trajectories cannot occur.
(b) Are there escaping trajectories on the tiling? If so, explain how to construct one and sketch it; if not, prove that escaping trajectories cannot occur. (An *escaping* trajectory eventually leaves a disk of any finite radius.)

DD

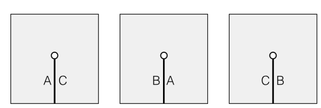
3.24. The picture to the right shows many trajectories of slope $1/2$ on the square torus. As usual, we care about when a given trajectory crosses a horizontal or vertical edge, and we record such crossings with an A or B, respectively. In this picture, I've added a diagonal of the square, and colored it on both sides: on the bottom side to indicate whether an incoming trajectory comes from a red or blue side, and on the top side to indicate whether an outgoing trajectory will hit a red or blue side. Show how to use just the diagonal (copied larger below) to record the edge crossings of the indicated trajectory.



Above, we transformed a problem about trajectories on the square torus into a problem about moving intervals around on a line segment. This sort of problem is called an *interval exchange transformation.* Jean-Christophe Yoccoz worked on dynamical systems, and in particular he studied interval exchange transformations. In joint work with Pierre Arnoux (§ 2.8), Yoccoz came up with the Arnoux-Yoccoz interval exchange transformation, which has particularly interesting properties and led to much further research. Yoccoz is pictured with the author in 2014 at a conference at Oberwolfach.

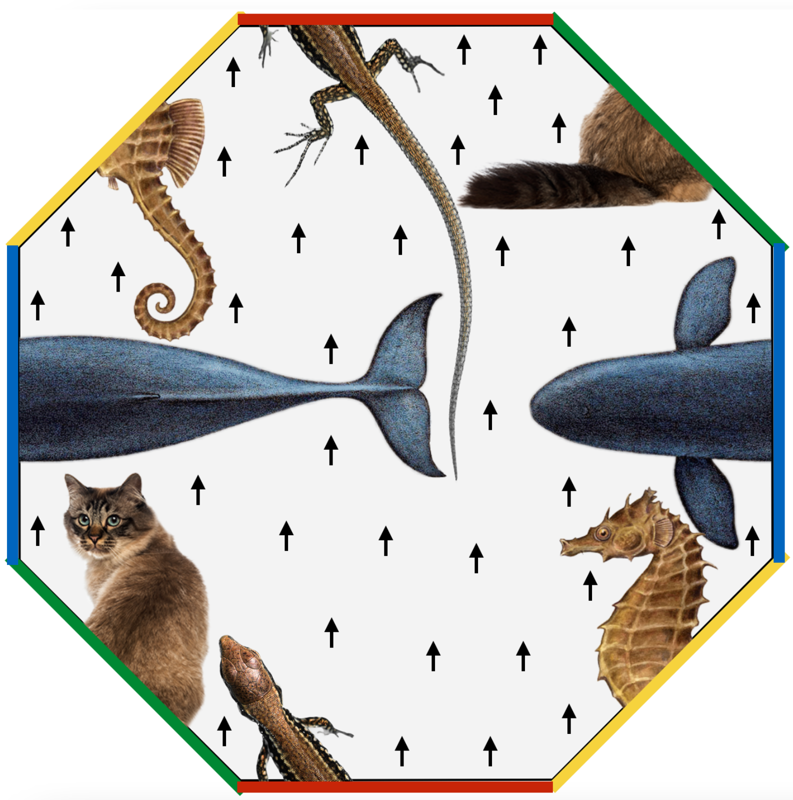
DD

3.25. We saw that the octagon and double pentagon surfaces each have just one cone point, with 6π of angle around it. What does this even mean? What does it look like?



Choose one of the following activities and do it (yes, do it! yes, you!):
(a) Cut slits in three sheets of paper, and tape the edges together as shown above. The vertex angle at the cone point is now that of three planes, which is 6π. Observe. Discuss.
(b) Cut out the double pentagon on the next page (which includes Dr. Libby Stein, Brown '15, dancing on this surface). Tear off each corner, keeping each

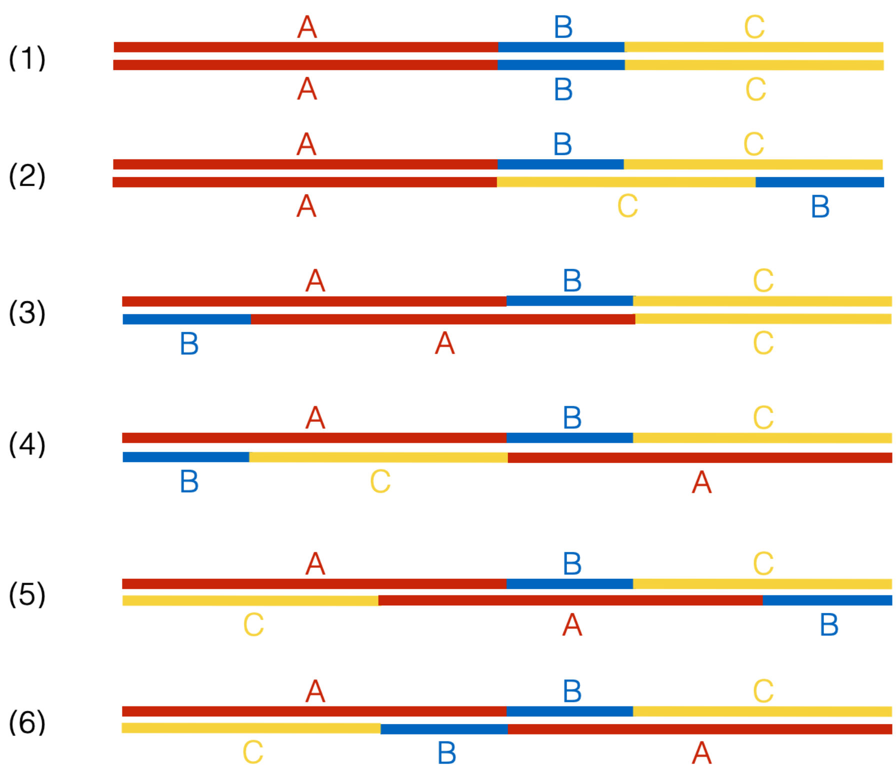
piece as large as you can. Tape them together according to which edges are identified. The angle at the cone point is now 6π. Observe. Discuss.
(c) Cut out the octagon on the next page. Do as described in **(b)**.



3.6 Interval exchange transformations

DD

3.26. The construction in Problem 3.24 showed how to represent a trajectory on a surface via the motion of a point on an *interval exchange transformation*.
(a) Explain why the dynamics of the interval exchange transformation (IET) in Problem 3.24 are identical to those of the rotation in Problem 3.19.
(b) Explain why *every* 2-interval IET is equivalent to a rotation.
(c) The figure below shows the six possible ways of rearranging three intervals. (1) is the identity, and (2) and (3) are the identity on part of the interval and 2-IETs (rotations) on the rest of the interval. Of the remaining three, two of these are also rotations, leaving just one true 3-IET. Which one?



DD

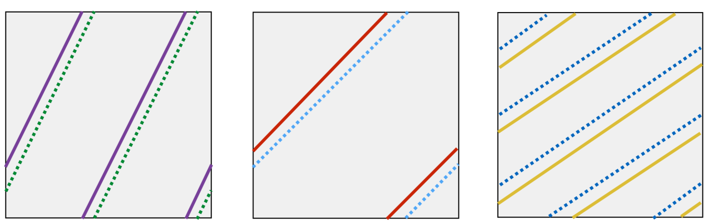
3.27. For the 3-IET that you identified in the previous problem:
(a) Choose a point, mark all the places it goes (its *orbit*), and find the period of its orbit. Does the orbit of every point have the same period?
(b) The interval lengths for the IETs above are $1/2, 1/6, 1/3$. What if they were irrational?

Everything I (the author) know about interval exchange transformations, I learned from Vincent Delacroix (left). I was studying tiling billiards with the group pictured in § 3.1, and (spoiler alert) we had figured out that tiling billiards on triangle tilings were equivalent to orbits of points on certain IETs – but I knew very little about IETs. At a conference in Marseille in 2017, after dinner and stretching into the early hours of the morning, Vincent explained to me some key tools (Rauzy diagrams) that were essential for studying these IETs. This illustrates a key principle, which is that many of the essential ideas in the field are passed down by oral tradition, one on one, people explaining things to each other and taking notes.

In addition to educating colleagues and writing research papers, Vincent writes and maintains a lot of python code related to exploring billiards, translation surfaces, IETs, and other aspects of dynamical systems.

DD

3.28. Each picture below shows a trajectory on the square torus surface (solid line), and a parallel trajectory that is slightly shifted, or "perturbed," from the original (dotted line). Let's consider them to be in the same "family."
(a) For each picture, draw another trajectory that is slightly perturbed from the given ones, and is also in the same family.
(b) If you perturb a trajectory enough, it will eventually hit a vertex. A "singular trajectory" that hits a vertex is not allowed, and forms the boundary of the family of trajectories. Draw in these boundaries for each of the pictures.
(c) The union of a family of trajectories is called a *cylinder*. Can you guess why this name was chosen?



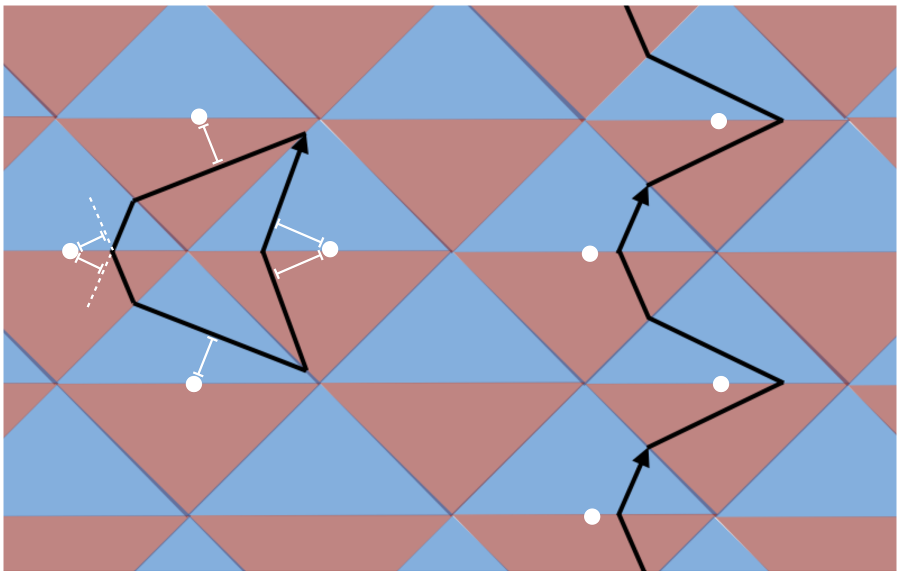
DD

3.29. *Stability under perturbation, part I*
Consider a billiard trajectory in the square billiard table.
(a) If you keep the direction the same, and change your starting point a little, what happens? Does the trajectory change a lot, or it essentially the same?
(b) How about the reverse – if you keep the starting point the same, and change your direction a little bit, what happens?

DD

3.30. Prove that, for a tiling billiards trajectory on a triangle tiling, the (perpendicular) distance between the circumcenter of a triangle in the tiling and a piece of trajectory in that tiling (or the extension of the trajectory) is the same for *every* triangle that the trajectory hits.

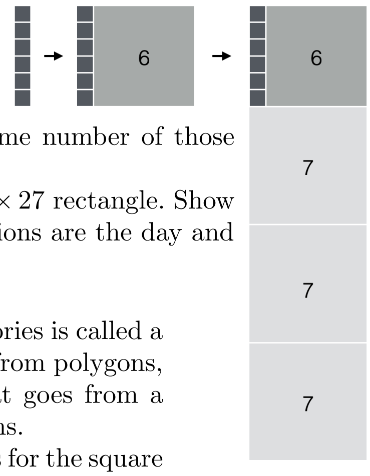
In the triangle tiling below, the circumcenters of the triangles are marked, with some example measurements from the trajectory to the circumcenter.



3.7 We do some computer experiments

DD

3.31. Here is a new game: make some number of 1×1 squares going vertically (here, six). Then make a big square that goes across all of them, and make some number of those going horizontally (here, one). Then make a big square that goes across all of *them*, and make some number of those going vertically (here, three), and so on.



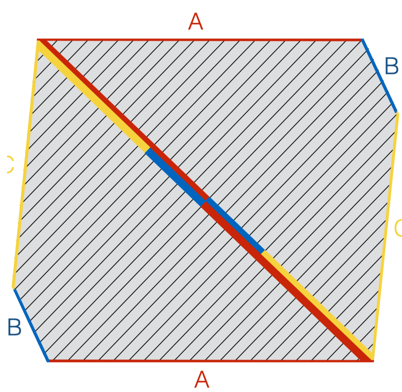
The picture shows how to do this to end up with a 7×27 rectangle. Show how to do this to end up with a rectangle whose dimensions are the day and month of your birth.

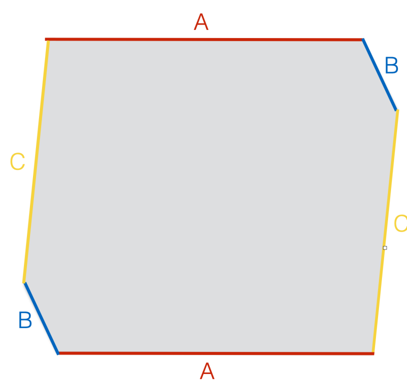
DD

3.32. Recall that the union of a family of parallel trajectories is called a *cylinder* (Problem 3.28). For a translation surface made from polygons, a *cylinder direction* is a direction of any trajectory that goes from a vertex to another vertex, possibly crossing many polygons.
(a) Explain why slopes $2/3$ and $5/7$ are cylinder directions for the square torus.
(b) What are *all* of the cylinder directions for the square torus?

DD

3.33. The pictures below show a surface made from a non-regular hexagon.
(a) The first picture shows a family of trajectories in a given direction. Explain how any trajectory in this direction can be represented by the orbit of a point on a particular IET.
(b) In the second picture, draw a family of trajectories in a different direction of your choice. Sketch the corresponding IET.





Contextual note. In mathematics, we often care about the *dimension* in which we are working. For example, A torus is a 2D object, and if we look at it as the surface of a bagel, it is a 2D surface *embedded* in 3D space. The family of parallel trajectories in a given direction on the square torus looks like a 2D system, but the problem above shows that the behavior of each one can be reduced to the orbit of a point on an IET, which is a 1D system.

3.34. Let's experiment a bit with trajectories on triangle tilings. Go to the web site https://awstlaur.github.io/negsnel/, coded by Pat Hooper and hosted by Alexander St Laurent.
(a) Click on "Help" at the top and learn how to control the applet with keys.
(b) Move the starting point (green point) and the direction (red point) and see what sort of things you can get.
(c) Click "New" and make yourself a more interesting tiling. Try a new triangle tiling determined by angles of your choice. Try to find a really big periodic trajectory. Try to find a really interesting trajectory. Take a screenshot.
(d) Click on "New" and select some other kind of tiling. Try to find a really interesting trajectory. Write down the parameters you used. Take a screenshot.
(e) Use the w, a, s, d keys to slightly nudge the direction. Is your trajectory stable or unstable under small perturbations in the direction?
(f) Notice that you can click Edit > Set iterations. Once you get something interesting, increase to more iterations and see what happens when you allow more bounces. (Turn down the iterations when perturbing the trajectory.)

It turns out that programming can be really helpful for figuring out what is going on in a dynamical system. If you have a program that models the system you want to study, you can experiment and get a sense of what is going on. For example, in the problem above, you may have noticed that the dynamics on some tilings are really boring, while the dynamics on other tilings are rich and fascinating. You'd want to spend your time on the latter. Experimentation also leads to conjectures, which you might be able to prove, or to counterexamples, which can prevent you from trying to prove something false.

Pat Hooper (above, with the author at a dynamical systems conference at Stony Brook in 2015) has written a lot of code for studying billiards and translation surfaces. In collaboration with Vincent Delacroix (§ 3.6) and Julian Rüth, Pat has developed a python package called *sage-flatsurf* that allows people to experiment, compute, and understand far more about flat surfaces than they could with paper, pencil and brain alone.

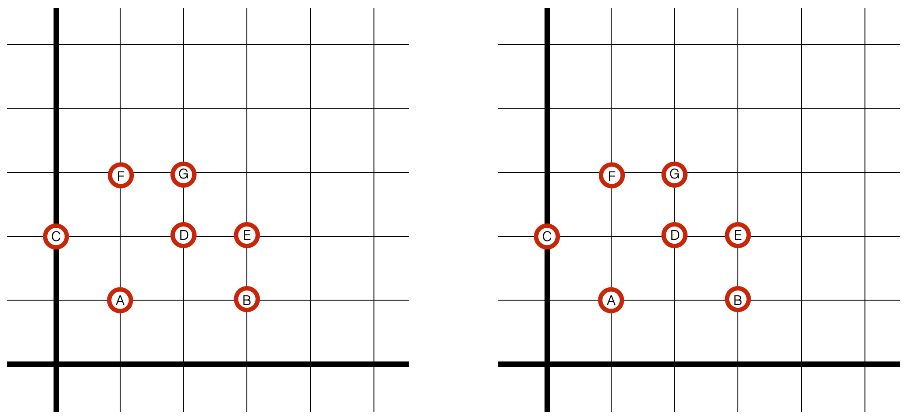
DD

3.35. *Stability under perturbation, part II*
Consider a trajectory on an *outer* billiard table.
(a) If you change your starting point a little, what happens? Does the trajectory change a lot, or it essentially the same?
(b) Explain why the starting point determines the trajectory – we don't get to choose a starting point *and* a direction, as we did on the square billiard table.

3.8 We unveil the deep hidden meaning

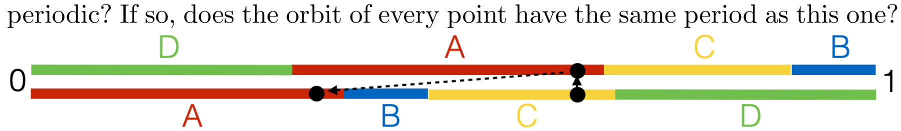
DD

3.36. Let's make sure your shearing skills are sharp. (Local sheep, beware!)
(a) In the left picture, draw the image of each of the identified lattice points under the horizontal shear $\left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]$.
(b) In the right picture, do the same for the vertical shear $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$.

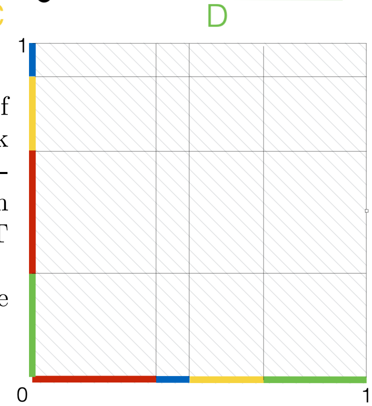


DD

3.37. Here is a 4-IET. Let's use the drawing convention, based on our experience from the translation surface construction in Problems 3.24 and 3.33, that a point flows directly up in the IET, and then shifts over to come down. One iteration for an example point is shown in the diagram.
(a) Find the orbit of this point for at least six more iterations. Is its orbit periodic? If so, does the orbit of every point have the same period as this one?



(b) An IET essentially cuts up an interval of points and reassembles them. So we can think of an IET as a function that maps points between 0 and 1 to points between 0 and 1. Graph the function corresponding to the above 4-IET on the axes to the right.
(c) Use the graph to find the orbit of the same point that you followed in part **(a)**.



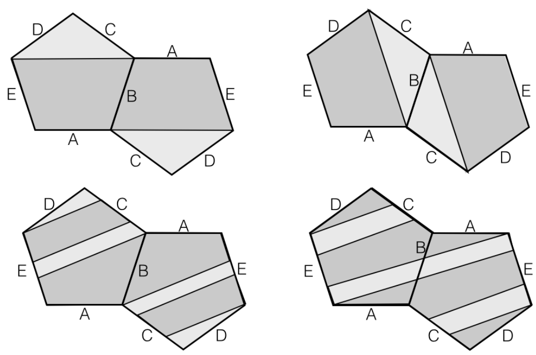
3.38. Show that if the length of every subinterval of an IET is rational, then the orbit of *every* point is periodic.

Interval exchanges are simple to define – just chop up an interval and rearrange the pieces – and even IETs with a small number of intervals can have interesting properties. We have shown that every 2-IETs is equivalent to a rotation (Problem 3.26, and that IETs with rational subinterval lengths have only periodic trajectories (Problem 3.38), but outside of these cases, things can get very interesting indeed.

Jon Chaika (left) has studied many properties of interval exchange transformations, particularly their ergodicity. A flow is *ergodic* if, roughly speaking, the amount of time that a point spends in each region is proportional to its size. For example, in the IET in Problem 3.37, if interval B has length $1/10$, a point should land in interval B, on average, $1/10$ of the time.

DD

3.39. For the square torus, in every cylinder direction there is only one cylinder. For surfaces made from other polygons, there can be multiple cylinders. The double pentagon surface has *two* cylinders in each cylinder direction. Here are cylinders on the double pentagon surface in four directions.

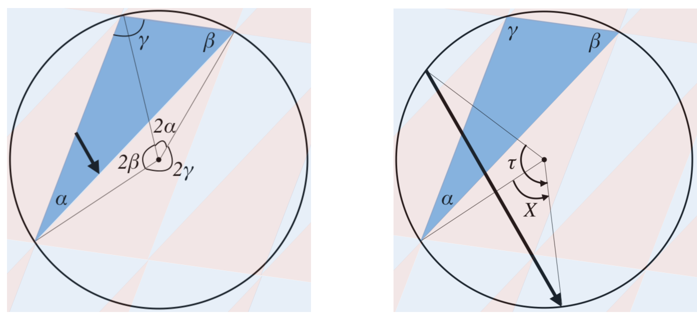


(a) For each set of cylinders shown, consider a trajectory on the surface, in the cylinder direction. Write down the cutting sequence for the trajectory in the light cylinder and for the trajectory in the dark cylinder. Think about similarities and differences with our work on the square torus.
(b) Construct a vertical cylinder decomposition of the surface.
(c) The two cylinder decompositions in the top line of the picture are equivalent under a rotation. Is the vertical decomposition equivalent to any of those shown?

The amazing result of this section will be that tiling billiards on triangle tilings are equivalent to the orbit of a point on a certain interval exchange transformation. Can you believe it? Let's work towards proving it.

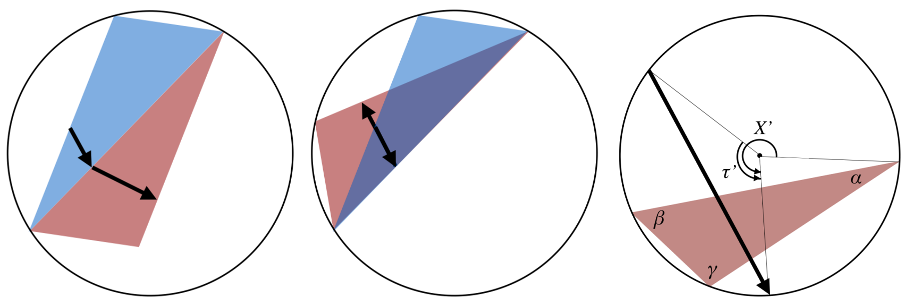
Let's define some notation. Given a tiling billiards trajectory on a triangle tiling, we choose a triangle, circumscribed by the unit circle, containing an oriented segment of the trajectory. We extend this oriented segment to a chord of the circle (see the pictures below).

Call the triangle's angles α, β, γ, listed in non-decreasing order, and reflect the triangle if necessary so that the angles α, β, γ are ordered counter-clockwise. Let X be the counter-clockwise angle from the vertex of angle α to the front end of the chord, and let τ be the central angle subtended by the trajectory chord.



3.40. Explain why the position of a trajectory within a given triangle is uniquely specified by the ordered pair (X, τ). In other words, if you know what triangle you're working with, and you know X and τ, you know exactly where the piece of trajectory is in the triangle.

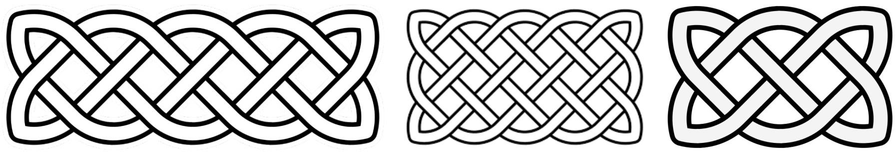
Consider two consecutive triangles that the trajectory crosses (see the pictures below). When the second triangle is folded onto the first along the shared edge, the segments of trajectory in each triangle align, with opposite orientations (Problem 3.8, and shown below).



3.41. Prove what is suggested by the third picture above:
(a) Show that the counterclockwise angle subtended by the chord trajectory in the second (red) triangle is $2\pi - \tau$.
(b) Show that, after reflecting the second triangle back to its correct orientation, the subtended angle of the chord trajectory is τ: in other words, $\tau' = \tau$.
(c) Is it also true that $X' = X$? Explain why or why not.

3.9 Hands-on activities for Chapter 3

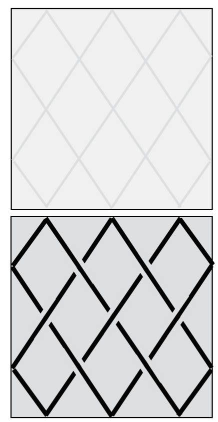
Celtic knots are a traditional form of decorative art associated with Ireland. They come in many different shapes, some of which are related to... periodic billiards on the square!



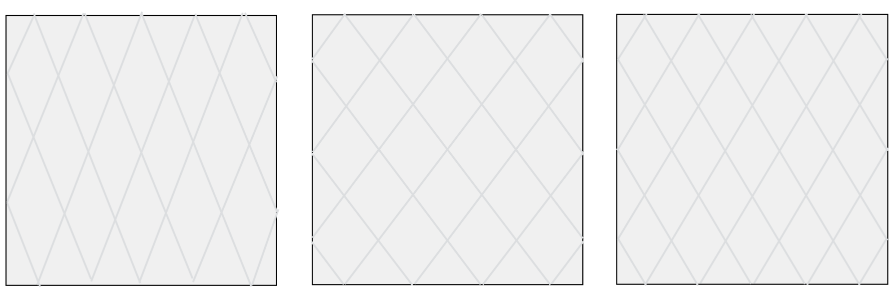
DD

3.42. Looking at the examples on this page, explain the relationship between billiards and Celtic knots.

DD



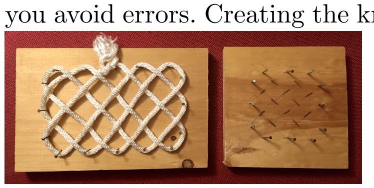
3.43. By now you should be pretty good at drawing diagrams of periodic billiard trajectories on the square. But how to turn it into a *knot*? All Celtic knots are *alternating*, meaning that if you follow a cord along its journey, it alternates over, under, over, under... as it crosses other parts of the cord.
(a) Follow along the knot diagrams on this page and convince yourself that they are all alternating.
(b) Draw Celtic knots based on the billiard trajectories below. An example is done on the right. (Consider drawing in the "crossings" first, following the path around to make it alternating, and then fill in the rest of the knot.)



DD

3.44. With a rope, create a Celtic knot based on periodic billiard trajectories.

Advice: Draw a picture of the desired knot, including the crossings, to help you avoid errors. Creating the knot is is easiest to do if you have a solid frame



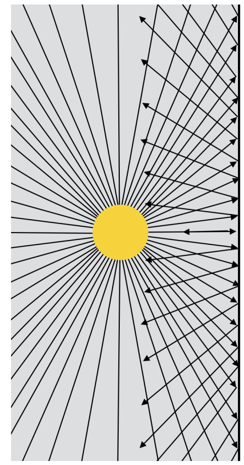
4

Broader horizons

In this chapter we expand our horizons to new questions (the illumination problem), we build trees of periodic directions, and we meet an eclectic menagerie of surfaces designed to do all sorts of interesting things.

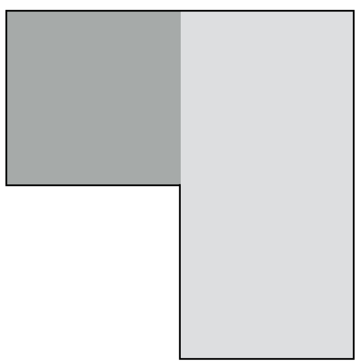
4.1 We prove that tiling billiards are related to IETs

DD



4.1. Let's get illuminated! The picture to the right shows what happens when you put a candle in a room: the light radiates out in every direction. Look closely at the right side of the picture: this room has a *mirror* on the wall, so the rays that hit the wall bounce off, following the billiard reflection law.

Suppose that you are in a room whose walls are *all* mirrored. You wish to illuminate your entire room with a single candle.
(a) Explain why this problem is easy when the room is convex.
(b) Suppose your room is an L-shape made of three squares, as shown below, and suppose you place the candle somewhere in the dark square. Does the candle illuminate the whole room? Explain why or why not.



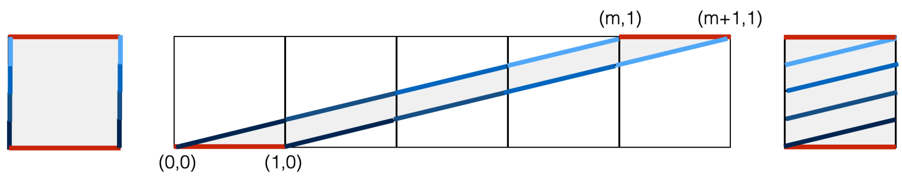
The *illumination problem* asks a generalization of the above: for which shapes of mirrored room can you put a candle *anywhere* in the room, and be sure that the light will reach every point? George Tokarsky constructed an example of a polygonal room that contains two points A and B that do not illuminate each other: a candle placed at A will illuminate every point *except* point B, and vice versa. Later, Samuel Lelièvre (§ 2.3), Thierry Monteil and Barak Weiss wrote a paper memorably titled "Everything is illuminated" where they proved that all polygons are basically like that: every point illuminates every other point, except possibly for a finite collection of points that don't illuminate each other.

To prove their result, Lelièvre, Monteil and Weiss used the "Magic Wand Theorem," the colloquial name for a collection of powerful results from a paper of Alex Eskin, Maryam Mirzakhani and Amir Mohammadi. Eskin (center) is shown with his wife Anna Smulkowska (left) and mathematician Ursula Hamenstädt (right).

Eskin received the Breakthrough Prize in 2020 for that work.

DD

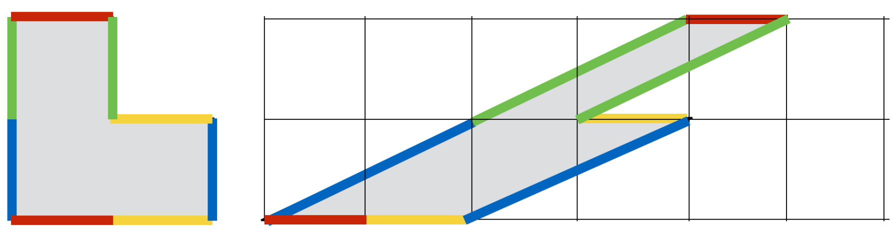
4.2. In our earlier work, we sheared the square torus by the matrix $\left[\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right]$, which transformed it into a parallelogram, and then we reassembled the pieces back into a square, which was a twist of the torus surface. Below is another way of shearing the square torus (left), this time via the matrix $\left[\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right]$, and reassembling the pieces (right) in such a way that the reassembly respects the edge identifications. The edge identifications are indicated with shades of blue. Explain what is going on.



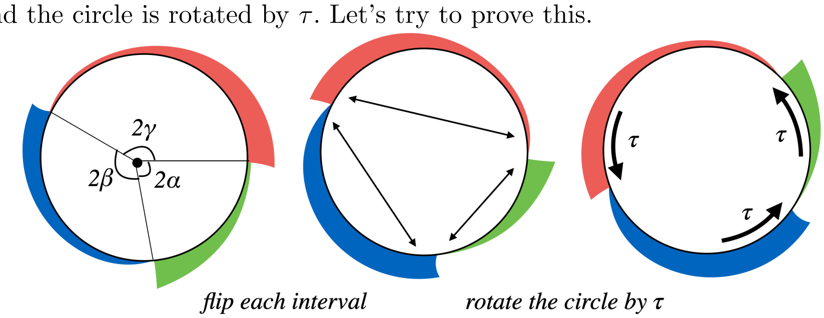
DD

4.3. Consider the L-shaped surface made of three squares, with edge identifications as shown in the left picture below. We shear it by the matrix $\left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right]$, as shown.

Show how to reassemble the sheared surface back into the L surface. Make sure your reassembly respects the edge identifications.

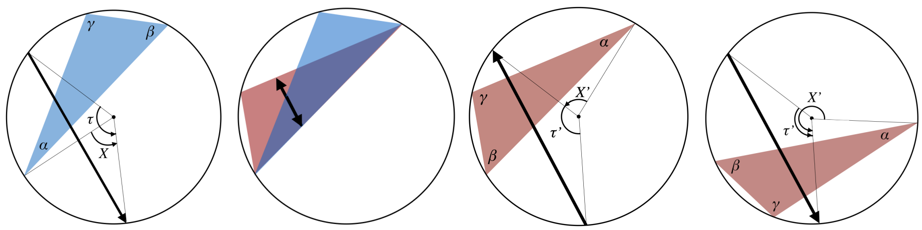


Finally, we'll see what we've all been waiting for: that a tiling billiards trajectory is equivalent to the orbit of a point on a certain IET. It turns out that movement of a tiling billiards trajectory whose chord subtends angle τ in a triangle tiling with angles α, β, γ is described by the orbit of a point on an orientation-reversing *circle exchange transformation* (see below): the unit circle is cut into intervals of length $2\alpha, 2\beta, 2\gamma$, each interval is flipped in place, and the circle is rotated by τ. Let's try to prove this.

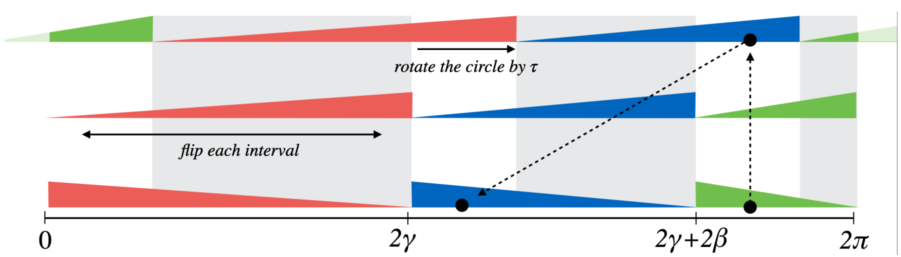


4.4. Refer to the pictures below. Show that, for a trajectory in a triangle tiling with parameters (X, τ), passing to the next triangle gives new parameters (X', τ'), where $\tau' = \tau$ (already shown in Problem 3.41), and

$$X' = \begin{cases} \tau + 2\gamma - X, & \text{if the side crossed is } C \ (0 < X < 2\gamma); \\ \tau - 2\beta + 2\gamma - X, & \text{if the side crossed is } A \ (2\gamma < X < 2\gamma + 2\alpha); \\ \tau - 2\beta - X, & \text{if the side crossed is } B \ (2\gamma + 2\alpha < X < 2\pi). \end{cases}$$



It's a bit more comfortable to work with IETs that look like intervals. In the picture below, the circle is represented by the interval from 0 to 2π; you have to remember that the two ends are identified. The intervals are shown as triangles, so that their orientation is clear: they are *flipped*. The Tiling Billiards IET specified by the angles of the tiling triangle: α, β, and τ. The three subintervals are each flipped, and the entire interval is shifted to the right by τ modulo 2π.



4.5. Use the previous result to prove the following:
Theorem (Tiling Billiards IET). Given a triangle tiling with angles α, β, γ, and a trajectory with associated parameters (τ, X), passing to the next triangle transforms X according to the following 3-IET:

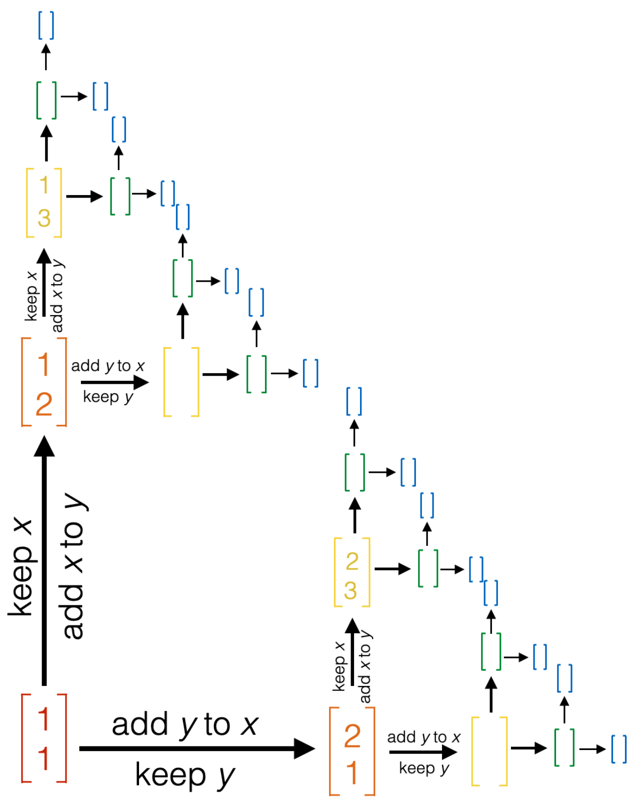
The interval $(0, 2\gamma)$ maps to $(\tau, \tau + 2\gamma)$, with the opposite orientation. The interval $(2\gamma, 2\gamma + 2\alpha)$ maps to $(\tau + 2\gamma, \tau + 2\gamma + 2\alpha)$, with the opposite orientation. The interval $(2\gamma + 2\alpha, 2\pi)$ maps to $(\tau + 2\gamma + 2\alpha, \tau + 2\pi)$, with with the opposite orientation. These transformations are all taken modulo 2π.

4.6. The parts of the IET where an interval overlaps itself (with a flip) would mean that the same side is hit twice in a row. In the picture above, they are shown shaded in grey.
(a) Explain why hitting the same side twice in a row is impossible.
(b) Show that, in fact, these regions are not in the domain of the trajectory system, because they correspond to chords that are disjoint from the triangle.

4.2 We meet our first exotic surface

DD

4.7. The picture below shows a way of starting with simple vectors and generating more complicated vectors. Here is how we construct this tree (called the *Farey tree*): start with the vector $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ in the lower left. At each step, choose to either add the entries together to get a new x-value (moving right), or choose to add the entries together to get a new y-value (moving up). Fill in as many entries as you can.



The picture shows the first five levels of an infinite binary tree. A *binary tree* means that at each *node* of the tree, you have two choices of where to go – in this case, right or up. I made each level smaller than the previous one so that five levels would fit on the page.

DD

4.8. (Continuation) Let's explore this tree a bit.
(a) Find $\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}3\\2\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}3\\5\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}8\\5\end{smallmatrix}\right]$ in the tree. Comment on any patterns.
(b) What vectors appear in this tree? Does your birthday vector $\left[\begin{smallmatrix}\text{month}\\\text{day}\end{smallmatrix}\right]$ appear in the tree? If so, at what level?

ST

4.9. *Counting periodic trajectories, part I*

One way to count periodic billiard trajectories in the square is to ask how many periodic trajectories it has with length less than L. Of course, periodic trajectories occur in parallel families (which form cylinders); we will count the number of such families.

(a) How long is the trajectory of slope 2? The trajectory of slope 3/4?

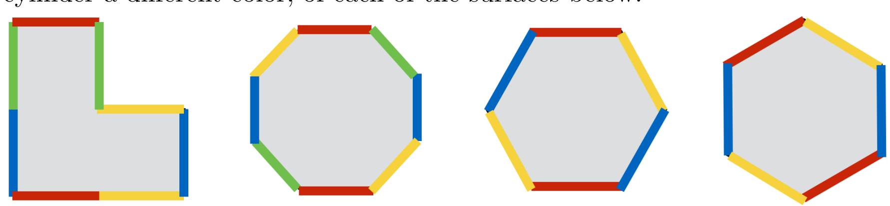
(b) Explain why the number of lattice points inside a disc of radius L is approximately πL^2, especially when L is large.

(c) Use the above to show that the number of periodic families of length less than L is approximately $\pi L^2/8$.

DD

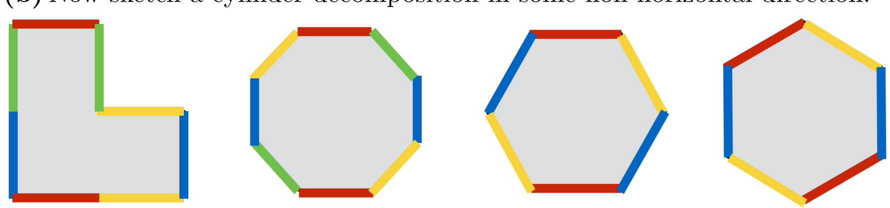
4.10. We have seen that we can can partition a polygon surface into *cylinders*. The boundary of a cylinder is a cylinder direction, and there are no vertices inside a given cylinder. To construct the cylinders, draw a line in the cylinder direction through each vertex of the surface, which might pass through many polygons before it reaches its ending vertex. These lines cut the surface up into strips, and then you can follow the edge identifications to see which strips are glued together. Recall Problem 3.39, where we saw several examples of cylinder decompositions for the double pentagon.

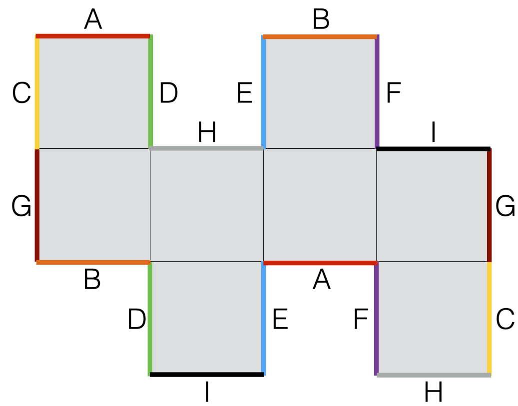
(a) Sketch the *horizontal cylinder decomposition*, by shading each horizontal cylinder a different color, of each of the surfaces below.



Notice that the regular hexagon surface has *one* cylinder in one cylinder direction, and *two* cylinders in another direction.

(b) Now sketch a cylinder decomposition in some non-horizontal direction.





Speaking of cylinders, let's meet a creature whose cylinders are legendary: the Eierlegendre Wollmilchsau, shown to the left. This surface is an example of a *square-tiled surface*, meaning that it is created by gluing together several – in this case, eight – unit squares, edge to edge. This surface is interesting because while it is not the square torus, it has a lot of properties in common with the square torus. We'll explore some of those now.

4.11. (a) Show that the surface has two horizontal cylinders and two vertical cylinders, and in each case the cylinder's width is 4 times its height. (We say that the cylinders have *modulus* 4.)

We have previously shown that the square torus has three types of automorphisms: rotations, reflections, and shears. The *group* consisting of all of the automorphisms of a surface is called the *Veech group* of the surface. If we think of the automorphisms in terms of the 2×2 matrices that perform them, we can say that the Veech group of the square torus is all 2×2 matrices with integer entries and determinant 1. This group is known as the "special (determinant 1) linear group of order 2 (2×2 matrices) with entries in $\mathbf{Z}$ (integers)," and denoted by $SL(2, \mathbf{Z})$.

(b) It turns out that the Veech group of the Eierlegendre Wollmilchsau[1] is also $SL(2, \mathbf{Z})$. Try to convince yourself that this is reasonable.

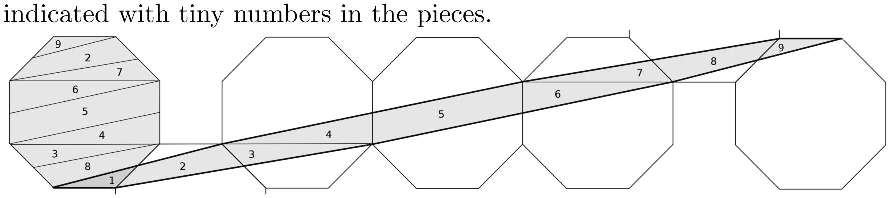

The Eierlegendre Wollmilchsau was independently discovered by Giovanni Forni in 2006 and Gabriela Weitze-Schmithüsen and Frank Herrlich in 2008. Schmithüsen and Herrlich (right) gave the surface its snappy name. It translates from German as "egg-laying wool-milk-sow" – an animal that provides eggs, woll, milk and meat, or in other words, everything a person could need. Similarly, this surface provides just about everything you could ever ask for in a surface.

[1] "EYE-ur-LEEG-un-duh VOLE-milsh-sow"

4.3 We finally meet Veech

DD

4.12. Amazingly, many surfaces made from regular polygons can be sheared, cut up and reassembled back into the original surface in the same way that we have done with the square and the L. One example is the regular octagon surface, shown below. The way to reassemble the sheared octagon pieces is indicated with tiny numbers in the pieces.

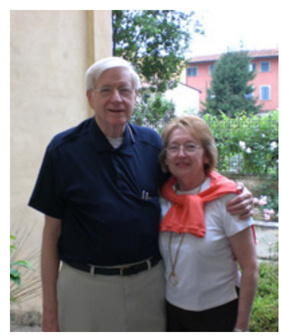


(a) By coloring each piece of each edge as in Problems 4.2–4.3, show that this reassembly respects the octagon surface's edge identifications. In other words, show that this shear is an automorphism of the octagon surface.

We say that a shear in a cylinder direction *twists* that cylinder, analogous to twisting the dough of a bagel. For example, in Problem 4.2 the shear $\left[\begin{smallmatrix}1 & m\\ 0 & 1\end{smallmatrix}\right]$ twists the square torus's single horizontal cylinder m times.

(b) In Problem 4.10, you identified this surface's two horizontal cylinders. In the shear above, how many times was each horizontal cylinder twisted?

(c) For each of the horizontal cylinders in the regular octagon surface, find its *modulus*, which is its width divided by its height. (You can think of the modulus as the "aspect ratio.")



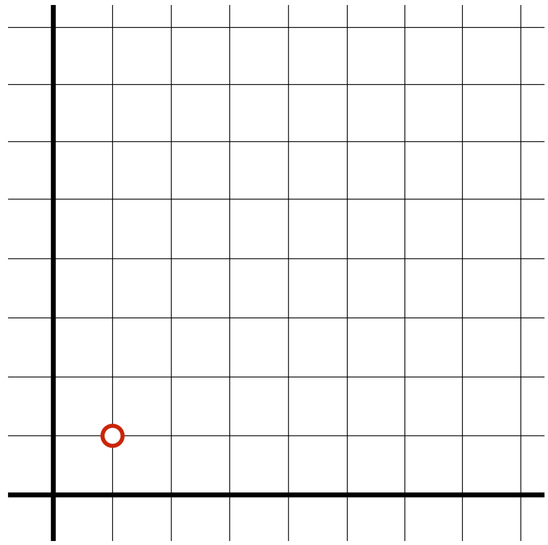
We previously said that the group of automorphisms of a surface is called its *Veech group* (Problem 4.11). When a surface has rotations, reflections, *and shears* in its Veech group, this means that its Veech group forms a lattice, so it is called a *lattice surface*.[2] Squares, regular octagons, and square-tiled surfaces are all examples of Veech surfaces. The key to being a Veech surface is that the set of cylinders in each periodic direction have *commensurable moduli*, meaning that they are all rational multiples of each other. For example, in the regular octagon surface above, one cylinder's modulus is twice the other's.

These notions are named for mathematician William Veech (above, with his wife Kay Veech), who really got this field going and then did a lot of tremendous work in it, including coming up with interval exchange transformations and Veech surfaces, and then proving results about all of their essential properties. One of his original examples of a Veech surface was the *double* regular octagon surface, chosen because its cylinders' moduli are equal.

---

[2] Historically such surfaces were called *Veech surfaces*, but people decided that too many things were named "Veech," and changed the terminology.

DD



4.13. You've built up rectangles from squares. You've filled in the binary tree of relatively prime vectors. Now let's look at a third way to generate all of the relatively prime vectors: *shears!*
(a) Start with $(1,1)$ as shown. If you apply the horizontal shear $\left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]$ to the red point, you'll get one new point, $(2,1)$ – draw this in orange. If you also apply the vertical shear $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$ to the red point, you'll get one new point $(1,2)$ – draw this in orange also.
(b) Now apply the horizontal shear $\left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]$ to the orange points, and draw these new points in yellow. Do the same for the vertical shear $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$, applying it to all of the orange points.
(c) Now apply both the horizontal and vertical shears to the yellow points. Draw these new points in green.
(d) Repeat the above for all the existing points. Draw the new points in blue. Continue. Mark all the points you get.

DD

4.14. *Making connections, again*
(a) Explain the connections between the three ways we have seen of generating new points: adding squares, adding vectors, and shearing the plane.
(b) Explain why every point we get in this way is *primitive*, meaning that the greatest common divisor of its components is 1.
(b) We can call this the set of *primitive* vectors, or the set of *visible points*: suppose that you are standing at the origin of an infinite orchard, and there is a tree at every lattice point. Then the points we generate above are the ones that you can see. Explain.

4.15. *Counting periodic trajectories, part II.* We can improve on our previous method of counting periodic trajectories (Problem 4.9) by counting primitive vectors, as these are the directions that give us different billiard trajectories.

Let P be the set of primitive vectors (Problems 4.13–4.14). For each natural number k, let kP be the set of primitive vectors multiplied by k, i.e. vectors $[a, b]$ where the greatest common divisor of a and b is k.

(a) Draw the set $2P$ on your picture in Problem 4.13.

(b) Explain why the union of all of the sets $P, 2P, 3P, \ldots$ is every lattice point in the first quadrant, and also show that the sets are disjoint (they have no elements in common). In other words, the sets form a *partition.*

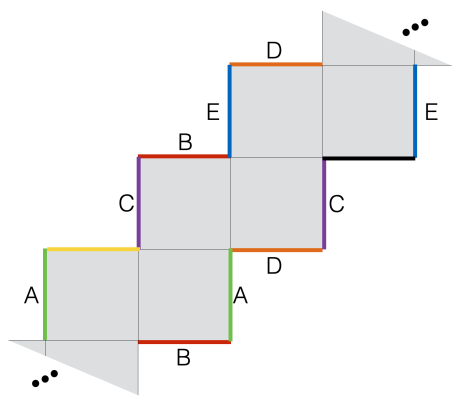
(c) We wish to know the proportion of the vectors in the first quadrant that are in P; let's call this proportion x. Show that the proportion of vectors in the first quadrant that are in in each set kP is $\frac{x}{k^2}$.

(d) Justify the equation $1 = \dfrac{x}{1^2} + \dfrac{x}{2^2} + \dfrac{x}{3^2} + \cdots = x\left(\dfrac{1}{1^2} + \dfrac{1}{2^2} + \dfrac{1}{3^2} + \cdots\right)$.

The latter sum is famous; it is known as $\zeta(2)$, as it is the value of the Riemann zeta function for exponent 2. It can be shown that the value is $\frac{\pi^2}{6}$, so the proportion of primitive vectors is $\frac{6}{\pi^2} \approx 61\%$.

Thanks to Juan Souto for explaining this proof to me, at a bar in Dublin.

4.16. *An infinite-area surface.* Consider the infinite staircase surface, shown to the right. It is a square-tiled surface, where edges are identified directly across, horizontally and vertically, as indicated. The same pattern continues forever in both directions.



(a) How many cone points does the surface have? What is the angle around each one? What is the genus of the surface?

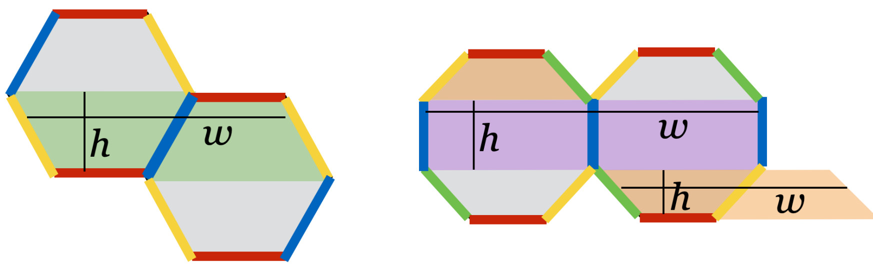
(b) Identify some periodic trajectories on the surface.

(c) Decompose the surface into cylinders in the direction of slope 1/2.

4.4 The Modulus Miracle

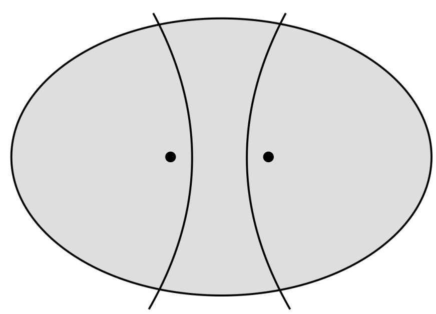
DD

4.17. Theorem *(Modulus Miracle)*. Every horizontal cylinder of a double regular n-gon surface the has same modulus ("aspect ratio"), which is $2\cot\pi/n$.
(a) Confirm this for the two surfaces shown, by calculating the modulus for each cylinder, and also the number $2\cot\pi/n$. If you can, prove it for all $n \geq 3$.
(b) Explain why this tells us that the horizontal shear $\left[\begin{smallmatrix}1 & 2\cot\pi/n\\ 0 & 1\end{smallmatrix}\right]$ is always an automorphism of the double regular n-gon surface.



(c) What are the benefits of using a double regular n-gon surface instead of a single one, like the familiar regular octagon surface? What happens with the moduli if you do use just a single n-gon?

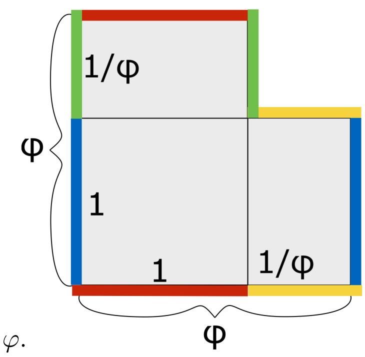
ST



4.18. An *ellipse* with foci F_1, F_2 and string length ℓ consists of all points X satisfying $|F_1X|+|XF_2| = \ell$. Similarly, a *hyperbola* with foci F_1, F_2 and "string length" ℓ consists of all points X satisfying $|F_1X|-|XF_2| = \pm\ell$.

In Problem 1.27, we showed that a trajectory *through* the foci always passes through the foci. In Problem 2.5, we showed that a trajectory *outside* the focal segment F_1F_2 stays outside and is tangent to an ellipse with the same foci. Show that every segment of a trajectory that passes *between* the foci is tangent to a hyperbola with the same foci.

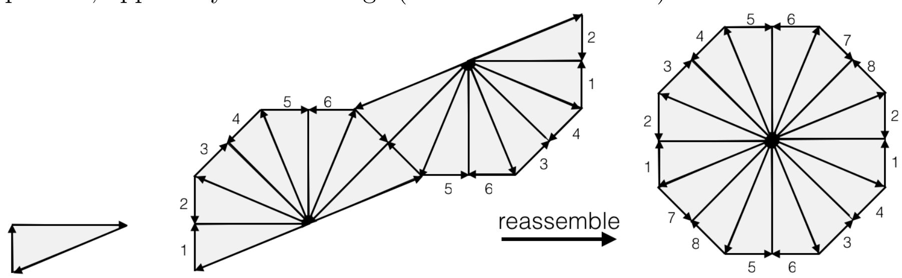
DD



4.19. A particularly nice flat surface is the "Golden L," whose opposite parallel edges are identified as shown, and whose edge lengths are as shown in the picture. The indicated number φ satisfies the property that when you cut off the largest possible square from a $1 \times \varphi$ rectangle, the leftover rectangle has the same proportions as the original.
(a) Show that this number satisfies the relation $\varphi = 1 + 1/\varphi$.
(b) Find the continued fraction expansion of φ.
(c) Two numbers are *commensurable* if they are rational multiples of each other. Are the moduli of the Golden L's cylinders commensurable?

DD

4.20. Our original motivation for studying the square torus was that it was the unfolding of the square billiard table. In fact, we can view *all* regular polygon surfaces as unfoldings of *triangular* billiard tables. We unfold the triangular billiard table with angles $(\pi/2, \pi/8, 3\pi/8)$ until every edge is paired with a parallel, oppositely-oriented edge (labeled with numbers):

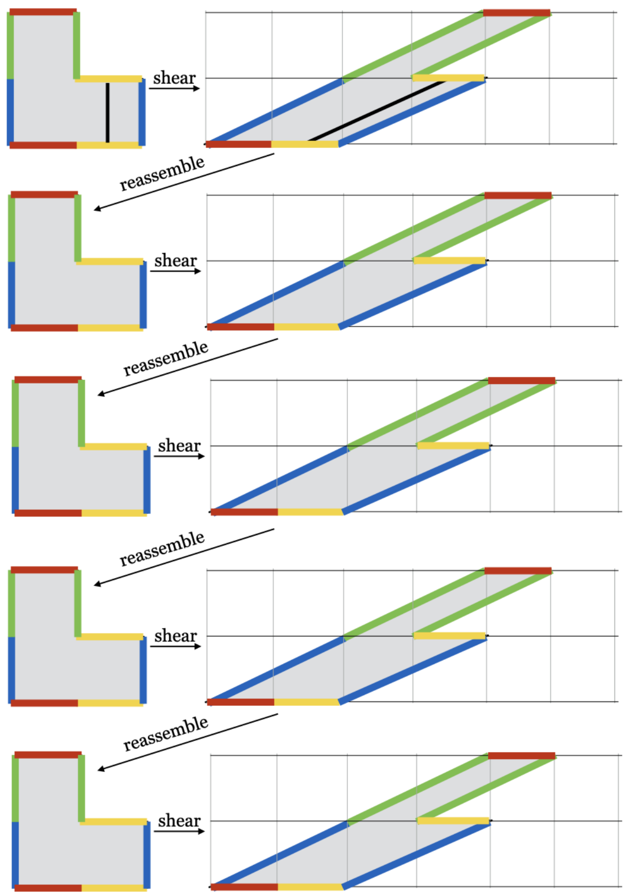


This gives us the regular octagon surface! So the regular octagon surface is the unfolding of the $(\pi/2, \pi/8, 3\pi/8)$ triangle.
(a) Draw the "shooting into the corner" period-6 trajectory in the triangular billiard table on the left above. Then unfold it to a periodic trajectory on the regular octagon surface.
(b) What triangle unfolds to the double regular pentagon surface? Dissect the double pentagon to figure it out, and then draw the unfolding as above.

A big question in the study of trajectories on surfaces is: “what happens to a trajectory on a surface when you apply an automorphism?”

DD

4.21. Let’s see what happens when we apply the horizontal shear $\left[\begin{smallmatrix}1&2\\0&1\end{smallmatrix}\right]$ to the L-shaped table, with a short vertical trajectory on it. The diagram to the right provides a template for drawing the image of the trajectory under this shear. Fill it in and sketch the image of this trajectory under five applications of this shear. Then say what the trajectory would look like if you sheared the surface many times.



Barak Weiss (top left) introduced me to the idea of twisting a cylinder over and over to see what happens. As you can see above, sometimes something interesting happens: one cylinder fills up, while the other stays empty.

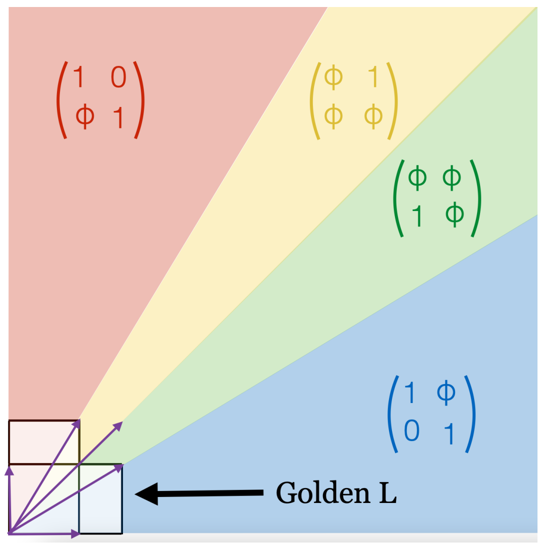
The picture at left shows some mathematicians on a hike in the mountains above Grenoble in 2019: Barak Weiss, Fernando Al Assal, Ben Dozier, and René Rühr, with the author.

4.5 The slit torus construction

We have seen that we can generate periodic directions on the square torus in three ways: adding vectors, adding squares, and applying shears. It turns out that applying shears (and more generally, applying automorphisms of the surface) is the method that best generalizes to other surfaces.

DD

4.22. The picture shows the first quadrant divided into four sectors, each corresponding to the sector created by consecutive diagonals of the golden L, shown with one corner at the origin.



(a) The dimensions of the golden L are given in Problem 4.19. Check that the purple vectors shown spanning diagonals of the golden L are $[0, 1]$, $[\phi, 1]$, $[\phi, \phi]$, $[1, \phi]$, $[0, 1]$.

(b) Explain why the blue matrix takes the entire first quadrant to the blue sector. Repeat for the three other colors.

Each of the matrices shown is an automorphism of the golden L. The blue and red matrices are horizontal and vertical shears, respectively. They are known as *parabolic* automorphisms. The green and yellow matrices act similarly to shears in a diagonal direction, but they tend to mix things up more than shears; they are known as *hyperbolic* automorphisms.

In Problem 4.13, we repeatedly applied horizontal and vertical shears to generate *all* of the periodic directions on the square torus. Similarly, to generate the set of *all* periodic directions on the golden L, we start with the vector $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and repeatedly apply the blue, green, yellow and red automorphism matrices.

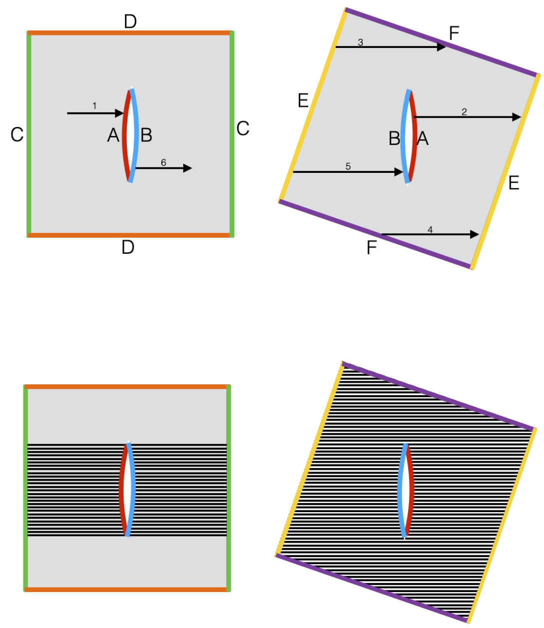
(c) Explain why this is reasonable.

(d) Compute the first two levels of the tree of periodic directions for the golden L (which contain 4 and 16 elements, respectively), using symmetry and the relation $\phi^2 = \phi + 1$ when possible to reduce your work.

So far, we have seen a lot of beautiful surfaces that do beautiful things. We have seen that a billiard trajectory with rational slope on a square table is periodic, and a billiard table with irrational slope is aperiodic. It turns out that every aperiodic trajectory on the square billiard table fills up the table evenly – its behavior is *ergodic*. That's because the square billiard table unfolds to a lattice surface. The *Veech dichotomy* (proved independently by William Veech and Howie Masur) says that a trajectory on a lattice surface is either periodic or ergodic.

When I first learned this, I thought it was obvious. After all, what other options are there? It turns out that there are surfaces where a trajectory can be dense in one region and not touch another region at all, or be half as dense in one region as in another region – or just about anything you can imagine. One nice demonstration of the first possibility is the slit torus.

DD



4.23. The *slit torus* surface is created by joining two square tori along a slit (see the top row of the picture to the right). One of the tori has horizontal and vertical edges, as usual, and the other one is rotated so that its edges have an irrational slope. We cut a vertical *slit* in each one, and identify the left and right edges A and B of one slit to the right and left edges B and A of the other, as shown in the top picture to the left.

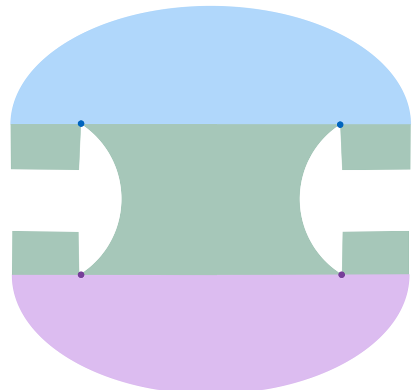
Edges A and B are vertical, but in the picture I have pulled them apart a little bit so that you can see that there is a slit between them.

(a) In the top picture, I have drawn the first six pieces of a horizontal trajectory. Draw the next 10 pieces. Do you expect this trajectory to be periodic?
(b) Explain why, over time, a horizontal trajectory will end up looking like the bottom picture.

Moon Duchin (center, in red shirt) explained the slit torus construction to me when I was a graduate student. Moon started out working in translation surfaces, and now works on identifying gerrymandering and creating fair districting practices. The picture shows Jane Wang, Viveka Erlandsson, Justin Lanier, Moon, Solly Coles, Madeline Elyze, Aaron Calderon, Felipe Ramírez, Andre Oliveira, Chandrika Sadanand and the author during a 2017 summer program that Moon organized.

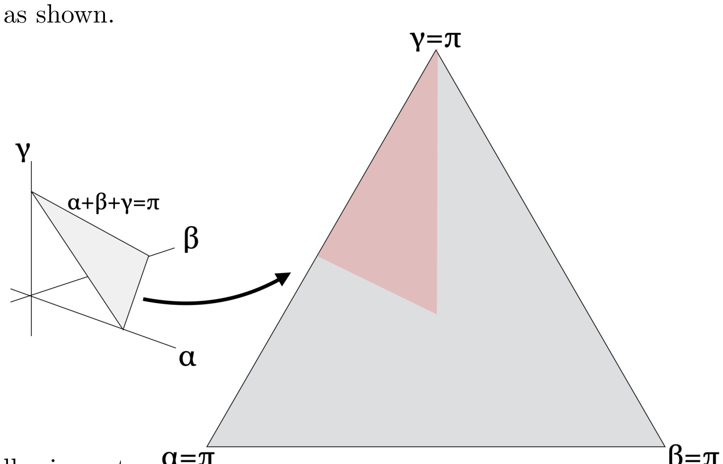
ST

4.24. *The Penrose unilluminable room.* One way to pose the illumination problem is: "Is every mirrored room illuminable from *some* point in the room?" Here is a counterexample, a room that cannot be illuminated from *any* point inside, shown to the right. The top and bottom are half-ellipses, whose foci are at the indicated points. Explain why this example works, by explaining which part of the room is illuminated when the candle is placed **(a)** in the interior of a half-ellipse, **(b)** in the middle part, and **(c)** in one of the rectangular parts.



DD

4.25. Up to symmetry, a triangle can be uniquely specified by its three angles α, β, γ. There are two restrictions on the angles: $\alpha+\beta+\gamma=\pi$ and $\alpha, \beta, \gamma > 0$. So we can represent the space of all possible triangles (up to similarity) by the triangular part of the plane $x+y+z=\pi$ that lies in the first octant, as shown. In this picture, each *point* of the space represents a *triangle*. So the space of triangles is itself a triangle! It's easier to see the picture if we lay the triangle flat, as shown.



Sketch the following sets:
(a) the set of right triangles (green),
(b) the set of isosceles triangles (blue),
(c) all triangles with angles $0.12\pi, 0.35\pi, 0.53\pi$ (black dots),
(d) the set of all acute triangles (shaded).

4.26. In this representation of the space of all triangles, the angles are *marked* – we keep track of which angle is α and which is β, so the $(0.12\pi, 0.35\pi, 0.53\pi)$ triangle is different from the $(0.35\pi, 0.53\pi, 0.12\pi)$ triangle. This is clearly redundant, so we can instead represent the space of triangles with *unmarked* angles. This takes advantage of the *symmetries* of the space of triangles to "fold up" the space so that each triangle is only represented once. DD

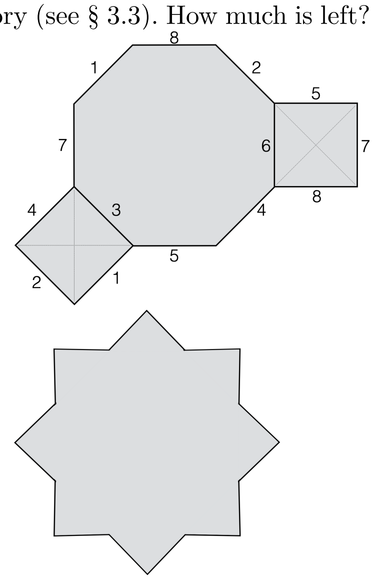
(a) Explain why the space of unmarked triangles is represented by just the red shaded part.

(b) Imagine folding up the space of triangles (grey) along all of its lines of symmetry. Explain why this gives you just the red shaded figure. Triangles with the most symmetry lie at the edges of this space. Explain.

(c) We have seen that right triangles with a vertex angle of π/n unfold to regular polygon surfaces. Sketch the set R of these triangles on your picture.

(d) Explain why the set R t is *discrete* in the space of triangles: for each point t of R, it is possible to find a little part of the space of triangles containing t, that does not contain any other point of R.

4.27. In the space of triangles, shade in the points that represent triangles that we *know* have a periodic billiard trajectory (see § 3.3). How much is left?



4.28. As mentioned above, surfaces with lots of symmetry are rare and precious. For some time, regular polygon surfaces and square-tiled surfaces were the only known lattice surfaces (surfaces whose automorphism group forms a lattice: it has rotations, reflections and shears). Then Veech's student, Clayton Ward, discovered a larger family of such surfaces, now known as *Ward surfaces*. One way to describe a Ward surface is as a regular $2n$-gon with two regular n-gons, where alternating edges of the $2n$-gon are glued to one of the n-gons, and the remaining edges of the $2n$-gon are glued to the other n-gon. DD

For $n = 4$, the Ward surface is an octagon and two squares, with edges identified as shown.

(a) Decompose this surface into horizontal cylinders, and check that their moduli are commensurable.

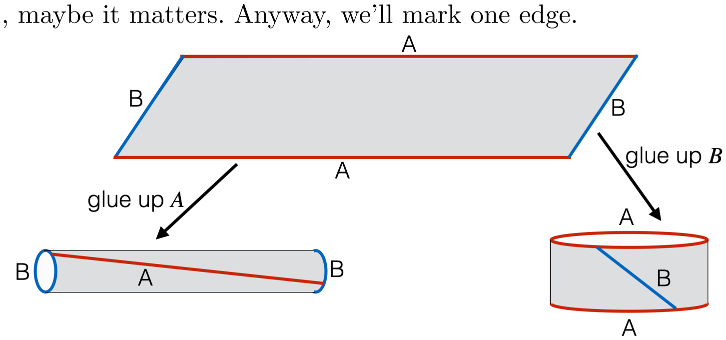
Ward actually represented this surface as a "flower": cut each of the squares into four pieces as shown in the top picture, and glue the eight "petals" around the octagon as shown in the bottom picture.

(b) Use the top picture to figure out which edges are identified in the bottom picture, and write in edge labels to record it.

4.6 Renormalization

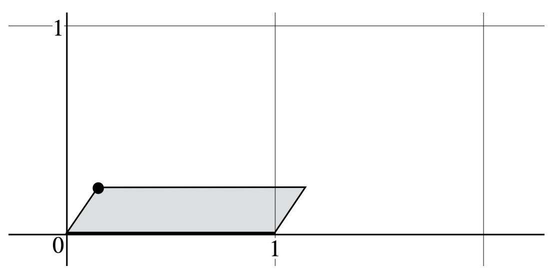
In Problem 4.25, we explored the space of all triangles. Now we'll explore the space of all tori made by gluing opposite parallel edges of parallelograms. This exposition follows § 20 of *Mostly Surfaces* by Rich Schwartz.

In particular, we will consider the space of *marked* parallelograms, meaning parallelograms with one edge identified. You can think of this as the marked edge being edge A, and the unmarked edge is B. When constructing the torus, we can think of first identifying the A edges to make a tube, and then identifying the B edges to close up the surface. When considering trajectories on the surface, it makes no difference which edges were identified "first," but when considering whether the surface is a skinny torus or a fat torus (see below), maybe it matters. Anyway, we'll mark one edge.



DD

4.29. Given any parallelogram, mark one of its edges. Translate, rotate, reflect and dilate the parallelogram in the plane until (1) the marked edge coincides with the interval $[0, 1]$ on the x-axis, and (2) the parallelogram lies in the *upper halfplane* (the region above the x-axis). Then the endpoint of the other edge that spans the parallelogram (the marked point in the picture) uniquely determines the parallelogram. Explain. In this way, we can say that the upper halfplane represents the space of all tori. Explain.



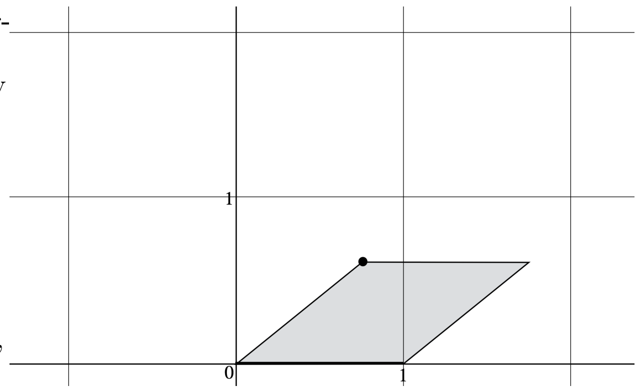
When working with the *square* torus, we used its Veech group, which is $SL(2, \mathbf{Z})$: the special (determinant 1) linear group with integer entries. Now we will work with $SL(2, \mathbf{R})$, the special linear group with real-valued entries. The *Iwasawa decomposition* says that any matrix in $SL(2, \mathbf{R})$ can be written as a product of matrices of the form K (compact), A (abelian) and N (nilpotent):

$$K = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad A = \begin{bmatrix} \lambda & 0 \\ 0 & \frac{1}{\lambda} \end{bmatrix} \qquad N = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$$

We wish to understand the effect of elements of $SL(2,\mathbf{R})$ on tori. In particular, we want to know how $SL(2,\mathbf{R})$ acts on the space of all tori that we created in Problem 4.29. Since every matrix in $SL(2,\mathbf{R})$ can be written as a product of matrices of the form K, A and N, the problem reduces to understanding the effects of these types of actions.

DD

4.30. Consider the parallelogram shown at right, which is uniquely determined by its marked point. Sketch the image of this marked point under *all possible* matrices of the form
(a) K (rotations),
(b) A (*horocycle flow*), and
(c) N (*geodesic flow*).



Marina Ratner (right) was a key figure in the field of dynamical systems and ergodic theory. She proved several powerful theorems that are key tools and inspired lots of further work. Several of her results were about the flows that we explored above. In her Ph.D. thesis, she studied geodesic flows, and in her later work, she proved several key "rigidity" results about horocycle flows.

4.31. Recall the slit torus surface in Problem 4.23. Show that, for the horizontal direction and also for the vertical direction, the left part of the surface has a cylinder decomposition but the right part does not. This is another example of behavior that fails to satisfy the Veech dichotomy.

4.32. *Renormalization and the Rauzy gasket.* Consider a triplet of numbers (a,b,c), where $a,b,c>0$ and $a+b+c=1$. You can think of these points as living on the same triangular piece of the plane $x+y+z=1$ as the space of triangles (Problem 4.25). Repeatedly perform the following algorithm:

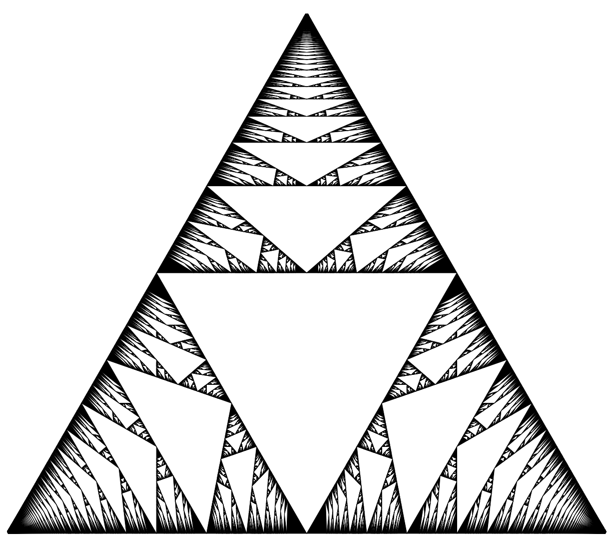
1. If
   $a>b+c$, subtract $b+c$ from a so that $(a,b,c)\mapsto(a-b-c,b,c)$.
   $b>a+c$, subtract $a+c$ from b so that $(a,b,c)\mapsto(a,b-a-c,c)$.
   $c>a+b$, subtract $a+b$ from c so that $(a,b,c)\mapsto(a,b,c-a-b)$.
   and if none of these are true, STOP.
2. Rescale the values so that they sum to 1.

(a) Show that $(7/12,4/12,1/12)\mapsto(2/7,4/7,1/7)\mapsto(2/4,1/4,1/4)$.
(b) Let $\alpha\approx 0.54369$ be the real solution to the equation $x+x^2+x^3=1$. Show that

$$(\alpha,\alpha^2,\alpha^3)\mapsto(\alpha^3,\alpha,\alpha^2),$$

so that this is a *fixed point* under permutation of the coordinates.

For most points, their iterated images eventually fail the condition that one element is greater than the sum of the other two, so the algorithm stops. But there are infinitely many points that can keep going in the algorithm forever; these points form a fractal set known as the *Rauzy gasket*, shown below.

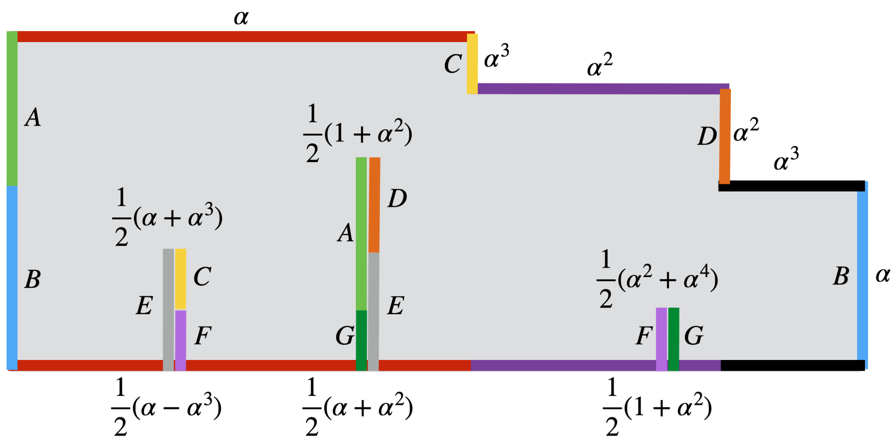


Whoa.

The algorithm above is considered a *renormalization* algorithm, because at the end of each step, you "normalize" so that the sum of the coordinates is 1. Similarly, in our algorithm for simplifying trajectories on the square torus, when a slope falls below 1 we "normalize" by flipping the trajectory so that the slope is above 1 again. Renormalization algorithms are a powerful tool.

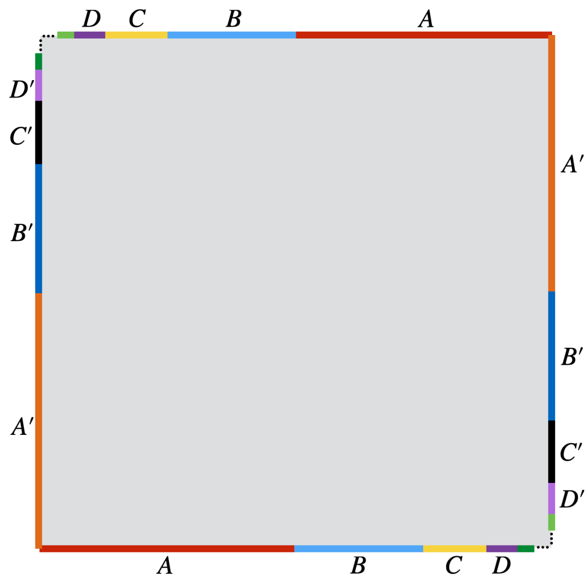
4.33. *The zippered rectangle construction.* The figure below shows how to create a translation surface out of "zippered" rectangles. The idea is that you glue together some rectangles, and you also make some vertical cuts, like a zipper. As in the slit torus construction, the two edges of the zipper are glued to different places.

The dimensions in the picture are in terms of α, the number from the previous problem. The horizontal edges are identified directly down, and the vertical edges are identified as shown. Pierre Arnoux (§ 2.8) created this surface to give a geometric realization of an IET. Show that it has genus 3.



4.7 Not all automorphism groups are $SL(2,\mathbb{Z})$

DD



4.34. *A wild translation surface.* The figure shows the *Chamanara surface.* The edges have lengths $1/2, 1/4, 1/8, \ldots$. Parallel edges of the same length are identified, as shown. The pattern continues all the way into the corners.

(a) Use vertex chasing to show that the surface has 2 vertices.

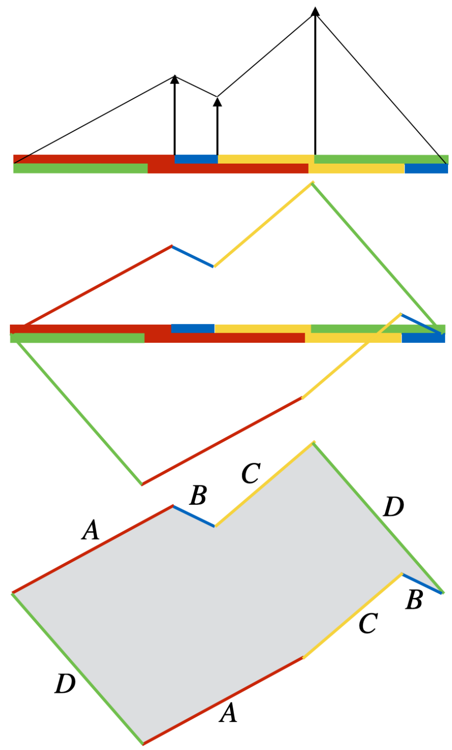
(b) But wait – how far apart are the two vertices? We can see that at the corners, the distance between the two vertices goes to 0. So in fact, there is only one vertex. Explain. This sort of behavior is what makes the surface *wild.*

(c) Show that the Chamanara surface has a cylinder decomposition in the direction of slope 4. If you can, show that all of the cylinders in this direction have the same modulus (51/4). Indeed, show that it has a cylinder decomposition in the direction of *every* slope of the form $1/2^n$.

To find cylinders, connect vertices with line segments. A line segment that connects two vertices *and* has no vertices in its interior is called a *saddle connection.*

Anja Randecker (left, with the author at Heidelberg University in summer 2022) studied wild translation surfaces for her Ph.D. thesis. She determined that wild translation surfaces had an important property that no one had identified before, so she studied it, and named it *xossiness*: *e*x*istence *o*f *s*hort *s*addle connections *i*ntersected *n*ot by *e*ven *s*horter *s*addle connections. The content of the problem above comes from Anja's thesis.

DD



4.35. In Problem 3.33, we found that the family of trajectories in a given direction (known as a *foliation*) on a particular translation surface has exactly the same behavior as a certain 3-IET. You might wonder: given *any* IET, can you find a translation surface, and a foliation direction, that matches the IET's behavior? Yes, you can, using a *suspension*.

Given any IET, do the following:

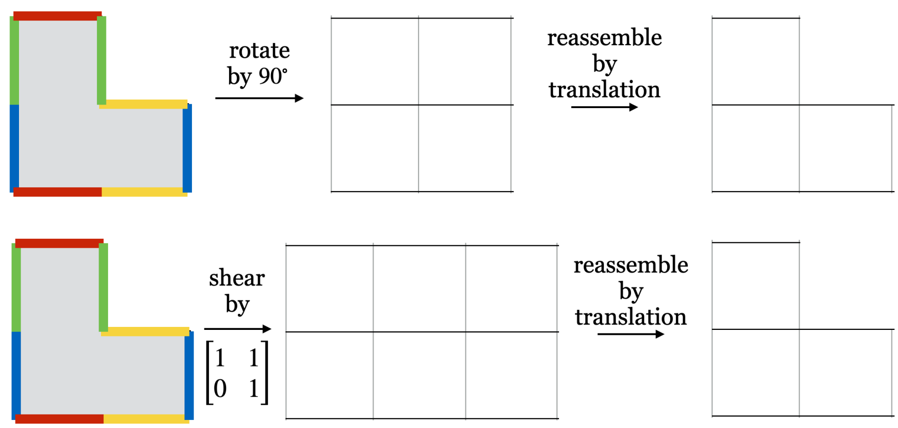
1. First, for each break point in your IET, choose a "height," and draw edges from the endpoints of the IET that attain each of the heights (top picture).
2. Color-code your edges and translate copies of them corresponding to the bottom part of the IET (middle picture).
3. Finally, make it into a translation surface.

Ta-da! You have a translation surface whose vertical foliation has exactly the same behavior as your IET.

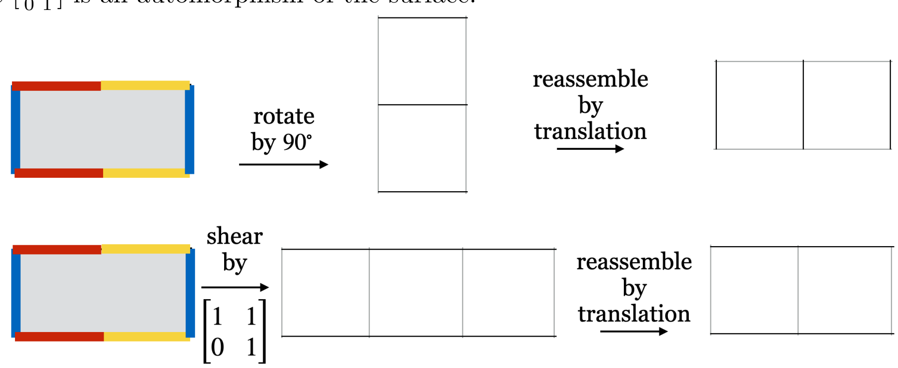
Make up an IET of your choice with at least 4 intervals. Suspend it to create a corresponding translation surface, as described above.

4.36. In Problem 4.3, we showed that the shear $\left[\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right]$ is an automorphism of the L-shaped surface made from three squares: You can apply this automorphism, and then rearrange the resulting pieces by translation, while respecting edge identifications, to get back the same surface you started with.

What about 90° rotations? What about the simpler shear $\left[\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right]$? Apply these transformations to the surface and show that the rotation is an automorphism of the surface, while the shear is not.



Do the same for the 2×1 rectangle to show that neither the rotation nor the $\left[\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right]$ is an automorphism of the surface.



DD

4.37. The biggest open problem in the study of lattice surfaces is: *Can we find more lattice surfaces?* and the related question, *Have we found them all yet?*

Here are the families of lattice surfaces we know about so far:

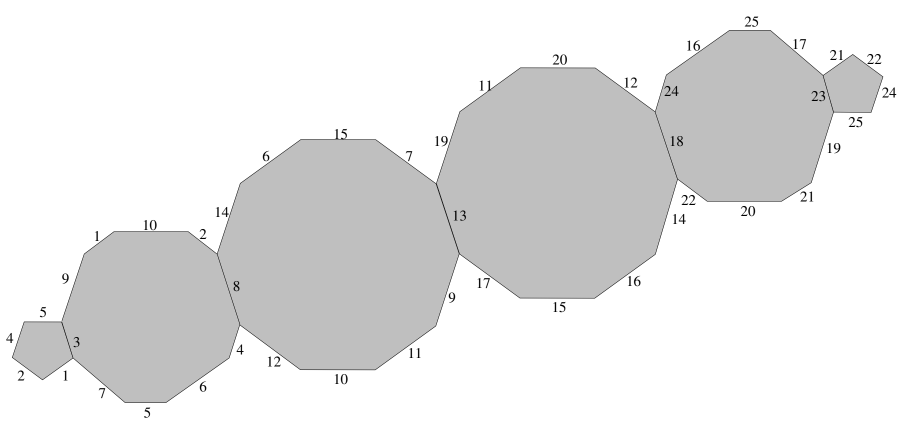
- square-tiled surfaces;
- regular polygons: double regular n-gons for any n, and single regular n-gons for even n;
- Ward surfaces: a regular $2n$-gon, with two regular n-gons glued to it along alternating edges.

In 2006, Irene Bouw and Martin Möller showed that double regular n-gon surfaces (made from 2 polygons) and Ward surfaces (made from 3 polygons) are the simplest examples in a larger family of lattice surfaces, now called Bouw-Möller surfaces in their honor. This family includes surfaces with any number $m \geq 2$ of polygons. Bouw and Möller gave an algebraic description of the surfaces, and later, Pat Hooper (§ 3.7) found a polygon decomposition for the surfaces, which we present here.

For any $m \geq 2$, and any $n \geq 3$, the (m, n) Bouw-Möller surface is created by identifying opposite parallel edges of m semi-regular $2n$-gons. A *semi-regular polygon* is an equiangular polygon with an even number of sides, whose edge lengths alternate between two values. (The lengths may be equal, and may be 0.) In particular, the k^{th} semi-regular $2n$-gon has edge lengths alternating between $\sin \frac{k\pi}{n}$ and $\sin \frac{(k+1)\pi}{n}$.

(a) Explain why a semi-regular $2n$-gon, half of whose edge lengths are 0, is a regular n-gon.

(b) The $m = 6$, $n = 5$ Bouw-Möller surface is shown below. Edge identifications are indicated by numbers. As you might have guessed, the edge lengths for these surfaces are carefully chosen so that the cylinders all have the same modulus. Shade each horizontal cylinder a different color. Does it seem plausible that all of the cylinders have the same modulus? Can you prove it?



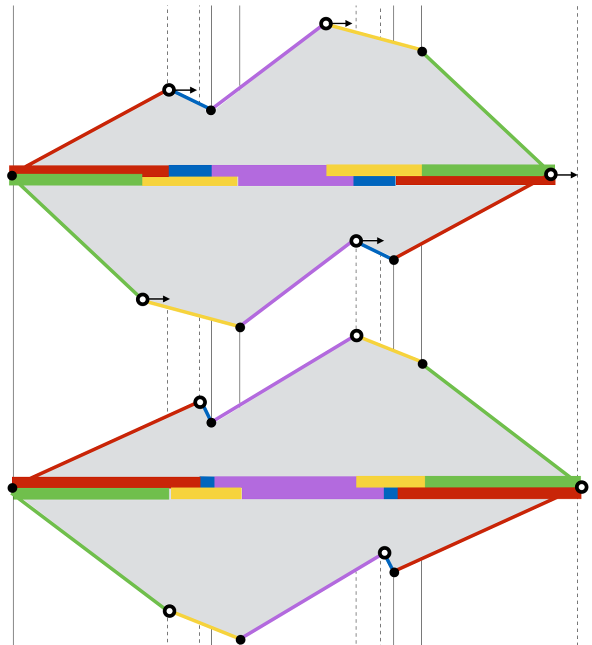
Irene Bouw and Martin Möller's paper describing their eponymous surfaces was published in the *Annals of Mathematics*, the most prestigious math journal. Accordingly, I thought that the two of them were untouchable gods of mathematics. So when I met Martin at a conference at ICERM in fall 2013, I thought it was the coolest thing, and I asked to take a picture with him (left picture below).

Come to find out, mathematicians are just people, and often very nice people at that. That's the idea I'm trying to get across in this book, with all of the pictures and profiles of active mathematicians working in billiards. Less than a year after that first picture was taken, Martin and I were running together for hours in the Black Forest in Germany, along with Ferràn Valdez and Ronen Mukamel. We have subsequently run together at many conferences, in several additional countries. The right picture below shows mathematician Erwan Lanneau, the author, and Martin Möller running in the Calanques at CIRM in France in 2017.

4.8 Moving around in the space of surfaces

DD

4.38. *The rel deformation.* Consider the translation surface to the right, created by suspending an IET as in Problem 4.35.

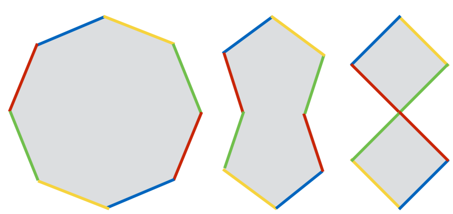


(a) Confirm that the surface has two cone points, indicated with black and white dots.

One way to get a new translation surface "near" the original one is to deform the surface by moving one cone point *relative* to the other. This is known as a *rel deformation.* The arrows in the top picture indicate that we will shift the white point slightly to the right. The bottom picture shows the surface after this deformation, along with the associated deformed IET.
(b) Explain what it means to be a "nearby" surface.
(c) How far can you push the white point to the right, and still create a valid translation surface? What about moving the white point in other directions – left, up, down, diagonally, etc.?
(d) Show that both of the above surfaces are in the stratum $\mathcal{H}(1,1)$. The rel deformation is thus a way to move continuously among a family of surfaces in $\mathcal{H}(1,1)$. Explain.

DD

4.39. The picture to the right shows the regular octagon surface (left) and the double pentagon surface (center).



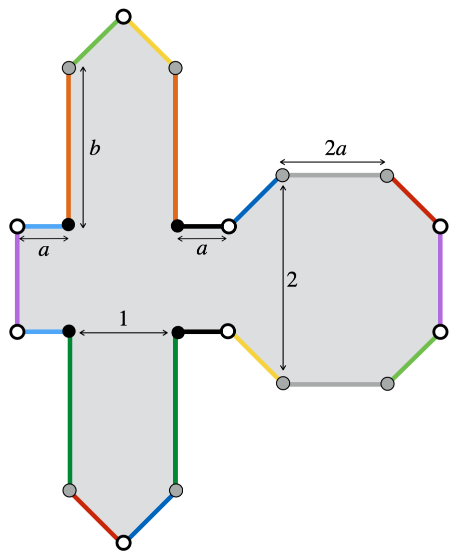
(a) Show how to smoothly deform the regular octagon surface into the double pentagon surface.

In Problem 3.13, you showed that both of these surfaces are in $\mathcal{H}(2)$.
(b) Suppose that we further deform the double pentagon surface into the double square surface (right). Explain why this surface is on the *boundary* of $\mathcal{H}(2)$. What kind of surface is it?

The above examples show that we can move around the space of surfaces in a given stratum. As we move around, most of the surfaces we encounter will be like the one in Problem 4.38: “random” surfaces with no nontrivial automorphisms, or in other words, no rotations, reflections or shears that preserve the structure of the surface.

On the other hand, the three surfaces in Problem 4.39 are lattice surfaces, which *do* have rotations, reflections and a shear as automorphisms. But as we move in $\mathcal{H}(2)$ to get from one to the other, the surfaces we encounter in between are *not* lattice surfaces. As mentioned in Problem 4.26, lattice surfaces are discrete – there is no way to move continuously among a family of lattice surfaces. This makes them difficult to find. When someone discovers a new family of lattice surfaces, it is a big deal.

DD **4.40.** In 2016, Curt McMullen (§ 2.7), Ronen Mukamel (below) and Alex Wright (§ 3.2) discovered the *gothic* family of lattice surfaces, so named because they look like the floor plan of a gothic cathedral. Its edges have slope 0, ∞, and ± 1, and are identified as shown. Its dimensions are as indicated; the lengths a and b determine the surface.



Show that each such surface has **(a)** 5 horizontal cylinders and 5 vertical cylinders; **(b)** 3 cone points, as indicated; and **(c)** genus 4.

The real key in showing that members of the gothic family are lattice surfaces is to carefully choose the measurements of a and b. It turns out that it is possible to choose rational numbers x, y and an integer $d \geq 0$ such that when

$$a = x + y\sqrt{d} \quad \text{and} \quad b = -3x - 3/2 + 3y\sqrt{d},$$

the surface is a lattice surface. The proof is beyond the scope of this text.

4.41. What stratum are the gothic surfaces in? Explain why it is not possible to find a *continuous* family of gothic surfaces in this stratum.

Ronen Mukamel (left, doing some math with the author in the Frankfurt airport in 2014) coauthored the result described above. He subsequently took a job working on computational biology and genetics.

4.9 The Rauzy fractal is everywhere

4.42. In a *tribonacci* sequence, each term is equal to the sum of the previous *three* terms. Find the first 12 terms of the tribonacci sequence beginning $0, 0, 1, \ldots$.

4.43. *Substitutions.* Consider a sequence of words made out of two letters, a and b. We use the following substitutions:

$$a \mapsto ab, \qquad b \mapsto a.$$

(a) Compute the first 10 terms of the sequence a, ab, aba, $abaab$, ...
(b) Show that the sequence of *lengths* of words is the Fibonacci sequence.
(c) Comment on any patterns you notice.
(c) Using the longest word you created above, plot a "broken line" in the following manner: start in the lower-left corner, and when you read an a, step to the right, and when you read a b, step up. Plot the resulting walk.

Notice that the points stay close to a line of slope $1/\phi$.

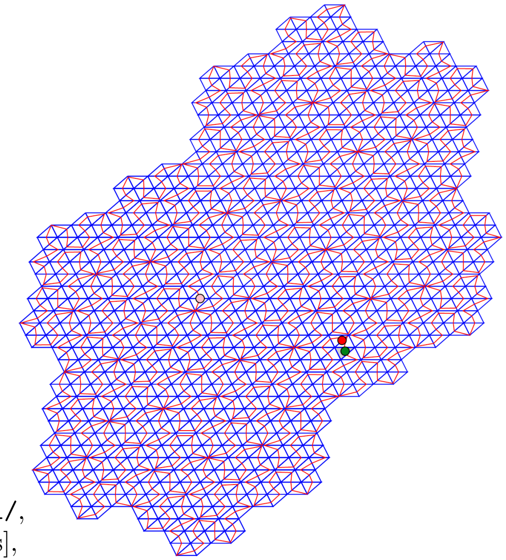
4.44. Now consider a sequence of words made out of a, b and c, with the substitutions

$$a \mapsto ab, \qquad b \mapsto ac, \qquad c \mapsto a.$$

Find the first 10 terms of the sequence a, ab, $abac$, ... and comment on any patterns.

Suppose that you use the resulting sequence to take a three-dimensional "walk" similar to the one in the previous problem, where a, b and c tell you to take steps in the x-, y- and z-directions, respectively. It turns out that, as in the previous problem, the points on this walk stay close to a line, now in 3D space. If we project these points in the direction of the line, onto a plane perpendicular to the line, we get a cluster of points that approach the *Rauzy fractal*, shown here.

DD

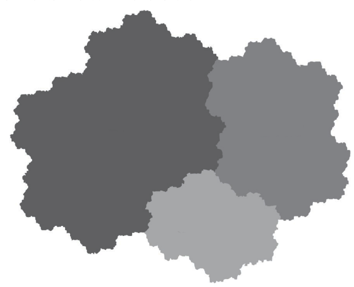


4.45. *Finding the Rauzy fractal in tiling billiards trajectories.* As before, let $\alpha \approx 0.54369$ be the real solution to $x + x^2 + x^3 = 1$. Consider tiling billiards on a triangle tiling with angles

$$\frac{\pi(1-\alpha)}{2} \approx 41.0679888577^\circ$$
$$\frac{\pi(1-\alpha^2)}{2} \approx 63.396203173^\circ$$
$$\frac{\pi(1-\alpha^3)}{2} \approx 75.535807969^\circ$$

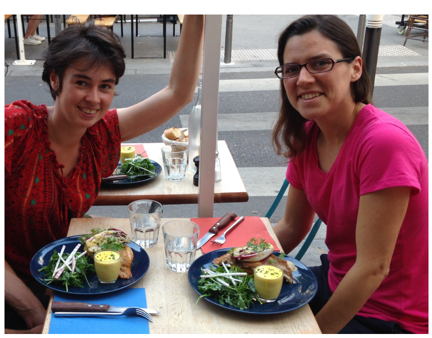
(a) Fire up the applet `https://awstlaur.github.io/negsnel/`, select New Triangle Tiling [angles], and type in two of the above angles. *Note:* things only get interesting when the angles are irrational, so enter all the digits listed above, to make the angles as irrational as possible.
(b) Move the green dot to the circumcenter of the triangle. You will have to approximate this as best you can. You will know when you are doing well because the trajectory will suddenly become very long.
(c) Move the red and green points to make a trajectory that is as long as you can. If your path does not close up, remember to increase the iterations! Can you find a periodic trajectory larger than the one shown above?

DD



4.46. As you find longer and longer trajectories, their appearance approaches that of the Rauzy fractal. Can you believe it?

The Rauzy fractal is a "tribonacci shape," in that three smaller copies of it join together to make one large copy of the same shape, as shown to the right. Explain.



Our number $\alpha \approx 0.54369$ is just one point in the Rauzy gasket (Problem 4.32). It turns out that when you make a triangle tiling based on a point in the Rauzy gasket, trajectories passing near the circumcenter *always* give you this type of fractal behavior. Olga Paris-Romaskevich (left, in Lyon in 2018 with the author) proved this, and many other results about tiling billiards.