Notes on Billiards

Rich Schwartz

November 21, 2019

These notes are taken from Chapters 17 and 18 of my book *Mostly Surfaces*. I omitted some of the material so that what is here is more focused on the points I bring up in the lecture.

1 Flat Cone Surfaces

1.1 Sectors and Euclidean Cones

A *sector* in $\boldsymbol{R}^2$ is the closure of one of the 2 components of $\boldsymbol{R}^2 - \rho_1 - \rho_2$, where ρ_1 and ρ_2 are two distinct rays emanating from the origin. For example, the nonnegative quadrant is a sector. The *angle* of the sector is defined as the angle between ρ_1 and ρ_2 as measured from inside the sector. For instance, the angle of the nonnegative quadrant is $\pi/2$.

Two sectors in $\boldsymbol{R}^2$ can be glued together isometrically along one of their edges. A *Euclidean cone* is a space obtained by gluing together, in a cyclic pattern, a finite number of sectors. The *angle* of the Euclidean cone is the sum of the angles of the sectors. The *cone point* is the equivalence class of the origin(s) under the gluing. The cone point is the only point which potentially does not have a neighborhood locally isometric to $\boldsymbol{R}^2$.

Note that two isometric Euclidean cones might have different descriptions. For instance, $\boldsymbol{R}^2$ can be broken into 4 quadrants or 8 sectors of angle $\pi/4$.

Exercise 1. Prove that two Euclidean cones are isometric if and only if they have the same angle.

Exercise 2. Define the unit circle in a Euclidean cone to be the set of points which are 1 unit away from the cone point. On the cone of angle 4π find the shortest path between every pair of points on the unit circle. This problem breaks down into finitely many cases, depending on where the points are located.

Exercise 3. Let C be a Euclidean cone, with cone point x. Say that a vector field on $C - x$ is parallel if an isometry carrying any open set of $C - x$ into $\boldsymbol{R}^2$ carries the vector field to a constant vector field. Prove that $C - x$ has a parallel vector field in a neighborhood of x if and only if the cone angle of C is a multiple of 2π. (*Hint*: Unroll C into the plane and watch the vector field as you go once around the cone point.)

1.2 Translation Surfaces

Say that a compact oriented surface Σ is a *Euclidean cone surface* if it has the following two properties:

- Every point $p \in \Sigma$ has a neighborhood which is isometric to a neighborhood of the cone point in a Euclidean cone of angle $\theta(p)$.
- We have $\theta(p) = 2\pi$ for all but finitely many points.

The points p, where $\theta(p) \neq 2\pi$, are called the *cone points*. The quantity

$$\delta(p) = 2\pi - \theta(p)$$

is called the *angle deficit*. So, there are only finitely many points with nonzero angle deficit, and these deficits could be positive or negative.

Exercise 4. Prove that every Euclidean cone surface has a triangulation.

A *translation surface* is a Euclidean cone surface which admits a parallel vector field which is defined everywhere except at the cone points. By Exercise 3 above, the cone angles of a translation surface are all integer multiples of 2π. At first it might seem that a Euclidean surface whose cone angles are all integer multiples of 2π must admit a parallel vector field, but this is not so. As Rick Kenyon pointed out to me, M. Troyanov constructed some counter examples. See "Les surfaces euclidienne a singularites coniques", by M. Troyanov, published in Enseign. Math (2) 32 (1986), 76-94. You might like to try to find some examples yourself without looking up Troyanov's article.

Recall that a *gluing diagram* for a surface is a list of finitely many polygons, together with a recipe for gluing together the sides of the polygon in pairs.

Lemma 1.1 *Suppose that S is a flat cone surface obtained from a gluing diagram in which the two sides in each glued pair are parallel. Then S is a translation surface.*

Proof: Once we show that S is orientable, we will know that S is a cone surface. On each polygon, we consider the standard pair of vector fields V_1 and V_2. Here V_j consists of vectors parallel to the basis vector e_j. Given the nature of the gluing maps, the vector fields piece together across the edges to give parallel vector fields V_1 and V_2 defined on the complement of finitely many points.

We first show that S is orientable. If S is not orientable, then S contains a Möbius band M. By shrinking M if necessary, we can arrange that M lies entirely in the region where both V_1 and V_2 are defined. But then we can define a continuous pair of linearly independent vector fields on a Möbius band. This is easily seen to be impossible. Hence S is oriented.

It now follows from definition that S is a translation surface. ♠

In light of Lemma 1.1, the surface obtained by gluing (with translations) the opposite sides of a regular $2n$-gon is a translation surface.

Translation Principle. Whenever we consider gluing diagrams for translation surfaces, in which more than one polygon is involved, we always think of the polygons in the plane as being pairwise disjoint. How the polygons sit in the plane is really not so important, in the following sense. Suppose that $P_1, \ldots, P_n$ are the polygons involved in a gluing diagram for some surface. Suppose that $Q_1, \ldots, Q_n$ are new polygons, such that Q_k is a translation of P_k for all k, and the pattern of gluing for the Q's is the same as the pattern of gluing for the Ps. Then the two resulting surfaces are canonically isometric. The canonical isometry is obtained by piecing together the translations that carry each P_k to Q_k. We mention this rather obvious principle because it guarantees that certain constructions, which seem based on arbitrary choices, are actually well defined independent of these choices.

1.3 Billiards and Translation Surfaces

Let P be a Euclidean polygon. A *billiard path* in P is the motion taken by an infinitesimal frictionless ball as it rolls around inside P, bouncing off the walls according to the laws of inelastic collisions: the angle of incidence equals the angle of reflection; see Figure 17.1 below. We make a convention that a path stops if it lands precisely at a vertex. (The infinitesimal ball falls into the infinitesimal pocket.)

The billiard path is *periodic* if it eventually repeats itself. Geometrically, a periodic billiard path corresponds to a polygonal path Q with the following properties:

- $Q \subset P$ (that is, the solid planar region).
- The vertices of Q are contained in the interiors of the edges of P.
- Q obeys the angle of incidence rule discussed above.

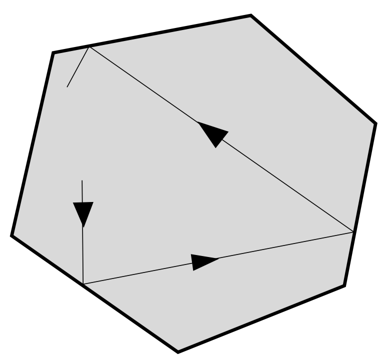

Figure 17.1. Polygonal billiards

Exercise 5. Find (with proof) all the examples of periodic billiard paths in a square which do not have self-intersubsections. So, the path Q has to be embedded.

The polygon P is called *rational* if all its angles are rational multiples of π. For instance, the equilateral triangle is a rational polygon. In this section I will explain how to associate a translation surface to a rational polygon. This is a classical construction, attributed by some people to A. Katok and A.N. Zemylakov. The geometry of the translation surface encodes many of the features of billiards in the polygon.

For each edge e of P there is a reflection R_e in the line through the origin parallel to e. Like all reflections, R_e has order 2. That is, $R_e \circ R_e$ is the identity map. Let G be the group generated by the elements $R_1, \ldots, R_n$. Here R_j stands for R_{e_j} and $e_1, \ldots, e_n$ is the complete list of edges. If e_i and e_j are parallel, then $R_i = R_j$. If P is a rational polygon then, after we suitably rotate P, there is some N such that e_j is parallel to some Nth root of unity. But then G is a group of order at most $2N$. In particular, G is a finite group. For each $g \in G$, we define a polygon

$$P_g = g(P) + V_g. \tag{1}$$

Here V_g is a vector included so that all the polygons $\{P_g |\ g \in G\}$ are disjoint. Thanks to the Translation Principle, the surface we will produce is independent of the choices of the translation vectors.

To form a gluing diagram, we declare that every two edges of the form

$$e_1 = g(e) + V_g, \qquad e_2 = gr(e) + V_{gr}, \qquad r = R_e. \tag{2}$$

are glued together by a translation. Here e is an arbitrary edge of P. Since $gr(e) = g(e)$, the edges e_1 and e_2 are parallel. Hence, it makes sense to glue them by a translation. Note also that $(gr)r = g$. So, our instructions tell us to glue e_1 to e_2 if and only if they tell us to glue e_2 to e_1. Let $\widehat{P}$ be the space obtained from the gluing diagram. Since the edges are glued in pairs, $\widehat{P}$ is a surface. By Lemma 1.1, $\widehat{P}$ is a translation surface.

Here we work out the example where P is an isosceles triangle with small angles $\pi/8$. In this case, the group G has order 16 and our surface will be made from 16 isometric copies of P.

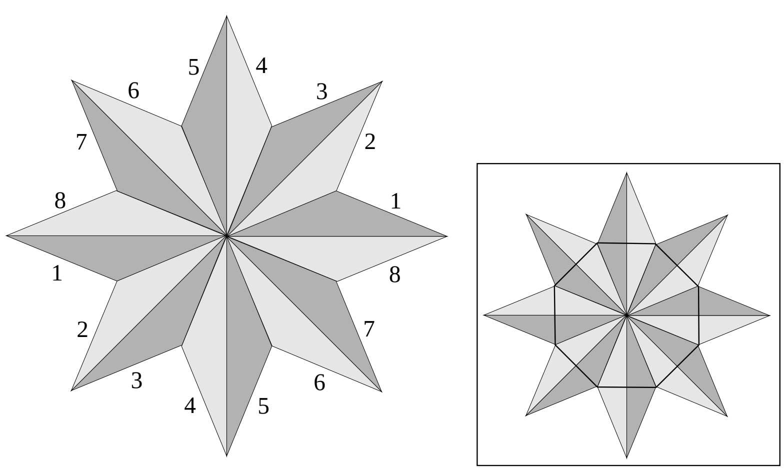

Figure 17.2. Gluing diagram for a translation surface

Figure 17.2 shows the resulting gluing diagram. We have chosen the translations so that all the long sides have already been glued together. Also, we have colored the triangles alternately light and dark so as to better show the pattern. The numbers around the outside of the figure indicate the gluing pattern for the short edges.

The gluing pattern in Figure 17.2 has an alternate description. Take two regular Euclidean octagons and glue each side of one to the opposite side of the other. The smaller inset picture in Figure 17.2 shows one of the two octagons. The other octagon is splayed open, and made by gluing together the pieces that are outside the octagon shown.

Let $\widehat{P}$ be the translation surface constructed above. A path $\gamma \in \widehat{P}$ is called *straight* if every point $p \in \gamma$ has a neighborhood U with the following property: Any isometry between U and a subset of $\boldsymbol{R}^2$ maps $\gamma \cap U$ to a straight line segment. (For concreteness we can always take U to be a little Euclidean ball centered at p.) There is an obvious map $\pi : \widehat{P} \to P$. We just forget the group element involved. This forgetting respects the way we have done the gluing and so π is a well-defined continuous map from $\widehat{P}$ to P. The map π is somewhat like a covering map, except that it is not locally a homeomorphism around points on the edges or vertices.

Lemma 1.2 *Suppose $\widehat{\gamma}$ is a straight path on $\widehat{P}$ which does not go through any vertices of $\widehat{P}$. Then $\gamma = \pi(\widehat{\gamma})$ is a billiard path on P.*

Proof: By construction γ is a polygonal path whose only vertices are contained in the interiors of edges of P. We just have to check the angle incidence condition at each vertex. You can see why this works by building a physical model: Take a piece of paper and make a crease in it by folding it in half (and then unfolding it.) Now draw a straight line on the paper which crosses the crease. This straight line corresponds to a piece of $\widehat{\gamma}$ which crosses an edge. When you fold the paper in half you see the straight line turn back at the crease and bounce like a billiard path. This folded path corresponds to γ. ♠

The converse is also true:

Lemma 1.3 *Suppose that γ is a billiard path on P. Then there is a straight path $\widehat{\gamma}$ on $\widehat{P}$ such that $\pi(\widehat{\gamma}) = \gamma$.*

Proof: We use the fact that the map π is almost a covering map. Think of γ as a parametrized path $\gamma : \boldsymbol{R} \to P$, with $\gamma(0)$ contained in the interior of P. We define $\widehat{\gamma}(0)$ to be the corresponding interior point of P_g, where $g \in G$ is any initial element of G we like. We can define $\widehat{\gamma}(t)$ until the first value $t_1 > 0$ such that $\gamma(t_1)$ lies on an edge, say e_1, of P. But then we can define $\widehat{\gamma}$ in a neighborhood of t_1 in such a way that $\widehat{\gamma}(t_1 - s) \in P_g$ and $\widehat{\gamma}(t + s) \in P_{rg}$ for $s > 0$ small, where r is reflection over side e_1. If you think about the folding construction described in the previous lemma, you will see that the straight path $\widehat{\gamma}(t_1 - \epsilon, t_1 + \epsilon)$ projects to $\gamma(t_1 - \epsilon, t_1 + \epsilon)$. Here ϵ is some small value which depends on the location of $\gamma(t_1)$. We can define $\widehat{\gamma}$ for $t > t_1$ until we reach the next time t_2 such that $\gamma(t_2)$ lies in an edge of P. Then we repeat the above construction for parameter values in a neighborhood of t_2. And so on. This process continues indefinitely, and defines $\widehat{\gamma}$ for all $t \geq 0$. Now we go in the other direction and define $\widehat{\gamma}$ for all $t < 0$. ♠

Note that $\widehat{\gamma}$ is a closed loop in $\widehat{P}$ if and only if γ is a periodic billiard path. Thus, the closed straight loops in $\widehat{P}$ correspond, via π, to periodic billiard paths in P.

Exercise 6. Suppose that P is the regular 7-gon. What is the Euler characteristic of $\widehat{P}$? As a much harder problem, can you find a formula for the Euler characteristic of $\widehat{P}$ as a function of the angles of P?

Exercise 7. The same construction can be made when P has some irrational angles. What do you get if P is a right triangle with the two small angles irrational multiples of π?

1.4 Affine Automorphisms

Recall that an *affine map* of $\boldsymbol{R}^2$ is a map of the form $x \to Ax + B$, where A is a 2×2 invertible and orientation-preserving matrix and B is another vector. If $B = 0$, then the map is linear. Note that the set of affine maps of $\boldsymbol{R}^2$ forms a group under composition.

Suppose that Σ is a translation surface. An affine automorphism of Σ is a homeomorphism $\phi : \Sigma \to \Sigma$ such that the following hold:

- ϕ permutes the nontrivial cone points of Σ.

- Every ordinary point of Σ has a neighborhood in which ϕ is an affine map.

The second condition needs a bit more explanation. Let $p \in \Sigma$ be an ordinary point. This is to say that there is a small disk Δ_p about p and an isometry I_p from Δ_p to a small disk in $\boldsymbol{R}^2$. The same goes for the point $q = \phi(p)$. The map $I_q \circ \phi \circ I_p^{-1}$ is defined on the open set $U = I_p(\Delta_p) \subset \boldsymbol{R}^2$ and maps it to another open set $I_q(\Delta_q) \subset \boldsymbol{R}^2$. The second condition says that this map is the restriction of an affine map to U.

We denote the set of all affine automorphisms of Σ as $A(\Sigma)$. It is easy to see that the composition of two affine automorphisms of Σ is again an affine automorphism. Likewise, the inverse of an affine automorphism of Σ is an affine automorphism of Σ. In short, $A(\Sigma)$ is a group.

Exercise 8. This exercise is an important one. Let Σ be the square torus. You can think of Σ as $\boldsymbol{R}^2/\boldsymbol{Z}^2$. That is, we say that two points of $\boldsymbol{R}^2$ are equivalent if their difference is an integer vector, and Σ is the space of equivalence classes. Let $[p] \in \Sigma$ denote the equivalence class of $p \in \boldsymbol{R}^2$. Let A be a 2×2 matrix with integer entries and determinant 1. Let B any vector. Let ϕ be the map $\phi([x]) = [Ax + B]$. Prove that ϕ is an affine automorphism of Σ. Thus, the square torus has a huge affine automorphism group.

Exercise 9. Give an example of a translation surface which has no nontrivial affine automorphisms.

1.5 The Differential Representation

Let $SL_2(\boldsymbol{R})$ denote the group of 2×2 matrices having real entries and determinant 1. Here we explain a canonical representation $\rho : A(\Sigma) \to SL_2(\boldsymbol{R})$. The basic property of Σ we use is that there are canonical identifications between any pair of tangent planes $T_p(\Sigma)$ and $T_q(\Sigma)$, defined as follows: By definition of translation surfaces, there exists a parallel vector field on $\Sigma - C$, where C is the set of cone points. Given $p, q \in \Sigma - C$, we can find an isometry I from a neighborhood of p to a neighborhood of q such that $I(p) = q$. If we insist that I preserves both the orientation and the parallel field, then I is unique. Moreover, I is independent of the choice of parallel field. The differential dI isometrically maps $T_p(\Sigma)$ to $T_q(\Sigma)$. We set $\phi_{pq} = dI$. So, in short

$$\phi_{pq} : T_p(\Sigma) \to T_q(\Sigma) \tag{3}$$

is a canonical isometry. One immediate consequence of our definition is that

$$\phi_{pr} = \phi_{qr} \circ \phi_{pq}, \qquad \phi_{qp} = \phi_{pq}^{-1}. \tag{4}$$

Now, given an element $f \in A(\Sigma)$ we choose and ordinary point $p \in \Sigma$, and let $q = f(p)$. Let df_p be the differential of f at p. This means that df_p is a linear map from $T_p(\Sigma)$ to $T_q(\Sigma)$. Note that the composition

$$M(f,p) = \phi_{qp} \circ df_p$$

is a linear isomorphism from $T_p(\Sigma)$ to itself. Using the isometry I_p from a neighborhood of p to a neighborhood of the origin in $\boldsymbol{R}^2$, we can identify $T_p(\Sigma)$ with, say, the tangent plane to $\boldsymbol{R}^2$ at the origin. We let $\rho(f)$ be the linear transformation of $\boldsymbol{R}^2$ which corresponds to $M(f,p)$ under the identification.

We claim that $\rho(f)$ is independent of the choice of point p. To see this, we note that the map $\rho(f)$ has the following alternate description. Using the coordinate charts I_p and I_q discussed above, the map $\rho(f)$ is just the linear part of

$$dI_q \circ df_p \circ dI_p^{-1}.$$

The linear part of an affine map does not depend on the point. Hence $\rho(f)$ has the same definition independent of which point we use inside our local coordinate chart. But the surface is connected, so $\rho(f)$ does not depend on the choice of point at all.

The determinant of $\rho(f)$ measures the factor by which f increases area in a neighborhood of any point. Since the whole surface has finite area and $\rho(f)$ is an automorphism, $\rho(f)$ must have determinant 1. Hence we can interpret $\rho(f)$ as an element of $SL_2(\boldsymbol{R})$. The map $f \to \rho(f)$ is a homomorphism because of the chain rule: The linear differential of a composition of maps is just the composition of the linear differentials of the individual maps. And composition of linear maps is the same thing as matrix multiplication in $SL_2(\boldsymbol{R})$.

We have now constructed the representation $\rho : A(\Sigma) \to SL_2(\boldsymbol{R})$. We let $V(\Sigma) = \rho(A(\Sigma))$. The matrix group $V(\Sigma)$ is sometimes called the *Veech group*.

1.6 Connection to Hyperbolic Geometry

Recall that $\boldsymbol{H}^2$ is the hyperbolic plane. We work in the upper half plane model. Every element of $SL_2(\boldsymbol{R})$ acts on $\boldsymbol{H}^2$ isometrically. In particular, the Veech group V acts isometrically on $\boldsymbol{H}^2$. The group V is said to act *properly discontinuously* on $\boldsymbol{H}^2$ if, for every metric ball $B \subset \boldsymbol{H}^2$, the set

$$\{g \in V| \ g(B) \cap B \neq \emptyset\}$$

is a finite set. In other words, all but finitely elements of V have such a drastic action on $\boldsymbol{H}^2$ that they move the ball B completely off itself. In this case, $\boldsymbol{H}^2/V$ is a hyperbolic orbifold.

Theorem 1.4 *If V is the Veech group of a surface, then V acts properly discontinuously on $\boldsymbol{H}^2$.*

We first take care of a trivial case of Theorem 1.4.

Exercise 10. Suppose that Σ is a translation surface with no cone points. Prove that Σ is isometric to a flat torus.

Exercise 11. Prove Theorem 1.4 in the case when the surface has no cone points.

From now on, we consider the case when Σ has at least one cone point. In this case, Σ is homeomorphic to a surface having negative Euler characteristic. Let C be the set of cone points of Σ. We call a map $\gamma : [0, 1] \to \Sigma$ a *saddle connection* if the following hold.

- $\gamma(t) \in C$ if and only if $t = 0, 1$.
- The restriction of γ to $(0, 1)$ is locally a straight line.

Exercise 12. Prove that Σ has a pair of non-parallel saddle connections that intersect at a point of $\Sigma - C$.

Lemma 1.5 *Let f be an affine automorphism of Σ. Let γ_1 and γ_2 be a pair of saddle connections, as in Exercise 8. Suppose f preserves the endpoints of γ_1 and γ_2, and $f(\gamma_j) = \gamma_j$ for $j = 1, 2$. Then f is in the kernel of the differential representation ρ.*

Proof: The restriction of an affine map to a straight line that is mapped to itself is just a dilation. Hence, the restriction of f to γ_j is just a dilation. Since $f(\gamma_j) = \gamma_j$, the dilation factor must be one: the total length is preserved. So f is the identity on γ_j.

Let p be an intersubsection point of γ_1 and γ_2. We know that $f(p) = p$. Since γ_1 and γ_2 are nonparallel, we see that df_p fixes two independent directions at p. Hence df_p is the identity. But then $\rho(f)$ is the identity. ♠

We will assume that the Veech group $V = V(\Sigma)$ is not properly discontinuous and we will derive a contradiction. Suppose that there is some ball $B \subset \boldsymbol{H}^2$ and an infinite collection $\{g_i\} \subset V$ such that $g_i(B) \cap B \neq \emptyset$. Thinking of $SL_2(\boldsymbol{R})$ as a subset of $\boldsymbol{R}^4$ by stringing out the coordinates of the matrices, we note that the elements in our set $\{g_i\}$ comprise an infinite bounded subset of $\boldsymbol{R}^4$. An infinite bounded subset of a Euclidean space always has an accumulation point. But this means that we can find a sequence of compositions of the form $g_i g_j^{-1} \in V$ which converge to the identity element.

What this means in terms of Σ is that we can find an infinite sequence $\{f_j\}$ of affine automorphisms such that $\rho(f_i)$ is not the identity but $\rho(f_i)$ converges to the identity as $i \to \infty$. All these elements permute the set of cone points somehow. So, by taking suitable powers of our elements, we can assume that each f_i fixes each cone point of Σ.

Let γ_1 and γ_2 be the saddle connections from Exercise 8. The segment $f_k(\gamma_1)$ is another saddle connection that connects the same two cone points as does γ_1. For k large, $f_k(\gamma_1)$ and γ_1 nearly point in the same direction and nearly have the same length. If they do not point in exactly the same direction, they cannot connect the same two endpoints. The two paths start out at the same cone point but then slowly diverge, so that one of them misses the cone point at the other end. Figure 18.1 shows what we mean.

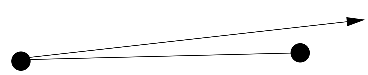

Figure 18.1. Nearly parallel paths

This means that $f_k(\gamma_1)$ and γ_1 point in exactly the same direction for k large. But then $f_k(\gamma_1) = \gamma_1$. The same argument shows that $f_k(\gamma_2) = \gamma_2$ for k large. But then, by the previous result, $\rho(f_k)$ is the identity for large k. This contradiction finishes the proof.

1.7 Triangle Groups

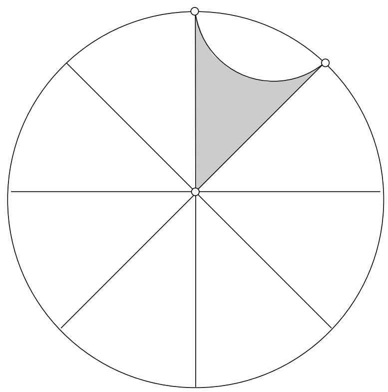

Figure 18.2. The hyperbolic triangle of interest

Recall that a geodesic hyperbolic triangle is a triangle in $\boldsymbol{H}^2$ whose sides are either geodesic segments, geodesic rays, or geodesics. The case of interest to us is the geodesic triangle with 2 ideal vertices and one other vertex having interior angle $2\pi/8$. Figure 18.2 shows a picture of the triangle we mean, drawn in the disk model. This triangle is known as the $(4, \infty, \infty)$ triangle.

Lemma 1.6 *Let γ be any geodesic in $\boldsymbol{H}^2$. Then there is an order 2 hyperbolic isometry which fixes γ.*

Proof: Thinking of $\boldsymbol{H}^2$ as the upper half-plane, the map $z \to -\overline{z}$ fixes the imaginary axis, which is a geodesic. We have already seen that any two geodesics are isometric to each other. If g is an isometry taking the geodesic γ_1 to the geodesic γ_2 and I is an order 2 isometry fixing γ_1, then gIg^{-1} is the desired order 2 isometry fixing γ_2. Thus, we can start with the one reflection desribed above and construct all the others by conjugation. ♠

The order 2 hyperbolic isometry fixing γ is called a *hyperbolic reflection* in γ. Given any geodesic triangle Δ, we can form the group $G(\Delta) \subset SL_2(\boldsymbol{R})$ as follows. We let I_1, I_2, I_3 be hyperbolic reflections fixing the 3 sides of Δ and then we let $G(\Delta)$ be the group generated by words of even length in I_1, I_2, I_3. For instance, I_1I_2 and $I_1I_2I_1I_3$ all belong to G but $I_1I_2I_3$ does not. All the elements in G are orientation preserving and it turns out that we can find matrices in $SL_2(\boldsymbol{R})$ for the elements I_1I_2, I_2I_3, and I_3I_1. This is enough to show that G actually comes from a subgroup of $SL_2(\boldsymbol{R})$.

1.8 Linear and Hyperbolic Reflections

As preparation for the Veech group example we will work out, we discuss how to convert between certain linear maps as they act on $\boldsymbol{R}^2$ and the corresponding linear fractional actions on $\boldsymbol{H}^2$.

We will say that a *linear reflection* is a linear transformation $T : \boldsymbol{R}^2 \to \boldsymbol{R}^2$ such that $T(v) = v$ and $T(w) = -w$ for some basis $\{v, w\}$ of $\boldsymbol{R}^2$. The corresponding hyperbolic isometry acting on $\boldsymbol{H}^2$ is a hyperbolic reflection. To make this correspondence explicit, we first consider the matrix representation of T in the standard basis of $\boldsymbol{R}^2$. Necessarily, the matrix will have a negative determinant since linear reflections are orientation reversing. Then, given a linear reflection T and its matrix representative

$$M = \begin{matrix} a & b \\ c & d \end{matrix},$$

the corresponding hyperbolic isometry is

$$z \to \frac{a\bar{z} + b}{c\bar{z} + d}.$$

This can be seen by considering the special case when $v = (1, 0)$ and $w = (0, 1)$ which corresponds to the hyperbolic reflection $z \mapsto -\bar{z}$. All other cases are conjugate to this one.

The map T is determined by the pair (v, w), but more than one basis determines T. The basis $(C_1 v, C_2 w)$ also determines T, where C_1 and C_2 are any 2 nonzero constants. For this reason, it is really the pair (L_1, L_2) that determines T, where L_1 is the line through v and L_2 is the line through w. The map T fixes L_1 pointwise and reverse L_2.

The map $-T$ fixes L_2 pointwise and reverses L_1. For this reason, the unordered pair $\{L_1, L_2\}$ determines the pair of maps $\{T, -T\}$. The map $\pm T$ corresponds to a hyperbolic reflection, and each hyperbolic reflection corresponds to a pair $\pm T$ of maps. In short, each hyperbolic reflection is determined by an unordered pair $\{L_1, L_2\}$ of lines through the origin. We call such a pair of lines a *cross*.

Let us first consider the case when L_1 and L_2 are perpendicular. In this case, we call $\{L_1, L_2\}$ a *plus*, because the two lines look like a + symbol, up to rotation. If we work in the disk model Δ of the hyperbolic plane, we can normalize so that the hyperbolic reflections corresponding to pluses all fix some geodesic through the origin in $\boldsymbol{C}$. Figure 18.3 shows two examples.

On the left-hand side of Figure 18.3 we show two pluses, one drawn thickly and one drawn thinly. On the right hand side of Figure 18.3, we show the geodesics in Δ fixed pointwise by the corresponding hyperbolic reflections.

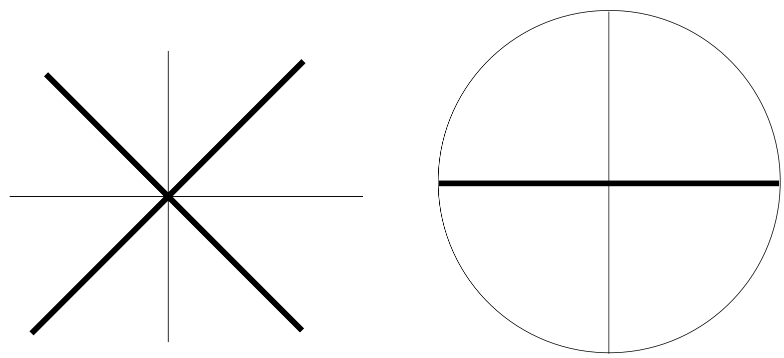

Figure 18.3. Euclidean and hyperbolic reflections

Exercise 12. This is a 3-part exercise. Let Δ denote the disk model of the hyperbolic plane. (Actually, the results of the exercise do not depend on which model is used, but we use the disk model for concreteness.)
(a) Let θ be the smallest angle between the lines of one plus and the lines of another. Prove that the corresponding geodesics in Δ meet at an angle of 2θ.
(b) Prove that an orientation preserving linear transformation which pointwise fixes a line through the origin corresponds to a hyperbolic isometry which fixes a point on $\partial\Delta$ and no point in Δ. Such hyperbolic isometries are called *parabolic*.
(c) Prove that the geodesics fixed by a pair of hyperbolic reflections have an ideal endpoint on $\partial\Delta$ in common if and only if the product of the hyperbolic reflections is parabolic.

In light of Exercise 12, we can draw 3 crosses whose corresponding geodesics in Δ are three sides of the $(4, \infty, \infty)$ triangle shown in Figure 18.2. Two of the crosses are pluses and one is not. The crosses are drawn thickly, and the thin lines are present for reference. The thin lines are evenly spaced in the radial sense.

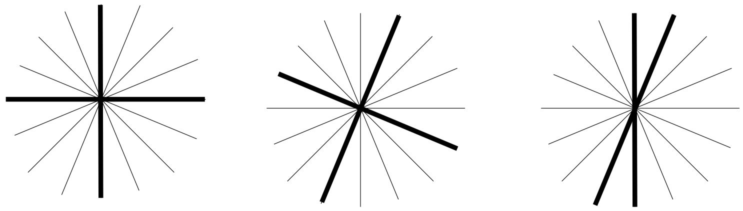

Figure 18.4. Three special crosses

To explain why this works, we first identify $\boldsymbol{H}^2$ with the disk model Δ, as above. Let $\pm T_1$, $\pm T_2$, and $\pm T_3$ be the (pairs of) linear reflections corresponding to each of the three crosses. Let R_j be the hyperbolic reflection corresponding to $\pm T_j$. Let γ_j be the geodesic fixed by R_j. By construction, γ_1 and γ_2 are geodesics through the origin in Δ. By Exercise 9a, γ_1 and γ_2 meet at an angle of $2\pi/8$. Thus, if we suitably rotate Δ, then γ_1 and γ_2 are exactly the geodesics through the origin in Figure 18.2.

To finish the argument, we just have to see that γ_3 has an ideal endpoint in common with each of γ_1 and γ_2. Consider γ_1 and γ_3. If the signs are chosen appropriately, the element T_1T_3 pointwise fixes the vertical line through origin, the line common to the two relevant crosses. By Exercise 9b, the product R_1R_3 is parabolic. By Exercise 9c, γ_1 and γ_3 have a a common ideal endpoint. The same argument works for R_2 and R_3.

We have gone through all this trouble because we want to recognize the $(4,\infty,\infty)$ triangle group as a subgroup of the group of all affine automorphisms of a certain translation surface. We will work this out in the next subsection.

1.9 Cylinders and Dehn Twists

A *flat cylinder* (or just *cylinder* for short) is a surface-with-boundary that is isometric to a quotient of an infinite strip in $\boldsymbol{R}^2$ by translation. Cylinders are not quite translation surfaces, because they have boundaries, but otherwise they are pretty close. In the next subsection we will see cylinders arising naturally as subsets of translation surfaces.

Even though a cylinder has a boundary, we can define what it means to have an affine automorphism of a cylinder. This is a homeomorphism of the cylinder which is locally affine on the interior. Likewise, we can say that it means for there to be an affine isomorphism between two different cylinders. Such a map is a homeomorphism between the two cylinders which is locally affine on the interior. There is an affine isomorphism from any one cylinder to any other.

Two cylinders C_1 and C_2 are called *similar* if there is an affine isomorphism $T : C_1 \to C_2$ such that dT is a similarity at each point. Not all cylinders are similar to each other. Informally speaking, you can have a "long and thin" cylinder and it will not be similar to a "short and fat" cylinder.

A *Dehn twist* of a cylinder C is a nontrivial affine automorphism $T : C \to C$ which fixes the boundary pointwise. At first glance, it is amazing that such a map could exist. Here is the prototypical example: Let C be the quotient of the horizontal strip $\boldsymbol{R} \times [0,1]$ by the equivalence

$$(x, y) \sim (x + n, y), \qquad \forall\, n \in \boldsymbol{Z}.$$

In other words, C is the quotient of a horizontal strip by integer horizontal translations. The map

$$T(x, y) = (x + y, y)$$

preserves $\boldsymbol{R} \times [0,1]$ and respects the equivalence relation. So does T^{-1}. Hence T induces an affine automorphism $[T]$ of C. Since T fixes the line $\boldsymbol{R} \times \{0\}$ pointwise, $[T]$ fixes the bottom boundary component of C. On the line $\boldsymbol{R} \times \{1\}$ we have $T(x,1) = (x+1,1) \sim (x,1)$. Hence $[T]$ also pointwise fixes the top boundary component of C. On the other hand, $[T]$ obviously acts in a nontrivial way, because dT is not the identity.

Call two linear transformations T_1 and T_2 *similar* if $T_2 = ST_1S^{-1}$ for some similarity S. Two similar cylinders C_1 and C_2 admit Dehn twists $g_j : C_j \to C_j$ whose differentials dg_1 and dg_2 are similar. If we know that dg_1 and dg_2 fix a common line through the origin in $\boldsymbol{R}^2$ we can say more strongly that $dg_1 = dg_2$.

1.10 Behold, The Double Octagon!

We will compute the Veech group of the translation surface associated to the Euclidean isosceles triangle having small angle $\pi/8$. As we saw in §1.3, this surface is obtained from a gluing diagram involving two regular Euclidean octagons. Each side of one octagon is glued to the opposite side of the other. Let Σ be this surface.

Theorem 1.7 *$V(\Sigma)$ is the even subgroup of the $(4, \infty, \infty)$ reflection triangle group.*

The $(4, \infty, \infty)$ triangle group is the group generated by the three hyperbolic reflections R_1, R_2, R_3 considered in the previous subsection. The *even subgroup* consists of elements made from composing an even number of these elements. Equivalently, the even subgroup is the subgroup consisting of orientation preserving elements. The even subgroup has index 2 in the whole group.

We will sketch a proof of Theorem 1.7. To make things work well, we define an *anti-affine automorphism* to be a homeomorphism of Σ which is locally anti-affine, meaning that the map locally has the form $x \to L(x) + C$, where L is an orientation-reversing linear map and C is some constant vector. The linear reflections considered in the previous subsection are of this form.

Let $\widehat{A}(\Sigma)$ be the group of these maps, and let $\widehat{V} = \rho(\widehat{A})$, where ρ is the differential representation as above. We will show that $\widehat{V}$ coincides with the group $\widehat{G}$ generated by the reflections in the sides of the $(4, \infty, \infty)$ triangle. The even elements of $\widehat{A}$ are orientation preserving and the rest are orientation reversing. So, the Veech group corresponds to the images of the even elements.

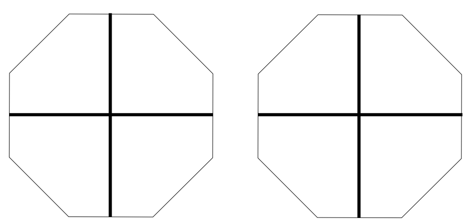

Figure 18.5. The first cross

Figure 18.5 shows the octagons involved in the gluing diagram for Σ. Again, each side of the left octagon is glued to the opposite side of the right octagon by a translation. Simultaneous reflection in the vertical sides of Σ induces an element T_1 of $\widehat{A}$. The differential of this map, evaluated at the center of the first octagon, fixes the vertical line through the center and reverses the horizontal line. The element $\pm dT_1$ therefore corresponds to the first plus in Figure 18.4. Hence $\rho(\pm T_1) = R_1$. Figure 18.6 does for R_2 what Figure 18.5 does for R_1. Here we take T_2 to be simultaneous reflection in the diagonals of positive slope.

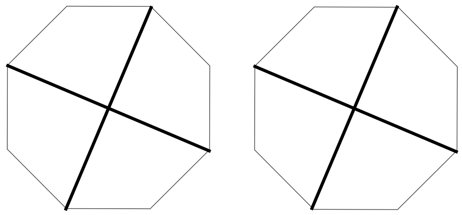

Figure 18.6. The second cross

So far we have used fairly trivial symmetries of our surface. Now we have to do something nontrivial to see the anti-affine automorphism that corresponds to the third cross. Figure 18.7 shows the cross $\{L_1, L_2\}$ we are aiming for, drawn on one of the octagons. The extra line L_3 will be explained momentarily.

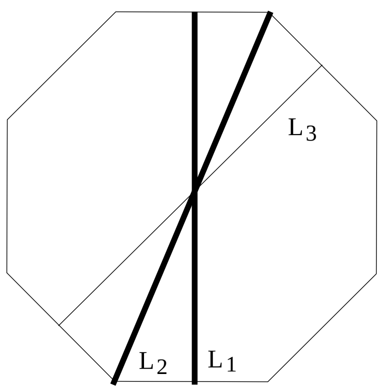

Figure 18.7. The third cross

We will produce an (anti-affine) automorphism $g : \Sigma \to \Sigma$ such that g fixes L_2 pointwise and $g(L_1) = L_3$ in a length-preserving and height-reversing way. That is, g maps the top vertex of L_1 to the bottom vertex of L_3. At the same time, the map T_2 fixes L_2 pointwise and maps L_3 to L_1 in a length-preserving way and height-preserving way. But then the composition $T_3 = T_2 \circ g$ fixes L_2 pointwise and reverses L_1. By construction, the maps $\pm T_3$ correspond to our third cross. We set $R_3 = \rho(\pm T_3)$, and we have the desired map.

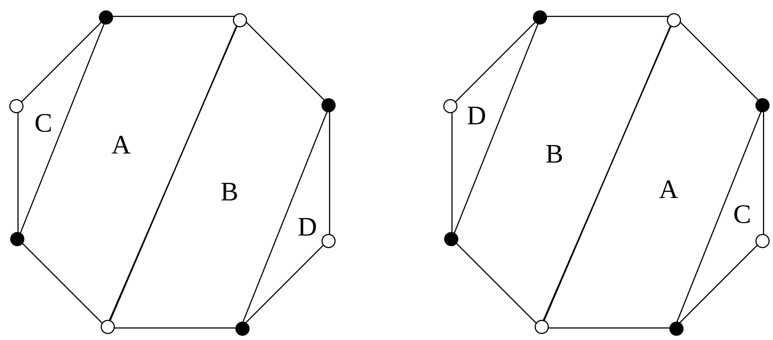

Figure 18.8. Cylinder decomposition

Now we turn our attention to the construction of the map g. Figure 18.8 shows a decomposition of Σ into 4 flat cylinders, labelled A, B, C, D. Remember, each side of the left octagon is glued to the opposite side of the right octagon. Thus, for instance, the two A pieces on the left and right glue together to make the A cylinder. The A and B cylinders are isometric to each other and the C and D cylinders are isometric to each other. Here is the miracle that makes everything work.

Exercise 13. Prove that the A and C cylinders are similar to each other. Hence, all 4 cylinders are similar to each other.

For starters, we say in advance that g commutes with the symmetry which swaps the two octagons. Figure 18.9 shows how g acts on one of the octagons, at least in a neighborhood of the line L_2. We will arrange that g maps the points labelled x to the points labelled y, in the manner suggested by the arrows. These points are at the midpoints of the relevant edges.

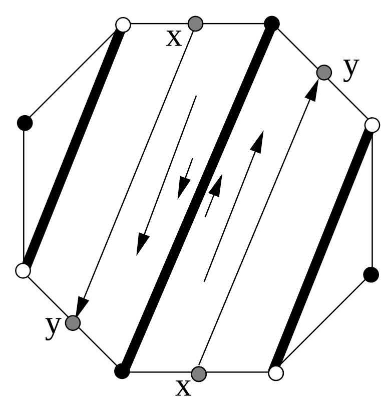

Figure 18.9. Action of the automorphism

Assume for the moment that there really is a locally affine automorphism of Σ that has this action. That is, assume that g really exists. By construction g fixes L_2 pointwise and g maps L_1 to L_3 in a length-preserving and height-reversing way. The point is that L_1 connects the two x points shown in Figure 18.9 and L_3 connects the two y points shown in Figure 18.9.

It only remains to show that g actually exists. We will construct g by doing Dehn twists on each cylinder and showing that they patch together correctly. First of all, we define g in a neighborhood of the "centerline" L_2. We start extending g outward until it is defined on the A cylinder. The lines connecting the x points to the y points glue together to form the central loops of the A and B cylinders. By construction g shifts these central loops half way around. The other boundary component of A is twice as far from the centerline as is the central loop of A. Arguing by proportionality, we see that g shifts the other boundary component of A "all the way around", which is to say that g pointwise fixes the other boundary component of A. In short, g is a Dehn twist of A. The same goes for B.

Since C is similar to A, we know that C admits a Dehn twist g' such that dg' and dg are similar. Since A and C share a common boundary component, we can say more strongly that $dg' = dg$. In short, we can say that C admits a Dehn twist having the same differential as dg. But this is just the same as saying that g extends continuously to C. The same goes for D. The key is that g pointwise fixes the common boundaries of all these cylinders. This establishes the existence of g.

Now we know that $\widehat{V}(\Sigma)$ contains the $(4,\infty,\infty)$ reflection triangle group. Hence, the Veech group $V(\Sigma)$ contains the even subgroup of the $(4,\infty,\infty)$ reflection triangle group. To finish our proof, we will show that $\widehat{V}(\Sigma)$ is precisely the reflection triangle group. Let Y denote the $(4,\infty,\infty)$ triangle. Let $\widehat{G}$ be the group generated by hyperbolic reflections in the sides of Y.

Exercise 11 (Challenge). Suppose that Γ is a group acting properly discontinuously on $\boldsymbol{H}^2$ and $\widehat{G} \subset \Gamma$. Prove that either $\Gamma = \widehat{G}$ or else Γ is the group generated by the reflections in the sides of the geodesic triangle obtained by bisecting Y in half.

If $\widehat{V}$ does not equal $\widehat{G}$, then Σ has an extra isometric symmetry which fixes the centers of the octagons. (This corresponds to the extra element, reflection in the bisector of Y.) But the octagons do not have any line of symmetry between the two drawn in our figures above. Hence, this extra symmetry does not exist. Hence $\widehat{V}(\Sigma) = \widehat{G}$. This is what we wanted to prove.

Exercise 12 (Challenge). Do all the same things as above for the translation surface associated to the isosceles triangle having small angles π/n for $n = 4, 6, 8, \ldots$.