Diana Davis

Billiards, Surfaces and Geometry: a problem-centered approach

July 17, 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. One exception is that the ability to multiply 2×2 matrices, use them to transform vectors, and geometrically interpret the result, is key throughout. Some familiarity with writing proofs is also helpful. This book is suitable for an undergraduate course, or for advanced high school students. Working through this book is also a great introduction for a graduate student who wishes to do research in this area.

The first four chapters will likely suffice for a one-semester or one-trimester course. If not covered in the course, the fifth chapter could be used as special topics e.g. for a final project. The hands-on activity sections don't contain necessary material for later in the book, but they do contain fun.

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 § 1.1 problems in class. For homework, students do § 1.2 problems.
- Second day of class: students spend class time discussing their solutions to § 1.2 problems. For homework, students do § 1.3 problems.
- Third day of class: students spend class time discussing their solutions to § 1.3 problems. For homework...

Depending on your students, you may need to assign more or less than this.

The problems teach the material on their own; no lectures are necessary. I wrote the problems to be hard enough that most students will not be able to solve all of the problems on their own, so that students have something to discuss with each other when they get to class. Since the need for discussion is built in, this book may not be ideal for an independent study.

For more guidance on how to run such a class, see the Appendix.

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 pictures out of this book)
- tape (for reassembling the cut-up surfaces)
- string (for making an ellipse)
- bagel, thin cord, board, hammer and nails (for hands-on activities)

Note that several problems ask you to draw certain things in specified colors, e.g. "draw these points in blue, and draw these other points in red." This occurs for two reasons: sometimes, it is so that you can compare notes with other people, without having to translate color meanings. Other times, it is so that the color coding matches up with diagrams in the book.

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 each class understanding the previous night's homework problems, as the next set of problems usually builds directly on them.

My goal for this book

There are so many beautiful things in the study of billiards and flat surfaces, and I've put all of my favorites in this book: the continued fraction algorithm, the folded-up flat torus, the Arnoux-Yoccoz IET... and the *people* in the billiards and flat surfaces community. If you work through these problems, draw all the pictures, compute all the orbits, and fold up all the toys, you'll be able to start doing research in this area. And if you read about all the people in this book, then when you read a paper or go to a conference, you'll feel like these people are already your friends. Welcome.

Bures-sur-Yvette, France *Diana Davis*
July 2023

Acknowledgements

Thank you to my department chairs, Frank Morgan at Williams College, and Gwyn Coogan at Phillips Exeter Academy, who allowed me to teach this course, once and twice respectively. Thank you to all of the students who have been in this course over the years, whose insightful questions and ideas improved this book a great deal.

Thank you to Samuel Lelièvre, who read through every problem in this book and provided innumerable suggestions for improvements, including (among many others) outlining a complete rewrite of the section on the modular group, giving me the diplotorus layout, and recreating the figures for the Rauzy gasket and the zippered rectangle surface. Thanks also to Yash Chandra and Barak Weiss for helpful suggestions.

Thank you to Thi Dang, Magali Jay and Samuel Lelièvre, who are working on a French translation.

And thank you to my Ph.D. advisor, Rich Schwartz, who suggested way back in 2009 that I read John Smillie and Corinna Ulcigrai's paper about linear trajectories on the regular octagon, and guided me to learn this beautiful subject and join this wonderful community.

NOTES ON THIS DRAFT:

I would love any feedback or suggestions that you have on this draft. Please send me an email or let me know in some other way.

For the photographs of people, I will get everyone's permission to have their likeness appear in the book, and I will obtain all necessary permissions and licenses, before including any image in any published book.

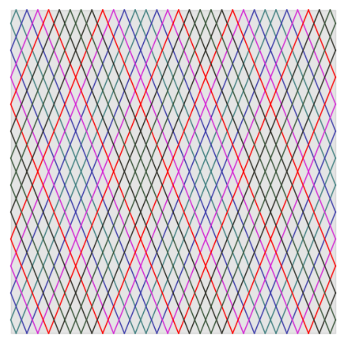
- If *you* appear in this book and you don't want to, let me know!
- If you *don't* appear in this book and you want to, send me a picture and let me know which section you suggest for yourself to appear!
- I am looking for casual group photos including:
  - Maryam Mirzakhani
  - William Veech

  - Marina Ratner
  - Irene Bouw
- . . . and I am also seeking better (casual, group) photos of Jon Chaika, Vincent Delecroix, Frank Herrlich, Irene Pasquinelli, Noelle Sawyer, Rich Schwartz, Caroline Series, Sasha Skripchenko, Gabriela Weitze-Schmithüsen, and Amie Wilkinson. If you have any, please send them my way.
- Additionally, I would love to have more (casual, group) photos of people in the book that do not contain *me*. If you have better photos of people in this book, again, please send them my way.

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 of understanding 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; in most cases, you can draw right on the picture that's on the page.

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 trajectories 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 we'll study more complicated surfaces.

Another powerful tool is transforming 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 paths and where can we find them?

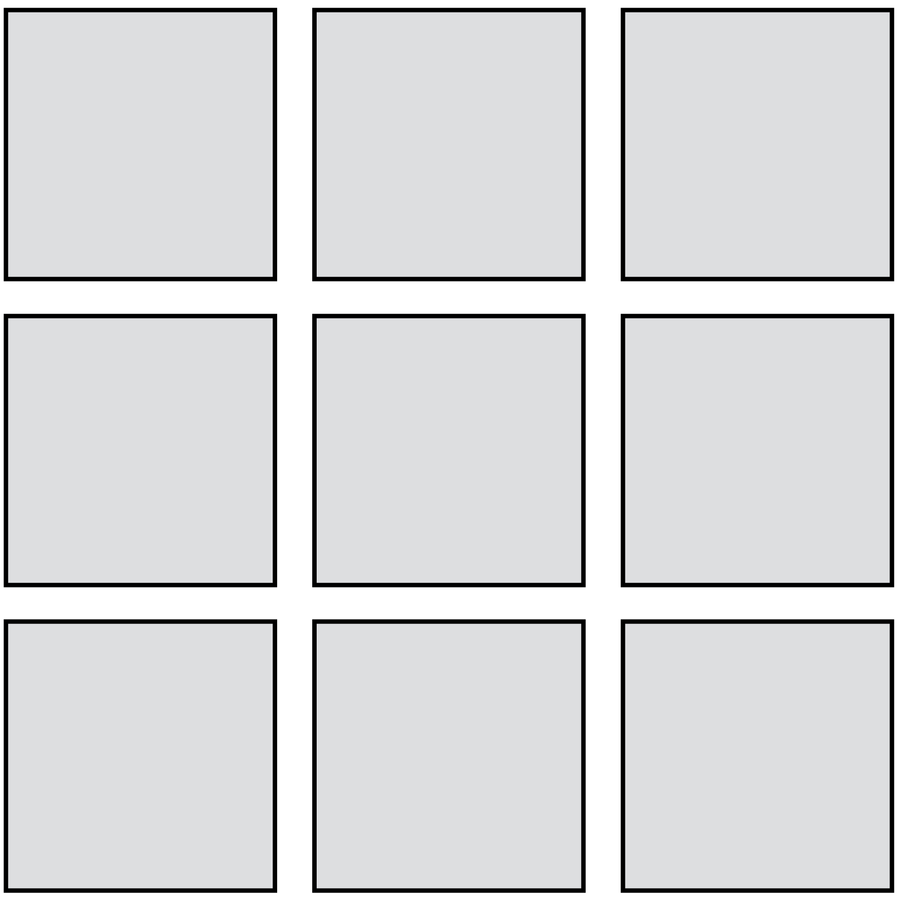
First day of class: in-class problems

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?

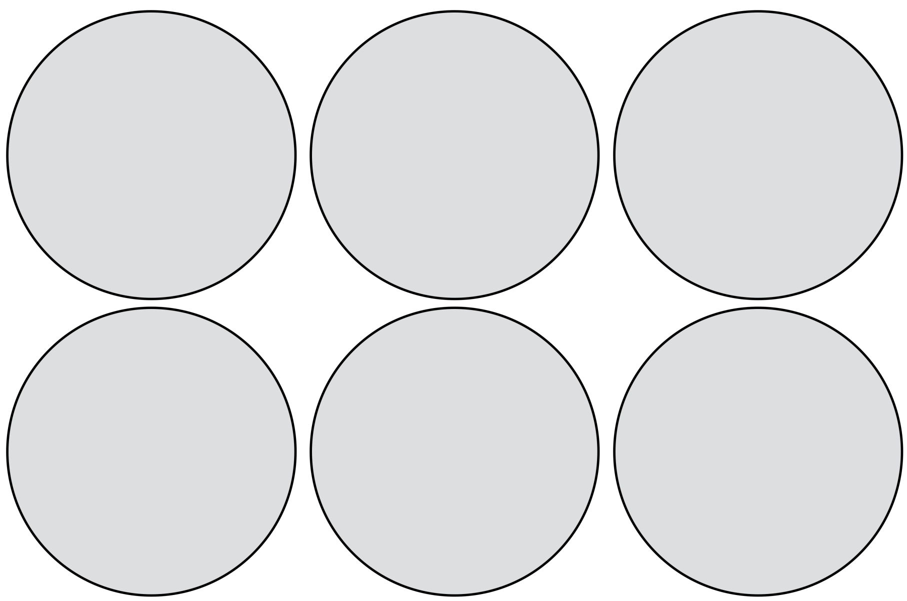


DD

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.

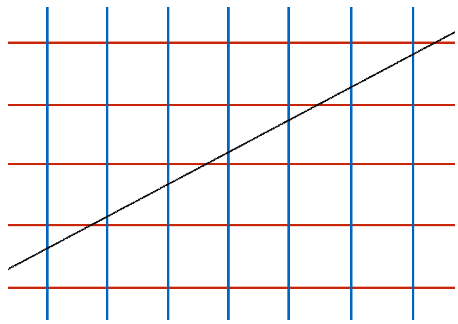
(b) Consider paths that close up (*periodic* paths), and also paths that don't (*aperiodic* paths). What is the probability that a billiard path in the circular table is periodic?
(c) Describe the behavior in general.



They did the math: Hello! I'm Diana. It's nice to meet you. I love billiards, especially on polygonal billiard tables, and especially on regular polygons. Here you can see that I have a regular octagon billiard table in my office, and I'm pretty excited about it, though when it comes down to it, I'm more of a regular pentagon aficionado. I hope you enjoy this book as much as I do.

1.2 We billiard outside of the box

DD



1.3. 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:

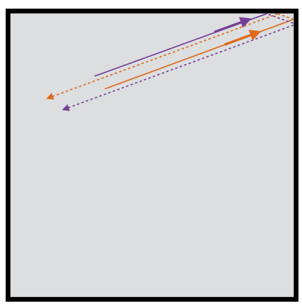
- *Authoritarian*: Trajectories are not allowed to hit vertices.
- *Minimalist*: If a trajectory hits a vertex, it stops.
- *Indecisive*: The vertex belongs to both sides, so it's ambiguous.
- *Optimistic*: If the ball hits the pocket in the corner, you win!

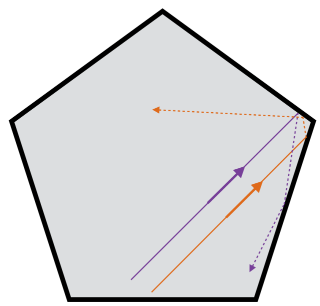
In any case, we generally consider trajectories that do not hit vertices.

They did the math: Caroline Series (pictured to the left) 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. We will see a little bit of hyperbolic geometry in § 5.1.

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!"

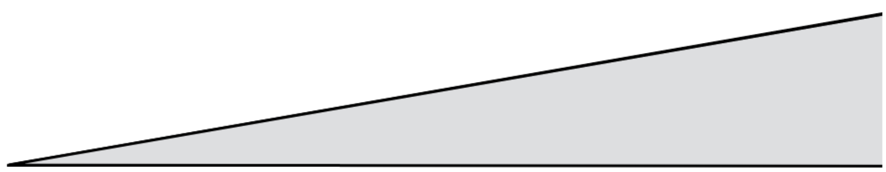
To be precise... In fact, while the *cutting sequence* corresponding to a trajectory that hits a vertex is ambiguous, the forward *trajectory* itself is not necessarily ambiguous. For example, on the square billiard table, nearby parallel trajectories continue to be nearby and parallel after two reflections (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 π, the behavior is like the square, and otherwise, the behavior is like the pentagon. 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. Thanks to Barak Weiss (§ 4.1) for pointing this out.





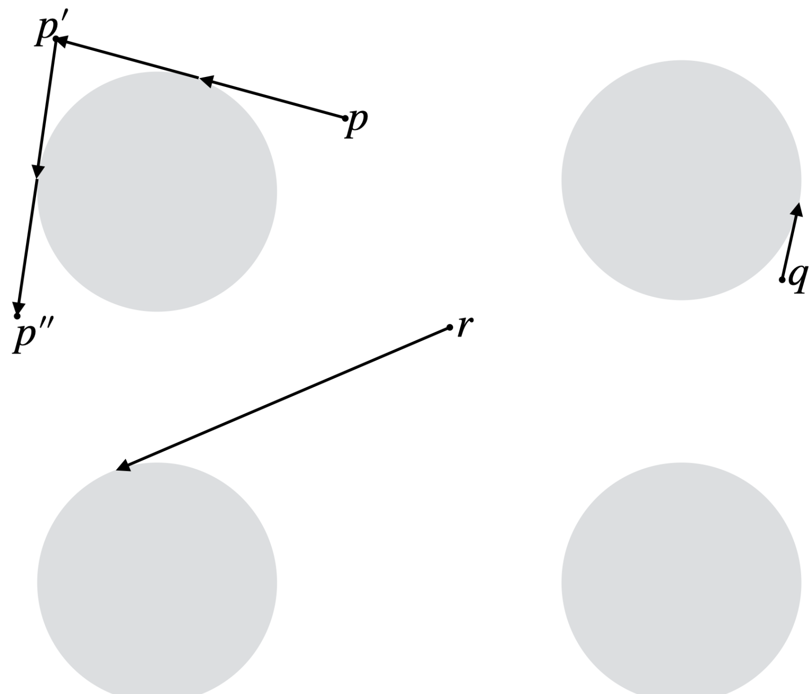
DD

1.4. 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.

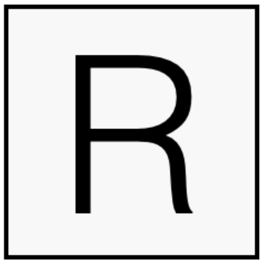


DD

1.5. *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



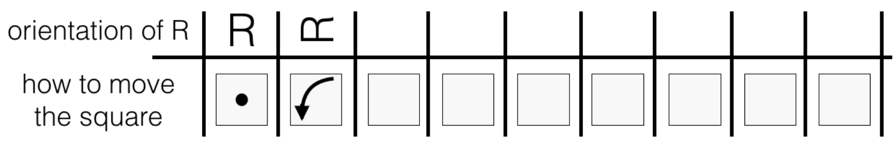
1.6. *Symmetries of the square.* If you turn a square 90° counter-clockwise, it looks the same as before. We call a 90° counter-clockwise rotation a *symmetry* of the square, because after you do it, you have a square just like the original.

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.

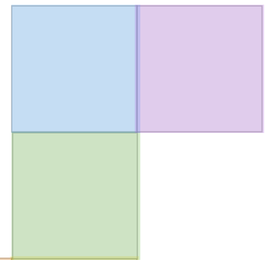


(d) Do you have all of them? If so, explain how you know.

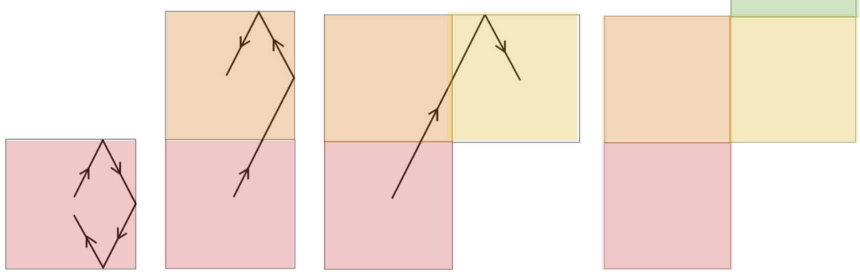
DD

1.3 We unfold

1.7. 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.



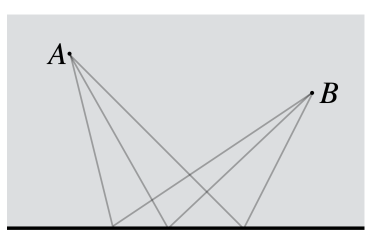
(a) Draw some more steps of the unfolding.
(b) Draw the complete billiard path in the original square: keep going until it closes up.



(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 1.6 will do for us in the future (Problem 2.10).

1.8. *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 (or "unfolding") 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 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 a little later.

1.9. 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 continued fraction expansion of $3.14 = 157/50$.

(b) 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?

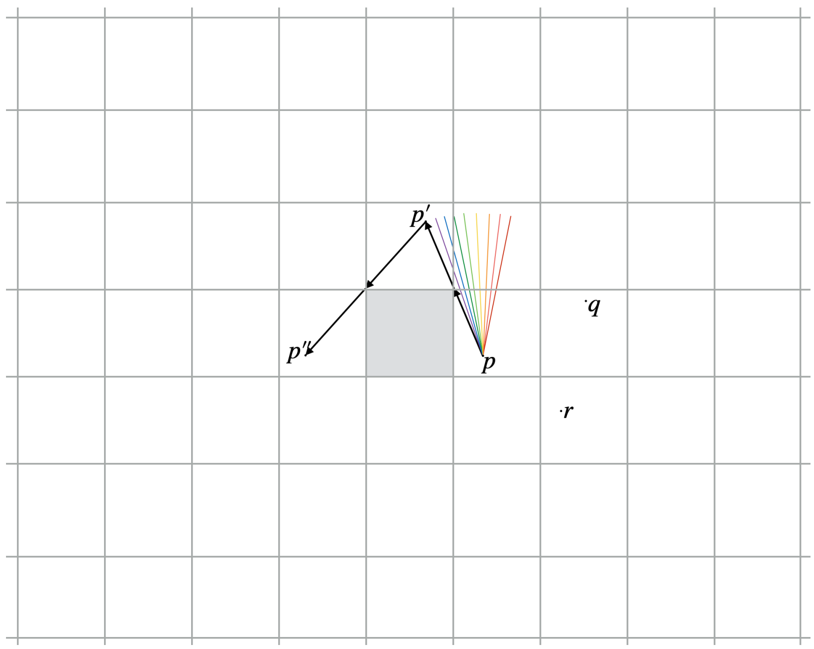
(c) Find a rational approximation of the number whose continued fraction expansion is $[1; 1, 1, 1, \ldots]$. This number, known as the golden ratio φ, is sometimes called the "most irrational number." Explain.

In part **(a)** 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, part **(b)** shows that we can find a better rational approximation by using the continued fraction expansion, and truncating it at a convenient point. Indeed, such *convergents* of the continued fraction expansion give the best rational approximations for a given size of denominator.

DD

1.10. 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.



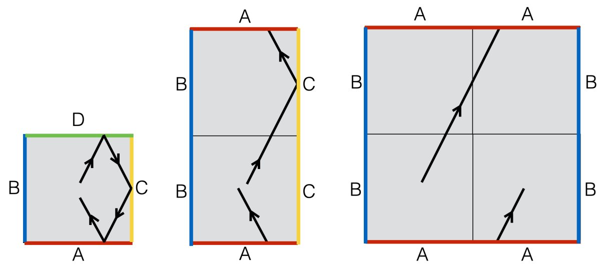
They did the math: The outer billiards system was proposed as a toy model for planetary motion: the table is the sun, and the point is the planet bouncing around it. It is easier to analyze a *discrete* dynamical system, where a planet jumps from place to place, than a *continuous* dynamical system where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

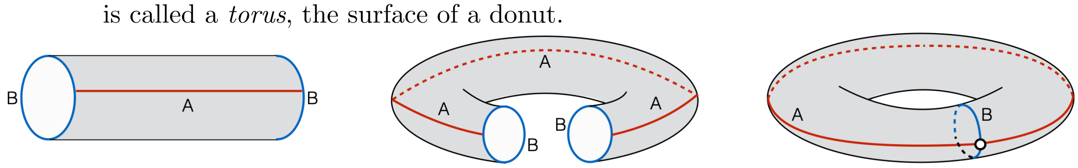
1.11. Here's another way that we can unfold the square billiard table. First, unfold across the top edge of the table, creating another copy in which the ball keeps going. The new top edge is just a copy of the bottom edge, so we now label them both A to remember that they are the same. Similarly, we 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 the corresponding *cutting sequence* for the trajectory on the surface on the right part of the figure?

(b) When we unfolded the trajectory to a line in Problem 1.7, 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 the tube around 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

1.12. 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.

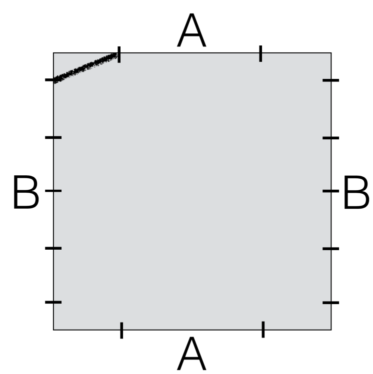
DD

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

They did the math: 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 working on flat surfaces, spread across the globe and particularly concentrated in France and the United States. 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 1.11, 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



1.14. In problem 1.11, 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.

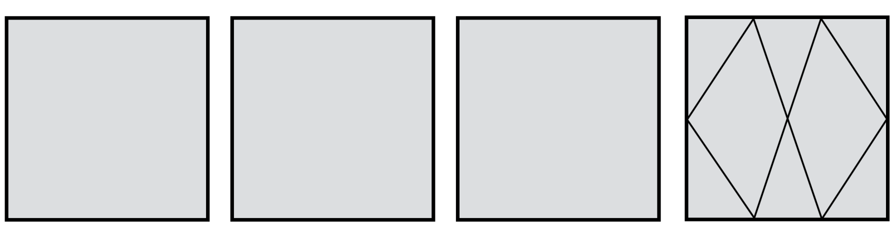
(a) Explain why, on the torus surface, these line segments connect up to form a continuous trajectory. Follow the trajectory along, and write down the corresponding cutting sequence of As and Bs.

(b) Exactly where should you place the tick marks so that all of the segments have the same slope? Prove your claim.

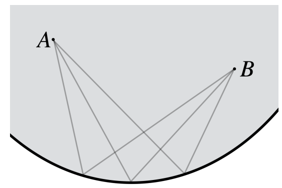
(c) Create an accurate picture for a trajectory of slope 1/2 and then 3/2. *Hint*: make sure that all of your segments look parallel.

(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



1.15. *The billiard reflection law, curved case.* We proved this law for linear boundaries in Problem 1.8; 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.

A note on terminology. In this book, I use the words "path" or "trajectory" to refer to linear motion on a billiard table or square torus. Other authors use the word "geodesic" to describe the same thing. On a surface, a *geodesic* is the (locally) shortest path between two points. For example, on a sphere, the geodesic between any two points is part of a great circle. On a flat surface, geodesics are lines.

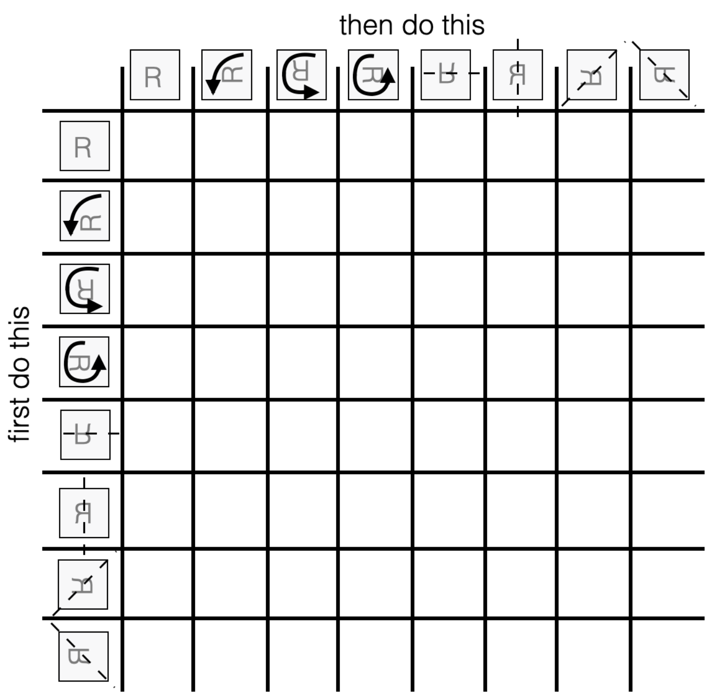
1.5 We do a little bit of group theory

DD

1.16. In Problem 1.6, 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. Use the arrow or dashed line!



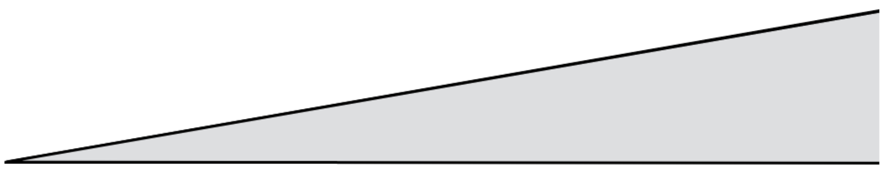
(c) Does this group of symmetries commute, i.e. is $ab = ba$ true for every pair of symmetries a, b? If not, is there *any* pair of symmetries that commutes?

They did the math: In the previous section, I mentioned that on flat surfaces, geodesics (locally shortest paths) are always linear, and this is true – *away from cone points*. We have been ignoring trajectories that hit cone points, but if you allow your trajectory to hit a cone point and want it to continue thereafter, the way to proceed is to require that the *turning angle* between the trajectory lines before and after hitting the cone point must be at least π on each side. Noelle Sawyer (left) studies these "singular" trajectories, and their geodesic continuations.

DD

1.17. Consider again (following Problem 1.4) 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. $\lceil \pi \rceil = 4$.



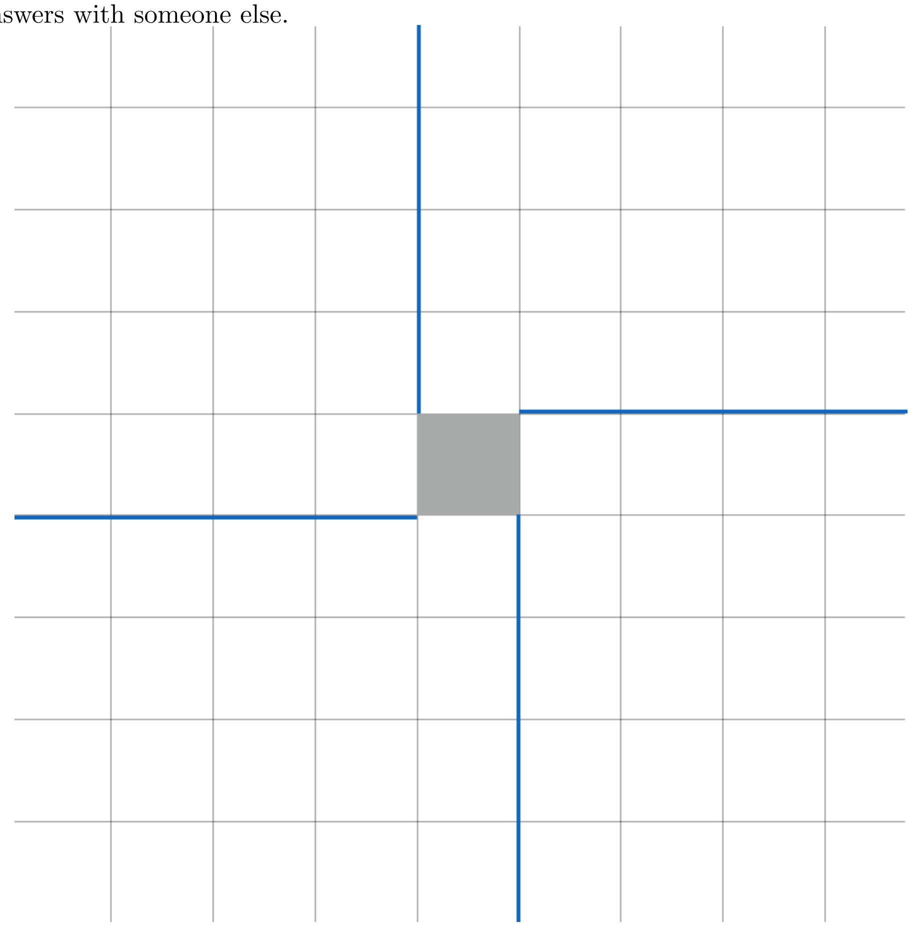
DD

1.18. Let's gather some data and make some conjectures.
(a) Construct an *accurate* (see Problem 1.14) 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

1.19. 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.



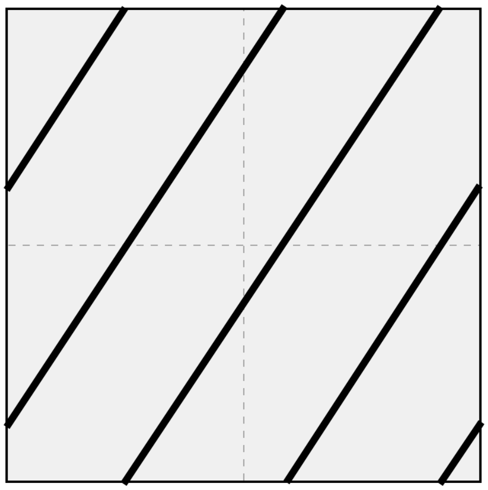
Note: The resemblance of the (incomplete) diagram to a swastika is unfortunately impossible to avoid. This symbol was first used 12,000 years ago; it is a natural construction that has become synonymous with an odious regime.

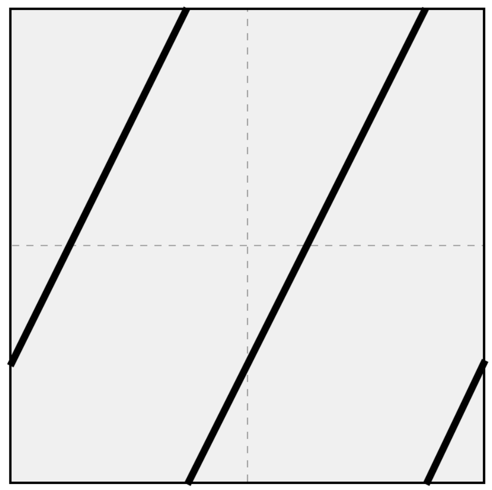
1.6 We fold up torus trajectories into billiards

DD

1.20. 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.





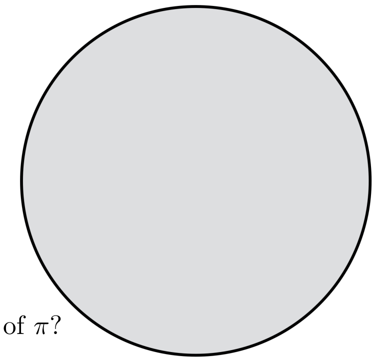
As previously explained, the study of flat surfaces is a very hot field these days, and many people are proving results about them. Sometimes, people are perfectly satisfied with results about flat surfaces, and they don't fold up their surfaces to get a billiard table back. You will not be one of these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

DD

1.21. In Problem 1.19 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.

ST

1.22. Consider a billiard trajectory in the unit circle, where at each impact the trajectory makes angle α with the (tangent line to the) circle.

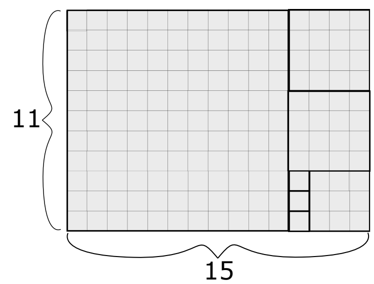


(a) Find the central angle θ from the circle's center, between each impact point and the next one, as a function of α.
(b) Prove that if $\theta = 2\pi p/q$ for integers p and q, then every billiard orbit is q-periodic and makes p turns around the circle before repeating.
(c) What happens if θ is *not*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

DD

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

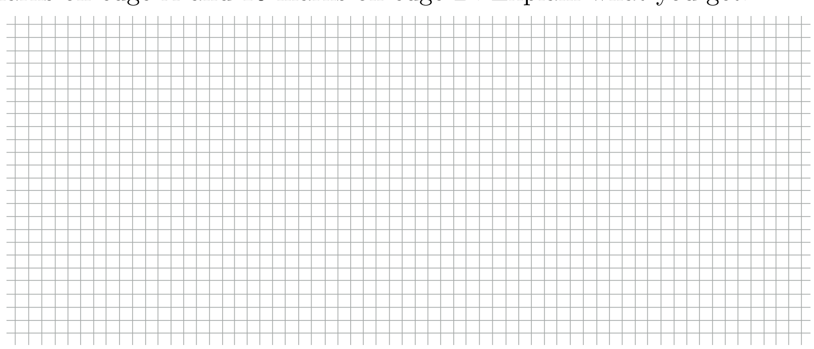
1. Begin with a $1\times x$ rectangle (or $p\times q$ if $x = p/q$).
2. Cut off the largest possible square, as many times as possible. Count how many squares you cut off; this is a_1.
3. With the remaining rectangle, cut off the largest possible squares; the number of these is a_2.
4. Continue until 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.
(b) Compute the continued fraction expansion for 5/7 in the way explained in Problem 1.9, 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.

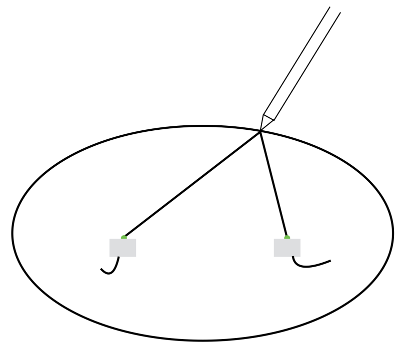
DD

1.24. In Problem 1.14, 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



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

1.26. 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.

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!

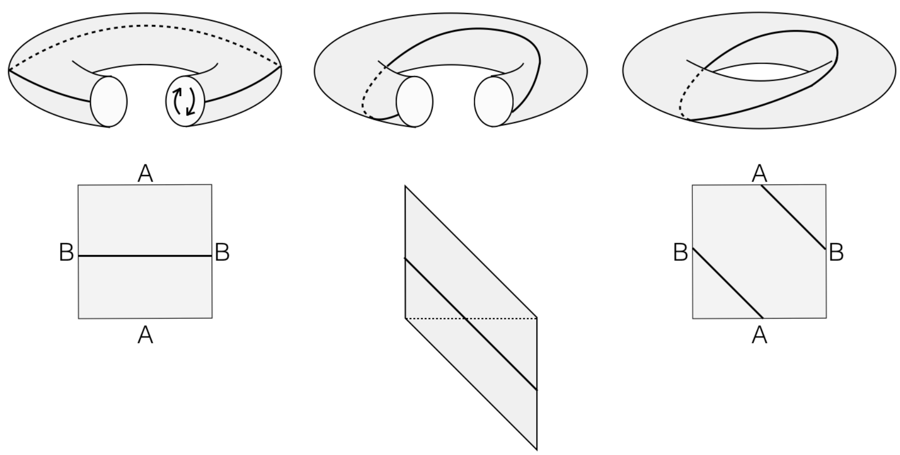
An accessible and impressive example of this is in Grand Central Station in New York City (above), where although the background noise is very loud, if you speak into one column, someone on the opposite column can hear you.

DD

1.27. 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 1.6 and 1.16.
(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?

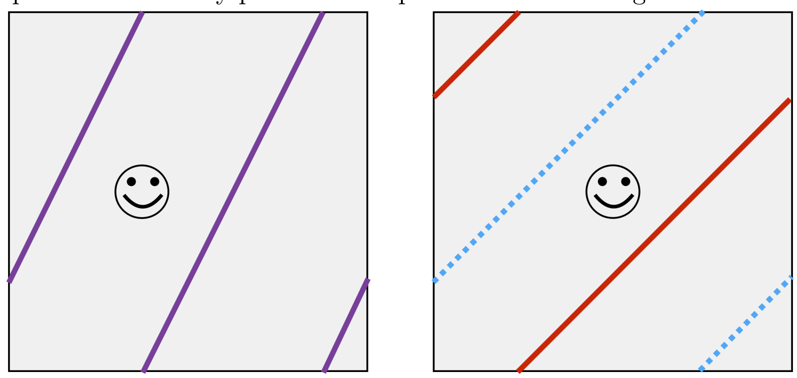
DD

1.28. 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?



1.8 Hands-on activities for Chapter 1

1.29. 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.



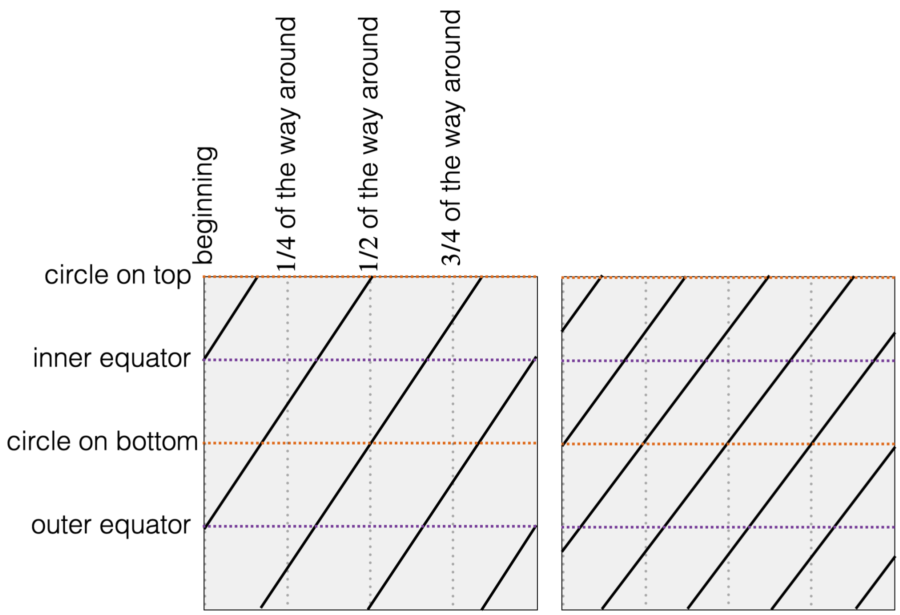
1.30. *Cutting a bagel into two linked rings.* **You will need: bagel with a large hold in the middle; serrated knife; tray to catch the crumbs.**
1. Draw the red and blue trajectories on your bagel.
2. Cut the bagel: The pointy end of the knife should follow the red trajectory, while the handle follows 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:
`https://www.georgehart.com/bagel/bagel.html`

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

DD

1.31. You will need: bagel with a large hole in it, marker. Choose a periodic trajectory, 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!



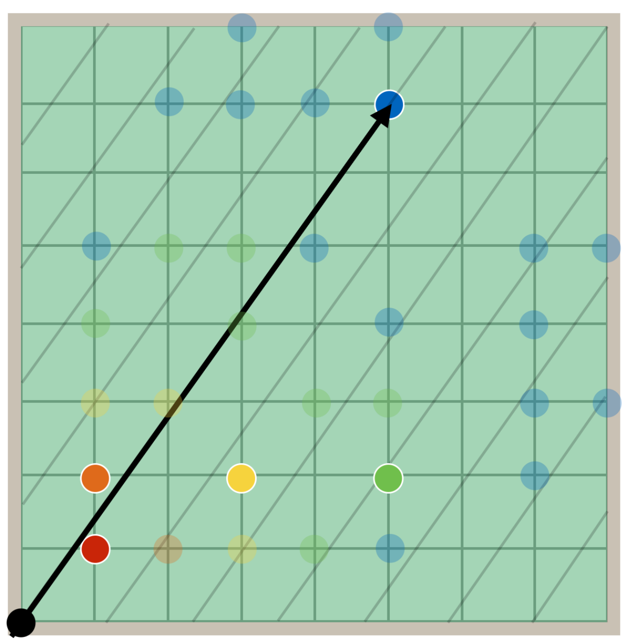
Connection to knot theory: Imagine that the bagel disappears, and all that is left is the trajectory corresponding to slope p/q, now made out of a piece of string. It turns out that if p and q are relatively prime, then you get a *knot* – a knotted-up loop that you can't untangle into a circle. The trajectory 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.

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.

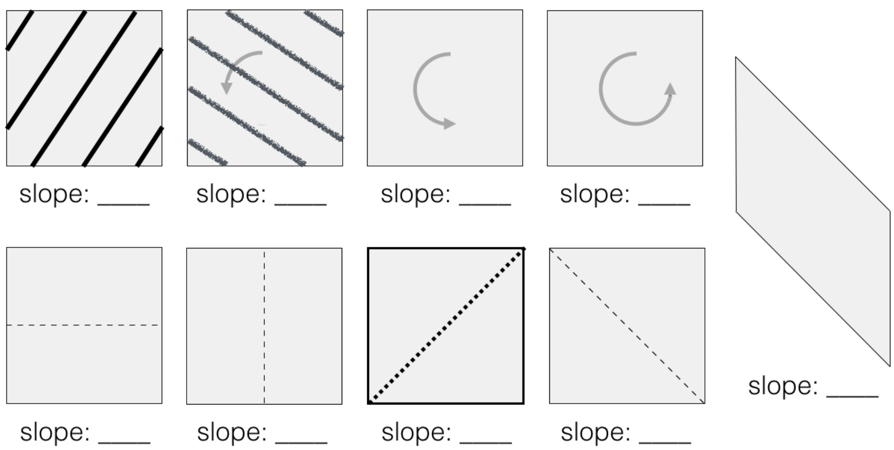


Thanks to Jaden Sides for the idea behind this picture.

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!

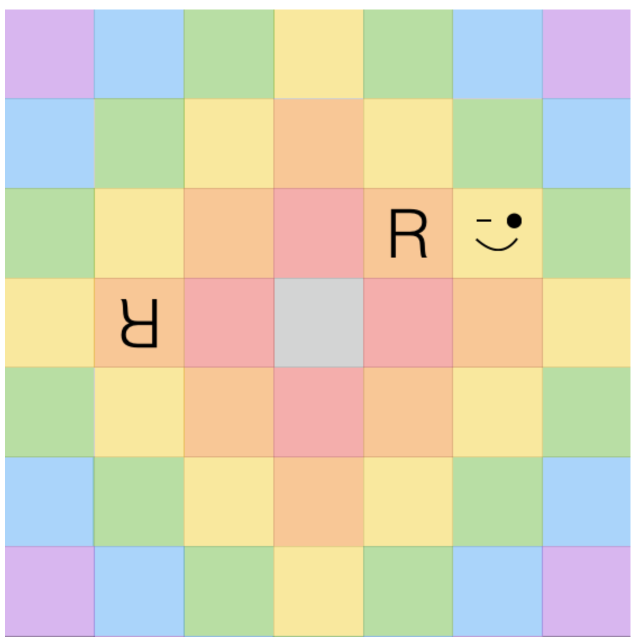
DD

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.

They did the math: 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 in the Cutting Sequence Characterization Theorem (Problem 3.6) that the same is true for cutting sequences on the square.

DD

2.4. In Problem 1.21, 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 end up on orange squares, and the winky faces 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 in the picture.

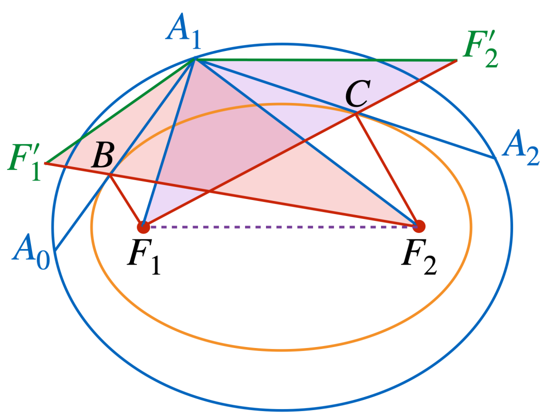
(a) (blue) Consider the billiard trajectory $A_0A_1A_2$ in the larger ellipse E shown in the figure. Explain why $\angle A_0A_1F_1 = \angle A_2A_1F_2$.

(b) (green) 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 $\angle A_2A_1F_2' = \angle A_2A_1F_2$.

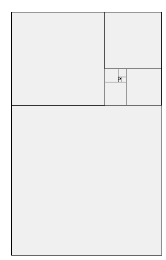
(c) Show that $\Delta F_1'A_1F_2$ and $\Delta F_1A_1F_2'$ are congruent.

(d) (red) 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.

They did the math: 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.

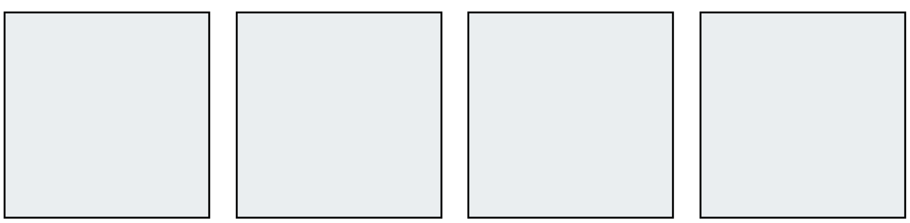
A much more efficient way to write down the effect of the automorphism is to record how it affects the *cutting sequence* corresponding to a trajectory. Then we could act on the cutting sequence – an operation on symbols, not on pictures! – and get the cutting sequence corresponding to the transformed trajectory. Irene Pasquinelli (left) figured out how to do this for a large class of surfaces, in her master's thesis. We'll do this 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.

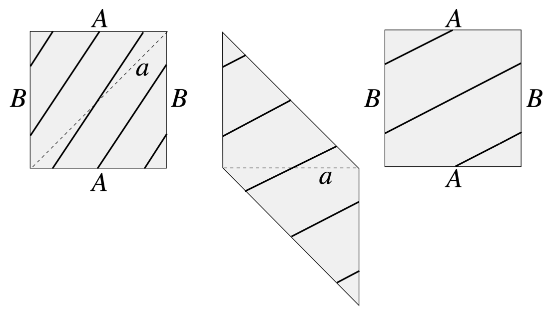
[1] τ is spelled *tau* and rhymes with "cow."

2.3 The dream comes true

In these problems, we will determine the effect of the shearing automorphism from Problem 1.28 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 (in Problems 3.11 and 3.16) 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 $\overline{BAA}$:

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

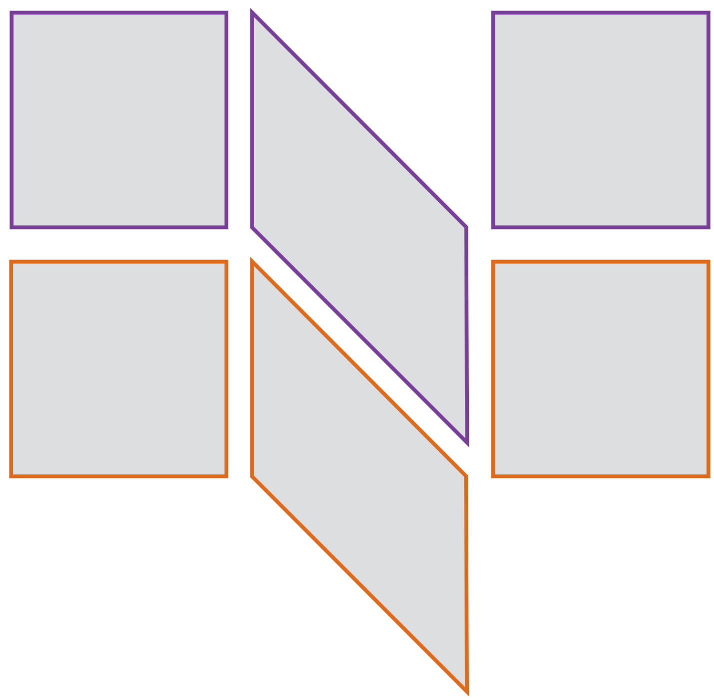
Explain.[2]

[2] The idea of augmented cutting sequences described here comes from John Smillie and Corinna Ulcigrai's paper `https://arxiv.org/abs/0905.0871`; see their § 1.2 and Figure 3.

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 1.14 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.

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 $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ 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 $\left[\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right]$ 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 $2(p+q)$.

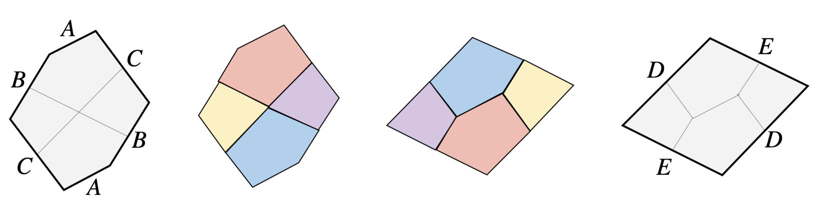
DD

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

DD

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

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.

They did the math: 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 and died in 2017.

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?

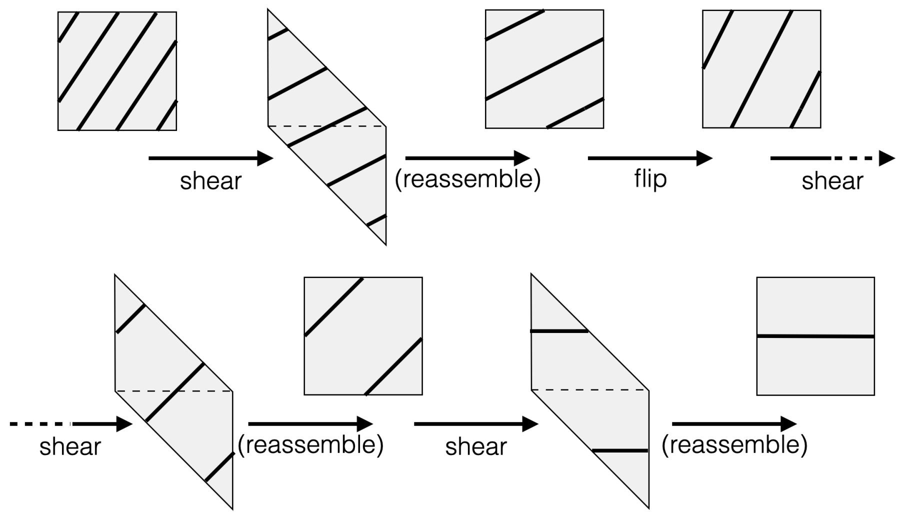
2.5 A grand unifying theory emerges

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 $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$.
2. If the slope is between 0 and 1, apply the flip $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.
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.

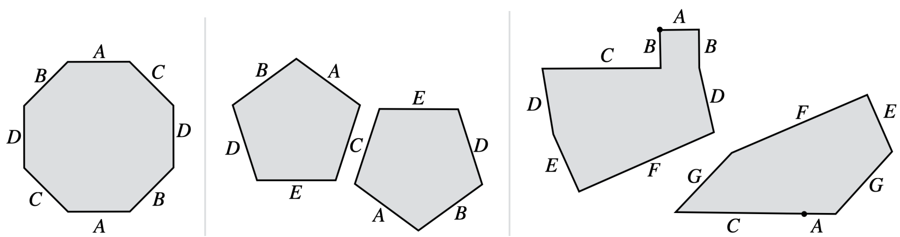
DD

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[3] from 26 to 29 minutes, which shows how to wrap the flat octagon surface (far left) into a curvy surface embedded 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?

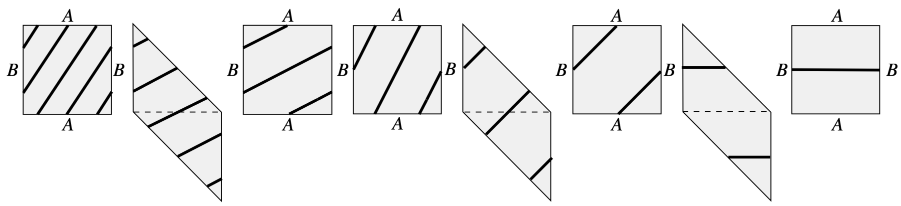
They did the math: Some people love translation surfaces, and other people *really* love translation surfaces. Jayadev Athreya (left, with the author in Marseille in 2015) has made many contributions to the field, but his most unique contribution just might be having a double pentagon tattooed on his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

[3] 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 simplifies a trajectory on the square torus with slope ≥ 1, by un-twisting it step by step, until it is 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.



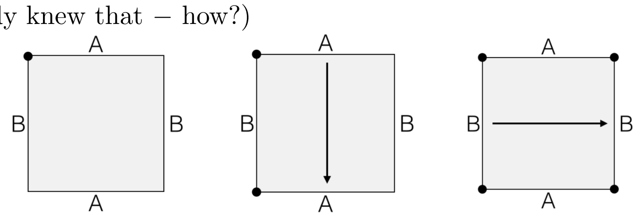
DD

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

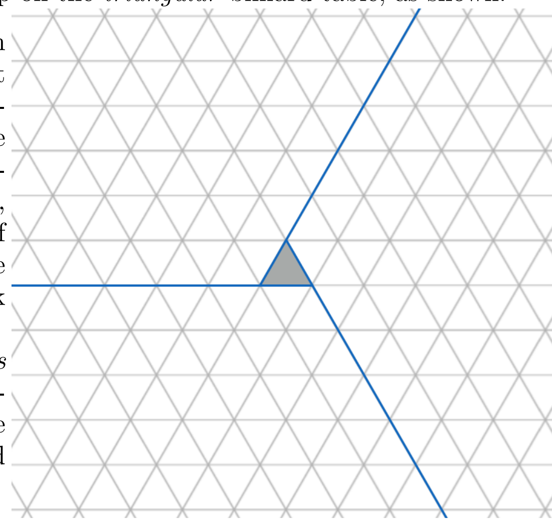
(a) a hexagon with opposite parallel edges identified (Problem 2.20);

(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 counterclockwise 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.

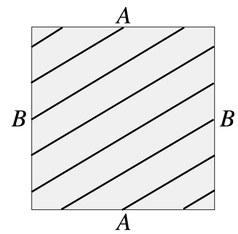
2.31. Let P be a convex quadrilateral that has a 4-periodic inner billiard trajectory that reflects consecutively in all four sides. Prove that P is cyclic: there is a circle containing all four of its vertices.

THEY DID THE MATH: Is the converse true – given a cyclic quadrilateral, must it have a 4-periodic inner billiard trajectory that reflects consecutively in all four sides? Katherine Knox, a 7$^{\text{th}}$-grade student participating in the Girls' Angle program in Boston, proved that the converse holds only when the quadrilateral contains its circumscribing circle's center. (American Mathematical Monthly article, to appear.)

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).

They did the math: 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):

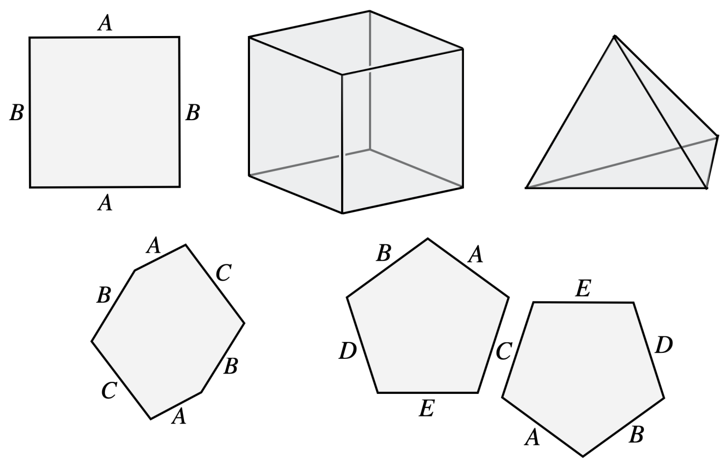
Given a surface S made by identifying edges of polygons, with V vertices, E edges, and F faces, its Euler characteristic $\chi(S)$ is[4]

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

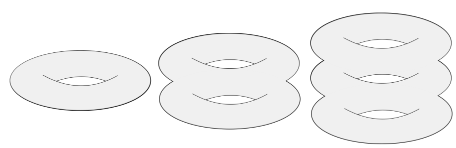
Note that a "face" must be a *simply connected* polygon, without holes.

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.

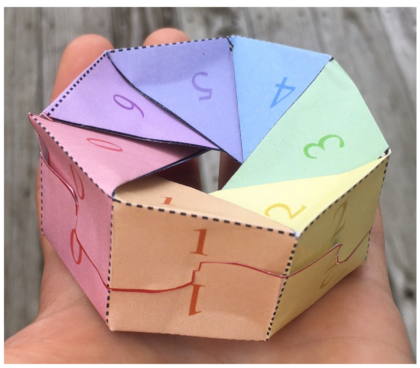
[4] *Euler* is pronounced "oiler." χ is spelled *chi* and is pronounced "kye."

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.

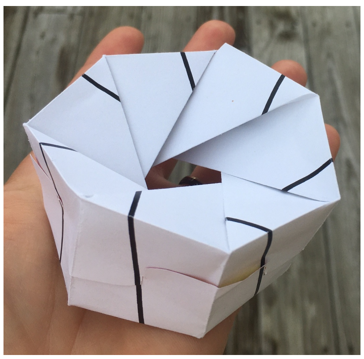


They did the math: The layout for this object was designed by Pierre Arnoux, (§ 4.7) Samuel Lelièvre (§ 4.5) and Alba Málaga Sabogal (above). They call it a *diplotorus*. The idea has been around for a while, and the path it took to get to them was a long and winding road:

- 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.
- In 2020 Samuel 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*!

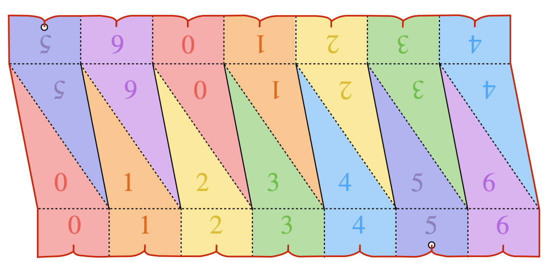
2.39. You will need: scissors, perseverance. Cut out the diplotorus layout on the next page, cutting on all of the red lines and curves. Then crease it along the indicated lines: the dotted lines should be mountain folds, and the solid lines valley folds. (*Hint*: Spend a long time making very strong creases on all of the lines. If you have good strong creases everywhere, putting the model together will be doable; if your creases are weak or inaccurate, it will be almost impossible for you to put the model together.) Then bring the edges with the same numbers and colors together, and attach the flaps via the red slits. The two white disks should coincide at the same point.

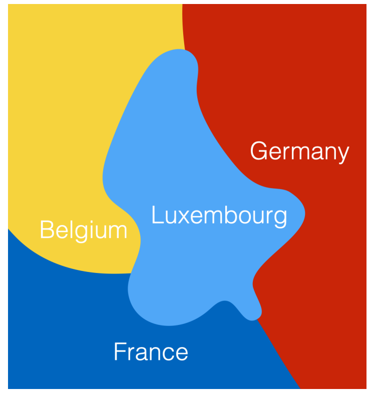
Behold, a torus that encloses positive volume, and is *flat* (2π of angle around each point) everywhere!



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

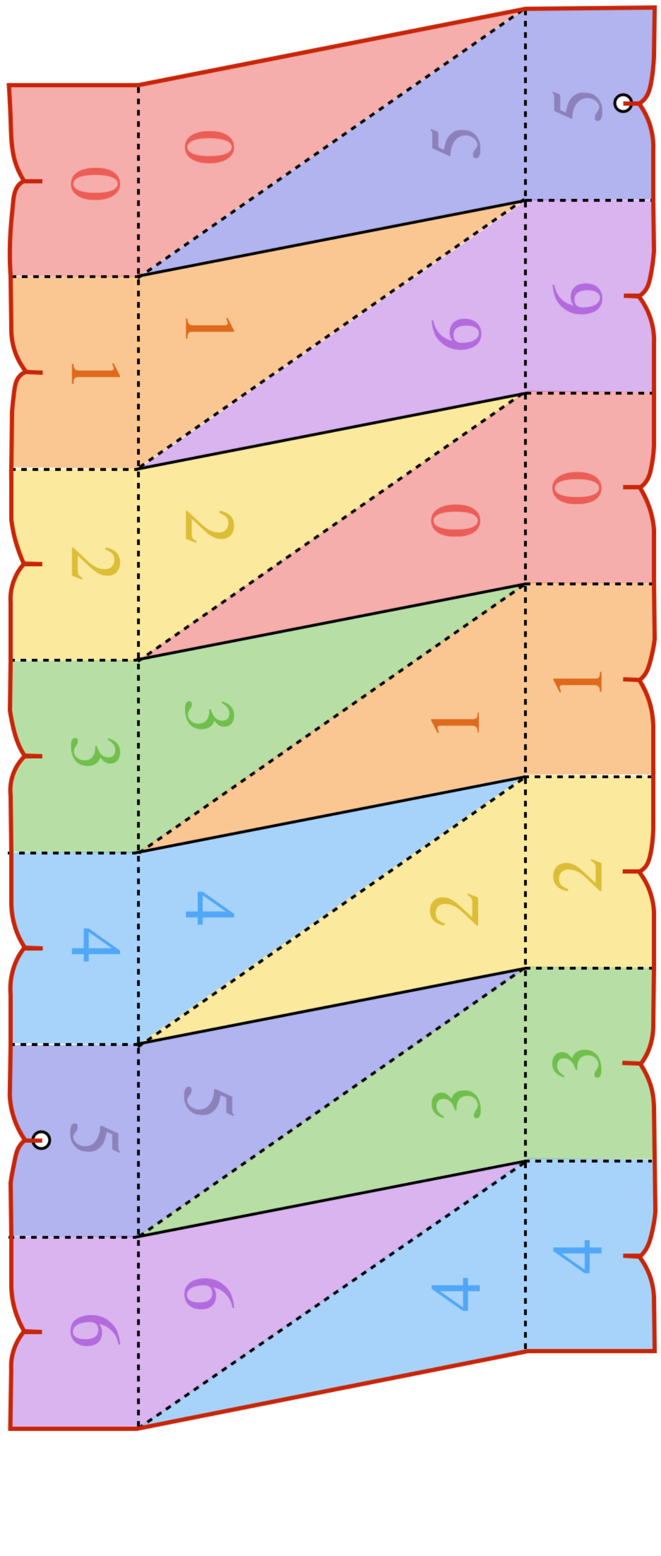
2.41. The diplotorus layout looks roughly like a parallelogram, and when you fold it up, you bring the short edges together, and you bring the long edges together. But you do not identify the numbered edges directly across; there is a *twist*. On the picture below, show how to cut and paste the diplotorus translation surface into a parallelogram whose opposite parallel edges are identified.





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. It is possible for 4 regions to all border each other (e.g. Luxembourg, Belgium, France, Germany), but not 5. For the torus, the minimum number of colors is 7; the coloring on the torus on the next page is such an example of 7 mutually adjacent regions. Looking at the flat layout, you can check that the red region 0 touches regions $1, 2, 3, 4, 5, 6$, and the same for each of the other colors.

Thanks to Moira Chas for suggesting this line of inquiry to Samuel, who suggested it to me.

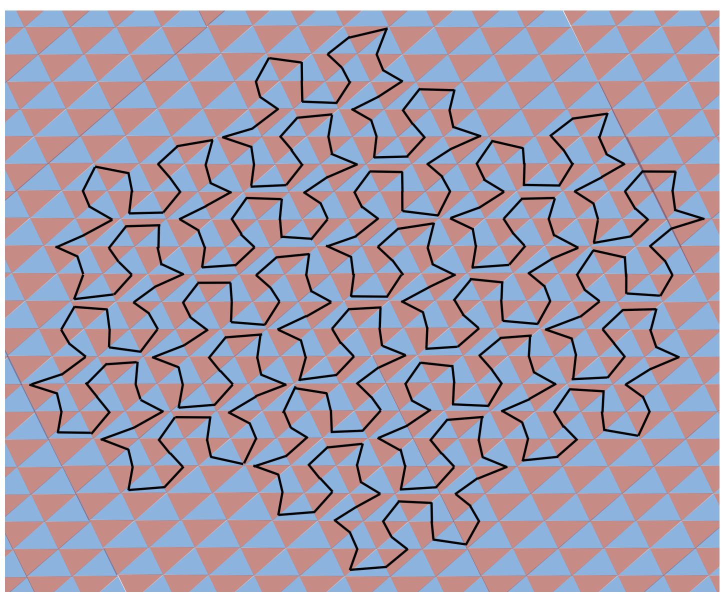


3

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.

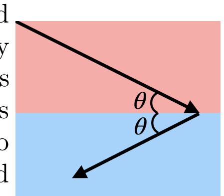
In Chapter 3, we will use the tools and insights we gained in our study of the square, as we more further afield to many different types of billiards. We will do billiards on non-square tables, and we will meet yet another kind of billiards. For billiards on non-square tables, the situation is often "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!



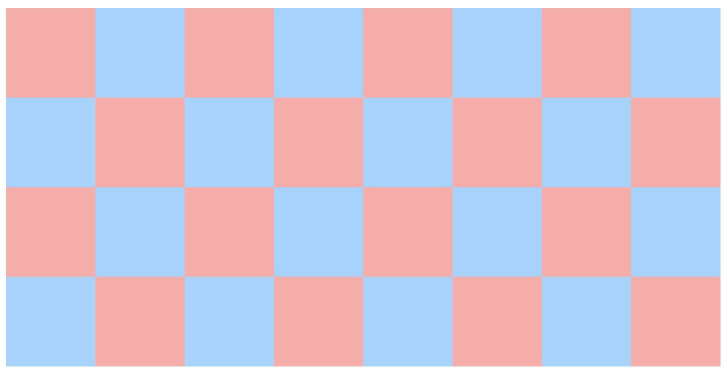
3.1 We meet tiling billiards

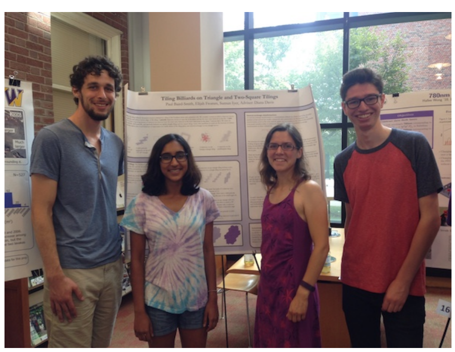
Now that we have explored the simplest case of classical billiards – inner billiards on the square – in detail, and understood it deeply, we will expand our view to other types of billiards.

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.[1]



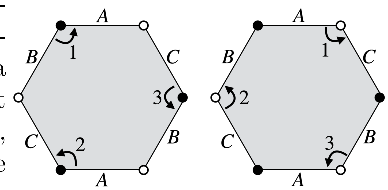


They did the math: Tiling billiards is motivated by the existence of *metamaterials*, solids that have a negative index of refraction. Typical materials such as water and glass have a positive index of refraction; you have likely worked with these in physics, with *Snell's Law*. The idea here is to create a two-colorable tiling out of materials with opposite indices of refraction, and see what happens as a laser beam refracts around it. The results about tiling billiards

[1] For a beautifully artistic dynamic rendering of tiling billiards, and a preview of § 5.3, see the video "Refraction Tilings" on YouTube: `https://www.youtube.com/watch?v=t1r1c01V35I`.

that we will explore in this chapter come from the work of three undergraduate students: Elijah Fromm, Sumun Iyer, and Paul Baird-Smith (shown at their poster session in 2016, with the author).

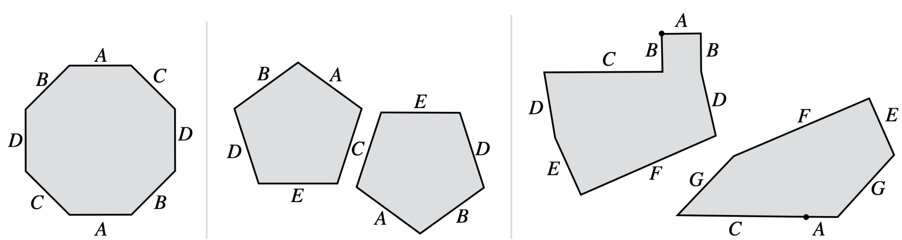
Walking around a vertex. We can determine the angle around a vertex by "walking around" it, as shown in the figure for a hexagon surface. The left picture shows that the angle around the black vertex is $3 \cdot \frac{2\pi}{3}$, and the right picture shows the same for the white vertex.



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.



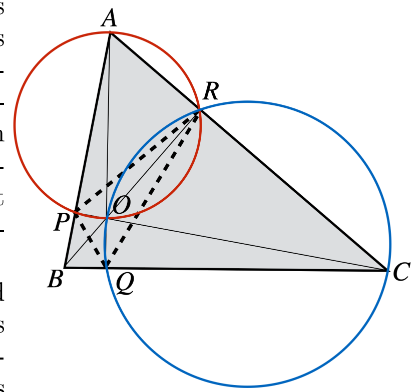
DD

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.

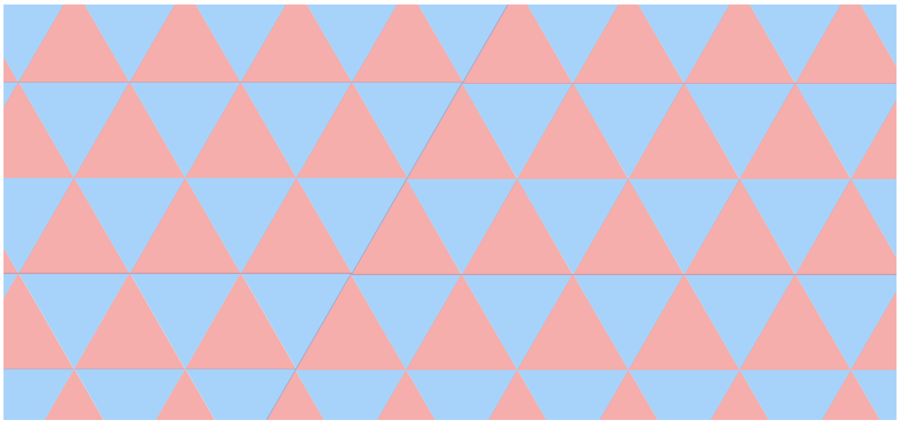
DD

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 Earlier, we unfolded; now, we fold

DD

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?

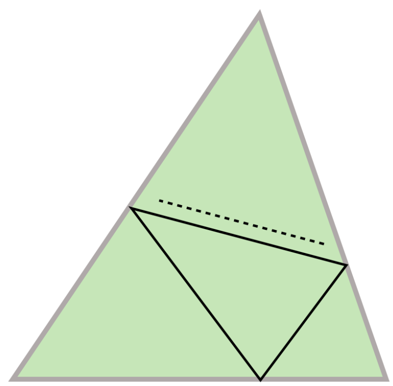


In billiards on the square, we *unfolded* a billiard trajectory into a line on the square grid, and onto a linear trajectory on the square torus. In an analogous way, *folding* is a powerful technique for understanding tiling billiards trajectories:

DD

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?

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.

THEY DID THE MATH: 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:

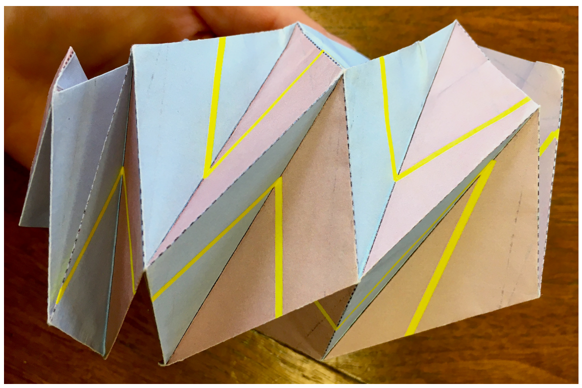
- If the slope is greater than 1, apply a vertical shear, and
- if the slope is less than 1, apply a *horizontal* shear.

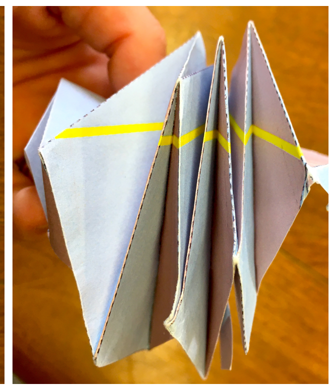
Explain.

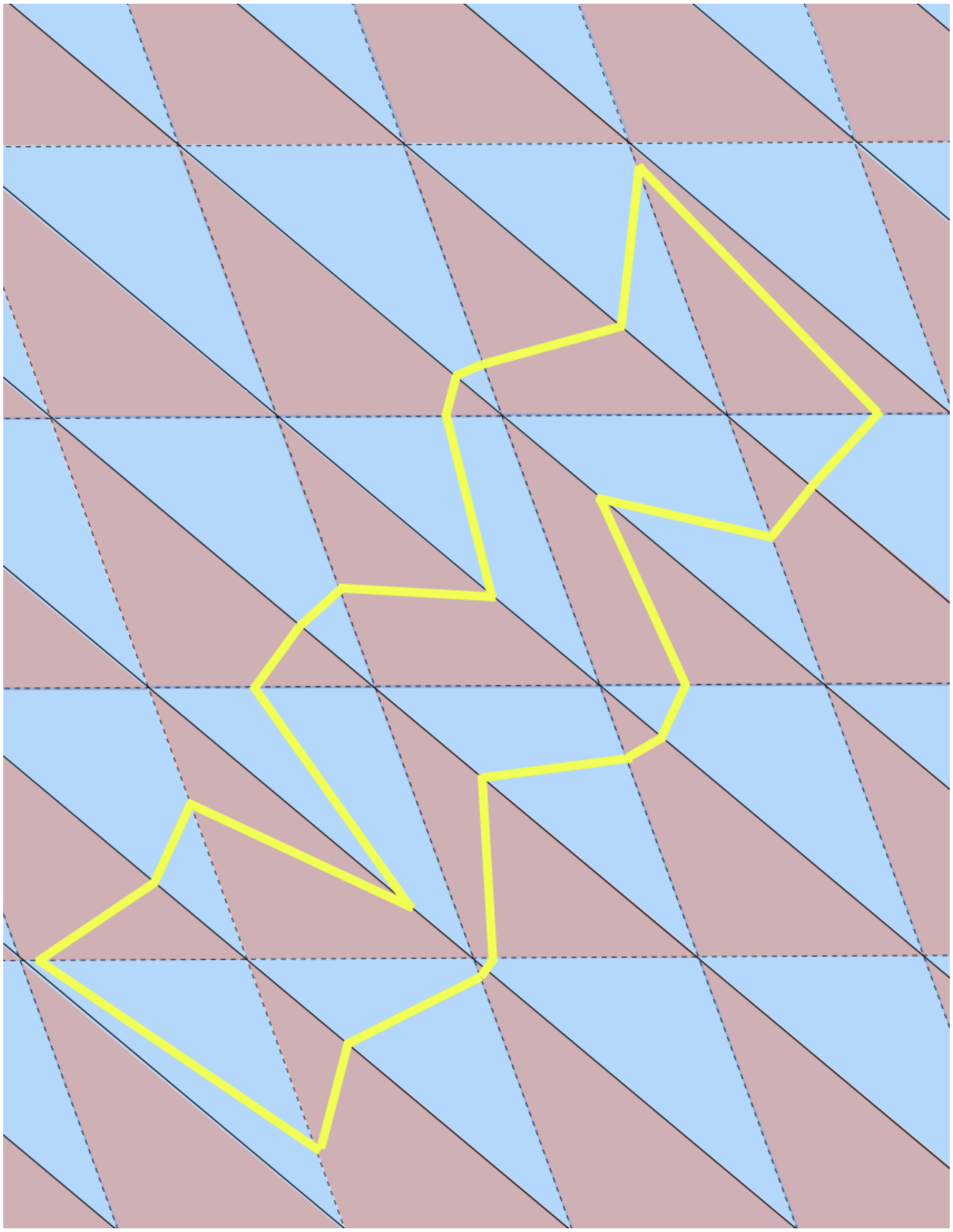
DD

3.12. 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*: Spend a long time making very strong creases on all of the folds. If you have good strong creases everywhere, getting this thing to fold flat will be doable; if your creases are weak or inaccurate, it will be more difficult for you to make it happen.

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.



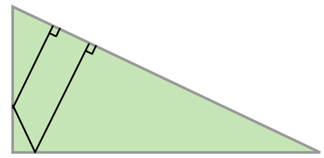




3.3 We meet the biggest open problem in billiards

DD

3.13. *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.)

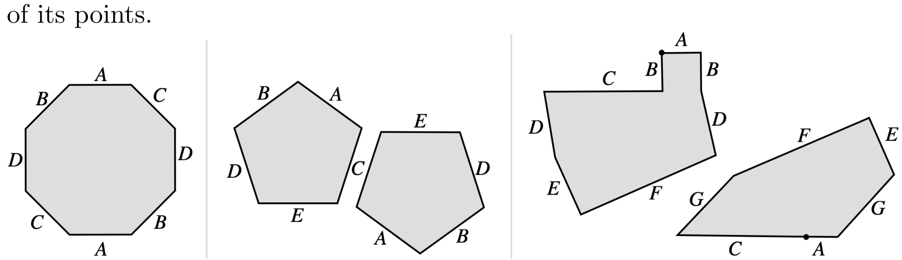
They did the math: The biggest open problem in billiards is: *does every triangular billiard table have a periodic trajectory?* The Fagnano trajectory shows that every *acute* triangle has a periodic billiard trajectory, and the construction above shows that every *right* triangle has one.

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 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 of a given type. 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.

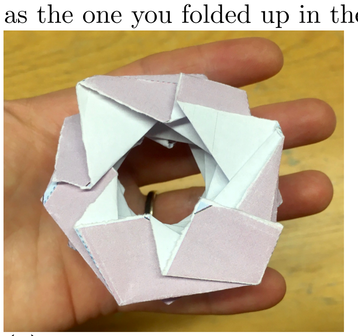
(Recall that in Problem 3.3, we proved that the angle at a cone point of a translation surface is always a multiple of 2π.) We say that the double pentagon surface is in the stratum $\mathcal{H}(2)$ because it has one cone point, with two extra multiples of 2π around it: 6π total, so $2 \cdot 2\pi$ extra. 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.



3.14. For each of the remaining surfaces above, identify which stratum it belongs to. Then come up with an example of a surface in $\mathcal{H}(1,1)$.

Note that a vertex 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.

3.15. You will need: your folded triangles from Problem 3.12. 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 Problem 3.12.



(a) Given two adjacent triangles in the tiling, prove that, if you fold along their shared edge, the circumcenters of the triangles coincide, and thus the two 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, in the folded state, all the pieces of trajectory are contained in a single chord of the circumscribing circle.

DD

3.16. 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):
(1) There are three types of automorphisms: rotations, reflections and shears. We understood the effects of rotations and reflections in Problems __________.
(2) 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 ______.
(3) 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

3.17. (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, given the matrix $\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}$, we can decompose it as

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

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 $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$.

3.4 Families of parallel trajectories

DD

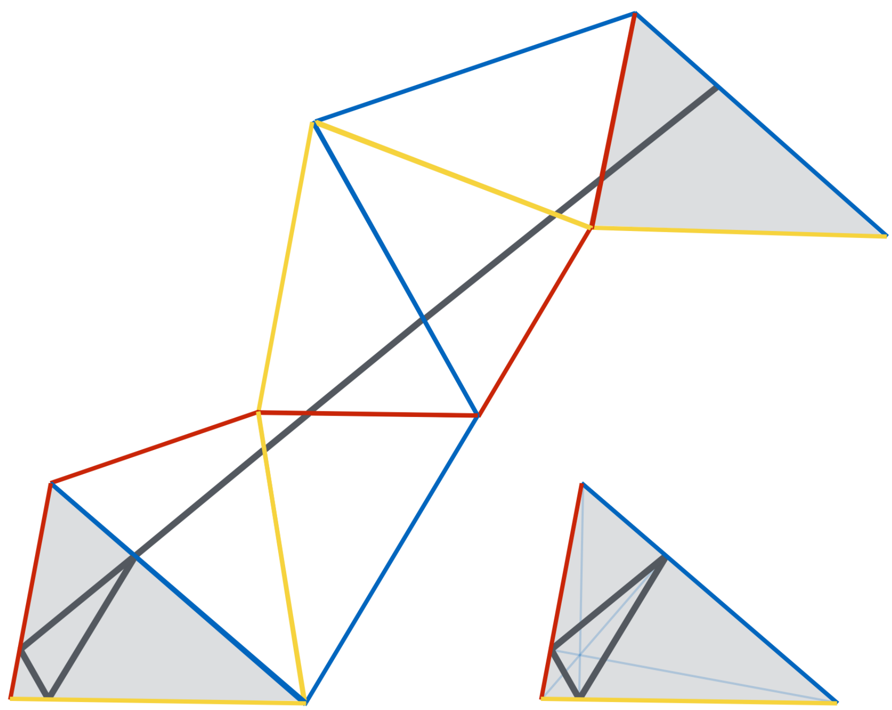
3.18. The figure below shows the Fagnano trajectory in the 40-60-80 triangle. In Problem 3.9 we showed that there are nearby parallel billiard trajectories of period 6.

(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.

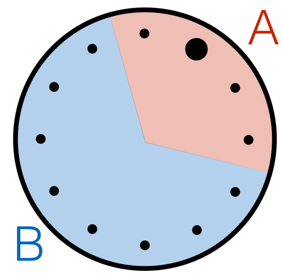
(c) Sketch one of your period-6 trajectories in the copy of the triangle that is in the lower left of the picture. Then draw the "unfolding" of your trajectory. The unfolding of the Fagnano trajectory is given.

(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

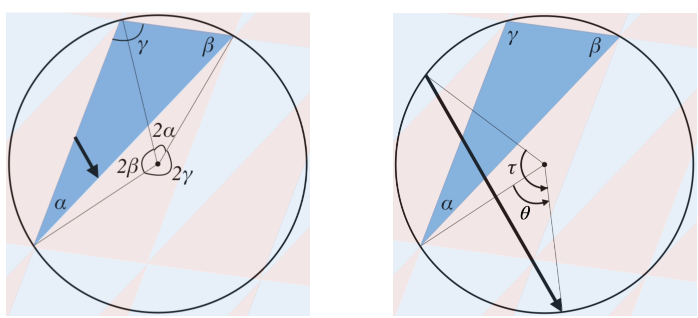
3.19. Consider a circle broken into a red arc and a blue arc, taking up 1/3 and 2/3 of the circle respectively, as shown. The game is to start with any point on the circle, repeatedly rotate it by a 1/3 turn, and each time note down which part of the circle it lands in – say, an A if it lands in the 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.



The amazing result of this chapter 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 θ 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.



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

DD

3.21. 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

3.22. The *Gauss-Bonnet Theorem* says that the total (Gaussian) curvature K of a closed surface S is[2]

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

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.

They did the math: Bill Thurston (left, with the author at the Cornell Topology Festival in 2012) was hugely influential in 20^{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.

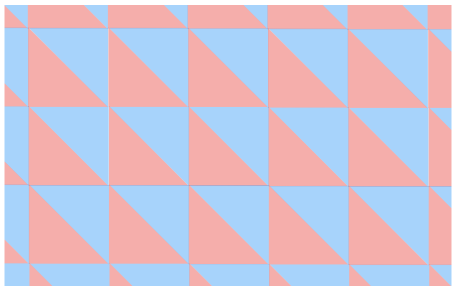
Bill received the Fields Medal in 1982. One of his later projects was working to smooth out the angle defect in polyhedra.

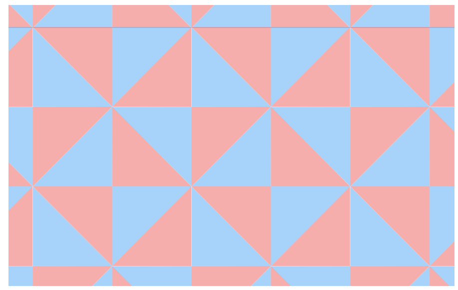
[2] If you don't know calculus, just skip part **(a)**. We'll only use part **(b)**.

3.5 A cone point with angle 6π

DD

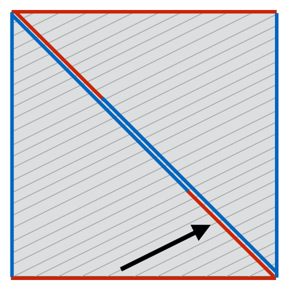
3.23. The pictures below show two different tilings of the plane by isosceles right triangles. Consider tiling billiards on each of them.





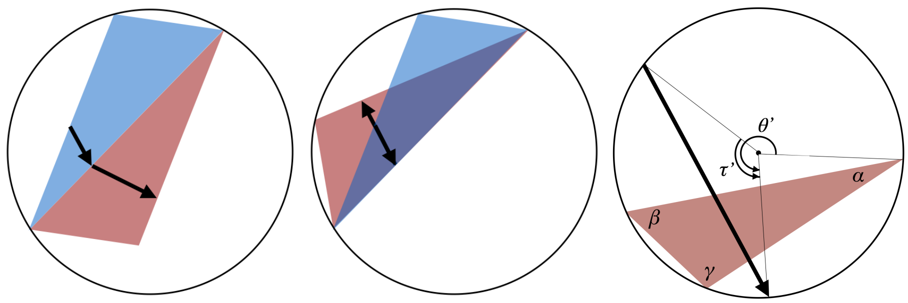
(a) For each tiling, consider: are there periodic trajectories on the tiling? If so, explain how to construct one and sketch it; if not, prove that periodic trajectories cannot occur.
(b) Are there escaping trajectories on the tiling? If so, explain how to construct one and sketch it; if not, prove that 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.

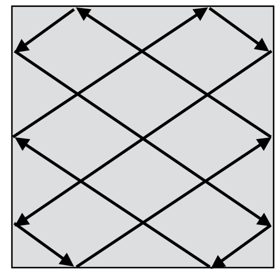
In the previous section, we defined some notation for tiling billiards trajectories in a triangle tiling, with their circumscribing circles. Consider two consecutive triangles that the trajectory crosses (see the pictures below). When the second triangle is folded onto the first along the shared edge, the segments of trajectory in each triangle align, with opposite orientations (Problem 3.8, and shown below).



3.25. 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 $\theta' = \theta$? Explain why or why not.

They did the math: 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 system 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 (§ 4.7), Jean-Christophe came up with the Arnoux-Yoccoz interval exchange transformation, which has particularly interesting properties and led to much further research. Jean-Christophe won the Fields Medal in 1994, and died in 2016. He is pictured with the author in 2014 at a conference at Oberwolfach.

3.26. 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.



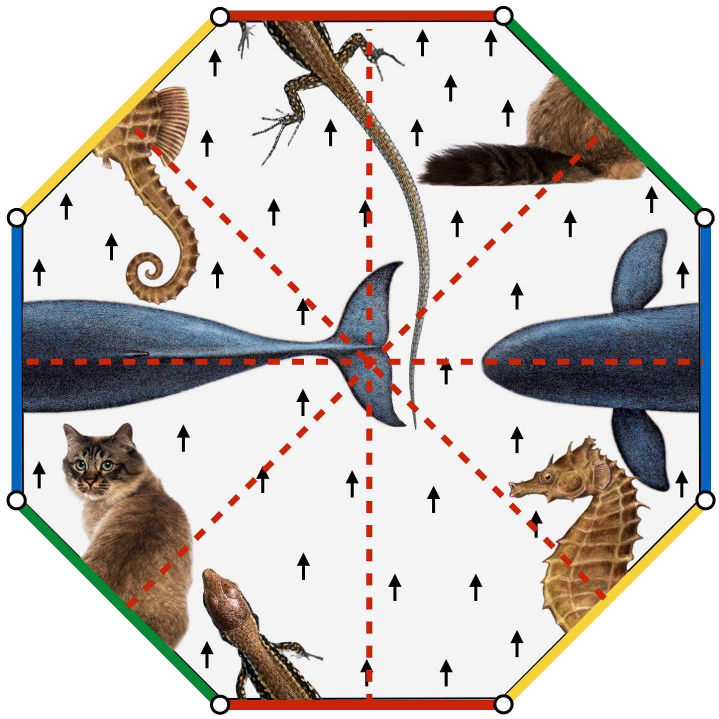
Based on the above, we say that there is a *1-parameter family* of billiard trajectories in a given direction on the square table. The idea is that, once you've chosen the direction, the only other thing left to choose is the starting point – say, along the bottom horizontal edge. There is only one dimension,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

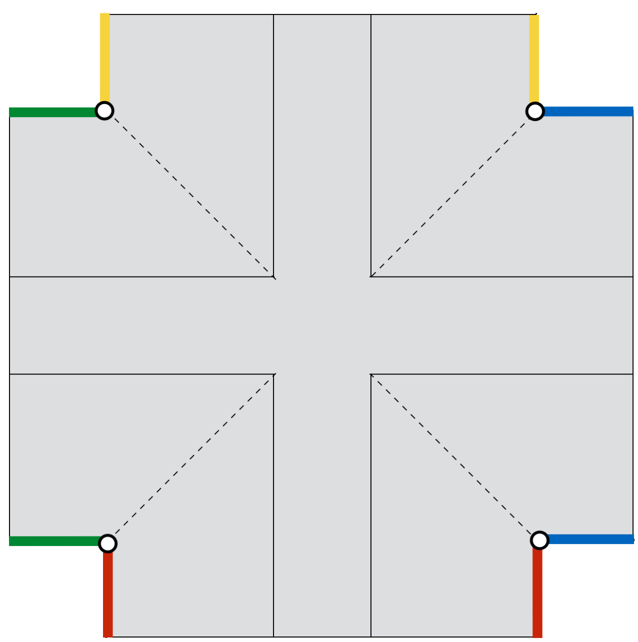
DD

3.27. You will need: scissors, tape. 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?

If your birthday is in the first half of the year, do **(a)**; otherwise, do **(b)**:
(a) Cut out the octagon on the next page. As usual, we wish to identify parallel edges of the same color. To make this happen, first tear along the dashed lines. Then tape the pieces together along edges with the same color, using the animal pictures to remember what is glued to what, and keeping the arrows pointing in the same direction to keep a consistent orientation. Notice that the angle at the white point is 6π! Discuss.
(b) Cut out the cross on the next page. Fold along the indicated edges: solid lines are "valley folds" and dashed lines are "mountain folds." Tape the same-color edges together. Notice that the angle at the white point is 6π! Discuss.

The latter construction comes from Florent Tallerie,. This is a small piece of his layout for a genus-2 flat surface that you can make out of folded paper. (Work in progress, to appear.)





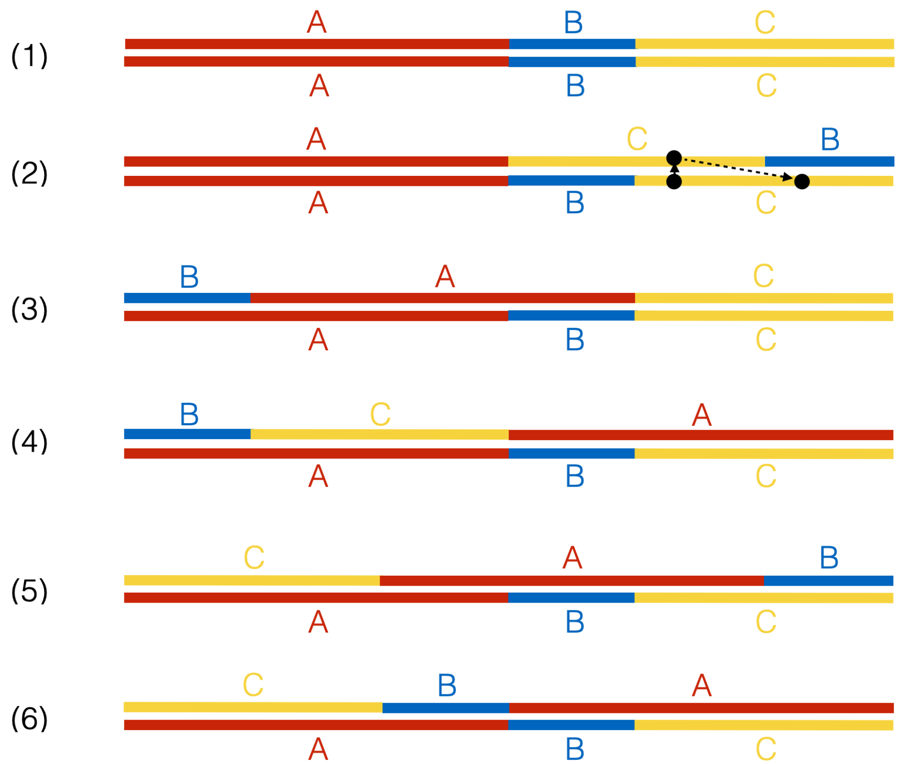
3.6 Interval exchange transformations

3.28. 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* (IET). Let's use the convention, based on that experience, that a point flows directly up in the IET, then shifts over to come down. An example is shown. DD

(a) Explain why the dynamics of the 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. Is this still the case when the interval lengths are irrational?

(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 *irredicible* 3-IET. Which one?



3.29. For the 3-IET that you identified in the previous problem: DD

(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 $1/2, 1/6, 1/3$. Show that the orbit of *every* point is periodic.

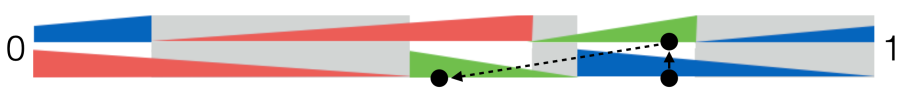
THEY DID THE MATH: Just about everything I (the author) know about interval exchange transformations, I learned from Vincent Delecroix (left). I was studying tiling billiards with the group pictured in § 3.1, and (spoiler alert) we had figured out that tiling billiards on triangle tilings were equivalent 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, such as Rauzy diagrams (see § 5.7) that were essential for studying these IETs. This illustrates a key principle, which is that many of the essential ideas in mathematics 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 software related to exploring translation surfaces, IETs, and other aspects of dynamical systems.

DD

3.30. So far, we have been studying IETs where you chop up an interval and then rearrange the pieces. Now, suppose that you chop up an interval, *flip each piece*, and then rearrange the pieces. The IET below shows the interval $[0, 1]$ chopped into three pieces (colored red, green and blue).

Each piece is flipped; this is indicated in the picture by drawing each interval as a triangle, so that you can tell which part of it is which. Then the pieces are reassembled. An IET where every interval is flipped is called a *fully flipped* IET. An example of the image of one point is shown.

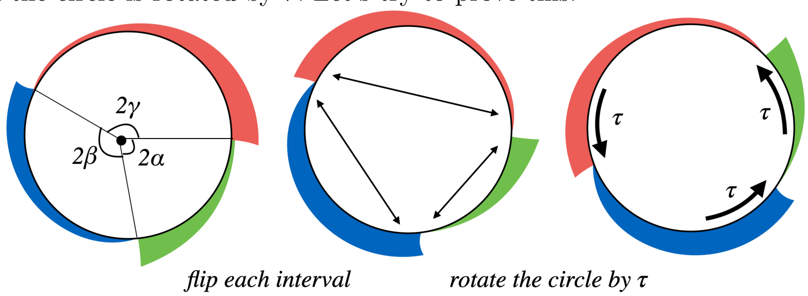


In this picture, we are thinking of 0 and 1 as being equivalent, just like $0 = 2\pi$ on a circle, so that the blue interval is not truly chopped in two, but is just overlapping the break point. So really, this is a *circle exchange transformation* (CET).

(a) The image of one point is shown in the figure. Draw its full orbit.

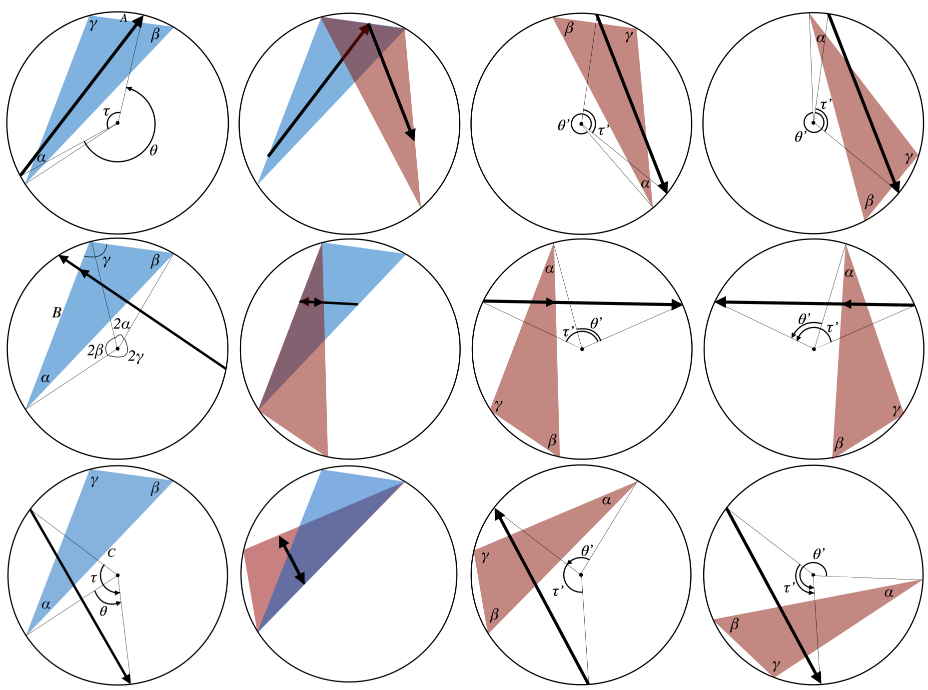
(b) Notice that some of the intervals overlap their images (with a flip, of course). These regions are shaded in grey. Prove that, for each point in an overlapping region, the period of its orbit is 2.

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.



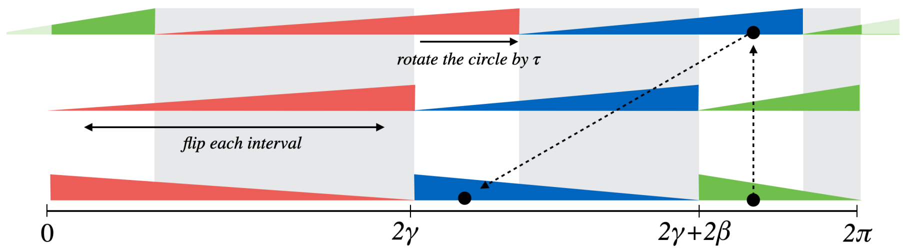
3.31. Refer to the pictures below. Show that, for a trajectory in a triangle tiling with parameters (θ, τ), passing to the next triangle gives new parameters (θ', τ'), where $\tau' = \tau$ (already shown in Problem 3.25), and

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



3.7 We put it all together

In the pictures below, the circles from Problem 3.31 are represented by the interval from 0 to 2π; you have to remember that the two ends are identified. As before, 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. (As usual, start at the bottom and read *up*.)



3.32. Use the result of Problem 3.31 to prove the following:
Theorem (Tiling Billiards IET). Given a triangle tiling with angles α, β, γ, and a trajectory with associated parameters (θ, τ), passing to the next triangle transforms θ 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 the opposite orientation.

  These transformations are all taken modulo 2π.

3.33. Use the previous Theorem (even if you were not able to prove it) to explain why our big result is true: "The orbit of a trajectory under tiling billiards is equivalent to the orbit of a point on an orientation-reversing IET."

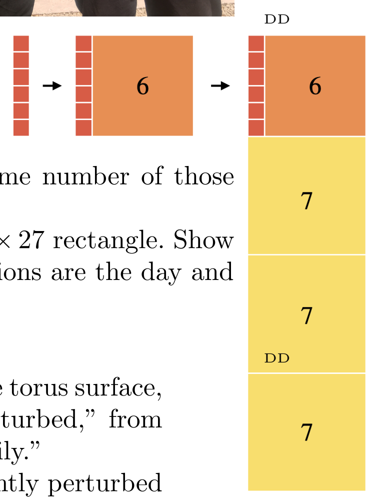
3.34. The parts of the IET where an interval overlaps itself (with a flip) would mean that the same side is hit twice in a row. In the picture above, they are shown shaded in grey.
(a) Explain why hitting the same side twice in a row is impossible.
(b) Show that, in fact, these regions are not in the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

When tiling billiards first came onto the scene, most people were not interested. When the students introduced in § 3.1 showed that the dynamics of tiling billiards are equivalent to an interval exchange transformation, a few more people got interested!

Two of them were Pascal Hubert (in blue shirt, with Nicolas Bédaride at CIRM in Marseille in 2023) and Olga Paris-Romaskevich (§ 5.3), who proved many new results about tiling billiards, including proving the students' conjectures, extending the triangle ideas to quadrilaterals, and applying substantial previous work on IETs by Arnaldo Nogueira.

Many people had been studying the dynamics of IETs, a 1-dimensional system that doesn't have many pictures. Tiling billiards gives a new 2-dimensional representation of that system, which is exciting.

3.35. 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.



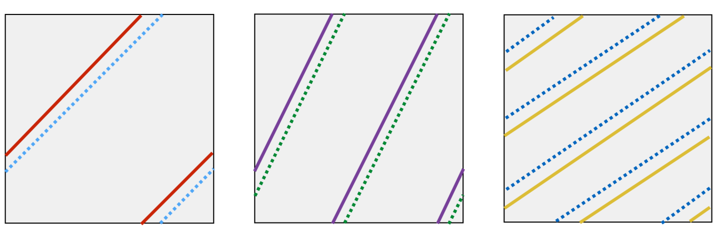
The picture shows how to do this to end up with a 7×27 rectangle. Show how to do this to end up with a rectangle whose dimensions are the day and month of your birth. Does every birthday work?

3.36. Each picture below shows a trajectory on the square torus surface, and a parallel trajectory that is slightly shifted, or "perturbed," from the original. 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.

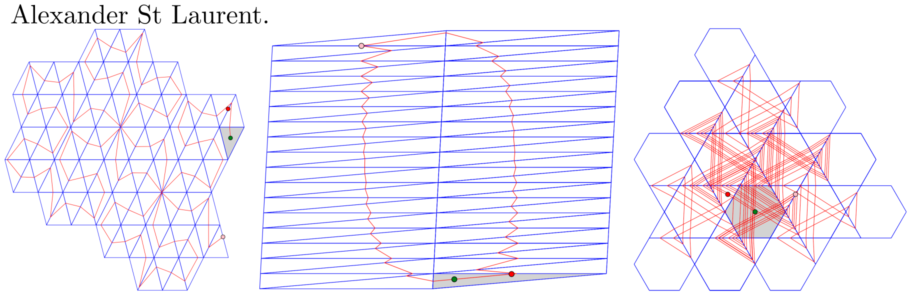
(b) If you perturb a trajectory enough, it will eventually hit a vertex. A "singular trajectory" that hits a vertex on both ends is called a *saddle connection*, and forms the boundary of the family of trajectories. Draw in these boundaries for each of the pictures.

(c) The union of such a family of periodic trajectories is called a *cylinder*. Can you guess why this name was chosen?



3.8 We do some computer experiments

3.37. 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) Move the starting point (green point) and the direction (red point) and see what sort of things you can get.
(b) Click on "Help" at the top and learn how to control the applet with keys.
(c) Click "New" and create a new triangle tiling determined by angles of your choice. Find a really big periodic trajectory. Find a really interesting trajectory. Take a screenshot.
(d) Click on "New" and select some other kind of tiling. 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.)

They did the math: 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 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 stop 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 Delecroix (§ 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.

3.38. Recall that the union of a family of parallel periodic trajectories is called a *cylinder*, and cylinders are separated by *saddle connections* between cone points (Problem 3.36). For a translation surface made from polygons, the set of *cylinder directions* and the set of saddle connection directions coincide. Both cylinders and saddle connections can cross many polygons.
(a) Explain why slopes 2/3 and 5/7 are cylinder directions for the square torus. (Note that the "corner" of the square torus is not a true cone point; we call it a *removable singularity*.)
(b) What are *all* of the cylinder directions for the square torus?

DD

3.39. *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 is 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?
(c) If you keep the starting point and direction the same, and perturb the *table* a little bit so that it is not quite a square, what happens to the trajectory?

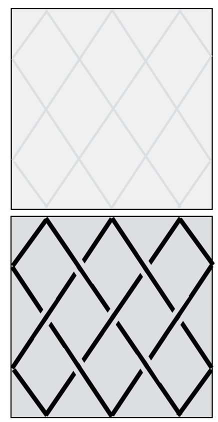
3.9 Hands-on activities for Chapter 3

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!

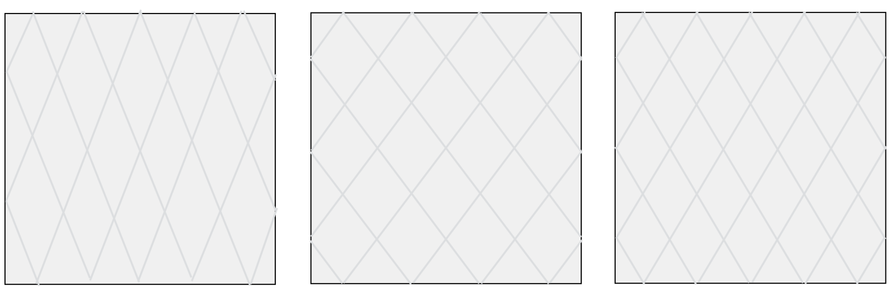
DD

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

DD

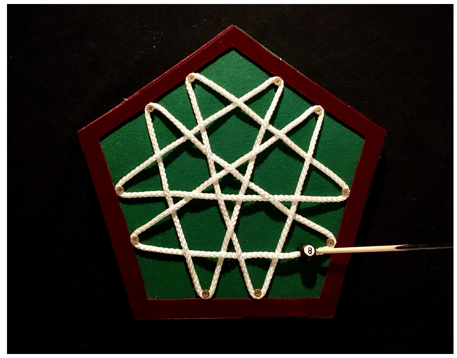


3.41. By now you should be pretty good at drawing diagrams of periodic billiard trajectories on the square. But how to turn it into a *knot*? All Celtic knots are *alternating*, meaning that if you follow a cord along its journey, it alternates over, under, over, under... as it crosses other parts of 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.)

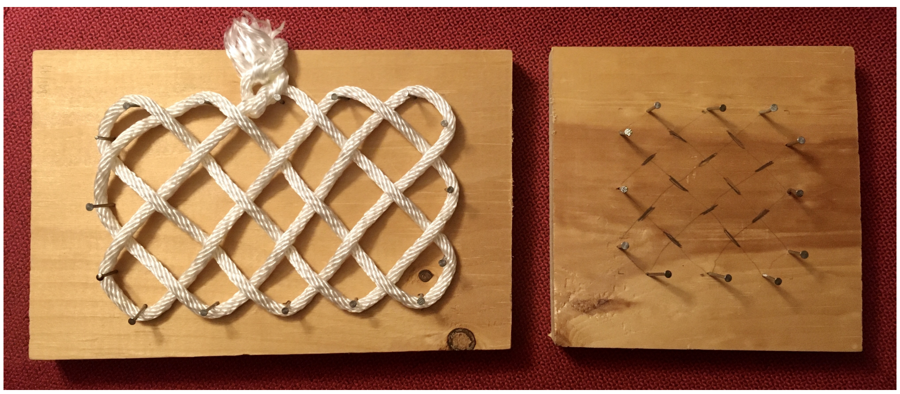


DD

3.42. You will need: rope. Optional: wooden board, hammer, nails. With a rope, create a Celtic knot based on periodic billiard trajectories. Some examples are below.



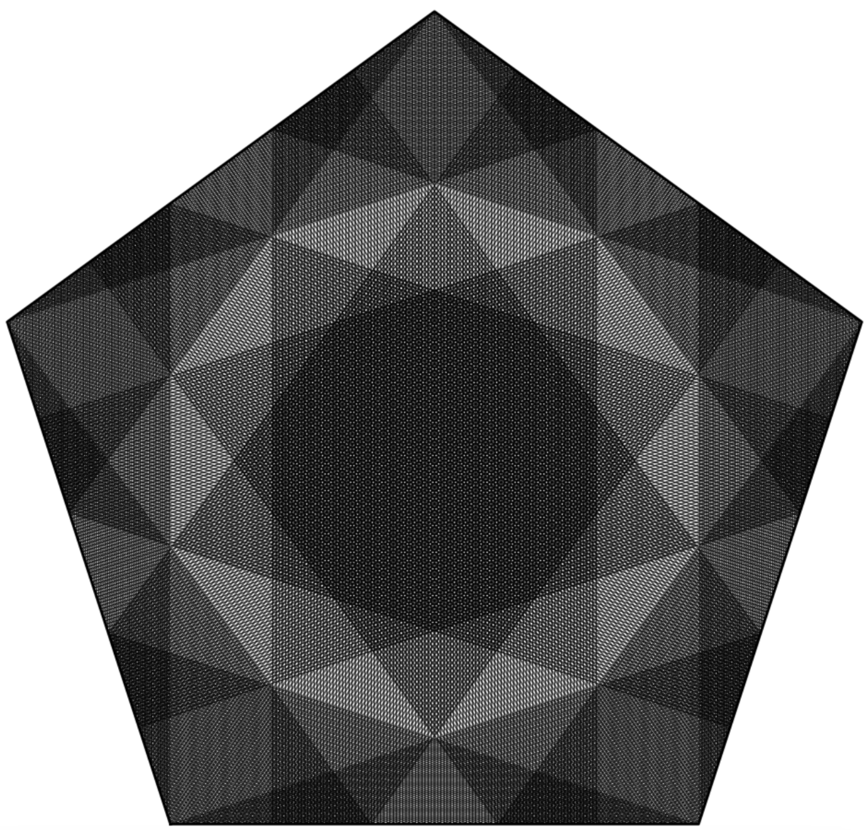
Advice: Draw a picture of the desired knot, including the crossings, to help you avoid errors. (Can you weave an alternating knot *without* looking at a diagram of the proper crossings? I have tried many times, but I have always made at least one mistake.) Creating the knot is is easiest to do if you have a solid frame, such as a board with nails in it as shown here, to hold the cord in place.



4

Cylinders and automorphisms

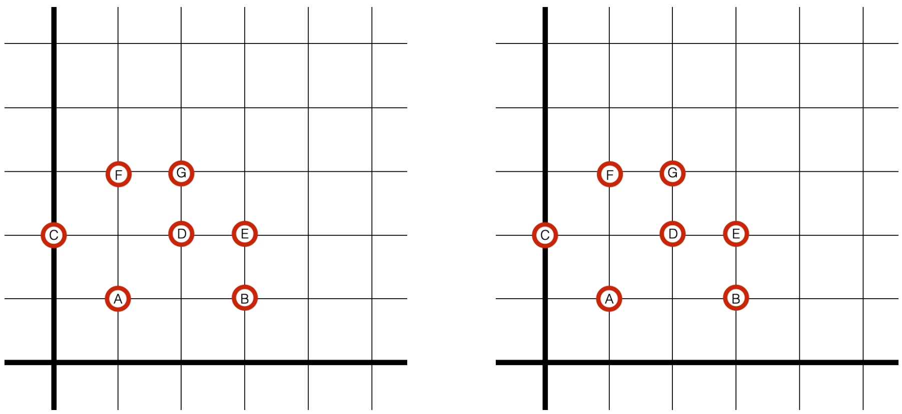
In this chapter we build up the full power of cylinders, so that we can understand *all* periodic trajectories on our surfaces by understanding how automorphisms act on surfaces and their cylinders. We also meet an eclectic menagerie of surfaces designed to do all sorts of interesting things.



4.1 Twisted cylinders

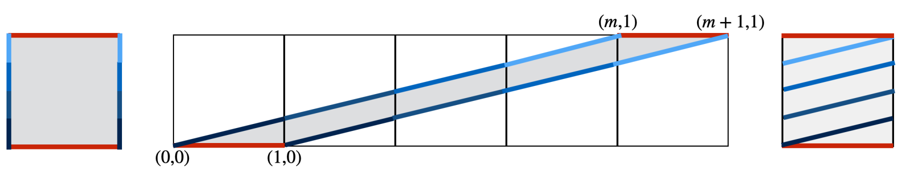
DD

4.1. 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

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.



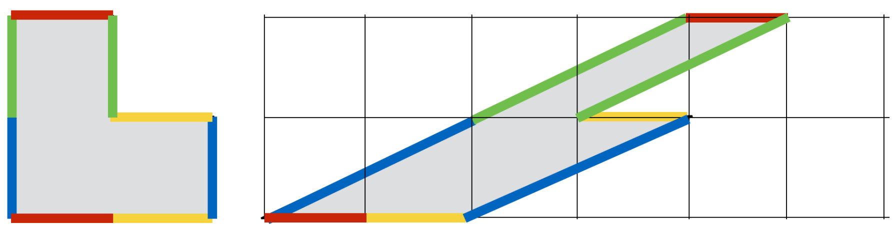
They did the math: Barak Weiss (top left) introduced me to the idea of twisting a cylinder over and over to see what happens. As we will see in Problem 4.25, sometimes something interesting happens.

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

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 that your reassembly respects the edge identifications.



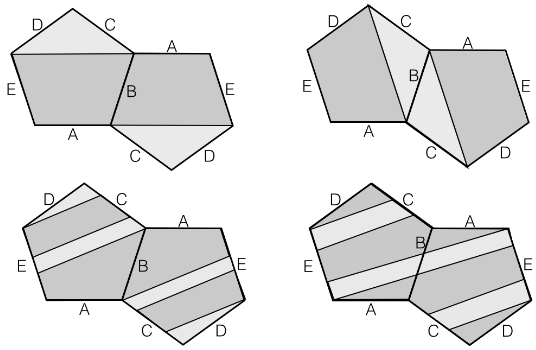
Since we get the same surface back – and thus the sheared version of the L-shaped surface differs from the original only by a cut-and-paste equivalence – we say that $\left[\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right]$ is an *automorphism* of the L-shaped surface.

DD

4.4. *Stability under perturbation, part II*
Consider a trajectory on a square *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 point *and* a direction, as we did on the square billiard table.

DD

4.5. 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.

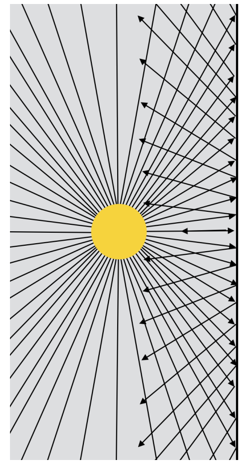


(a) For each set of cylinders shown, consider a trajectory on the surface, in the cylinder direction. Write down the cutting sequence for the trajectory in the light cylinder and for the trajectory in the dark cylinder. Think about similarities and differences with our work on the square torus.
(b) Construct 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?

4.2 Let's get illuminated!

DD

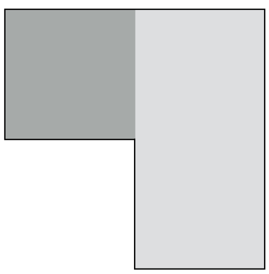
4.6. The picture to the right shows what happens when you put a candle in a room: the light radiates out in every direction. Look closely at the right side of the picture: this room has a *mirror* on the wall, so the rays that hit the wall bounce off, following the billiard reflection law.

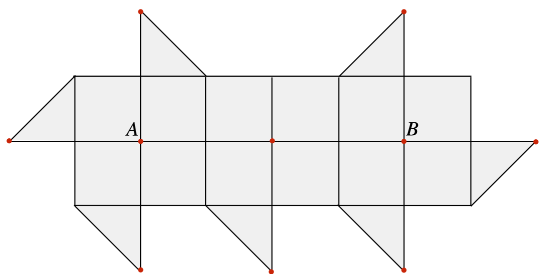


Suppose that you are in a room whose walls are *all* mirrored. You wish to illuminate your entire room with a single candle.

(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.





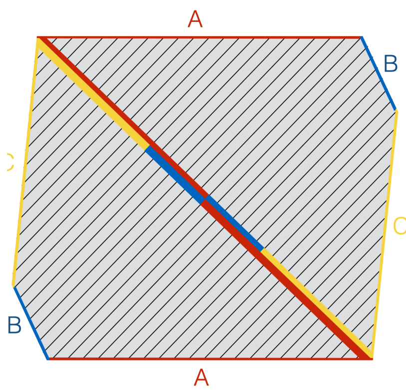
The *illumination problem* asks a generalization of the above: for which shapes of mirrored room can you put a candle *anywhere* in the room, and be sure that the light will reach every point? George Tokarsky constructed an example of a polygonal room, shown above, that contains two points A and B that do not illuminate each other: a candle placed at A will illuminate every point *except* point B, and vice versa. Later, Samuel Lelièvre (§ 4.5), Thierry Monteil and Barak Weiss (§ 4.1) wrote a paper memorably titled "Everything is illuminated" that, in conjunction with a paper by Barak's student Amit Wolecki, shows 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.

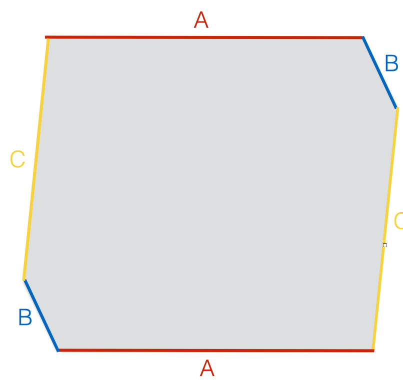
They did the math: To prove their result, Samuel, Thierry and Barak used the "Magic Wand Theorem," the colloquial name for a collection of powerful results from a paper of Alex Eskin, Maryam Mirzakhani (§ 2.4) and Amir Mohammadi. Alex (center) is shown with his wife Anna Smulkowska and mathematician Ursula Hamenstädt.

Alex received the Breakthrough Prize in 2020 for that work.

DD

4.7. 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. Draw in the diagonal that is closest to perpendicular to your trajectories, and use it to sketch the corresponding IET.
(c) Show that the "top" and "bottom" segments on the diagonal corresponding to a given edge (e.g. edge A) are always the same size.

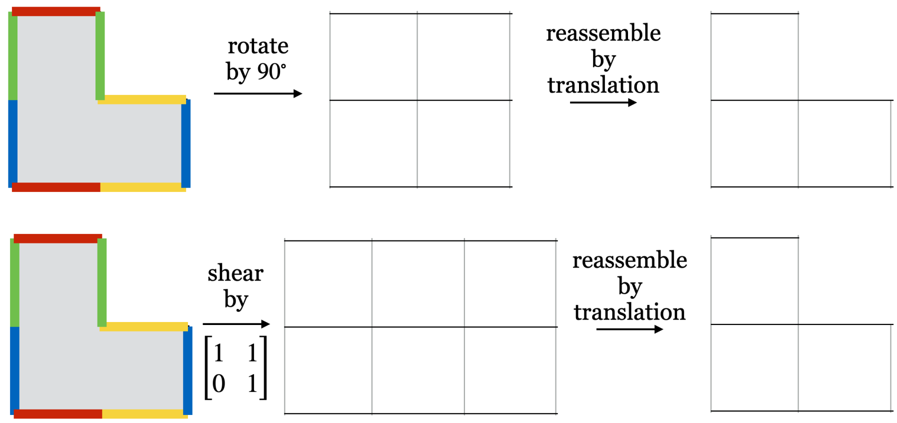




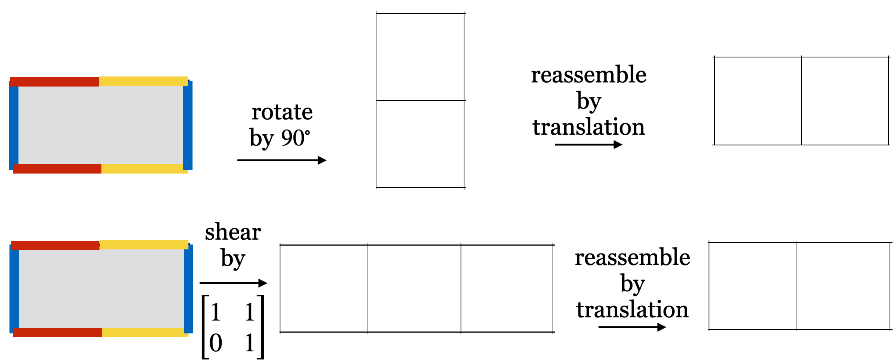
Contextual note. In mathematics, we often care about the *dimension* in which we are working. For example, a torus is a 2D object, and if we look at it as the surface of a bagel, it is a 2D surface *embedded* in 3D space. The family of parallel trajectories in a given direction on the square torus looks 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.

4.8. 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 determine whether the rotation and the shear are automorphisms of the surface.



4.9. Do the same for the 2×1 rectangle.



The surfaces above are called *square-tiled surfaces*, meaning that they are created by gluing together unit squares, edge to edge. The group of 2×2 matrices with integer entries and determinant 1 is known as the "special (determinant 1) linear group of order 2 (2×2 matrices) with entries in $\mathbf{Z}$ (integers)," and is denoted by $\mathrm{SL}(2, \mathbf{Z})$.

Given a square-tiled surface and a matrix in $\mathrm{SL}(2, \mathbf{Z})$, what do you think is the probability that the matrix is an automorphism of the surface?

THEY DID THE MATH: Jane Wang (right front, orange shirt) and Sunrose Shrestha (right back, green shirt) studied the statistics of square-tiled surfaces: what proportion of the square-tiled surfaces in each stratum have various interesting properties. They are shown here with billiards enthusiasts (front row) Chandrika Sadanand, the author, Aaron Calderon, Jane; (back row) Michael Wan, Solly Coles, Samuel Lelièvre, and Sunrose.

As an interesting aside, just behind this happy group is the only place in the world where an airplane can fly over a car that is driving over a train that is passing over a boat, located near Boston University.

ST

4.10. *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. (Here by *length* we mean e.g. by using a ruler to measure the distance along a trajectory in one period.) 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?

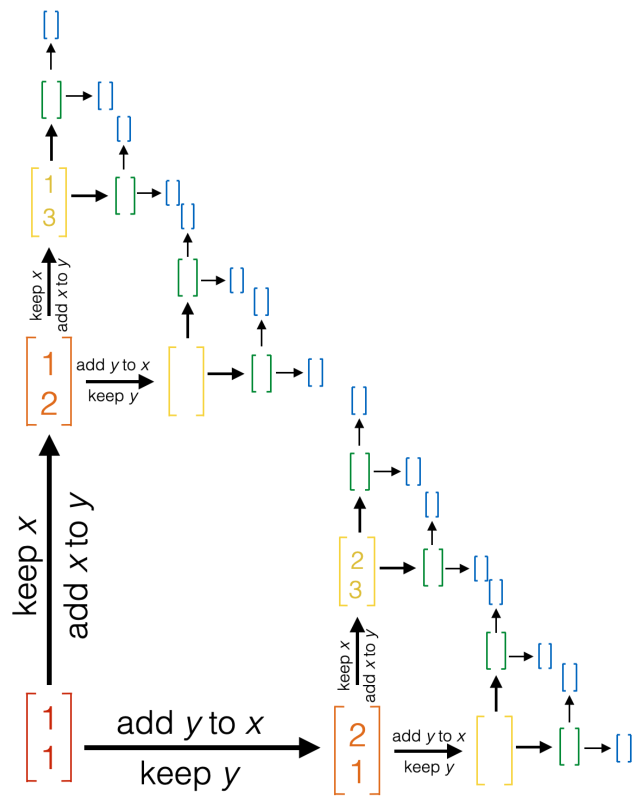
(b) Explain why the number of lattice points inside a disc of radius L is approximately πL^2, especially when L is large.

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

4.3 The tree of periodic directions

DD

4.11. 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 either to add the entries together to get a new x-value (moving right), or 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.

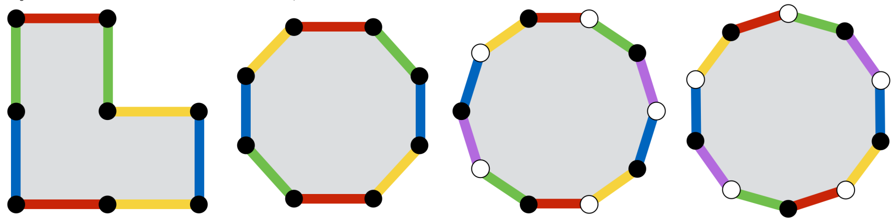
DD

4.12. (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?
(c) For a vector $\left[\begin{smallmatrix}p\\q\end{smallmatrix}\right]$, the continued fraction expansion of q/p tells you how to move in the tree to reach $\left[\begin{smallmatrix}p\\q\end{smallmatrix}\right]$. Explain.

DD

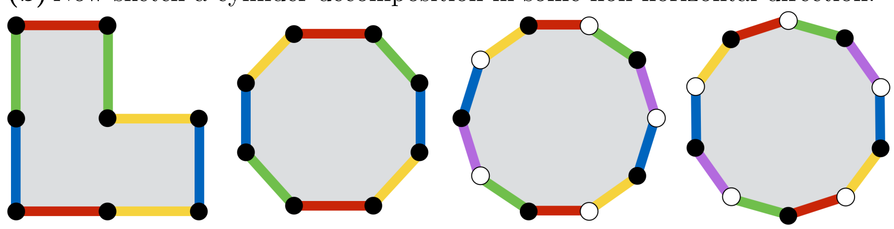
4.13. We have seen that we can can partition a polygon surface into *cylinders*. The boundary of a cylinder is a *saddle connection* – a line segment connecting two cone points with no cone points in its interior – and there are no vertices inside a given cylinder. To construct the cylinders, draw a 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 4.5, 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 decagon surface has *one* cylinder in one cylinder direction, and *two* cylinders in another direction.

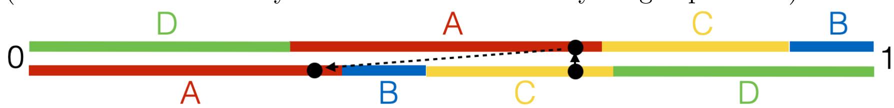
(b) Now sketch a cylinder decomposition in some non-horizontal direction.



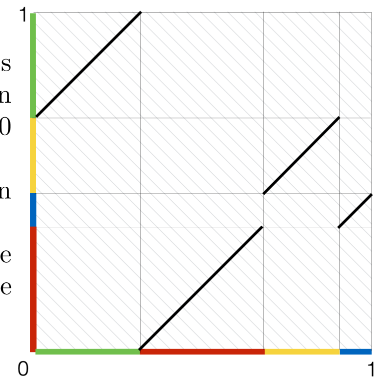
Recall that a *saddle connection* is a line segment connecting two cone points, with no cone points in its interior. In the decagon surface, some saddle connections connect a vertex to itself (black to black or white to white), while others connect two different vertices. An automorphism *cannot* map a same-vertex saddle connection to a different-vertex saddle connection, or vice versa.

DD

4.14. Here is a 4-IET. An example of the image of one point is shown.
(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? (*Hint:* measure carefully! Don't *assume* that everything is periodic!)



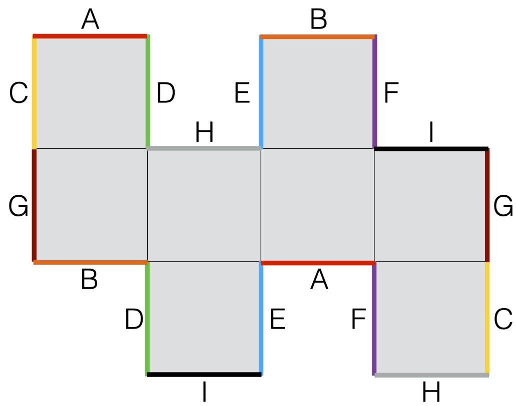
An IET essentially cuts up an interval of points and reassembles them. So we can think of an IET as a function that maps points between 0 and 1 to points between 0 and 1.
(b) The picture to the right shows the function corresponding to the above 4-IET. Explain.
(c) Use the graph to find the orbit of the same point that you followed in part **(a)**. Note the helpful foliation by lines with slope -1!



4.15. 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-IET is equivalent to a rotation (Problem 3.28), and that IETs with rational subinterval lengths have only periodic behavior (Problem 4.15), but outside of these cases, things can get very interesting indeed.

THEY DID THE MATH: Jon Chaika (above) 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 the region's size. For example, in the IET in Problem 4.14, if interval B has length $1/10$, a point should land in interval B, on average, $1/10$ of the time.



Speaking of cylinders and automorphisms, let's meet a creature that is legendary in these areas: the Eierlegende Wollmilchsau (left). This surface is interesting because while it is clearly not the square torus, it has a lot of properties in common with the square torus. We'll explore some of those now.

4.16. (a) Show that the surface has two horizontal cylinders and two vertical cylinders, and in each case the cylinder's width (in the cylinder direction) is 4 times its height (perpendicular to the cylinder direction). We say that the cylinders have *modulus* 4. You can think of the cylinder's modulus as its "aspect ratio."

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 $\mathrm{SL}(2, \mathbf{Z})$.

(b) It turns out that the Veech group of the Eierlegende Wollmilchsau[1] is also $\mathrm{SL}(2, \mathbf{Z})$. Check this by transforming the surface via each of the generators[2] $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ of $\mathrm{SL}(2, \mathbf{Z})$ and checking that you get the same surface back, as we practiced in Problems 4.3 and 4.8.

They did the math: The Eierlegende Wollmilchsau was independently discovered by Giovanni Forni in 2006 and Gabriela Weitze-Schmithüsen and Frank Herrlich in 2008. Gabi and Frank (right) gave the surface its snappy name. It translates from German as "egg-laying wool-milk-sow" – an animal that provides eggs, wool, 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.

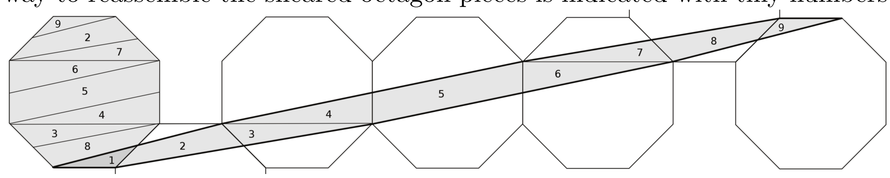
[1] "EYE-ur-LEEG-un-duh VOLE-milsh-sow"

[2] Recall that in Problem 3.17, you proved that these matrices generate $\mathrm{SL}(2, \mathbf{Z})$.

4.4 We finally meet Veech

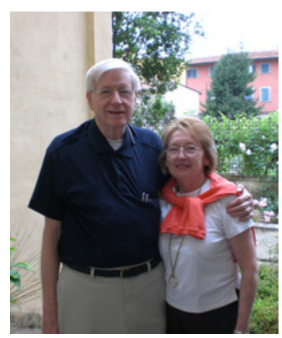
DD

4.17. 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, the L, and the Wollmilchsau. One example is the regular octagon surface, shown below sheared by the matrix $\begin{bmatrix} 1 & 2(1+\sqrt{2}) \\ 0 & 1 \end{bmatrix}$. The way to reassemble the sheared octagon pieces is indicated with tiny numbers.



(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 4.13, you identified this surface's two horizontal cylinders. In the shear above, show that the top/bottom cylinder is twisted once, while the middle cylinder is twisted twice.
(c) For each of the horizontal cylinders in the regular octagon surface, find its modulus (recall Problem 4.16.) How is the modulus related to the number of twists?



We previously said that the group of automorphisms of a surface is called its *Veech group* (Problem 4.16). When a surface has rotations, reflections, *and shears* in its Veech group, this means that its Veech group forms a lattice; such a surface is called a *Veech surface*, or *lattice surface*.[3] 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 the moduli are all rational multiples of each other. For example, in the regular octagon surface above, one cylinder's modulus is twice the other's.

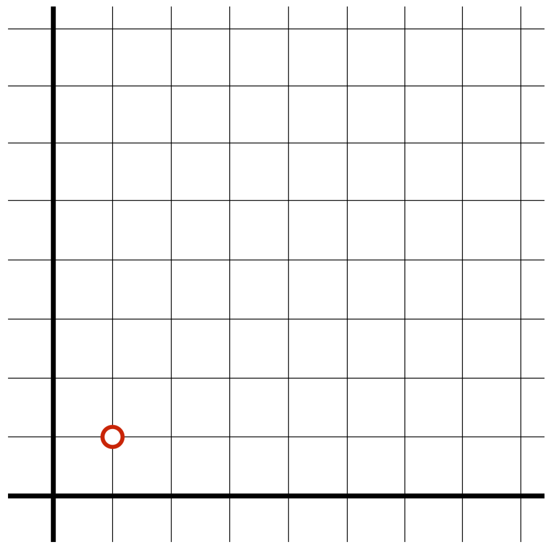
They did the math: These notions are named for mathematician William Veech (above, with his wife Kay Veech), who really got this field going and then did a lot of tremendous work in it, including coming up with interval

[3] They are traditionally called Veech surfaces, but some people think that too many things are named "Veech," and are trying to change the terminology.

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.

DD

4.18. 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 to get new yellow 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 green points. Draw the new points in blue. Continue in purple. Mark all the points you get.

DD

4.19. *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.
(c) 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.
(d) Notice that to reach the point $(5,7)$ in the picture above, you applied the transformations

$$(1,1) \xrightarrow{\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]} (1,2) \xrightarrow{\left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]} (3,2) \xrightarrow{\left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]} (5,2) \xrightarrow{\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]} (5,7).$$

Explain how to use horizontal and vertical shears to implement the continued fraction algorithm for $7/5$. Done in reverse, this is *Euclid's algorithm* for finding the greatest common factor of two numbers: here, 5 and 7.

4.20. *Counting periodic trajectories, part II.* We can improve on our previous method of counting periodic trajectories (Problem 4.10) by counting primitive vectors, as these are the directions that give us different billiard trajectories.

Let P be the set of primitive vectors (Problems 4.18–4.19). 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 $2P$ in black on your picture in Problem 4.18.

(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.*

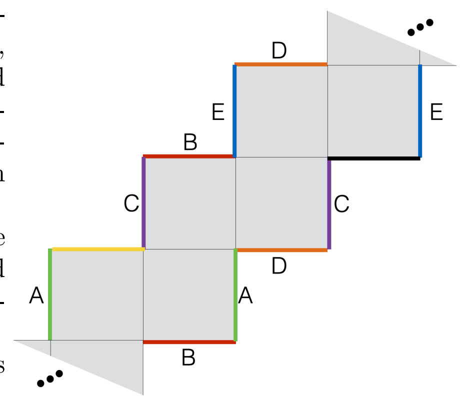
(c) We wish to know the proportion of the vectors in the first quadrant that are in P; let's call this proportion x. Show that the proportion of vectors in the first quadrant that are in each set kP is $\frac{x}{k^2}$.

(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.

4.21. *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?

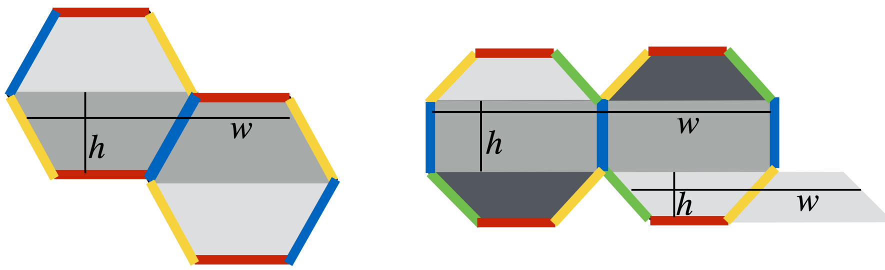
(b) Identify some periodic trajectories on the surface.

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

4.5 The Modulus Miracle

DD

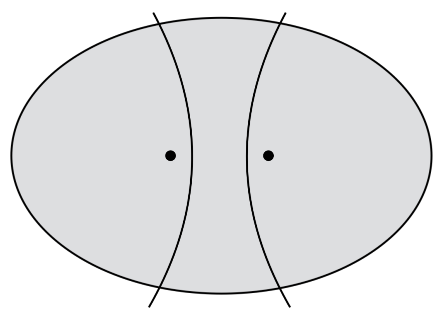
4.22. Theorem *(Modulus Miracle)*. Every horizontal cylinder of a double regular n-gon surface the has same modulus ("aspect ratio"), which is $2\cot\pi/n$. **(a)** Confirm this for the two surfaces shown, by calculating the modulus for each cylinder, and also the number $2\cot\pi/n$. If you can, prove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



Thus *all* surfaces made from a double regular polygon are Veech surfaces! In other words, they have rotation, reflection and shearing symmetries, like the square torus.

The benefit of using a double regular polygon surface instead of a single one is that all of the cylinder moduli are equal. If you do use just a single polygon, like our familiar regular octagon surface, then the moduli of some of the cylinders are double the others (see Problem 4.17).

ST



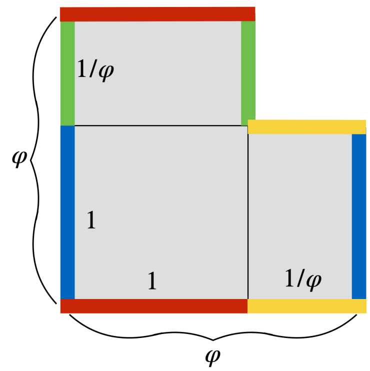
4.23. An *ellipse* with foci F_1, F_2 and string length ℓ consists of all points P satisfying $|F_1P| + |PF_2| = \ell$. Similarly, a *hyperbola* with foci F_1, F_2 and "string length" ℓ consists of all points P satisfying $|F_1P| - |PF_2| = \pm\ell$.

In Problem 1.26, 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. Conclude that *every* segment of such a trajectory will pass between the foci.

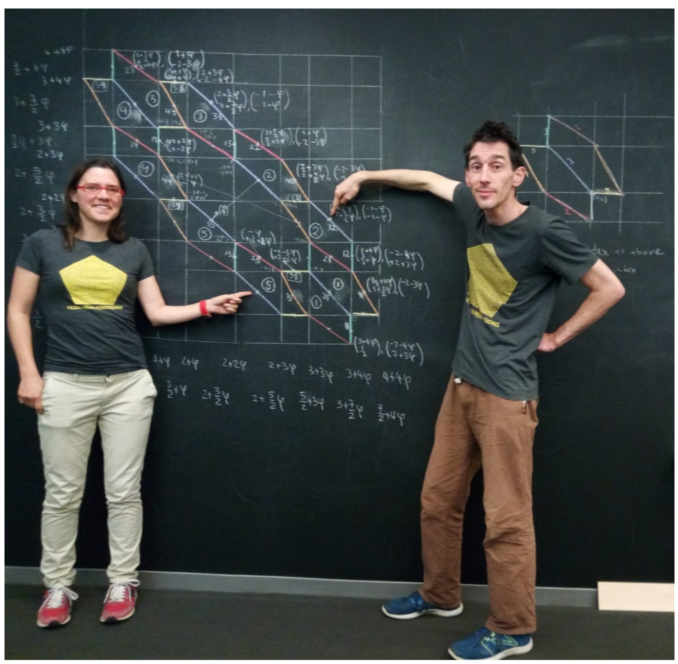
This construction will enable us to construct an unilluminable room.

DD

4.24. A particularly nice flat surface is the "Golden L," whose opposite parallel edges are identified as shown, and whose edge lengths are indicated in the picture. The number $\varphi \approx 1.618$ 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. Explain why φ is sometimes called "the most irrational number."
(c) Two numbers are *commensurable* if they are rational multiples of each other. Are the moduli of the Golden L's cylinders commensurable?

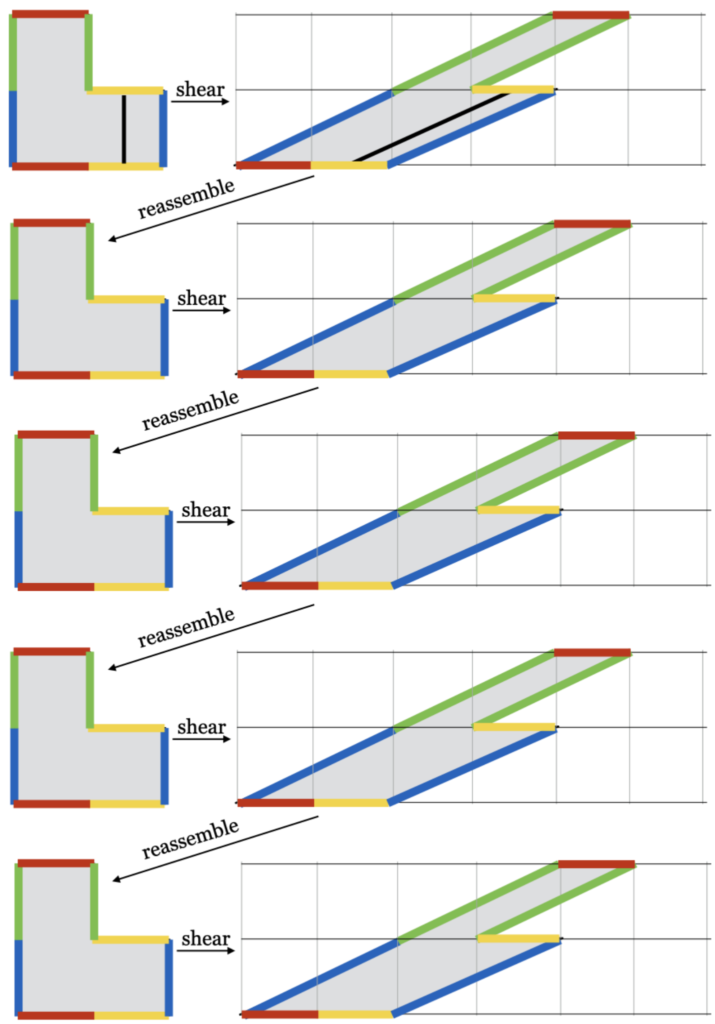


They did the math: Samuel Lelièvre has studied the golden L surface in detail. In joint work with Jayadev Athreya (§ 2.5) and Jon Chaika (§ 4.3), he studied the *gaps* between slopes of cylinder directions in the golden L. It turns out that the golden L and regular pentagon surfaces are closely related; the picture to the right shows Samuel collaborating with the author on a project about cylinders on a translation surface made from ten sheared regular pentagons, while wearing coordinating T-shirts.

A big question in the study of trajectories on surfaces is: "what happens to a trajectory on a surface when you apply an automorphism?" For example, in Problem 2.1 we explored the effects of rotations, reflections and the shear on a trajectory on the square torus, and determined the effect of each automorphism on the trajectory's slope.

DD

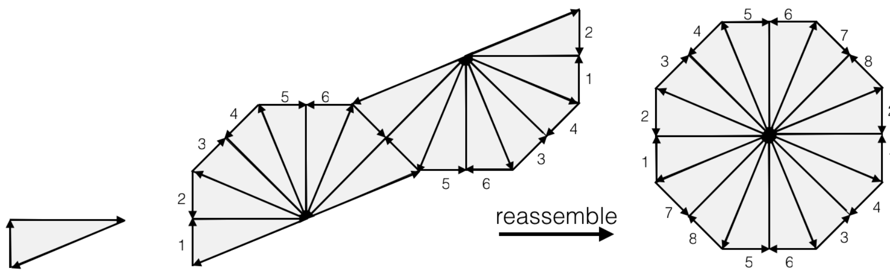
4.25. Let's see what happens when we apply the horizontal shear $\left[\begin{smallmatrix}1&2\\0&1\end{smallmatrix}\right]$ to the L-shaped table, with a short vertical trajectory on it. The diagram to the right provides a template for drawing the image of the trajectory under this shear. Fill it in 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 (§ 4.1) introduced me to the idea that interesting things can happen when you twist a surface many times in a cylinder direction. Above, one cylinder fills up, while the other stays empty. Samuel Lelièvre and I used a similar strategy to create the periodic billiard trajectory on the regular pentagon that appears on the first page of this chapter.

DD

4.26. 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. *Hint:* This trajectory has period 2 on the regular octagon surface, and passes through 6 triangles, including edges 1 and 7.

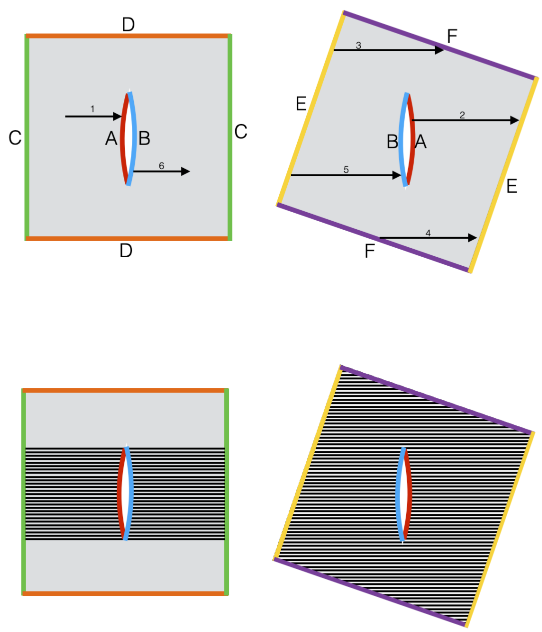
(b) What triangle unfolds to the double regular pentagon surface? Dissect the double pentagon to figure it out, and then draw the unfolding as above.

4.6 The slit torus construction

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 Veech surface. The *Veech dichotomy* (proved by William Veech) says that for a given direction on a Veech surface, every trajectory in that direction 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



4.27. 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 above.

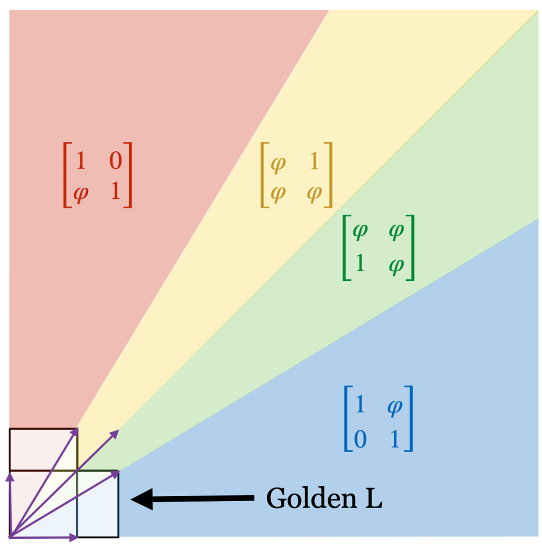
Edges A and B are vertical, but in the picture I have pulled them apart a little bit so that you can see that there is a slit between them.
(a) In the top picture, I have drawn the first six pieces of a horizontal trajectory. Draw the next 10 pieces. Do you expect this trajectory to be periodic?
(b) Explain why, over time, a horizontal trajectory through the slit will end up looking like the bottom picture.

They did the math: 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 billiards summer research program that Moon organized.

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

4.28. 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.



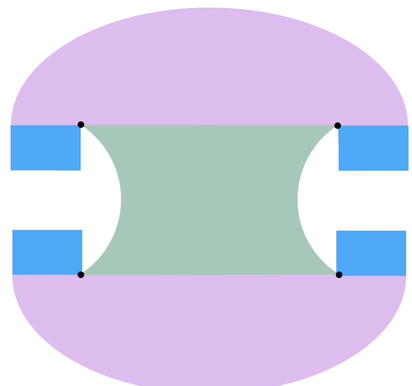
(a) The dimensions of the golden L are given in Problem 4.24. Check that the purple vectors shown spanning diagonals of the golden L are $[0,1]$, $[\varphi,1]$, $[\varphi,\varphi]$, $[1,\varphi]$, $[0,1]$.
(b) Explain why the blue matrix takes the entire first quadrant to the blue sector. Check that its determinant is 1, meaning that it preserves areas. 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.

4.29. In Problem 4.18, we repeatedly applied horizontal and vertical shears to generate *all* of the periodic directions on the square torus. In Problem 4.19, we explained how applying the two different shears is essentially the continued fraction algorithm in reverse. Similarly, to generate the set of *all* of the periodic directions on the golden L, we start with the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and repeatedly apply the blue, green, yellow and red automorphism matrices. People describe this as a "generalized continued fraction algorithm." Explain.

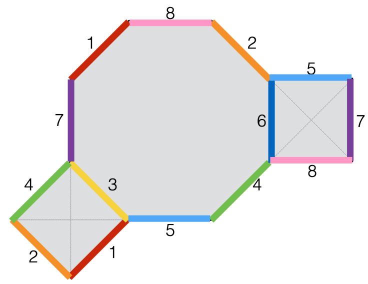
ST

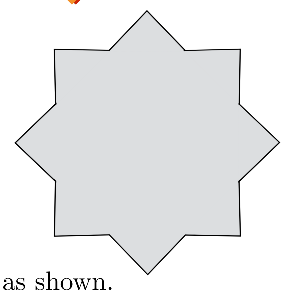
4.30. *The Penrose unilluminable room.* One way to pose the illumination problem is: "Is every mirrored room illuminable from *some* point in the room?" Here is a counterexample, a room that cannot be illuminated from *any* point inside, shown to the right. The top and bottom are half-ellipses, whose 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

4.31. 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 Veech surfaces (surfaces whose automorphism group forms a lattice: it has rotations, reflections and shears). Then William 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 (top picture).





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 (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.

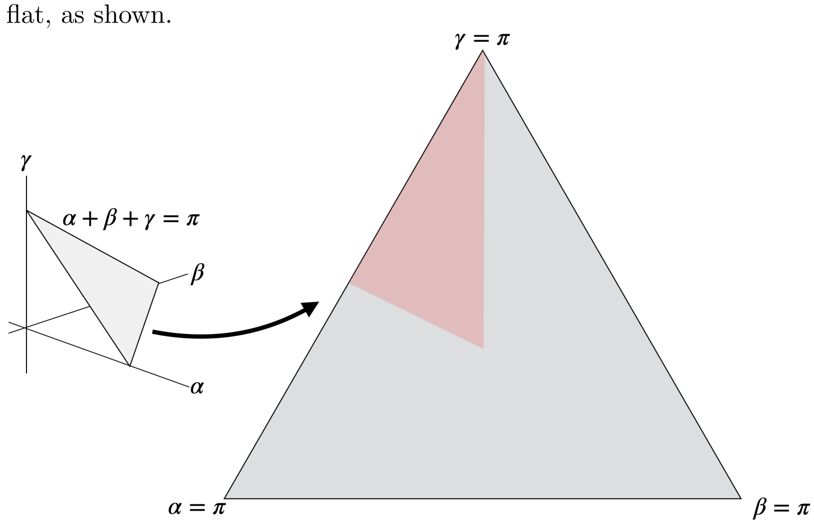
(c) For this surface's cylinders, are the moduli equal or merely commensurable?

4.7 The space of triangles

We have talked here and there about the space of all translation surfaces. The following problem builds understanding about what it means to have a space where each point represents an object.

DD

4.32. Up to similarity and isometries, 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 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.



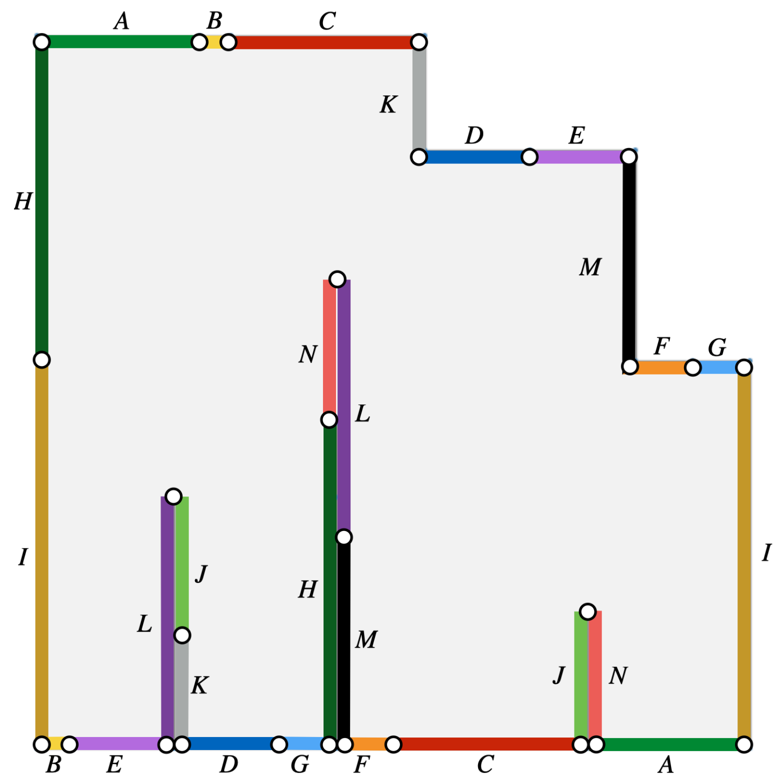
Sketch the following sets:
(a) the set of right triangles (green),
(b) the set of isosceles triangles (blue),
(c) all triangles with angles $0.12\pi, 0.35\pi, 0.53\pi$ (black dots),
(d) the set of all acute triangles (shaded).

DD

4.33. 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.
(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.

4.34. *Where can we find the triangles we love?*
(a) 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.
(b) The Ward surface given in Problem 4.31 is the unfolding of the triangle with angles $\pi/16, \pi/4, 9\pi/16$. Explain. What triangle unfolds to the Ward surface given by a decagon with two pentagons – which triangle unfolds to it? How about the surface with a dodecagon and two hexagons? Mark these triangles, and the rest of the Ward family W, on the diagram.

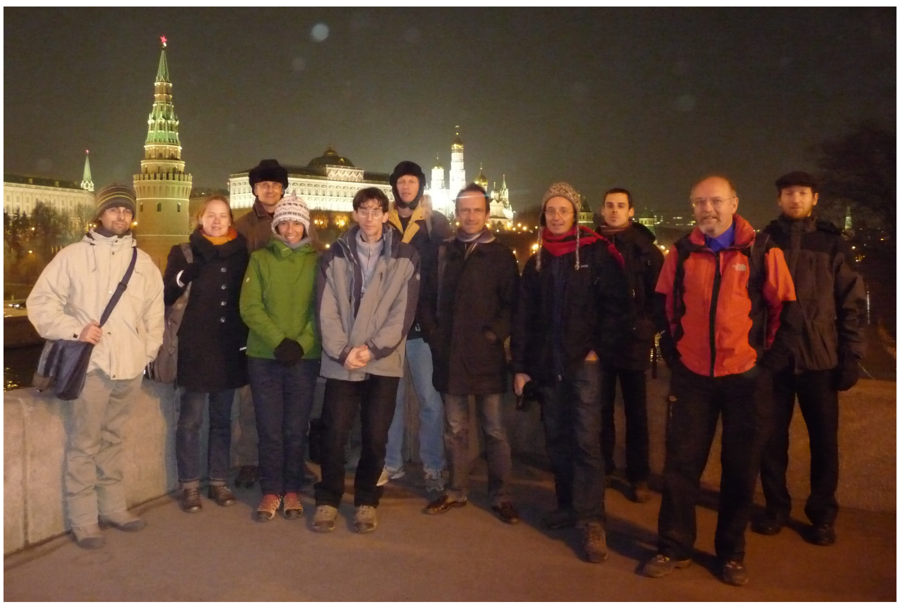
The sets R and W are *discrete* in the space of triangles: for each triangle t of R or W, it is possible to find a little region in the space of triangles containing t, that does not contain any other point of R or W. The fact that they are discrete makes Veech surfaces difficult to find!
(c) In the 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?



The zippered rectangle construction. The figure above 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 each "zipper" are glued to different places.

4.35. For the zippered rectangle surface above:
(a) Consider the vertical flow on this surface. Show that its behavior is described by a 7-IET.
(b) Show that the surface has nine vertices: two with 6π of angle around them, and the rest with 2π of angle around them. *Hint*: Notice that at the bottom of the zippers, the two flaps each have their own vertex point, to indicate that these are typically not identified with the same point.
(c) Show that the surface has genus 3.

They did the math: Pierre Arnoux (below, in orange jacket) created the above surface to give a geometric realization of his eponymous IET (see Problem 5.14). The picture below shows flat surfaces enthusiasts Corentin Boissy, Anna Lenzhen, Serge Troubetzkoy, Aslı Yaman, Samuel Lelièvre, Barak Weiss, Xavier Bressaud, Pascal Hubert, Luca Marchese, Pierre, and Alexey Glutsyuk in Moscow in 2012.



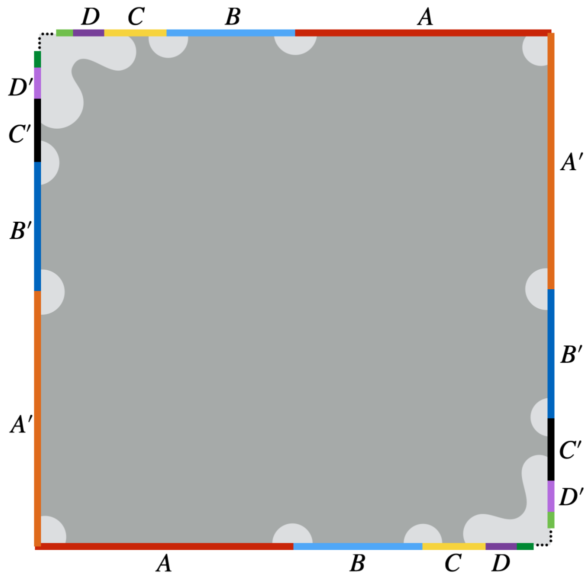
DD

4.36. Recall the slit torus surface in Problem 4.27. Show that, 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. What about the horizontal direction?

4.8 We get a little bit wild

DD

4.37. *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.



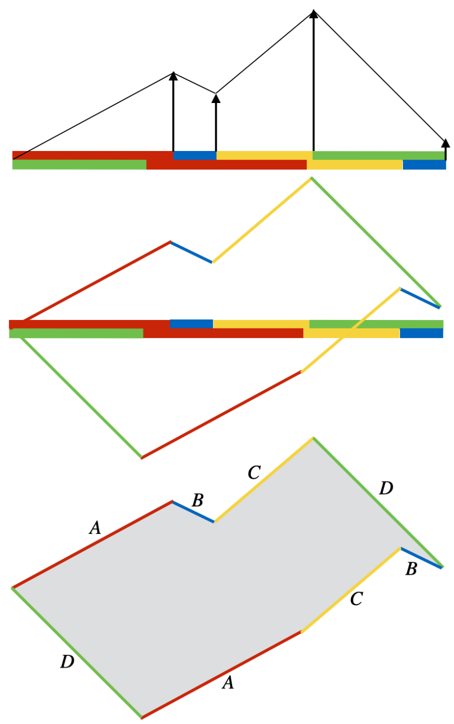
(a) Use vertex chasing to show that the surface has 2 vertices.

(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 does it have two vertices, or just one? Hmmm...

(c) Show that the Chamanara surface has a cylinder decomposition in the direction of slope 4, and 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 2^n for integer n.

(d) As you can see, this Chamanara surface has a dark blob that is gradually filling up the surface, avoiding but approaching the vertices, that is growing out towards the corners. Show that (contrary to appearances) the complement of such a blob is always *connected*. This makes the surface "wild."

They did the math: 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.



In Problem 4.7, 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 the top part of your IET, choose a "height" (possibly 0), and draw edges 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 (bottom picture).

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

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

DD

4.39. (For fun) Explain what is going on with the super laptop sticker pictured at right. Is it a flat surface? Is it a torus?

The biggest open problem in the study of Veech surfaces is: *Can we find more Veech surfaces?* and the related question, *Have we found them all yet?* Here are the families of Veech surfaces we have seen so far in this book:

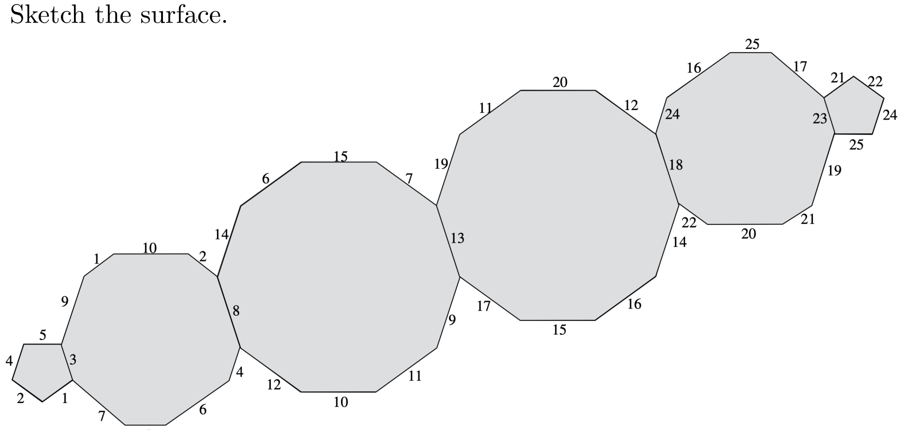
- square-tiled surfaces (Problems 4.8, 4.16);
- regular polygons: double regular n-gons for any n, and single regular n-gons for even n (Problem 4.22);
- Ward surfaces: a regular $2n$-gon, with two regular n-gons glued to it along alternating edges (Problem 4.31).

They did the math: 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 Veech surfaces, now called Bouw-Möller surfaces in their honor. This family includes surfaces with any number $m \geq 2$ of polygons. Irene and Martin gave an algebraic description of the surfaces, and later, Pat Hooper (§ 3.8) found a polygon decomposition for the surfaces, which we present here. Martin is shown above (in the vest) running in Marseille with mathematician Erwan Lanneau and the author.

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.) So that the cylinder moduli in Bouw-Möller surfaces are equal, the k^{th} semi-regular $2n$-gon has edge lengths alternating between $\sin \frac{k\pi}{n}$ and $\sin \frac{(k+1)\pi}{n}$.

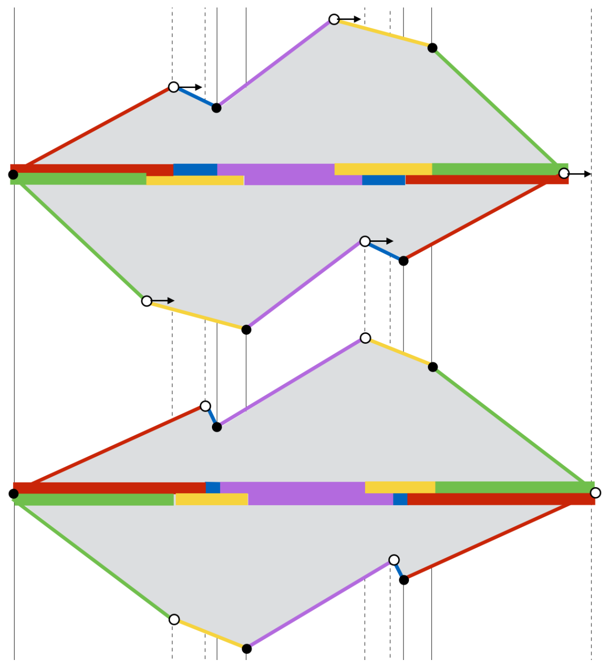
DD

4.40. (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 (for the reasoning behind the zig-zag edge-numbering system, see § 5.6). Shade each horizontal cylinder a different color. Does it seem plausible that all of the cylinders have the same modulus?
(c) For the $m = 4$, $n = 3$ Bouw-Möller surface: How many polygons does it have? How many edges does each polygon have, and what are their lengths? Sketch the surface.



4.9 Moving around in the space of surfaces

DD



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

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

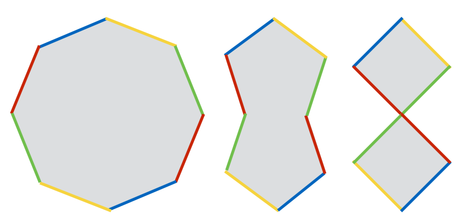
One way to get a new translation surface "near" the original one is to deform the surface by moving one cone point *relative* to the other. This is known as a *rel deformation.* The arrows in the top picture indicate that we will shift the white point slightly to the right. The bottom picture shows the 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



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

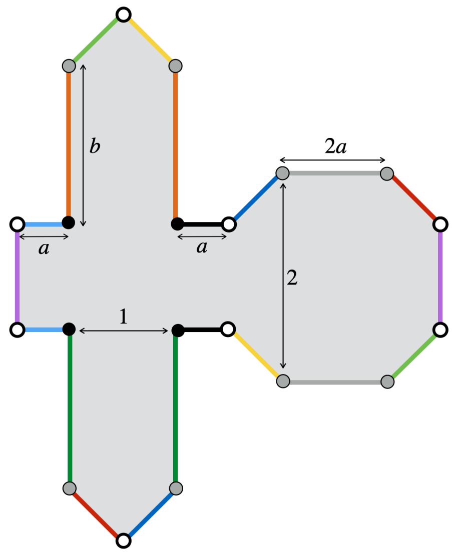
(a) Show how to smoothly deform the regular octagon surface into the double pentagon surface.

In Problem 3.14, 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 4.41: “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 4.42 are Veech 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* Veech surfaces. As mentioned in Problem 4.33, Veech surfaces are discrete – we cannot move continuously among a family of Veech surfaces. This makes them difficult to find. When someone discovers a new family of Veech surfaces, it is a big deal.



DD

4.43. In 2016, Curt McMullen (§ 2.7), Ronen Mukamel (below) and Alex Wright (§ 3.2) discovered the *gothic* family of Veech 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.

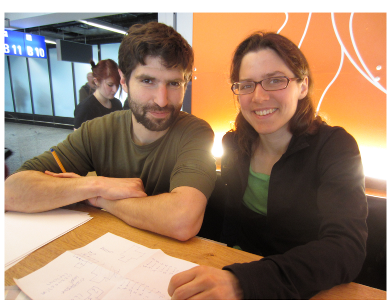
Show that each such surface has **(a)** 5 horizontal cylinders and 5 vertical cylinders; **(b)** 3 cone points, as indicated; and **(c)** genus 4.

The real key in showing that members of the gothic family are Veech 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 Veech surface.

4.44. What stratum are the gothic surfaces in? Explain why this construction does not give a *continuous* family of gothic lattice surfaces in this stratum.



They did the math: 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.

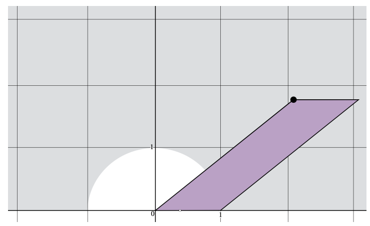
5

Further topics and tools

Each of the sections in this chapter is a self-contained set of problems that explores a single topic or tool. The best way to learn a new piece of mathematics is to work out some problems about it, so I've written some problems that explore each of these topics.

5.1 The modular group

In Problem 4.32, we explored the space of all triangles. Now we'll explore the space of all tori. We'll do this by considering the space of all surfaces made by gluing opposite parallel edges of parallelograms. In particular, if you can cut and paste one parallelogram surface into another while respecting the edge identifications, we'll consider those to be the same surface.

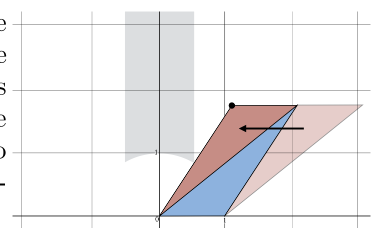


Given any parallelogram surface, do the following:

- Translate and rotate the parallelogram until its shortest edge is on the x-axis, and the parallelogram lies above the x-axis.
- Scale the parallelogram so that its shortest edge has length 1, and coincides with the segment $[0,1] \times \{0\}$.

DD

5.1. Given a parallelogram translated, rotated and scaled (or "normalized") as described above:
(a) Show that its upper-left corner uniquely determines its shape.
(b) Show that the upper-left corner always lies outside the unit circle.



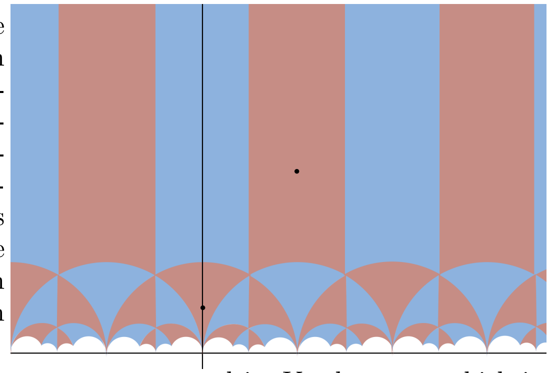
Now we want to mod out by cut-and-paste equivalence of parallelogram *surfaces*. To make this happen, we can cut and paste triangles (as suggested by the picture to the right), while respecting the surface's edge identifications, to yield an equivalent surface represented by a different parallelogram.

5.2. Show that, under cut and paste equivalence, every normalized parallelogram is equivalent to a parallelogram whose top left vertex lies in the infinite vertical strip $[-1/2, 1/2] \times [0, \infty]$. Justify the claim that every normalized parallelogram can be represented by a point in the shaded region shown above (which is meant to extend infinitely upward).

The shaded region in the picture above is known as a *fundamental domain*. In Problem 5.5, you will see that we have many choices for which region to take as our chosen fundamental domain; people traditionally choose the one pictured above.

5.3. It is possible that, after a cut-and-paste equivalence, the shortest side of your parallelogram is no longer the one on the x-axis, so you must switch edges and rescale. Give an example of such a parallelogram.

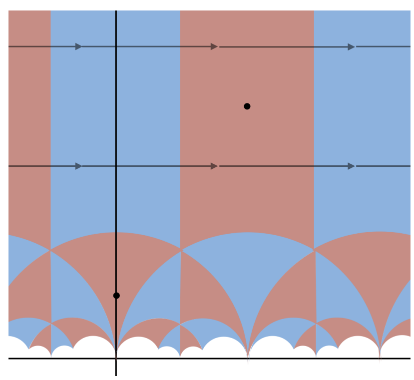
5.4. On the diagram to the right, mark all of the points in the upper halfplane that represent a 2×1 rectangle surface, or any surface equivalent to it under the actions described above. Two such points are marked for you. *Hint:* there is one corresponding point in each of the colored tiles. Can you explain why?



When working with the *square* torus, we used its Veech group, which is $\mathrm{SL}(2, \mathbf{Z})$: the special (determinant 1) linear group with integer entries. Now we will work with $\mathrm{SL}(2, \mathbf{R})$, the special linear group with real-valued entries. The *Iwasawa decomposition* says that any matrix in $\mathrm{SL}(2, \mathbf{R})$ 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 $\mathrm{SL}(2, \mathbf{R})$ on tori. In particular, we want to know how $\mathrm{SL}(2, \mathbf{R})$ acts on the space of all tori that we defined in Problem 5.1. Since every matrix in $\mathrm{SL}(2, \mathbf{R})$ 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. Our normalization requires that one edge must lie on the x-axis, so we ignore rotations, so we'll focus on A (*geodesic flow*) and N (*horocycle flow*).[1]



DD

5.5. Consider two of the points representing 2×1 rectangles, and the effect of horocycle flow and geodesic flow on them.

(a) Show that, for a point above $y = 1$, horocycle flow acts as indicated by the arrows.

(b) For each of the points, apply a tiny bit of horocycle flow, e.g. $\begin{bmatrix} 1 & 1/10 \\ 0 & 1 \end{bmatrix}$ and then normalize as described above. What happens to the points?

(c) Do the same for a tiny bit of geodesic flow, e.g. $\begin{bmatrix} 11/10 & 0 \\ 0 & 10/11 \end{bmatrix}$.

[1] For a beautiful animated view of the action of these flows on lattices, see the video Shape of Lattices on YouTube by Pierre Arnoux and Edmund Harriss: `https://www.youtube.com/watch?v=vLrliPt4Uc0`.

They did the math: Marina Ratner (right) proved several powerful theorems, which together are known as *Ratner's measure and orbit classification for unipotent flows on homogeneous spaces*. These are key tools in the fields of dynamical systems and ergodic theory, and inspired lots of further work. In her Ph.D. thesis, she studied geodesic flows, and in her later work, she proved several important "rigidity" results about horocycle flows: precisely the two types of flows that we analyzed above. She died in 2017.

Thanks to Samuel Lelièvre (§ 4.5) for helping to make my dreams for this section become reality.

5.2 Renormalization

5.6. *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 4.32). Repeatedly perform the following algorithm:

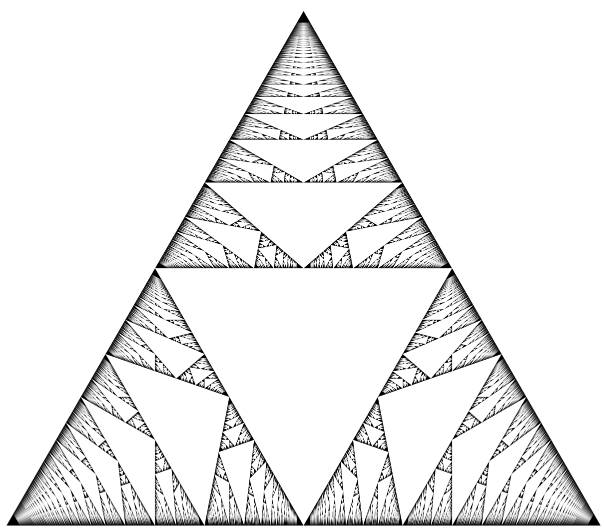
1. If
   $a > b + c$, subtract $b + c$ from a so that $(a, b, c) \mapsto (a - b - c, b, c)$.
   $b > a + c$, subtract $a + c$ from b so that $(a, b, c) \mapsto (a, b - a - c, c)$.
   $c > a + b$, subtract $a + b$ from c so that $(a, b, c) \mapsto (a, b, c - a - b)$.
   and if none of these are true, 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

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

so that this is a *periodic point*.

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.

5.7. The algorithms that we performed on square torus trajectories (Problem 2.22) and on their corresponding cutting sequences (Problem 2.27) are also renormalization algorithms. Explain.

THEY DID THE MATH: Sasha Skripchenko (right) studied the Rauzy gasket and, in joint work with Artur Avila and Pascal Hubert (§ 3.7), answered a question of Pierre Arnoux (§ 4.7) to show that its Hausdorff dimension is less than 2.

5.8. Let

$$M_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad M_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

(a) Given a triplet of numbers (a, b, c) where $a, b, c > 0$ and $a + b + c = 1$, show that you can implement the algorithm from the previous problem by multiplying the column vector $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ by suitable *inverses* of M_1, M_2 and M_3.
(b) Consider a finite product $\overline{M}$ of M_1, M_2 and M_3 that includes at least one copy of each of the three (e.g. $\overline{M} = M_1 {M_3}^2 M_2 M_1$), and suppose that (a, b, c) has the property that

$$\overline{M} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

for some real number λ.[2] Show that (a, b, c) is a point in the Rauzy gasket.
(c) Show that *every* point of the Rauzy gasket can be obtained in this way.

(d) Explain why at least one copy of each of M_1, M_2, M_3 is needed.

Thanks to Vincent Delecroix (§ 3.6) for explaining how to use these matrices to work with the Rauzy gasket.

[2] In other words, $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is an *eigenvector* of $\overline{M}$.

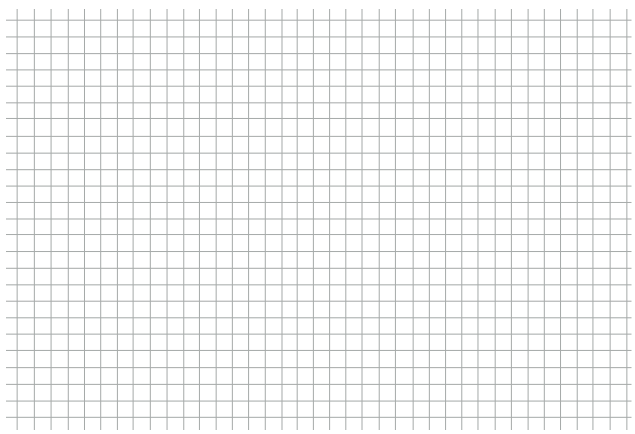
5.3 We find the Rauzy fractal in tiling billiards

5.9. 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 $0, 0, 1, \ldots$.

5.10. *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 8 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.



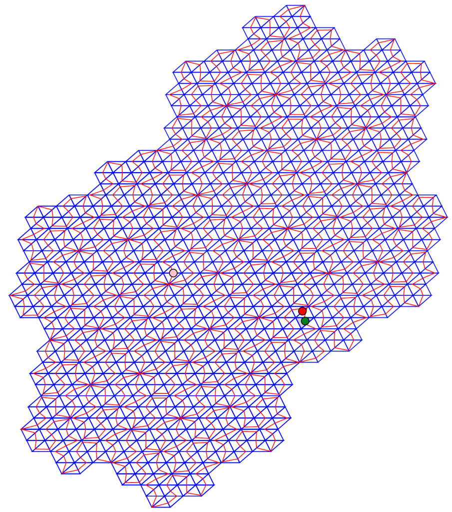
Notice that the points stay close to a line of slope $1/\varphi$.

5.11. 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 6 terms of the sequence a, ab, $abac$, ... and comment on any patterns.

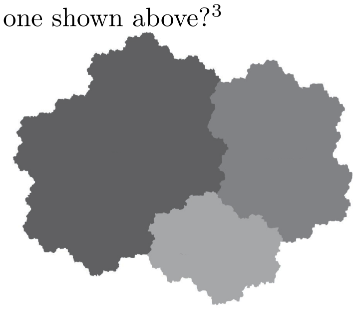
Suppose that you use the resulting sequence to take a three-dimensional "walk" similar to the one in the previous problem, where a, b and c tell you to take steps in the x-, y- and z-directions, respectively. It turns out that, as in the previous problem, the points on this walk stay close to a line, 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 **5.12.** *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

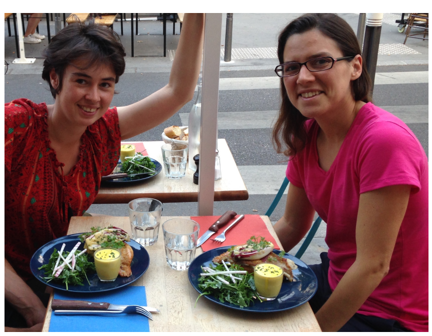
$$\frac{\pi(1-\alpha)}{2} \approx 41.0679888577^\circ$$
$$\frac{\pi(1-\alpha^2)}{2} \approx 63.396203173^\circ$$
$$\frac{\pi(1-\alpha^3)}{2} \approx 75.535807969^\circ$$

(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 close to 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?[3]



DD **5.13.** 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.



They did the math: Our number $\alpha \approx 0.54369$ is just one point in the Rauzy gasket (Problem 5.6). It turns out that when you make a triangle tiling based on *any* point in the Rauzy gasket, trajectories passing near the circumcenter *always* give you fractal behavior. Olga Paris-Romaskevich (left, in Lyon in 2018 with the author) and Pascal Hubert (§ 3.7) proved this, and many other tiling billiards results.

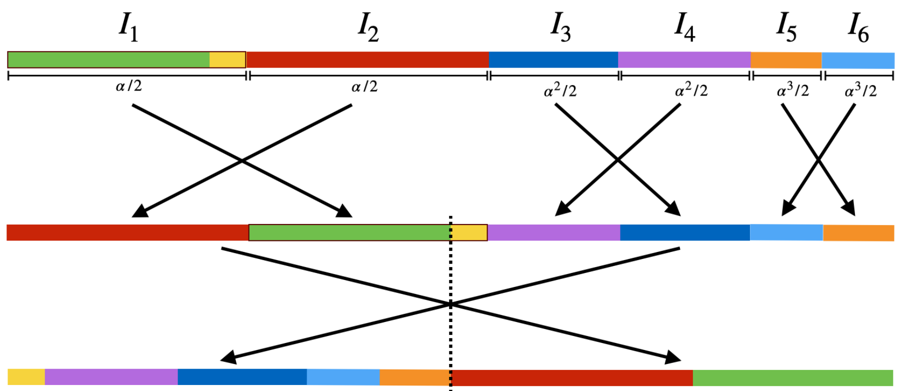
[3] See the second half of the video "Refraction Tilings" on YouTube.

5.4 The Arnoux-Yoccoz IET and arithmetic graphs

5.14. *The Arnoux-Yoccoz IET.* First, define $\alpha \approx 0.54369$ as in Problem 5.6, as the real solution to $x + x^2 + x^3 = 1$.

1. Divide the unit interval into 6 subintervals $I_1, I_2, \ldots I_6$ with consecutive lengths $\alpha/2$, $\alpha/2$, $\alpha^2/2$, $\alpha^2/2$, $\alpha^3/2$, $\alpha^3/2$.
2. Switch the pieces of the same length.
3. Cut the interval in half and switch the halves.

The construction is illustrated below. (Notice that I_1 gets broken into two pieces.) We have seen this IET before, in another form: the vertical flow on the zippered rectangle surface in Problem 4.35 is the same as this IET. This transformation is *ergodic*, meaning that the orbit of every point fills in the space evenly.



(a) Choose a point on the interval, and follow its orbit for 10 iterations. Make a note of the sequence of intervals $I_1, \ldots, I_6$ it ends up in, for use in the next problem. Does it seem plausible that the transformation is ergodic?

Written as a piecewise function, the Arnoux-Yoccoz IET is

$$f(x) = \begin{cases} f_1(x) = x - \frac{\alpha-1}{2} \bmod 1 & \text{if } x \in I_1 \approx [0, 0.27185] \\ f_2(x) = x + \frac{\alpha^3-1}{2} \bmod 1 & \text{if } x \in I_2 \approx [0.27185, 0.54369] \\ f_3(x) = x - \frac{\alpha^3-1}{2} \bmod 1 & \text{if } x \in I_3 \approx [0.54369, 0.69149] \\ f_4(x) = x + \frac{\alpha^2-1}{2} \bmod 1 & \text{if } x \in I_4 \approx [0.69149, 0.83929] \\ f_5(x) = x - \frac{\alpha^2-1}{2} \bmod 1 & \text{if } x \in I_5 \approx [0.83929, 0.91965] \\ f_6(x) = x + \frac{\alpha-1}{2} \bmod 1 & \text{if } x \in I_6 \approx [0.91965, 1] \end{cases}$$

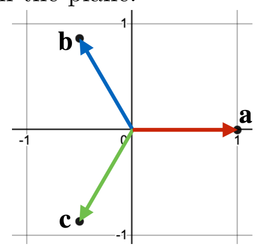
(b) Check that the orbit of your point under this function matches what you found in part **(a)**.

(c) The Arnoux-Yoccoz IET has seven intervals, but there are only six parts to the piecewise function above. Are you concerned?

We love being able to visualize the behavior of an IET in two dimensions. Efforts to do so that we have seen so far include graphing the associated piecewise function (Problem 4.14), tiling billiards (e.g. Problem 3.32), and suspending an IET (Problem 4.38) or making it part of a zippered rectangle (Problem 4.35). The form of the function above, with three pairs of related operations, suggests another method, of creating a "walk" in the plane. Let's do that.

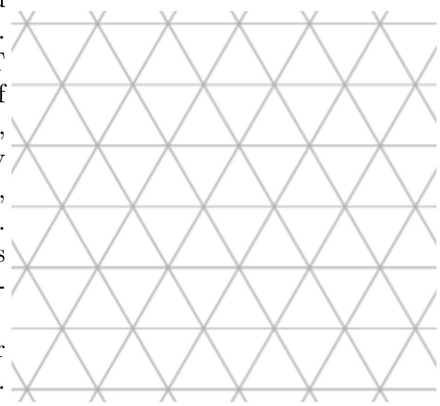
5.15. (Continuation) Let $\mathbf{a}, \mathbf{b}, \mathbf{c}$ be the vectors shown in the picture below right. Note that $\mathbf{a}+\mathbf{b}+\mathbf{c} = \mathbf{0}$. Recalling the piecewise function $f(x)$ from the previous problem, we define a related function $g(z)$ on the plane:

$$g(z) = \begin{cases} g_1(z) = z + \mathbf{a} \\ g_2(z) = z - \mathbf{c} \\ g_3(z) = z + \mathbf{c} \\ g_4(z) = z - \mathbf{b} \\ g_5(z) = z + \mathbf{b} \\ g_6(z) = z - \mathbf{a} \end{cases}$$



We choose any number $x \in [0,1]$ and we start at any point in the plane. We iterate the Arnoux-Yoccoz IET on x, by applying some sequence of functions f_i, for example $f_2, f_5, f_1, \ldots$, to x. At the same time, we apply the corresponding functions g_i, e,g, $g_2, g_5, g_1, \ldots$, to z and its images. This amounts to adding the vectors $\pm\mathbf{a}, \pm\mathbf{b}, \pm\mathbf{c}$, yielding a walk on the triangular grid.

Plot the walk corresponding to your point from Problem 5.14 on the grid. Does it close up?



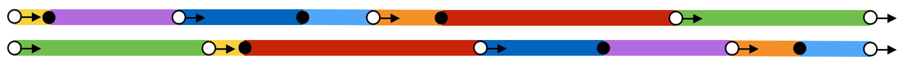
The "walk" shown above is known as an *arithmetic graph*.[4] Rich Schwartz (§ 1.3) has made significant use of arithmetic graphs in his exploration of the behavior of outer billiards on polygons, and in his proof that every triangle whose largest angle is less than 100° has a periodic billiard orbit.

[4] Since "arithmetic" is used as an adjective here, it is pronounced air-ith-MET-ic. This follows the same differential adjective/noun syllable stress pattern as e.g. "I'll reCORD a REcord, and disCOUNT it with a DIScount."

5.5 We find the Rauzy fractal in an arithmetic graph

Note that the content of the previous section is required for this one.

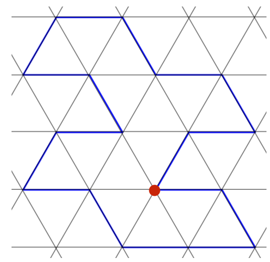
5.16. The Arnoux-Yoccoz IET is pictured below, now back to our usual convention of "flow up, then shift when you come down." As you can see, we are not content with simply the Arnoux-Yoccoz IET; we will apply a rel deformation (recall Problem 4.41). We will leave the black vertices fixed, and shift the white vertices to the right, as the picture suggests.
(a) Check that the picture is consistent, e.g. the red subinterval has a black vertex on its left end and a white vertex on its right end, for both copies.
(b) Also check that things are fine at the endpoints (0 and 1) of the interval.



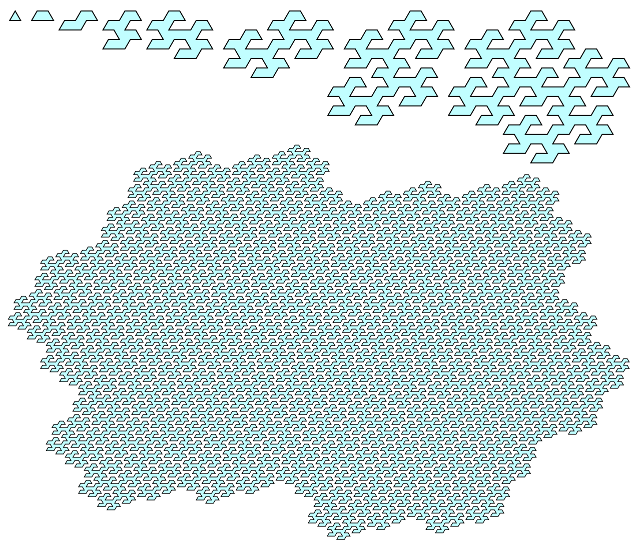
THEY DID THE MATH: The Rauzy fractal is named for Gérard Rauzy. In the picture above, he is in the front row in the red shirt, with a group of mathematicians at Pierre Arnoux's (§ 4.7) apartment in Marseille in 2006. He was especially interested in Fibonacci and tribonacci substitutions, and discovered his eponymous Rauzy fractal that we saw in § 5.3 and will imminently see again.

It so happens that any walk on the triangular grid corresponding to the orbit of a point on the Arnoux-Yoccoz IET, such as the one you computed in Problem 5.15, is unbounded. On the other hand, if you change the Arnoux-Yoccoz IET via a rel deformation by some tiny amount r, any walk corresponding to the orbit of a point on the rel-deformed IET is periodic, with larger periods as $r \to 0$.

5.17. (a) Let $r = 0.06$. Write out the rel-deformed Arnoux-Yoccoz IET, which is a modification of the six-part function in Problem 5.14.
(b) . Compute the orbit of the point $x = 0.6$ on the rel-deformed IET. Show that its period is 18.
(c) Show that the corresponding arithmetic graph to the orbit of $x = 0.6$ is as shown to the right. The red point is the starting point.

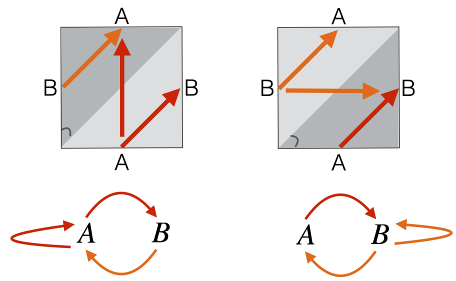


The eight shortest arithmetic graphs corresponding to this construction, and the 15^{th}-shortest, are shown below. It's the Rauzy fractal again!



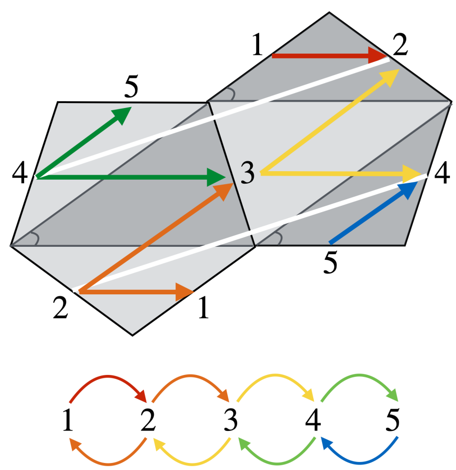
Thank you to Pat Hooper (§ 3.8) for helping to make my dreams for this section become reality.

5.6 Transition diagrams



Recall our friend the square torus, with horizontal and vertical edges labeled A and B, respectively. We have explored many ways of understanding linear trajectories on the square torus, including transforming the geometric problem about trajectories into a combinatorial problem about cutting sequences. *Transition diagrams* inject some geometry back into those cutting sequences, via a flow chart of which characters (edge labels) can follow which others.

In the left picture above, we have restricted trajectories to those that go left to right with slope ≥ 1. For such trajectories, A can be followed by A or B (red arrows), and B can only be followed by A (orange arrow). We represent this information using the transition diagram shown at the bottom of the figure below. For trajectories with slope between 0 and 1, the situation is similar: A can only be followed by B (red arrow), while B can be followed by either A or B (orange arrows).



For the double pentagon, the symmetries of the surface allow us to restrict our attention to trajectories with angle between 0 and $\pi/5$, indicated by the shaded sectors.

5.18. For trajectories within this sector of directions, edge 1 can only be followed by edge 2 (red arrow), because going to any other edge would require going down (e.g. to edge 4 in the right pentagon), or going too steeply upward (e.g. to edge 3 in the left pentagon). By similar logic, edge 2 can only be followed by edge 1 or edge 3 (orange arrows), and so on.
(a) Work through all five edges, and confirm that the transition diagram below the surface accurately reflects the allowed *transitions* for cutting sequences corresponding to such trajectories.
(b) Confirm that the cutting sequence $\overline{2343}$ corresponds to a valid periodic trajectory on the surface (white), and also that it corresponds to a periodic path on the transition diagram.

5.19. Draw the transition diagram for the double pentagon corresponding to the set of trajectories whose angle is between $\pi/5$ and $2\pi/5$.

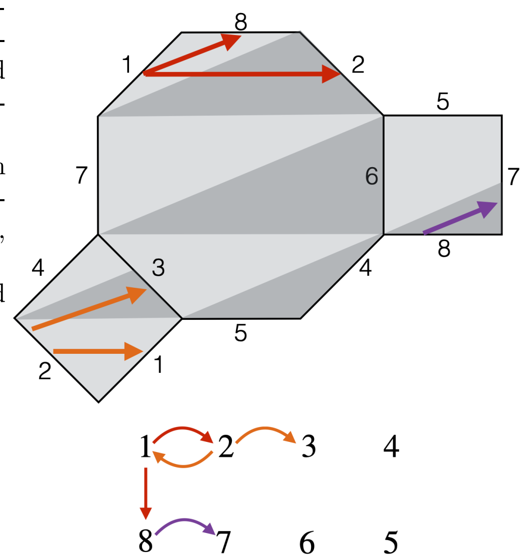
The idea of transition diagrams is that they allow us to determine the direction of a trajectory, without having to draw a picture. For the square torus, if a cutting sequence is valid on the first transition diagram, its slope is greater than 1, while if it is valid on the second transition diagram, the slope is less than 1.

Similarly, for a cutting sequence corresponding to a trajectory on the double pentagon, you can determine which of the five transition diagrams – one was given; you drew a second, and there are three more – it is valid on, and this will tell you the trajectory's direction. This reduces the geometric problem about trajectories and surfaces to a combinatorial problem about symbols, which is much easier to characterize and check.

5.20. Ooh, now things get interesting! Recall the Ward surface from Problem 4.31. We restrict to angles between 0 and $\pi/8$, as suggested by the darkened sectors in the picture.

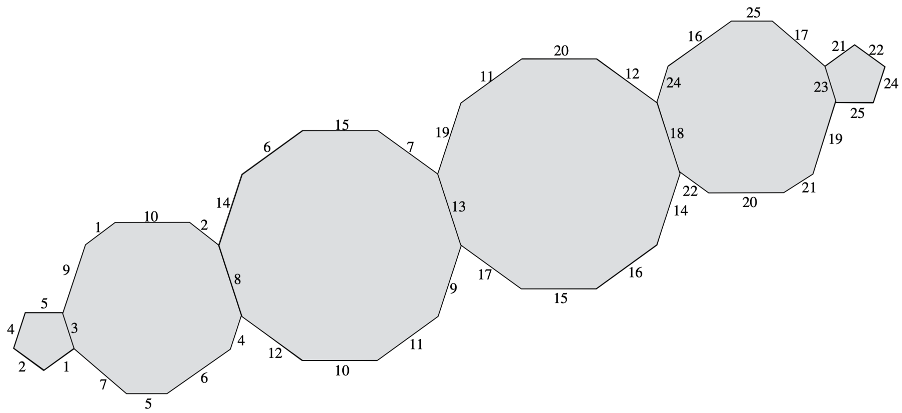
(a) A few of the arrows on the surface, and on the corresponding transition diagram, are given; you fill in the rest.

(b) Explain the logic behind the edge-numbering system.



Everything I know about transition diagrams, I learned from a paper by John Smillie and Corinna Ulcigrai (§ 2.1), *Symbolic coding for linear trajectories in the regular octagon*, which defined them as above and gave the example of the square torus; see their § 1.2 and Figure 2.

I came up with the "zig-zag" numbering scheme myself, and I think it's awesome.



5.21. (Challenge) Recall the $m = 6$, $n = 5$ Bouw-Möller surface from Problem 4.40, shown above.

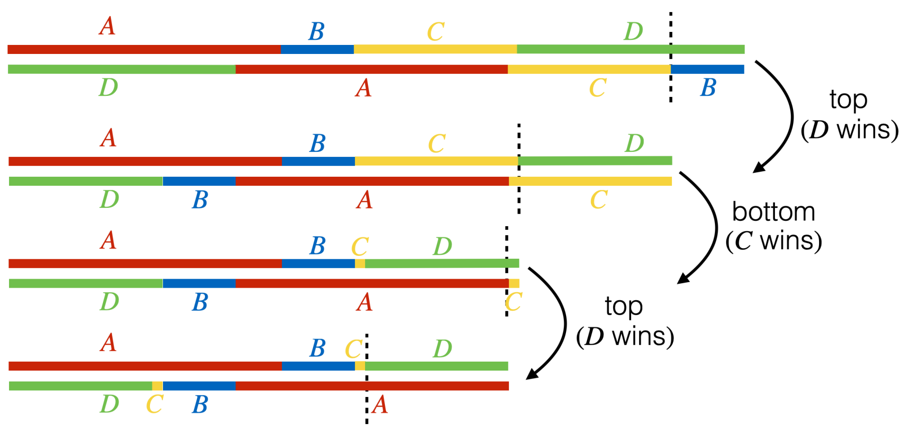
(a) What should the angle restriction be on trajectories for this surface, given its symmetries?

(b) Fill in the rest of the transition diagram. The locations of some of the nodes are given; you figure out where the rest should go, and also the arrows.

5.7 Rauzy-Veech induction

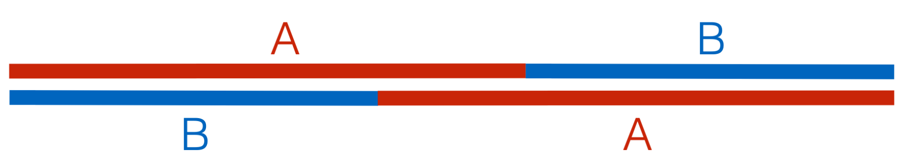
Rauzy-Veech induction is a tool for simplifying, and thus better understanding, the behavior of IETs. We'll use the example of the 4-IET from Problem 4.14, shown on the top line below.

- First, choose one end of the IET or the other. Typically, people choose the right end, as we do here.
- Of the two subintervals that are at the right end, one is longer than the other; this interval is said to "win," and the other one is said to "lose." For our IET, D wins and B loses.
- Chop off the end of the IET, the length of the losing interval. Follow the path of that losing interval for one iteration: Here, B becomes the right end of D. So in the next line, we replace the right end of D with B, to get a new IET.
- Repeat this process with the new IET for as long as you like.



5.22. Draw the diagrams for the next three steps of Rauzy induction on the above IET. Notice that the only intervals that change are the ones that are "fighting": for example, in the first step above, D and B are fighting, so the red and yellow intervals A and C stay exactly the same from the first picture to the second. Can you explain why?

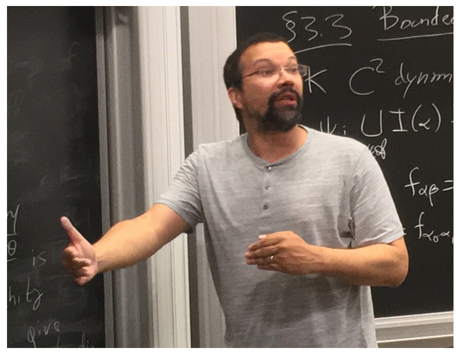
5.23. In the picture below, the length ratio of segments A to B is $7:5$. Perform Rauzy induction on this IET until both segments are the same length. Then compare with your work in Problems 1.9 and 1.23 and explain the relationship between Rauzy-Veech induction and the continued fraction algorithm.



We want to know if we get back to an IET that is combinatorially "the same" as the one we started with. For the 4-IET on the previous page, we can keep track of the Rauzy-Veech induction steps via the following notation:

$$\begin{pmatrix} A\ B\ C\ D \\ D\ A\ C\ B \end{pmatrix} \xrightarrow[\text{(D wins)}]{\text{top}} \begin{pmatrix} A\ B\ C\ D \\ D\ B\ A\ C \end{pmatrix} \xrightarrow[\text{(C wins)}]{\text{bottom}} \begin{pmatrix} A\ B\ C\ D \\ D\ B\ A\ C \end{pmatrix} \xrightarrow[\text{(D wins)}]{\text{top}} \begin{pmatrix} A\ B\ C\ D \\ D\ C\ B\ A \end{pmatrix}$$

5.24. Write down the notation for the next three steps of Rauzy induction, corresponding to your diagrams from Problem 5.22.



THEY DID THE MATH: Carlos Matheus (right) works on dynamical systems and number theory, including IETs, translation surfaces, orbits, strata, and many of the other objects introduced in this book. I took the content of Problems 5.23 and 5.25 from a talk he gave in July 2023. Thank you, Matheus!

5.25. Starting from the IET represented by $\left(\begin{smallmatrix} A & B & C & D \\ D & C & B & A \end{smallmatrix}\right)$, shown in the middle of the diagram below, there are two options: either the top wins, or the bottom wins. Then, starting from each of *those* IETs, there are two options: again, either the top or the bottom wins. We can work out the entire Rauzy-Veech diagram for all of the options, which ends up looking like the below "butterfly." The right side of the diagram is already done; fill in the grey blanks on the left side to complete it.