

**TRANSLATION SURFACES AND THEIR ORBIT
CLOSURES:
AN INTRODUCTION FOR A BROAD AUDIENCE**

ALEX WRIGHT

Translation surfaces can be defined in an elementary way via polygons, and arise naturally in the study of various basic dynamical systems. They can also be defined as differentials on Riemann surfaces, and have moduli spaces called strata that are related to the moduli space of Riemann surfaces. There is a $GL(2, \mathbb{R})$ action on each stratum, and to solve most problems about a translation surface one must first know the closure of its orbit under this action. Furthermore, these orbit closures are of fundamental interest in their own right, and are now known to be algebraic varieties that parameterize translation surfaces with extraordinary algebro-geometric and flat properties. The study of orbit closures has greatly accelerated in recent years, with an influx of new tools and ideas coming from diverse areas of mathematics.

Many areas of mathematics, from algebraic geometry and number theory, to dynamics and topology, can be brought to bear on this topic, and known examples of orbit closures are interesting from all these points of view.

This survey is an invitation for mathematicians from different backgrounds to become familiar with the subject. Little background knowledge, beyond the definition of a Riemann surface and its cotangent bundle, is assumed, and top priority is given to presenting a view of the subject that is at once accessible and connected to many areas of mathematics.

Acknowledgements: I am especially grateful to Ronen Mukamel for helpful conversations that shaped the presentation of real multiplication in genus 2, and to Curtis McMullen for helpful comments and corrections. I am also very grateful to Preston Wake, Clark Butler, Elise Goujard, Emre Sertoz, Paul Apisa, Ian Frankel, Benjamin Dozier, Zhangchi Chen, and Zhan Jiang for finding typos and making helpful comments on these notes.

This survey grew out of notes for a five lecture course at the Graduate Workshop on Moduli of Curves, held July 7–18, 2014 at the Simons Center in Stony Brook, NY, organized by Samuel Grushevsky, Robert

Lazarsfeld, and Eduard Looijenga. The author is grateful to the organizers and participants.

The author is partially supported by a Clay Research Fellowship.

Other resources. There are a number of very good surveys on translation surfaces and related topics, for example [Esk06, For06a, HS06, Mas06, MT02, Möl13, Möl09, SW07, Via06, Yoc10, Yoc09, Yoc06, Zor06]. See also the seminal paper [McM03a] of McMullen.

1. TRANSLATION SURFACES

1.1. Equivalent definitions. This subsection has been written in a fairly technical way, so that it may serve as a reference for anyone looking for details on foundational issues. Most readers will want to skip some of the proofs on first reading. Anyone who already knows what a translation surface is can skip this subsection entirely.

In these notes, all Riemann surfaces will be assumed to be compact and connected. (A Riemann surface is a manifold of real dimension two with an atlas of charts to \mathbb{C} whose transition maps are biholomorphic.) Thus the term “Riemann surface” will be synonymous to “irreducible smooth projective algebraic curve over \mathbb{C} .”

Definition 1.1. An *Abelian differential* ω on a Riemann surface X is a global section of the cotangent bundle of X .

A *translation surface* (X, ω) is a nonzero Abelian differential ω on a Riemann surface X .

Thus “nonzero Abelian differential” and “translation surface” are synonymous terms, but sometimes the notation is slightly different. Sometimes we might omit the word “translation,” and say “let (X, ω) be a surface” when we should say “let (X, ω) be a translation surface.”

The complex vector space of Abelian differentials on X will be denoted $H^{1,0}(X)$. We assume the following facts are familiar to the reader.

Theorem 1.2. *Let g denote the genus of X . Then $\dim_{\mathbb{C}} H^{1,0}(X) = g$. If $g > 0$, each Abelian differential ω on X has $2g - 2$ zeros, counted with multiplicity.*

Each nonzero Abelian differential is a 1-form, which is closed but not exact, and hence $H^{1,0}$ is naturally a subspace of the first cohomology group $H^1(X, \mathbb{C})$ of X .

The following result is key to how most people think about translation surfaces.

Proposition 1.3. *Let (X, ω) be a translation surface. At any point where ω is not zero, there is a local coordinate z for X in which $\omega = dz$. At any point where ω has a zero of order k , there is a local coordinate z in which $\omega = z^k dz$.*

Proof. We will work in a local coordinate w , and suppose that ω vanishes to order k at $w = 0$. Thus we can write $\omega = w^k g(w)$, where g is some holomorphic function with $g(0) \neq 0$. Note that

$$\int_0^w g(t)t^k dt$$

vanishes to order $k + 1$ at 0, and thus admits a $(k + 1)$ -st root. Define z by

$$z^{k+1} = (k + 1) \int_0^w g(t)t^k dt.$$

By taking d of each side, we find $z^k dz = \omega$ as desired. □

Let $\Sigma \subset X$ denote the finite set of zeros of ω . At each point p_0 of $X \setminus \Sigma$, we may pick a local coordinate z as above. This choice is unique if we require $z(p_0) = 0$, and otherwise it is unique up to translations. This is because if $f(z)$ is any holomorphic function with $df = dz$, then $f(z) = z + C$ for some constant C . This leads to the following.

Proposition 1.4. *$X \setminus \Sigma$ admits an atlas of charts to \mathbb{C} whose transition maps are translations.*

Proof. The atlas consists of all local coordinates z with the property that $\omega = dz$. □

In particular, this gives $X \setminus \Sigma$ the structure of a flat manifold, since translations preserve the standard flat (Euclidean) metric on \mathbb{C} . However, there is even more structure: for example, at every point there is a choice of “north” (the positive imaginary direction).

We will see that the flat metric on $X \setminus \Sigma$ does not extend to a flat metric on X . This should be reassuring, since for us typically X will have genus at least 2, and the Gauss-Bonnet Theorem implies that such surfaces do not admit (nonsingular) flat metrics.

The points of Σ are thus considered to be singularities of the flat metric. From now on the term “singularity” of (X, ω) will be synonymous with “zero of ω .” The singularity is said to be order k if ω vanishes to order k .

We are now left with the task of determining what the flat metric looks like in a neighbourhood of a singularity p_0 of order k . We may use a local coordinate z where $z(p_0) = 0$ and $\omega = (k + 1)z^k dz$ (this is a scalar multiple of the local coordinate constructed above). The 1-form

$(k+1)z^k dz$ is the pull back of the form dz via the branched covering map $z \mapsto z^{k+1}$, since $d(z^{k+1}) = (k+1)z^k dz$. Near every point near but not equal to p_0 , $w = z^{k+1}$ is a local coordinate in which $\omega = dw$. Thus the flat metric near these point is the pull back of the flat metric on \mathbb{C} under this map $z \mapsto z^{k+1}$.

This pull back metric may be thought of very concretely: take $(k+1)$ copies of the upper half plane with the usual flat metric, and $(k+1)$ copies of the lower half plane, and glue them along the half infinite rays $[0, \infty)$ and $(-\infty, 0]$ in cyclic order, as in Figure 1.1.

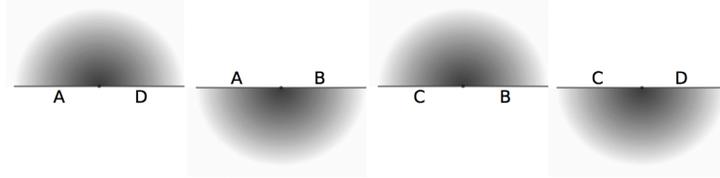


FIGURE 1.1. Four half planes glued in cyclic order. A neighbourhood of any singularity of order 1 is isometric to a neighbourhood of 0 in this picture.

Definition 1.5 (Second definition of translation surface). A *translation surface* is a closed topological surface X , together with a finite set of points Σ and an atlas of charts to \mathbb{C} on $X \setminus \Sigma$ whose transition maps are translations, such that at each point p_0 of Σ there is some $k > 0$ and a homeomorphism of a neighborhood of p_0 to a neighbourhood of the origin in the $2k+2$ half plane construction that is an isometry away from p_0 .

The singularity at p_0 is said to have cone angle $2\pi(k+1)$, since it can be obtained by gluing $2k+2$ half planes, each with an angle of π at the origin. The term “cone point” is another synonym for “singularity.”

Proposition 1.6. *The first and second definition of translation surface are equivalent.*

We have already shown that the structure in the first definition leads to the structure in the second definition, so it now suffices to show the converse.

Proof. Given such a flat structure on a surface X as in the second definition, we get an atlas of charts to \mathbb{C} away from the singularities, whose transition functions are translations. Since translations are bi-holomorphisms, this provides $X \setminus \Sigma$ with a complex structure, where Σ is the set of singularities. Furthermore, we get an Abelian differential

on $X \setminus \Sigma$, by setting $\omega = dz$ for any such local coordinate z . At each singularity p_0 of order k of the flat metric, we can find a unique coordinate z such that $z(p_0) = 0$ and the covering map $z^{k+1}/(k+1)$ is a local isometry (except at the point p_0) to a neighbourhood of 0 in $\mathbb{C} \setminus \{0\}$. In this coordinate z , the calculations above show that $\omega = z^k dz$ on a neighbourhood of p_0 minus p_0 itself (ω has not yet been defined at p_0).

As soon as we check that the remaining transition maps are biholomorphic, we will conclude that this atlas of charts (given by the coordinates z as above) on X gives X a complex structure. Setting $\omega = z^k dz$ at each singularity, in the local coordinate above, completes the definition of the Abelian differential ω .

The transition maps can be explicitly computed. Suppose z and w are local coordinates, such that $\omega = z^k dz$ in a neighborhood of a singularity p_0 , and $\omega = dw$ in a smaller open subset not containing the singularity. Then there is some constant C such that

$$w(z) = C + \int_0^z \eta^k d\eta.$$

This is evidently a local biholomorphism away from $z = 0$. □

The third definition is the most concrete, and is how translation surfaces are usually given.

Definition 1.7 (Third definition of translation surface). *A translation surface is an equivalence class of polygons in the plane with edge identifications: Each translation surface is a finite union of polygons in \mathbb{C} , together with a choice of pairing of parallel sides of equal length that are on “opposite sides.” (So for example two horizontal edges of the same length can be identified only if one is on the top of a polygon, and one is on the bottom. Each edge must be paired with exactly one other edge. These conditions are exactly what is required so that the result of identifying pairs of edges via translations is a closed surface.) Two such collections of polygons are considered to define the same translation surface if one can be cut into pieces along straight lines and these pieces can be translated and re-glued to form the other collection of polygons. When a polygon is cut in two, the two new boundary components must be paired, and two polygons can be glued together along a pair of edges only if these edges are paired.*

Proposition 1.8. *The third definition of translation surface is equivalent to the second.*

We will sketch the proof, but first some definitions are required.

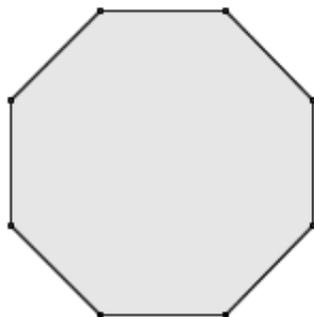


FIGURE 1.2. When opposite edges of a regular octagon are identified, the result is a translation surface with one cone point of angle 6π . (Generally the identifications are not drawn when opposite edges are identified—this situation is so common that it is the default.) An Euler characteristic computation shows that this has genus 2. ($2 - 2g = V - E + F$, where g is the genus, V is the number of vertices, E is the number of edges, and F is the number of faces. In this example, after identification of the edges there is 1 vertex, 4 edges, and 1 face, so $2 - 2g = 1 - 4 + 1 = -2$.)

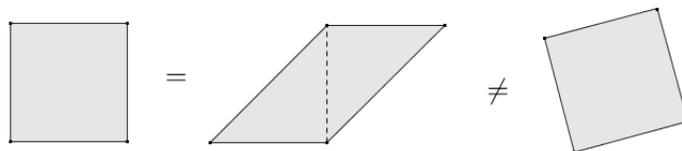


FIGURE 1.3. In each of these three polygons, opposite edges are identified to give a genus one translation surface. The first two are the same surface, since the second polygon can be cut (along the dotted line) and re-glued to give the first. However, the third translation surface is not equal to the first two, even though it is flat isometric. There is no flat isometry between them that sends “north” (the positive imaginary direction) to “north.”

Definition 1.9. A *saddle connection* on a translation surface is a straight line segment (i.e., a geodesic for the flat metric) going from a singularity to a singularity, without any singularities in the interior of

the segment. (The two endpoints can be the same singularity or different.) The *complex length* (also known as the *holonomy*) of a saddle connection is the associated vector in \mathbb{C} , which is only defined up to multiplication by ± 1 .

A *triangulation* of a translation surface is a collection of saddle connections whose interiors are disjoint, and such that any component of the complement is a triangle.

Remark 1.10. We will not discuss general geodesics for the singular flat metric, except to remark in passing that saddle connections are examples, and a general geodesic is either an isometrically embedded circle (i.e., the core curve of a cylinder) or a sequence of saddle connections such that each forms an angle of at least π with the next. In a certain definite sense, most flat geodesics contain more than one saddle connection.

Lemma 1.11. *Every translation surface (using the second definition) can be triangulated.*

Sketch of proof. In fact, any maximal collection of saddle connections whose interiors are disjoint must be a triangulation. \square

Sketch of proof of Proposition 1.8. In one direction, the lemma says that every surface as in the second definition can be triangulated. Cutting each saddle connection in the triangulation gives a collection of polygons (triangles) with edge identifications. Two edges are identified if they were the same saddle connection before cutting.

In the other direction, given a collection of polygons as in the third definition, the paired edges may be identified via translations. At each point on the interior of a polygon, the natural coordinate z of \mathbb{C} can be used. (The polygon sits in the complex plane \mathbb{C} .) At any point on the interior of an edge, the two polygons can be glued together, giving locally a coordinate. The structure at the singularities can be verified in an elementary way.

Indeed, after identifying pairs of edges, some of the vertices will become singular points of the flat metric. The main point is that the total angle around these singularities is an integer multiple of 2π . If the total angle were anything else, there would be no well defined choice of “north.” See Figure 1.4. \square

1.2. Examples. The flat torus $(\mathbb{C}/\mathbb{Z}[i], dz)$ is a translation surface. This is pictured in Figure 1.3.

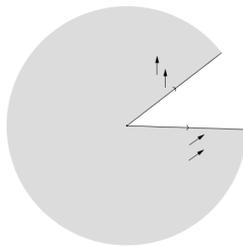


FIGURE 1.4. Here is a model of a cone angle that is not an integer multiple of 2π (here it is less than 2π), and hence cannot occur on a translation surface. It is obtained by identifying the two radial segments via rotation. On this picture, there is not a consistent choice of north: if a northward pointing vector is dragged across the radial segment, it no longer points north.

Definition 1.12. A translation covering $f : (X, \omega) \rightarrow (X', \omega')$ between translation surfaces is a branched covering of Riemann surfaces $f : X \rightarrow X'$ such that $f^*(\omega') = \omega$.

Translation coverings are, in particular, local isometries away from the ramification points. (By definition, the ramification points are the preimages of the branch points.) They must also preserve directions: for example, “north” must map to “north.” The ramification points must all be singularities. An unramified point is a singularity if and only if its image under the translation covering is. Branch points may or may not be singularities.

The fact that translation coverings are local isometries away from ramification points is especially clear if one notes that the flat length of a tangent vector v to the translation surface is $|\omega(v)|$.

Definition 1.13. A translation covering of $(\mathbb{C}/\mathbb{Z}[i], dz)$ branched over a single point is called a *square-tiled surface*.

Indeed, $(\mathbb{C}/\mathbb{Z}[i], dz)$ is a square with opposite sides identified, and the branch point can be assumed to be the corners of the square. The square-tiled surface will be tiled by d lifted copies of this square, where d is the degree.

The slit torus construction. In this construction, one starts with two genus one translation surfaces, and picks a parallel embedded straight line segment on each of them, of the same length. The two segments are cut open, and the resulting tori with boundary are glued together.

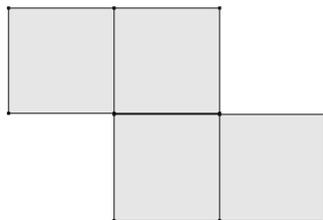


FIGURE 1.5. An example of a square-tiled surface. Opposite edges are identified. This translation surface is a degree 4 cover of $(\mathbb{C}/\mathbb{Z}[i], dz)$ branched over 1 point. It is genus 2, and has two singularities, each of cone angle 4π .

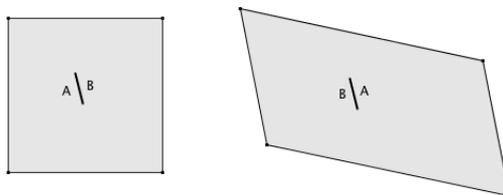


FIGURE 1.6. An example of the slit tori construction. Opposite edges are identified, giving two tori. In each, a slit is made, so that each torus has boundary consisting of two line segments, labelled A and B here. These are glued together to give a translation surface of genus two with two singularities of cone angle 4π , one at each end of the slit.

Unfolding rational billiards. Perhaps the original motivation for translation surfaces came from the study of rational billiards. This will be explained in Section 4. For now, begin with a polygon P in \mathbb{C} , all of whose angles are rational multiples of π . This restriction is equivalent to saying that the subgroup H of $O(2)$ generated by the derivatives of reflections in the sides of the polygon is a finite group. (It is a dihedral group.)

For each $h \in H$, we consider hP . We translate these if necessary so that the collection $\{hP : h \in H\}$ is a finite collection of disjoint polygons. We identify the edges in pairs: if $h'P$ is the reflection of hP in an edge of hP , this edge and the corresponding edge of $h'P$ are identified.

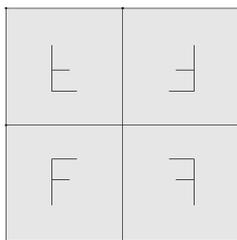


FIGURE 1.7. The unit square unfolds to four squares, glued together to give the flat torus $(\mathbb{C}/2\mathbb{Z}[i], dz)$. Here each square has been decorated by the letter F, to illustrate which squares are reflections of other squares.

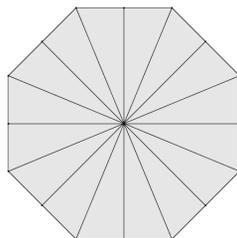


FIGURE 1.8. Unfolding the right angled triangle with smallest angle $\pi/8$ gives the regular octagon with opposite sides identified.

Proposition 1.14. *Suppose $\gcd(p, q, r) = 1$ and $p + q + r = n$. The triangle with angles $\frac{p}{n}\pi, \frac{q}{n}\pi, \frac{r}{n}\pi$ unfolds to the Abelian differential $\frac{ydx}{x(x-1)}$ on the normalization of the algebraic curve*

$$y^n = x^p(x-1)^q.$$

This curve is a cyclic cover of $\mathbb{C}P^1$ via the map $(y, x) \mapsto x$.

This proposition is not hard to prove. One proof uses Schwarz-Christoffel mappings, a topic in classical complex analysis. See for example [DT02].

1.3. Moduli spaces. Consider the problem of deforming the regular octagon. We may specify four of the edges as vectors in \mathbb{C} , and thus guess (correctly!) that the moduli space is locally \mathbb{C}^4 .

An Abelian differential in genus two can have either a double zero, or two zeros of order one. The octagon has a double zero, and deformations as above will always have a double zero. Figures 1.5 and 1.6 illustrate genus two translation surfaces with two zeros of order one.

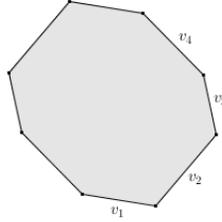


FIGURE 1.9. An octagon with opposite edges parallel may be specified by four complex numbers $v_1, v_2, v_3, v_4 \in \mathbb{C}$. (Not all choices give a valid octagon without self crossings, but there is an open set of valid choices.)

The collection of all Abelian differentials of genus g is of course a vector bundle over the moduli space of Riemann surfaces. However, this space is stratified according to the number and multiplicity of the zeros of the Abelian differentials.

Let $g > 1$ and let κ denote a partition of $2g - 2$, i.e. a nonincreasing list of positive integers whose sum is $2g - 2$. So if $g = 2$, the partitions are (2) and $(1, 1)$, and if $g = 3$ the partitions are $(4), (3, 1), (2, 2), (2, 1, 1), (1, 1, 1, 1)$.

Define the stratum $\mathcal{H}(\kappa)$ as the collection of genus g translation surfaces (X, ω) , where the multiplicity of the zeros of ω are given by κ . So $\mathcal{H}(2)$ denotes the collection of genus two translation surfaces with a double zero, and $\mathcal{H}(1, 1)$ denotes the collection of all genus two translation surfaces with two zeros of order one.

Proposition 1.15. *Each stratum $\mathcal{H}(\kappa)$ is a complex orbifold of dimension $n = 2g + s - 1$, where $s = |\kappa|$ denotes the number of distinct zeros of Abelian differentials in the stratum. Away from orbifold points (or on an appropriate cover without orbifold points) each stratum has an atlas of charts to \mathbb{C}^n with transition functions in $GL(n, \mathbb{Z})$.*

Thus each stratum looks locally like \mathbb{C}^n , and has a natural affine structure.

Sketch of formal proof. Let S be fixed topological surface of genus g , with a set Σ of s distinct marked points. Let us begin with the space $\tilde{\mathcal{H}}(\kappa)$ of translations surfaces (X, ω) equipped with an equivalence class of homeomorphisms $f : S \rightarrow X$ that send the marked points to the zeros of ω . The equivalence relation is isotopy rel marked points.

We will see that the map from $\tilde{\mathcal{H}}(\kappa)$ to $\mathcal{H}(\kappa)$ that forgets f is an infinite degree branched covering.

Fix a basis for the relative homology group $H_1(S, \Sigma, \mathbb{Z})$. If $\Sigma = \{p_1, \dots, p_n\}$, this is typically done by picking a (symplectic) basis

$$\gamma_1, \dots, \gamma_{2g}$$

for absolute homology $H_1(S, \mathbb{Z})$, and then picking a curve γ_{2g+i} from p_i to p_n for each $i = 1, \dots, n-1$. The map

$$\tilde{\mathcal{H}}(\kappa) \rightarrow \mathbb{C}^n, \quad (X, \omega, [f]) \mapsto \left(\int_{f_* \gamma_i} \omega \right)_{i=1}^{2g+s-1}$$

is locally one-to-one and is onto an open subset of \mathbb{C}^n . The easiest way to see this is via flat geometry: These integrals determine the integrals of every relative homology class, and in particular the complex lengths of the edges in any polygon decomposition for (X, ω) . The edges in this polygon decomposition of course determine (X, ω) .

The mapping class group of S (with s distinct unlabeled marked points) acts on $\tilde{\mathcal{H}}(\kappa)$ by precomposition of the marking. The induced action on these \mathbb{C}^n coordinates is via $GL(n, \mathbb{Z})$ (change of basis for relative homology). The quotient is $\mathcal{H}(\kappa)$. \square

In this survey we will ignore orbifold issues, and just pretend strata are complex manifolds rather than complex orbifolds. Given a translation surface (X, ω) , and a basis γ_i of the relative homology group $H_1(X, \Sigma, \mathbb{Z})$, we will refer to

$$\left(\int_{\gamma_i} \omega \right)_{i=1}^{2g+s-1} \in \mathbb{C}^n$$

as *period coordinates* near (X, ω) . Implicit in this is that the basis can be canonically transported to nearby surfaces in $\mathcal{H}(\kappa)$, thus giving a map from a neighbourhood of (X, ω) to a neighbourhood in \mathbb{C}^n . However, when X has automorphisms preserving ω , this is not the case canonically. This is precisely the issue we are ignoring when we pretend that $\mathcal{H}(\kappa)$ is a manifold instead of an orbifold.

It is precisely the period coordinates that provide strata an atlas of charts to \mathbb{C}^n with transition functions in $GL(n, \mathbb{Z})$. The transition functions are change of basis matrices for relative homology $H_1(X, \Sigma, \mathbb{Z})$.

Compactness criterion. Strata are never compact. Even the subset of unit area translation surface is never compact. (Here and throughout these notes we refer to the analytic topology, which is the weakest topology for which period coordinates are continuous.)

Masur's compactness criterion gives that a closed subset of the set of unit area surfaces in a stratum is compact if and only if there is

some positive lower bound for the length of all saddle connections on all translation surfaces in the subset.

Let us also remark that the map $(X, \omega) \mapsto X$ is not proper, even when restricting to unit area surfaces. For example, it is possible to have a sequence of translation surfaces (X_n, ω_n) of area 1 in $\mathcal{H}(1, 1)$ converge to $(X, \omega) \in \mathcal{H}(2)$ (so two zeros coalesce) in the bundle of Abelian differentials over the moduli space of Riemann surfaces. However, such a sequence (X_n, ω_n) will diverge in $\mathcal{H}(1, 1)$: there will be shorter and shorter saddle connections joining the two zeros.

Every Abelian differential in genus two. We now give flat geometry pictures of all Abelian differentials in genus 2. The discussion includes only sketches of proofs.

Proposition 1.16. *Every translation surface in $\mathcal{H}(2)$ can be obtained by gluing a cylinder into a slit torus, as in Figure 1.10. Every translation surface in $\mathcal{H}(1, 1)$ is obtained from the slit torus construction, as in Figure 1.11*

Lemma 1.17. *For any translation surface in genus 2, not every saddle connection is fixed by the hyperelliptic involution ρ .*

Proof. Triangulate the surface. If every saddle connection in this triangulation was fixed by the hyperelliptic involution, then each triangle would be mapped to itself. However, no triangle has rotation by π symmetry, so this is impossible. \square

Sketch of proof of proposition. Fix a translation surface in $\mathcal{H}(2)$, and a saddle connection c not fixed by ρ . Since ρ_* acts on homology by -1 , c and $\rho(c)$ are homologous curves. Cutting c and $\rho(c)$ decomposes the surface into two subsurfaces with boundary, one of genus one and one of genus zero. The genus zero component is a cylinder and the genus one part must be a slit torus.

Similarly for a translation surface in $\mathcal{H}(1, 1)$, consider a triangulation. This must contain at least one triangle with one vertex at one of the singularities, and the other two vertices at the other singularity (i.e., for at least one of the triangles, not all corners are at the same singularity). At most one of the edges of this triangle is fixed by the hyperelliptic involution, so the triangle must have at least one edge c that is not fixed by the hyperelliptic involution and goes between the two zeros.

Cutting along c and $\rho(c)$ decomposes the surface into two tori. In other words, every translation surface in $\mathcal{H}(1, 1)$ may be obtained from the slit torus construction. \square

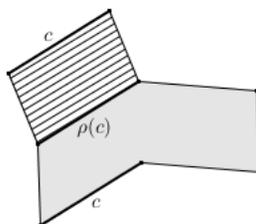


FIGURE 1.10. Every translation surface in $\mathcal{H}(2)$ admits a decomposition into a cylinder and a slit torus, and hence can be drawn as in this picture. Opposite sides are identified.

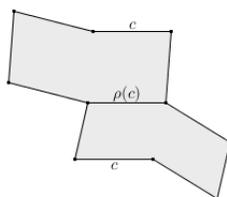


FIGURE 1.11. Every translation surface in $\mathcal{H}(1,1)$ comes from the slit torus construction, and thus can be drawn as in this picture.

Philosophical conclusions. Polygon decompositions and flat geometry provide a fundamentally different way to think about Abelian differentials on complex algebraic curves. In this perspective an Abelian differential is extremely easy to write down (as a collection of polygons), and it is easy to visualize deformations of this Abelian differential (the edges of the polygons change). However, some things become more difficult in this perspective. For example, given a plane algebraic curve it is often an easy exercise to write down a basis of Abelian differentials. However, given a translation surface given in terms of polygons, this is typically impossible to do. And it is typically a transcendently difficult problem to write down the equations for the algebraic curve given by some polygon decomposition.

The study of translation surfaces and algebraic curves is enriched by the transcendental connections between the two perspectives.

2. AFFINE INVARIANT SUBMANIFOLDS

2.1. Definitions and first examples. Fix a stratum, and a translation surface (X, ω) in the stratum. As always, let $\Sigma \subset X$ be the zeros of ω .

There is a linear injection $H^{1,0}(X) \rightarrow H_1(X, \Sigma, \mathbb{C})^* = H^1(X, \Sigma, \mathbb{C})$. Given $\omega \in H^{1,0}(X)$, the linear functional it determines in $H_1(X, \Sigma, \mathbb{C})^*$ is simply integration of ω over relative homology classes. Given a basis $\gamma_1, \dots, \gamma_n$ of $H_1(X, \Sigma, \mathbb{Z})$, we get an isomorphism

$$H^1(X, \Sigma, \mathbb{C}) \rightarrow \mathbb{C}^n, \quad \phi \mapsto (\phi(\gamma_i))_{i=1}^n.$$

The period coordinates of the previous section are just the composition of the map sending (X, ω) to ω considered as a relative cohomology class, followed by this isomorphism to \mathbb{C}^n . About half the time it is helpful to forget this coordinatization, and just consider period coordinates as a map to $H^1(X, \Sigma, \mathbb{C})$ sending (X, ω) to the relative cohomology class of ω .

In the next definition we fix a stratum $\mathcal{H} = \mathcal{H}(\kappa)$, where κ is a partition of $2g - 2$ and g is the genus.

Definition 2.1. An *affine invariant submanifold* is the image of a proper immersion f of an open connected manifold \mathcal{M} to a stratum \mathcal{H} , such that each point p of \mathcal{M} has a neighbourhood U such that in a neighbourhood of $f(p)$, $f(U)$ is determined by linear equations in period coordinates with coefficients in \mathbb{R} and constant term 0.

The difference between immersion and embedding is a minor technical detail, so for notational simplicity we will typically consider an affine invariant submanifold \mathcal{M} to be a subset of a stratum. The definition says that locally in period coordinates \mathcal{M} is a linear subspace of \mathbb{C}^n . The requirement that the linear equations have coefficients in \mathbb{R} is equivalent to requiring that the linear subspace of \mathbb{C}^n is the complexification of a real subspace of \mathbb{R}^n (or, in coordinate free terms, of $H^1(X, \Sigma, \mathbb{R})$).

The word “affine” does not refer to affine varieties; it refers to the linear structure on \mathbb{C}^n . (Perhaps “linear” would be a better term than “affine”, but this is the terminology in use.)

In these notes, when we refer to the dimension of complex manifolds or vector spaces, we will mean their complex dimension.

Note that affine invariant submanifolds must have dimension at least 2. This is because if (X, ω) is a point on an affine invariant submanifold, then the linear subspace must contain the real and imaginary parts of the period coordinates, and these cannot be collinear. That they

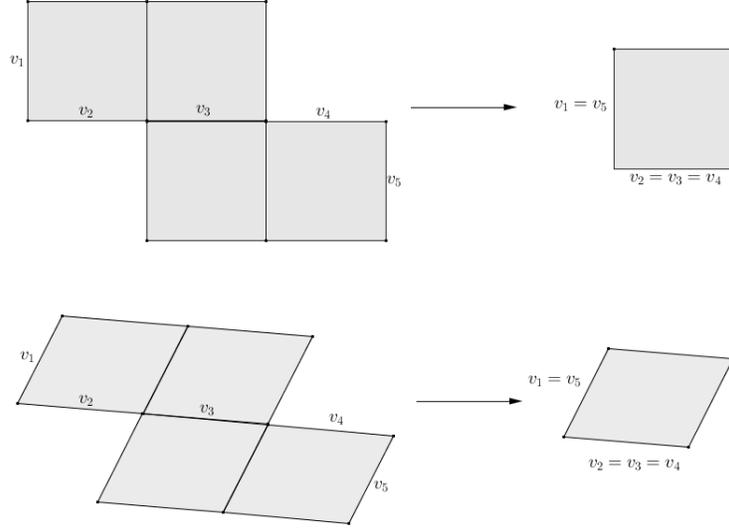


FIGURE 2.1. Period coordinates $(v_1, v_2, v_3, v_4, v_5) \in \mathbb{C}^5$ have been indicated on two surfaces in $\mathcal{H}(1, 1)$ (upper and lower left) that are degree four translation covers of tori (upper and lower right). The linear equations indicated locally cut out the locus of surfaces in $\mathcal{H}(1, 1)$ that are degree four translation covers of tori branched over 1 point.

cannot be collinear follows from the fact that $\operatorname{Re}(\omega)$ and $\operatorname{Im}(\omega)$ cannot be collinear real cohomology classes, because

$$\int_X \operatorname{Re}(\omega) \operatorname{Im}(\omega)$$

gives the area of the translation surface. (Locally $\omega = dz = dx + idy$, so $\operatorname{Re}(\omega) \operatorname{Im}(\omega) = dxdy$.)

Example 2.2. Let \mathcal{M} be a connected component of the space of degree d translation covers of a genus one translation surface, which is allowed to vary, branched over k distinct points. Then \mathcal{M} is an affine invariant submanifold of whichever stratum it lies in.

Indeed, \mathcal{M} is a $(k + 1)$ -dimensional immersed manifold: the moduli space of genus one translation surfaces is 2-dimensional, and after one branched point is fixed at 0, the other $k - 1$ are allowed to vary on the torus. So it suffices to see that locally in period coordinates, \mathcal{M} is contained in an $(k + 1)$ -dimensional linear subspace.

Let $(X, \omega) \in \mathcal{M}$, and say $f : (X, \omega) \rightarrow (X', \omega')$ is the translation covering, branched over k points. Let $\Sigma \subset X$ be the set of zeros of ω , and $\Sigma' \subset X'$ be the set of branch points of f , so $f(\Sigma) = \Sigma'$.

Let $\gamma \in \ker(f_*) \subset H_1(X, \Sigma, \mathbb{Z})$, and note that

$$\begin{aligned} \int_{\gamma} \omega &= \int_{\gamma} f^*(\omega') \\ &= \int_{f_*\gamma} \omega' = 0. \end{aligned}$$

Thus since $\ker(f_*)$ has dimension $\dim_{\mathbb{C}} H_1(X, \Sigma, \mathbb{C}) - \dim_{\mathbb{C}} H_1(X', \Sigma', \mathbb{C})$, we see that locally \mathcal{M} lies in an $\dim_{\mathbb{C}} H_1(X', \Sigma', \mathbb{C}) = (k+1)$ -dimensional linear subspace in period coordinates, as desired.

To rephrase the discussion in coordinate free terms, the relative cohomology class of ω lies in $f^*(H^1(X', \Sigma', \mathbb{C}))$.

Example 2.3. Fix a stratum \mathcal{H} and numbers k and d , and let \mathcal{M} be a connected component of the space of all degree d translation covers of surfaces in \mathcal{H} branched over $k \geq 0$ distinct points. Then similarly \mathcal{M} is an affine invariant submanifold.

Similarly, if \mathcal{M} is an affine invariant submanifold, and \mathcal{M}' is a connected component of the space of all degree d translation covers of surfaces in \mathcal{M} branched over $k \geq 0$ distinct points, then \mathcal{M}' is also an affine invariant submanifold.

Example 2.4. Fix a stratum \mathcal{H} and a number k , and let \mathcal{Q} be a connected component of the locus of $(X, \omega) \in \mathcal{H}$ where X admits an involution ι with $\iota^*(\omega) = -\omega$ and k fixed points. Then \mathcal{Q} is an affine invariant submanifold of \mathcal{H} (although it may be empty). It is locally defined by the equations

$$\int_{\iota_*\gamma} \omega + \int_{\gamma} \omega = 0,$$

for $\gamma \in H_1(X, \Sigma, \mathbb{Z})$.

A special case of this example is when ι is the hyperelliptic involution. In some strata, there is a whole connected component of hyperelliptic surfaces; in some, there are none; and in some, the hyperelliptic surfaces form a proper affine invariant submanifold.

2.2. Real multiplication in genus 2. We will begin with an elementary construction of some affine invariant submanifolds in genus 2. After we have constructed them, we will show that they parameterize Riemann surfaces X whose Jacobian admits real multiplication with ω as an eigenform.

These affine invariant submanifolds were independently discovered by McMullen and Calta [McM03a, Cal04] from very different perspectives. Our presentation is variant of McMullen's, and we suggest that the interested reader also consult [McM03a].

For the next proposition, recall that $H^{1,0}(X)$ can be considered as a subspace of $H^1(X, \mathbb{C})$, and that there is a natural symplectic pairing $\langle \cdot, \cdot \rangle$ on $H^1(X, \mathbb{Z})$. An endomorphism M of $H^1(X, \mathbb{Z})$ is called self-adjoint (with respect to this symplectic form) if $\langle Mv, w \rangle = \langle v, Mw \rangle$ for all $v, w \in H^1(X, \mathbb{Z})$. Note that different eigenspaces for a self-adjoint endomorphism must be symplectically orthogonal, since if $Mv = \lambda v$ and $Mw = \mu w$, then

$$\lambda \langle v, w \rangle = \langle Mv, w \rangle = \langle v, Mw \rangle = \mu \langle v, w \rangle.$$

Proposition 2.5. *Fix an integer $D > 0$ not a square. Consider the locus of (X, ω) in $\mathcal{H}(1, 1)$ or $\mathcal{H}(2)$ for which there is a self-adjoint endomorphism $M : H^1(X, \mathbb{Z}) \rightarrow H^1(X, \mathbb{Z})$ whose extension to $H^1(X, \mathbb{C})$ satisfies*

$$M\omega = \sqrt{D}\omega.$$

In $\mathcal{H}(1, 1)$ this locus is a finite union of 3-dimensional affine invariant submanifolds, and in $\mathcal{H}(2)$ it is a finite union of 2-dimensional affine invariant submanifolds.

In the proof, it is important to remember that because $\int_X \operatorname{Re}(\omega) \operatorname{Im}(\omega) > 0$, the restriction of the symplectic form to the subspace $\operatorname{span}(\operatorname{Re}(\omega), \operatorname{Im}(\omega))$ is symplectic.

Proof. First we will check that this locus is closed. (Recall that our default topology is the analytic topology, which we are referring to here.) Suppose that (X_n, ω_n) is a sequence of surfaces in this locus, converging to (X, ω) . We will first show that necessarily (X, ω) is in the locus, and hence the locus is closed.

Let M_n denote the endomorphism for (X_n, ω_n) . This endomorphism M_n has two \sqrt{D} -eigenvectors, $\operatorname{Re}(\omega_n)$ and $\operatorname{Im}(\omega_n)$. Since M_n is an integer matrix, its $-\sqrt{D}$ -eigenspace is the Galois conjugate of its \sqrt{D} -eigenspace, and hence must also be 2-dimensional. Since M_n is self-adjoint, the \sqrt{D} and $-\sqrt{D}$ -eigenspaces are symplectically orthogonal. In particular, the $-\sqrt{D}$ -eigenspace is the symplectic perp of the \sqrt{D} -eigenspace.

Note that ω_n converges to ω as cohomology classes. (The topology on strata is the topology on period coordinates, and period coordinates exactly determine the relative cohomology class of ω .) Hence M_n converges to an endomorphism M of $H^1(X, \mathbb{Z})$, which acts by \sqrt{D} on

$\text{span}(\text{Re}(\omega), \text{Im}(\omega))$ and by $-\sqrt{D}$ on the symplectic perp. Thus (X, ω) is in the locus also, and we have shown the locus is closed.

We must now show linearity. For this, it is important to note that above, since $\text{End}(H^1(X, \mathbb{Z}))$ is a discrete set, we see that M_n is in fact eventually constant. So any (X', ω') in the locus close to (X, ω) has the same endomorphism.

Suppose (X, ω) is in the locus, with endomorphism M . Consider the set of (X', ω') sufficiently close to (X, ω) for which $M\omega' = \sqrt{D}\omega'$. Since this equation is linear, we see that locally the locus is linear.

In other words, ω' must lie in the 2-dimensional \sqrt{D} -eigenspace in $H^1(X, \mathbb{C})$. In the 4-dimensional $\mathcal{H}(2)$ the set of such ω' is 2-dimensional, however in the 5-dimensional $\mathcal{H}(1, 1)$, this set is 3-dimensional. (Period coordinates for $\mathcal{H}(2)$ may be considered as a map to $H^1(X, \mathbb{C})$, and for $\mathcal{H}(1, 1)$ they may be considered as a map to $H^1(X, \Sigma, \mathbb{C})$. There is a natural map from $H^1(X, \Sigma, \mathbb{C})$ to $H^1(X, \mathbb{C})$, and we are requiring that the image of ω lies in a codimension 2 subspace there.) \square

We can be more explicit about the linear equations that define the locus in period coordinates. Indeed, suppose that $\gamma_1, \gamma_2 \in H_1(X, \mathbb{Z})$ are such that $\gamma_1, \gamma_2, M^*\gamma_1, M^*\gamma_2$ are a basis, where M^* denotes the dual linear endomorphism of $H_1(X, \mathbb{Z})$. Then the linear equations are

$$\int_{M^*\gamma_i} \omega = \sqrt{D} \int_{\gamma_i} \omega, \quad \text{for } i = 1, 2.$$

Examples. Let us now show these loci are nonempty in $\mathcal{H}(2)$, by giving explicit examples. A similar argument could show that these loci are nonempty in $\mathcal{H}(1, 1)$. (But in $\mathcal{H}(1, 1)$ there is a softer argument as well, which we will not present.)

Proposition 2.6. *The surfaces indicated in Figure 2.2 are in the loci described in Proposition 2.5. (But the D in the proposition might not be the same as the D in the figure.)*

Lemma 2.7. *For the translation surface (X, ω) in Figure 2.2,*

$$\text{span}(\text{Re}(\omega), \text{Im}(\omega))$$

is symplectically orthogonal to the Galois conjugate subspace of $H^1(X, \omega)$.

Proof. Since the periods of ω are in $\mathbb{Q}[\sqrt{D}, i]$, we can define a cohomology class ω' in $H^1(X, \mathbb{Q}[\sqrt{D}, i])$ that is Galois conjugate to ω . Note that ω' is not expected to be represented by a holomorphic 1-form. It can be described more concretely via the isomorphism

$$H^1(X, \mathbb{Q}[\sqrt{D}, i]) = \text{Hom}(H_1(X, \mathbb{Z}), \mathbb{Q}[\sqrt{D}, i]).$$

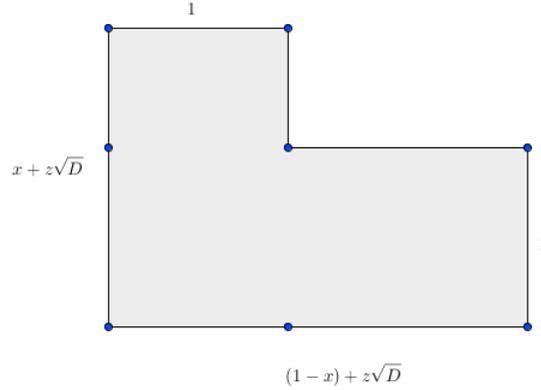


FIGURE 2.2. Opposite sides are identified, and all edges are vertical or horizontal. The numbers indicate signed real length in the direction indicated, and the parameters x, z are rational.

From this point of view, ω' is the composition of ω with the field endomorphism of $\mathbb{Q}[\sqrt{D}, i]$ that fixes i and sends \sqrt{D} to $-\sqrt{D}$.

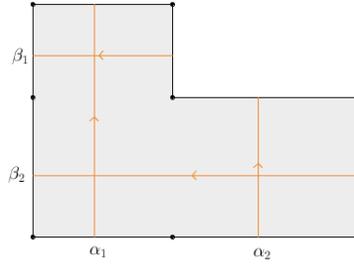


FIGURE 2.3. A basis of homology is indicated.

In Figure 2.3, note that $\alpha_1, \beta_1, \alpha_2, \beta_2 - \beta_1$ is a symplectic basis of homology. So, if $\phi, \psi \in H^1(X, \mathbb{R}) = \text{Hom}(H_1(X, \mathbb{R}), \mathbb{R})$, we can compute their symplectic pairing as

$$\begin{aligned} \langle \phi, \psi \rangle &= \phi(\alpha_1)\psi(\beta_1) - \phi(\beta_1)\psi(\alpha_1) + \\ &\quad \phi(\alpha_2)\psi(\beta_2 - \beta_1) - \phi(\beta_2 - \beta_1)\psi(\alpha_2). \end{aligned}$$

Suppose that we set $A_j = \frac{1}{i} \int_{\alpha_j} \omega$ and $B_j = \int_{\beta_j} \omega$. We are assuming that $B_1 = -1$ and $A_2 = 1$. Then we may compute

$$\begin{aligned} \langle \text{Im}(\omega), \text{Re}(\omega') \rangle &= A_1 B'_1 + A_2 (B'_2 - B'_1) \\ &= -(A_1 - B'_2 - 1). \end{aligned}$$

If $A_1 = x + z\sqrt{D}$ and $B_2 = y + w\sqrt{D}$, this quantity is zero if and only if $x - y = 1$ and $w = -z$.

Both $\langle \operatorname{Re}(\omega), \operatorname{Re}(\omega') \rangle$ and $\langle \operatorname{Im}(\omega), \operatorname{Im}(\omega') \rangle$ are easily seen to be automatically zero, and $\langle \operatorname{Re}(\omega), \operatorname{Im}(\omega') \rangle$ is the Galois conjugate of $-\langle \operatorname{Im}(\omega), \operatorname{Re}(\omega') \rangle$. Hence these conditions $x - y = 1$ and $w = -z$ are equivalent to the orthogonality of $\operatorname{span}(\operatorname{Re}(\omega), \operatorname{Im}(\omega))$ and $\operatorname{span}(\operatorname{Re}(\omega'), \operatorname{Im}(\omega'))$. \square

Proof of Proposition. We define M_0 to be the endomorphism of $H^1(X, \mathbb{R})$ that acts by \sqrt{D} on $\operatorname{span}(\operatorname{Re}(\omega), \operatorname{Im}(\omega))$ and by $-\sqrt{D}$ on the Galois conjugate subspace $\operatorname{span}(\operatorname{Re}(\omega'), \operatorname{Im}(\omega'))$. This matrix is rational, and it is self-adjoint if and only if $\operatorname{span}(\operatorname{Re}(\omega), \operatorname{Im}(\omega))$ and $\operatorname{span}(\operatorname{Re}(\omega'), \operatorname{Im}(\omega'))$ are symplectically orthogonal.

Since M_0 is a rational matrix, for some $m > 0$ we get that $M = m^2 M_0$ is integral. Thus (X, ω) is in the locus constructed in the Proposition 2.5, but the D value in the proposition is $m^2 D$ here. \square

Connection to real multiplication. Recall that the complex vector space $H^{1,0}(X)$ is canonically isomorphic as a real vector space to $H^1(X, \mathbb{R})$, via the map $\omega \mapsto \operatorname{Re}(\omega)$. The natural symplectic form on $H^1(X, \mathbb{R})$ pulls back to the pairing

$$\langle \omega_1, \omega_2 \rangle = \frac{1}{2} \operatorname{Re} \left(\int \omega_1 \overline{\omega_2} \right)$$

on $H^{1,0}(X)$. This symplectic form on $H^{1,0}(X)$ is compatible with the complex structure on $H^{1,0}(X)$, in that $\langle i\omega_1, i\omega_2 \rangle = \langle \omega_1, \omega_2 \rangle$ and $-\langle \omega, i\omega \rangle > 0$ for all $\omega \neq 0$.

We will let $H^{1,0}(X)^*$ denote the dual of the complex vector space $H^{1,0}(X)$. As a real vector space, $H^{1,0}(X)^*$ is canonically isomorphic to $H_1(X, \mathbb{R})$, via the usual integration pairing between homology classes and Abelian differentials. The space $H^{1,0}(X)^*$ inherits a dual compatible symplectic pairing, which we will also denote $\langle \cdot, \cdot \rangle$.

Definition 2.8. The *Jacobian* $\operatorname{Jac}(X)$ of a Riemann surface X is the complex torus $H^{1,0}(X)^*/H_1(X, \mathbb{Z})$, together with the data of the symplectic pairing $\langle \cdot, \cdot \rangle$ on its tangent space to the zero. (This tangent space is $H^{1,0}(X)^*$.) An *endomorphism* of $\operatorname{Jac}(X)$ is an endomorphism of the complex torus. An endomorphism is called self-adjoint if the induced endomorphism on the tangent space to the identity is self-adjoint with respect to the symplectic form.

Thus, endomorphisms of $\operatorname{Jac}(X)$ can be thought of either as complex linear endomorphisms of $H^{1,0}(X)^*$ that preserve $H_1(X, \mathbb{Z})$, or as linear endomorphisms of $H_1(X, \mathbb{Z})$ whose real linear extension to $H^{1,0}(X)^*$

happens to be complex linear. From our point of view, the second perspective is more enlightening, and the requirement of complex linearity is the deepest part of the definition. This is because the complex structure on $H^{1,0}(X)^*$ is determined by how $H^{1,0}(X)$ sits in $H^1(X, \mathbb{C})$, i.e., the Hodge decomposition. The Hodge decomposition varies as the complex structure on X varies, in a somewhat mysterious way. The complex linearity restriction is the only part of the data of an endomorphism of $\text{Jac}(X)$ that depends on the complex structure on X (the rest could be defined for a topological surface instead of a Riemann surface).

Recall that a *totally real field* is a finite field extension of \mathbb{Q} , all of whose field embeddings into \mathbb{C} have image in \mathbb{R} . Every real quadratic field is totally real, but there are cubic real fields that are not totally real, for example $\mathbb{Q}[2^{\frac{1}{3}}]$. An *order* in a number field is just a finite index subring of the ring of integers. The key example to keep in mind is that $\mathbb{Z}[\sqrt{D}]$ is an order in $\mathbb{Q}[\sqrt{D}]$.

Definition 2.9. The Jacobian $\text{Jac}(X)$ of a Riemann surface of genus g is said to have *real multiplication* by a totally real number field \mathbf{k} of degree g if there is some order $\mathcal{O} \subset \mathbf{k}$ that acts on $\text{Jac}(X)$ by self-adjoint endomorphisms. An Abelian differential $\omega \in H^{1,0}(X)$ is said to be an *eigenform* for this action if it is an eigenvector for the induced action of \mathcal{O} on the cotangent space of $\text{Jac}(X)$ at 0. (This cotangent space is $H^{1,0}(X)$.)

Genus two is special for real multiplication because of the following.

Lemma 2.10. *Fix a compatible symplectic structure on \mathbb{C}^2 , and let $M : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ be a self-adjoint real linear endomorphism. If M preserves a complex line $L \in \mathbb{C}^2$, then M is complex linear.*

Proof. First note, that the only self-adjoint real linear endomorphisms of \mathbb{R}^2 are scalars times identity. (This can be verified in coordinates, for a 2 by 2 matrix and the standard symplectic form on \mathbb{R}^2 .)

Since the symplectic form is compatible, L^\perp is a complex line. M leaves invariant the two complex lines L and L^\perp , and acts as a scalar on each, so it must be complex linear. \square

It follows from this that

Theorem 2.11 (McMullen). *The loci in $\mathcal{H}(2)$ and $\mathcal{H}(1, 1)$ defined in Proposition 2.5 in fact parameterize (X, ω) where $\text{Jac}(X)$ admits real multiplication by $\mathbb{Q}[\sqrt{D}]$ with ω as an eigenform.*

Proof. From the definition of the eigenform loci we get a self-adjoint action of $\mathbb{Z}[\sqrt{D}]$ on $H^1(X, \mathbb{Z})$, where \sqrt{D} acts by M . This gives a

dual action on $H_1(X, \mathbb{R})$, which preserves $H_1(X, \mathbb{Z})$. It suffices to show that this action is complex linear. This follows from the lemma, and the observation that the complex line consisting of the annihilator of $\text{span}(\text{Re}(\omega), \text{Im}(\omega))$ is preserved by the action. \square

Remark 2.12. For fixed D , the locus of (X, ω) where $\text{Jac}(X)$ admits real multiplication by $\mathbb{Q}[\sqrt{D}]$ with ω as an eigenform in fact has infinitely many connected components. This is related to the fact that there are infinitely many orders in $\mathbb{Q}[\sqrt{D}]$. If one fixes the maximal order that acts on $\text{Jac}(X)$ the situation is greatly ameliorated, for example the locus in $\mathcal{H}(1, 1)$ is connected and is closely related to a Hilbert modular surface. McMullen’s original treatment [McM03a] keeps track of the maximal order, because it is necessary to study connected components and give the relation to Hilbert modular surfaces. See also [KM].

Remark 2.13. In McMullen’s original work [McM03a], the study of real multiplication arose naturally from flat geometry and low dimensional topology in the following way. As we will discuss later in these notes, every (X, ω) in a two complex dimensional affine invariant submanifold must possess many affine symmetries. McMullen showed that in genus 2, certain affine symmetries naturally give rise to real multiplication on $\text{Jac}(X)$, and that ω is an eigenform.

Remark 2.14. It is important to note the proof of Proposition 2.5 doesn’t work in higher genus. This is because the M_n are not guaranteed to converge (they might diverge). (Here we fix a generator for an order in a totally real number field, and M_n is the action of this generator.)

Given that the M_n are automatically complex linear in genus 2, it is natural to impose this condition in higher genus, and try to see if loci of eigenforms as defined in Definition 2.9 give affine invariant submanifolds. With the complex linearity, it turns out the M_n must converge, which at least lets you show the locus is closed.

However, even with complex linearity the end result isn’t true, because even if M is complex linear at (X, ω) , there is no reason for this to be true at (X', ω') . So in higher genera eigenforms are locally contained in nice linear spaces, but the real multiplication can vanish as you move from an eigenform (X, ω) to a nearby translation surface (X', ω') whose periods satisfy the same linear equations.

2.3. Torsion, real multiplication, and algebraicity. There is however a still a very strong connection between endomorphisms of Jacobians and affine invariant submanifolds in higher genus, as discovered

by Möller in the case that the affine invariant submanifold has complex dimension 2.

Definition 2.15. Given (X, ω) , let $V \subset H^1(X, \mathbb{Q})$ be the smallest subspace such that $V \otimes \mathbb{C}$ contains ω , and such that $V \otimes \mathbb{C} = V^{1,0} \oplus V^{0,1}$, where

$$V^{1,0} = (V \otimes \mathbb{C}) \cap H^{1,0} \quad \text{and} \quad V^{0,1} = \overline{V^{1,0}}.$$

Set $V_{\mathbb{Z}}^* = \{\phi \in V^* : \phi(V \cap H^1(X, \mathbb{Z})) \subset \mathbb{Z}\}$. Define

$$\text{Jac}(X, \omega) = (V^{1,0})^* / V_{\mathbb{Z}}^*,$$

together with the data of the symplectic form on V^* . (This object $\text{Jac}(X, \omega)$ does not have a standard name, and this notation is not standard.)

V inherits a symplectic form by restriction from $H^1(X, \mathbb{Q})$. The restriction is automatically symplectic (i.e., nondegenerate) because of the condition $V \otimes \mathbb{C} = V^{1,0} \oplus V^{0,1}$. The symplectic form on V^* is the dual symplectic form.

$\text{Jac}(X, \omega)$ is a factor of $\text{Jac}(X)$ up to isogeny, and it is the “smallest factor containing ω ”.

Definition 2.16. Let p, q be two points of (X, ω) . We say that $p - q$ is *torsion* in $\text{Jac}(X, \omega)$ if, for any relative homology class $\gamma_{p,q}$ of a curve from p to q , there is a $\gamma \in H_1(X, \mathbb{Q})$ such that for all $\omega' \in V^{1,0}$ (including the most important one $\omega' = \omega$),

$$\int_{\gamma_{p,q}} \omega' = \int_{\gamma} \omega'.$$

If $\text{Jac}(X, \omega) = \text{Jac}(X)$, this is equivalent to $p - q$ being torsion in the group $\text{Jac}(X)$. For every affine invariant submanifold \mathcal{M} of complex dimension two, there is a natural algebraic extension of \mathbb{Q} called the trace field, which will be defined in the final section. The definition of real multiplication on $\text{Jac}(X, \omega)$ is exactly analogous to that for $\text{Jac}(X)$, except the degree of the field is required to be equal to the complex dimension of $\text{Jac}(X, \omega)$, which is sometimes less than g , and the order is required to act on $\text{Jac}(X, \omega)$ (instead of $\text{Jac}(X)$).

Theorem 2.17 (Möller [Möl06b, Möl06a]). *For every affine invariant submanifold \mathcal{M} of complex dimension two, and every $(X, \omega) \in \mathcal{M}$, $\text{Jac}(X, \omega)$ has real multiplication by an order in the trace field with ω as an eigenform, and furthermore if p and q are zeros of ω , then $p - q$ is torsion in $\text{Jac}(X, \omega)$.*

This result is actually two deep theorems, and their importance for the field is very great. One might expect that, given that an affine invariant submanifold is defined in terms of (X, ω) , the other holomorphic 1-forms on X would not be special, but this result says that they are (since the definition of torsion involves all or many ω' , and real multiplication produces other eigenforms for the Galois conjugate eigenvalues).

The converse of Möller’s result is true, and is much easier than the result itself.

Proposition 2.18 (Wright). *Let \mathcal{M} be a 2-dimensional submanifold of a stratum (not assumed to be linear), and suppose that for every $(X, \omega) \in \mathcal{M}$, $\text{Jac}(X, \omega)$ has real multiplication by the trace field with ω as an eigenform, and furthermore if p and q are zeros of ω , then $p - q$ is torsion in $\text{Jac}(X, \omega)$. Then \mathcal{M} is an affine invariant submanifold.*

Proof. We must show that \mathcal{M} is locally linear. For notational simplicity, we will assume $\text{Jac}(X, \omega) = \text{Jac}(X)$.

There are only countably many totally real number fields of degree g , and only countably many actions of each on $H^1(X, \mathbb{Q})$. Since the locus of eigenforms for each is closed, we may assume that the totally real number field is constant, and the action on $H^1(X, \mathbb{Q})$ is locally constant. (Again, this is a simple connectivity argument: The connected space \mathcal{M} is covered by disjoint closed sets, so there can only be one.)

Similarly, we can assume that for each pair of zeros p, q of ω , the rational homology class γ in Definition 2.16 is locally constant.

We will show that the span of $\text{Re}(\omega)$ and $\text{Im}(\omega)$ does not change in $H^1(X, \Sigma, \mathbb{C})$. This span gives the linear subspace that locally defines \mathcal{M} .

The real multiplication condition gives that the image of this span is locally constant in absolute homology (it is an eigenspace), and the torsion condition gives that each relative period $\int_{\gamma_{p,q}} \omega$ is equal to an absolute period $\int_{\gamma} \omega$, and hence the relative periods are determined by absolute periods.

That concludes the proof, but in closing we will write down the linear equations explicitly at $(X, \omega) \in \mathcal{M}$, in the case where $\text{Jac}(X, \omega) = \text{Jac}(X)$. Say the number field is \mathbf{k} and has \mathbb{Q} -basis r_1, \dots, r_g . Pick two absolute homology classes γ_1, γ_2 so that

$$\rho(r_i)\gamma_j, \quad i = 1, \dots, g, \quad j = 1, 2,$$

are a basis for $H_1(X, \mathbb{Q})$. Say the zeros of ω are p_1, \dots, p_s , and let α_i be a relative cycle from p_i to p_s , for $i = 1, \dots, s - 1$. For each i , let $\alpha'_i \in H_1(X, \mathbb{Q})$ be given from α_i by Definition 2.16. Then the equations

are

$$\int_{\rho(r_i)\gamma_j} \omega = r_i \int_{\gamma_j} \omega, \quad i = 1, \dots, g, \quad j = 1, 2,$$

and

$$\int_{\alpha_i} \omega = \int_{\alpha'_i} \omega, \quad i = 1, \dots, s - 1.$$

□

Recently, Simion Filip has generalized Möller's result to affine invariant submanifolds of any dimension. As in the previous proposition, this gives enough algebro-geometric conditions to characterize affine invariant submanifolds, and so Filip is able to conclude the following [Filb, Fila].

Theorem 2.19 (Filip). *All affine invariant submanifolds are quasi-projective varieties.*

Filip's proof crucially uses dynamics, and no other proof is known.

3. THE ACTION OF $GL^+(2, \mathbb{R})$

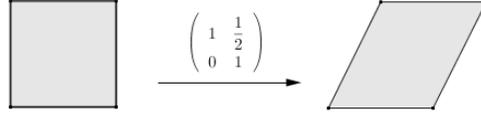
This section will begin to set the stage for the following driving theme in the study of translation surfaces.

The behavior of certain dynamical systems is fundamentally linked to the structure of affine invariant submanifolds.

The dynamical systems involved are the action of $GL^+(2, \mathbb{R})$ on each stratum, which is the topic of this section, and the straight line flow on each individual translation surface, which is the topic of the next section. It is very important to note that the connections go in both directions: Dynamical information often powers structural results on affine invariant submanifolds, and in the opposite direction results about the structure of linear manifolds are crucial in the study of the dynamical problems.

3.1. Definitions and basic properties. In this section, it is most helpful to think of a translation surface using the third definition (polygons). Let $GL^+(2, \mathbb{R})$ be the group of two by two matrices with positive determinant.

There is an action of $GL^+(2, \mathbb{R})$ on each stratum \mathcal{H} induced from the linear action of $GL^+(2, \mathbb{R})$ on \mathbb{R}^2 . If $g \in GL^+(2, \mathbb{R})$, and (X, ω) is a translation surface given as a collection of polygons, then $g(X, \omega)$ is the translation surface given by the collection of polygons obtained by acting linearly by g on the polygons defining (X, ω) .



Naively, one might think that $g(X, \omega)$ is always very different from (X, ω) when g is a large matrix, because g distorts any polygon a large amount. But because of the cut and paste equivalence, this is not the case. For example,

Proposition 3.1. *The stabilizer of $(\mathbb{C}/\mathbb{Z}[i], dz)$ is $GL^+(2, \mathbb{Z}) = SL(2, \mathbb{Z})$. The stabilizer of any square-tiled surface is a finite index subgroup of $GL^+(2, \mathbb{Z})$.*

Proof. Note that $g(\mathbb{C}/\mathbb{Z}[i], dz) = (\mathbb{C}/g(\mathbb{Z}[i]), dz)$. This gives that the stabilizer of $(\mathbb{C}/\mathbb{Z}[i], dz)$ is exactly the matrices $g \in GL^+(2, \mathbb{R})$ preserving $\mathbb{Z}[i] \subset \mathbb{C}$. Hence the stabilizer of $(\mathbb{C}/\mathbb{Z}[i], dz)$ is $SL(2, \mathbb{Z})$.

In general, the stabilizer preserves the periods of a surface. For a square-tiled surface, the periods are $\mathbb{Z}[i]$, so for any square-tiled surface the stabilizer is a subgroup of $SL(2, \mathbb{Z})$.

Now, suppose (X, ω) is a square-tiled surface. For any $g \in SL(2, \mathbb{Z})$, $g(X, \omega)$ is a square-tiled surface with the same number of squares. Hence, the stabilizer of (X, ω) is finite index in $SL(2, \mathbb{Z})$. \square

More basic properties are

Proposition 3.2. *The $SL(2, \mathbb{R})$ -orbit of every translation surface is unbounded. The stabilizer of every translation surface is discrete but never cocompact in $SL(2, \mathbb{R})$.*

Proof. If $g \in GL^+(2, \mathbb{R})$ is close enough to the identity, then $g(X, \omega)$ and (X, ω) are in the same coordinate chart. The coordinates of $g(X, \omega)$ are obtained from those of (X, ω) by acting linearly on $\mathbb{C} = \mathbb{R}^2$. This shows that for g sufficiently small enough and not the identity, $g(X, \omega) \neq (X, \omega)$, because they have different coordinates. Hence the stabilizer is discrete.

Set

$$r_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad \text{and} \quad g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}.$$

For every (X, ω) , there is some θ such that $r_\theta(X, \omega)$ has a vertical saddle connection. (Pick any saddle connection, and rotate so that it becomes vertical.) On $g_t r_\theta(X, \omega)$ this saddle connection is e^{-t} times as long, so as $t \rightarrow \infty$, we see that $g_t r_\theta(X, \omega)$ has shorter and shorter saddle connections and hence diverges to infinity in the stratum.

Let Γ be the stabilizer of (X, ω) , and suppose (X, ω) lies in the stratum \mathcal{H} . Consider the natural orbit map

$$SL(2, \mathbb{R})/\Gamma \rightarrow \mathcal{H}, \quad [g] \mapsto g(X, \omega).$$

By definition, Γ is cocompact if $SL(2, \mathbb{R})/\Gamma$ is compact. If $SL(2, \mathbb{R})/\Gamma$ were compact, then its image under this map would be compact also. However, the image is just the $SL(2, \mathbb{R})$ -orbit of (X, ω) , which must be unbounded and hence cannot be compact. \square

Definition 3.3. A *cylinder* on a translation surface is an isometrically embedded copy of a Euclidean cylinder $(\mathbb{R}/c\mathbb{Z}) \times (0, h)$ whose boundary is a union of saddle connections. The number c is the *circumference*, and the number h is the *height* of the cylinder. The ratio h/c is the *modulus* of the cylinder. The *direction* of the cylinder is the direction of its boundary saddle connections (so directions are formally elements of $\mathbb{R}P^1$). A translation surface (X, ω) is called *periodic* in some direction if (X, ω) is the union of the cylinders in that direction together with their boundaries.

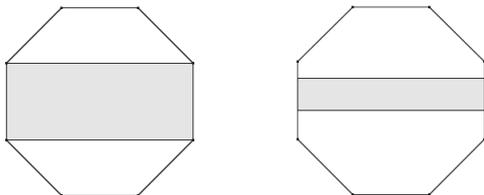


FIGURE 3.1. On the left octagon (with opposite sides identified), both the shaded part and its complement are horizontal cylinders. On the right octagon, the shaded rectangle is *not* a cylinder according to our definition, because its boundary is not a union of saddle connections. The effect of requiring the boundary to consist of saddle connections is equivalent to requiring the cylinders to be “maximal”, unlike this example on the right, whose height could be increased. The regular octagon is horizontally periodic, since it is the union of two horizontal cylinders (and their boundaries).

Proposition 3.4. A translation surface contains a matrix of the form

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

in its stabilizer if and only if the surface is horizontally periodic, and all reciprocals of moduli of horizontal cylinders are integer multiples of t .

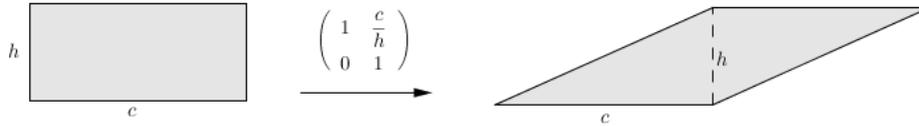


FIGURE 3.2. Each individual cylinder is stabilized by a parabolic matrix. In this picture, opposite edges are identified, except for the horizontal edges, which are the upper and lower boundary of the cylinder. The cylinder on the right can be cut along the dotted line and re-glued to give the cylinder on the left.

Proof. Assume the surface is horizontally periodic, and all reciprocals of moduli of horizontal cylinders are integer multiples of t . As is illustrated in Figure 3.2, each horizontal cylinder of modulus m is stabilized by the matrix

$$\begin{pmatrix} 1 & m^{-1} \\ 0 & 1 \end{pmatrix}.$$

The result follows.

For the converse, see [MT02]. □

Note that the previous proposition applies to any direction (not just horizontal) by first rotating the surface. It is stated for the horizontal direction only for notational simplicity.

3.2. Closed orbits and orbit closures.

Proposition 3.5. *Affine invariant submanifolds are $GL^+(2, \mathbb{R})$ invariant.*

Proof. It suffices to show that for each (X, ω) in the affine invariant submanifold \mathcal{M} there is a small neighbourhood U of the identity in $GL^+(2, \mathbb{R})$ such that for all $g \in U$ we have $g(X, \omega) \in \mathcal{M}$. This gives that the set of g for which $g(X, \omega) \in \mathcal{M}$ is open in $GL^+(2, \mathbb{R})$. Since \mathcal{M} is closed and the action is continuous, this set is also closed. Since $GL^+(2, \mathbb{R})$ is connected, any nonempty subset that is both open and closed must be everything.

If g is small enough, both $g(X, \omega)$ and (X, ω) are in the same coordinate chart. The coordinates of $g(X, \omega)$ are obtained by letting g act

linearly on the real and imaginary parts of each coordinate of (X, ω) , using the isomorphism $\mathbb{C} \cong \mathbb{R}^2$.

For example, g_t scales the real part of the coordinates by e^t and the imaginary part of the coordinates by e^{-t} . And u_t adds t times the imaginary part of the coordinates to the real coordinates.

Since \mathcal{M} is defined by linear equations with *real* coefficients and constant term 0, both the real and imaginary parts of the coordinates also satisfy the linear equations, as well as any complex linear combination of them. \square

That a converse is true is a recently established very deep fact, due to Eskin-Mirzakhani-Mohammadi [EM, EMM]. The proof is vastly beyond the scope of these notes. When we say “closed” below, we continue to refer to the analytic topology on a stratum.

Theorem 3.6 (Eskin-Mirzakhani-Mohammadi). *Any closed $GL^+(2, \mathbb{R})$ invariant set is a finite union of affine invariant submanifolds. In particular, every orbit closure is an affine invariant submanifold.*

This theorem is false if $GL^+(2, \mathbb{R})$ is replaced with the diagonal subgroup: there are closed sets invariant under the diagonal subgroup that are locally homeomorphic to a Cantor set cross \mathbb{R} . Determining to what extent the theorem holds for the unipotent subgroup

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

is a major open problem.

The context for these theorems comes from homogenous space dynamics, where Ratner’s Theorems give that orbit closures of a unipotent flow on a homogenous space must be sub-homogenous spaces.

Because of the work of Eskin-Mirzakhani-Mohammadi (and a converse that will we discuss below), the term “affine invariant submanifold” is synonymous with “ $GL^+(2, \mathbb{R})$ -orbit closure”, usually abbreviated “orbit closure”. Sometimes we will consider orbit closures for actions of subgroups of $GL^+(2, \mathbb{R})$, but in these cases we will make this clear by specifying the subgroup. The default is $GL^+(2, \mathbb{R})$ (or, for some other people, $SL(2, \mathbb{R})$).

Closed orbits. Let us now turn to the case of 2-dimensional affine invariant submanifolds \mathcal{M} . In this case $GL^+(2, \mathbb{R})$ acts transitively on \mathcal{M} . (Indeed, the real dimension of $GL^+(2, \mathbb{R})$ is equal to that of \mathcal{M} , and it is easily checked that the $GL^+(2, \mathbb{R})$ -orbit of any point is open in \mathcal{M} . The different $GL^+(2, \mathbb{R})$ -orbits in \mathcal{M} are open disjoint sets, so since \mathcal{M} is connected there must only be one. Hence \mathcal{M} is the $GL^+(2, \mathbb{R})$ -orbit of any translation surface in \mathcal{M} .)

Thus “closed $GL^+(2, \mathbb{R})$ -orbit” is synonymous with “2-dimensional affine invariant submanifold.”

Remark 3.7. We have already given a number of examples of 2-dimensional affine invariant submanifolds, namely the eigenform loci in $\mathcal{H}(2)$, and spaces of branched covers of genus 1 translation surfaces branched over exactly 1 point.

The following result has more than one proof [Vee95, SW04], but all known proofs follow the same basic outline and use dynamics in a nontrivial way.

Theorem 3.8 (Smillie). *Suppose that $(X, \omega) \in \mathcal{H}$ has closed orbit and stabilizer Γ . Then the orbit map*

$$GL^+(2, \mathbb{R})/\Gamma \rightarrow \mathcal{H}, \quad [g] \mapsto g(X, \omega)$$

is a homeomorphism, and Γ is a lattice in $SL(2, \mathbb{R})$.

The first conclusion is entirely expected and is standard (but technical to prove), and the second is deep and fundamentally important.

Remark 3.9. All square tiled surfaces are part of 2-dimensional affine invariant submanifolds and hence have closed orbit. The stabilizer of a square-tiled surface is a finite index subgroup of $SL(2, \mathbb{Z})$.

The eigenforms illustrated in Figure 2.2 must, by Smillie’s theorem, have lattice stabilizer. However, it is very hard to write down this stabilizer, even in specific examples. The orbifold type of the stabilizer was computed in [McM05, Bai07, Mukb], and an algorithm for computing the stabilizer was given in [Muka].

Finally, we note that if \mathcal{M} is a 2-dimensional orbit closure, then the projection to the moduli space of Riemann surfaces (via $(X, \omega) \mapsto X$) is 1-dimensional. One complex dimension is lost, since (X, ω) and $(X, r\omega)$ map to the same point, for any $r \in \mathbb{C}$. A corollary of Smillie’s theorem (and also Filip’s theorem, obtained much later) is that this projection of \mathcal{M} is in fact an algebraic curve.

Proposition 3.10. *The projection of closed orbit to the moduli space of Riemann surfaces is an algebraic curve, which is isometrically immersed with respect to the Teichmüller metric.*

Isometrically immersed curves in the moduli space of Riemann surfaces are called *Teichmüller curves*. Up to a “double covering” issue relating quadratic differentials to Abelian differentials, all Teichmüller curves are projections of closed orbits.

Royden has shown that the Kobayashi metric on the moduli space of Riemann surfaces is equal to the Teichmüller metric. Using this,

McMullen has shown that Teichmüller curves are rigid [McM09]. The study of Teichmüller curves is a fascinating area at the intersection of dynamics and algebraic geometry.

See [Wri13] for a list of known Teichmüller curves, and see [McM06b, Möl08, BM12, MW] for some finiteness results.

Stable and unstable manifolds for g_t . We will now give a flavor of the dynamics of the g_t action. We will not return to this explicitly in these notes, but the ideas we present now underlie most of the proofs that have been omitted from these notes.

Say the period coordinates of (X, ω) are $v_j = x_j + iy_j$ for $j = 1, \dots, n$. We can then think of (X, ω) in coordinates as a 2 by n matrix whose first row gives the real parts of period coordinates, and whose second row gives the imaginary parts.

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix}$$

The advantage of writing the coordinates this way is that any $g \in GL^+(2, \mathbb{R})$ close to the identity (so $g(X, \omega)$ stays in the same chart) acts by left multiplication. In particular, for small t we have that $g_t(X, \omega)$ is the matrix product

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix}.$$

However, for large t , we expect $g_t(X, \omega)$ will have left the coordinate chart. When it enters a different coordinate chart, a change of basis matrix can be used to compute the new coordinates from the old. This matrix is a n by n invertible integer matrix, which we will call $A_t(X, \omega)$. (This is an imprecise definition of $A_t(X, \omega)$, which of course depends on choices of coordinates, etc.)

We get that the coordinates of $g_t(X, \omega)$ are

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix} A_t(X, \omega).$$

The matrix $A_t(X, \omega)$ is called the *Kontsevich-Zorich cocycle*. It is a cocycle in the dynamical systems sense, which simply means

$$A_{t+s}(X, \omega) = A_t(g_s(X, \omega))A_s(X, \omega).$$

The Kontsevich-Zorich cocycle is the complicated part of the dynamics of g_t . However, its effect is usually beat out by the effect of the (e^t) 's on the left. In particular,

Theorem 3.11 (Masur, Veech, Forni [Mas82, Vee86, For06b]). *Fix an affine invariant submanifold \mathcal{M} . For almost every $(X, \omega) \in \mathcal{M}$, and*

every (X', ω') in the same coordinate chart with the same real parts of period coordinates, the distance between $g_t(X, \omega)$ and $g_t(X', \omega')$ goes to zero as $t \rightarrow \infty$.

Without the interference of the Kontsevich-Zorich cocycle, this would simply be the obvious fact that

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & 0 \\ y_1 - y'_1 & y_2 - y'_2 & \cdots & y_n - y'_n \end{pmatrix} \rightarrow 0.$$

The theorem says that even with the Kontsevich-Zorich cocycle added in on the left this still holds.

Overall, g_t expands the real parts of period coordinates exponentially, and contracts the imaginary parts exponentially. But the reader should be warned that there are technical complications arising from the fact that strata are not compact.

Flows or transformations that expand and contract complimentary directions exponentially are called *hyperbolic*. Using the hyperbolic dynamics of g_t , one can show that every affine invariant submanifold is a g_t -orbit closure. In particular,

Theorem 3.12. *Every affine invariant submanifold is a $GL^+(2, \mathbb{R})$ -orbit closure.*

4. THE STRAIGHT LINE FLOW

For much of this section, a good reference is the survey [MT02].

4.1. Definition and basic properties. Fix a unit length vector $v \in \mathbb{C}$. The straight line flow on a translation surface (X, ω) sends each point to the point obtained by starting at that point and moving in the direction v at unit speed for time t . This gives for each $t \in \mathbb{R}$ a map $\phi_t^v : (X, \omega) \setminus B_t \rightarrow (X, \omega)$, where B_t is the set of “bad points” whose straight line flow hits a singularity in time at most t . B_t consists of a finite union of line segments.

Let $B = \cup_t B_t$. Although this set might be dense, it has measure zero and hence it should be considered to be of negligible size. The straight line flow is defined on $(X, \omega) \setminus B$ for all time t .

One of the reasons straight line flow is important is because billiard trajectories in rational polygons “unfold” to orbits of straight line flow. This was the original motivation for unfolding rational polygons to translation surfaces; instead of bouncing off the edges of the polygon, the trajectory can continue straight into a reflected copy of the polygon.

The most basic question about straight line flow is: can the surface be cut into pieces invariant under the straight line flow? The easiest

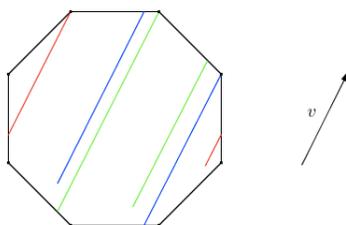


FIGURE 4.1. The colored line segments consist of (some of the) points whose orbit under straight line flow in direction v hits the singularity in finite time.

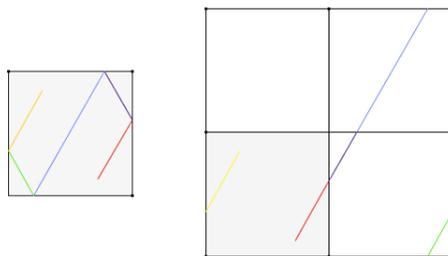


FIGURE 4.2. A billiard trajectory in the square (left) “unfolds” to a straight line flow trajectory in the associated translation surface (right).

thing to do is to cut out all saddle connections in direction v . Possibly this disconnects the surface, possibly not.

Definition 4.1. A straight line flow is called *minimal* if every orbit that is defined for all time is dense.

Let v have irrational slope, and let $(X, \omega) = (\mathbb{C}/\mathbb{Z}[i], dz)$. The straight line flow in direction v on (X, ω) is minimal. The proof of this will follow from a more general result below, but the reader is invited to try to convince themselves now that this flow is minimal.

Proposition 4.2 (Katok-Zemljakov [ZK75]). *After removing all saddle connections in a given direction on a translation surface (X, ω) , each connected component is either a cylinder or has minimal straight line flow in the given direction.*

This is a striking dichotomy: on each component, either every single orbit is periodic (the cylinder case), or every single orbit is dense.

Proposition 4.3. *On every (X, ω) , the straight line flow is minimal in all but countably many directions.*



FIGURE 4.3. Opposite sides are identified, and segment A on the left is identified to segment A on the right, etc., to give a translation surface in $\mathcal{H}(2)$ that we think of as a cylinder glued onto a slit torus. Then removing the saddle connections A and B disconnects the surface, giving a slit torus (left) and a cylinder (right). If the segment A has irrational slope and the torus is 1 by 1, then the flow in direction A will be minimal in the slit torus, and periodic on the cylinder.

Proof. There are only countably many saddle connections. In any direction that does not have saddle connections, the previous result says the straight line flow must be minimal. \square

4.2. Ergodicity. A measure preserving flow on a space Y with a probability measure m is a homomorphism ϕ from \mathbb{R} to the group of invertible measure preserving transformations from Y to itself. (These transformations are typically only considered to be defined on set of full measure.)

Definition 4.4. A measure preserving flow on a space Y with probability measure m is said to be *ergodic* if Y cannot be written as the disjoint union of two subsets that are invariant under the flow and each of positive measure.

The flow is *uniquely ergodic* if m is the unique invariant measure for the flow. (It follows that m is ergodic.)

Thus ergodicity is a basic indecomposability condition saying that the dynamics cannot be split into two smaller pieces (each of which could be studied separately). It is somewhat surprising that it implies the following strong restriction on the dynamics, whose proof is nontrivial but can be found in any book on ergodic theory.

Theorem 4.5 (Birkhoff Ergodic Theorem). *Let ϕ be a measure preserving ergodic flow on a space Y with probability measure m , and let $f \in L^1(Y, m)$. Then, for m -almost every $y \in Y$,*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\phi_t(y)) dt = \int_Y f dm.$$

You should think of the case where $f = \chi_A$ is the characteristic function of a set A . Then the theorem says that long orbit segments $\{\phi_t(y) : t \in [0, T]\}$ spend about $m(A)$ of their time in A . Thus the Birkhoff Ergodic Theorem says that almost every orbit is *equidistributed*, in that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \text{Leb}(\{t : t \in [0, T], \phi_t(y) \in A\}) \rightarrow m(A),$$

where Leb denotes Lebesgue measure on \mathbb{R} and $A \subset Y$ is any measurable set. (Technically, to be true as stated, some additional restriction must be placed on A , for example, A is open. Otherwise any given orbit can be removed from any measurable set A , giving a set A' often of the same measure of A , but which the given orbit does not intersect at all.)

Renormalization of straight line flow. Say that the g_t -orbit

$$\{g_t(X, \omega), t \geq 0\}$$

is *recurrent* if there is some compact set K of the stratum, such that $g_t(X, \omega) \in K$ for arbitrarily large t . This exactly says that $g_t(X, \omega)$ does not diverge to infinity.

Given an orbit segment of vertical straight line flow on (X, ω) of length L , it yields an orbit segment of vertical straight line flow on $g_t(X, \omega)$ of length $e^{-t}L$. In this way long orbit segments of vertical straight line flow become small under g_t , and we say that g_t *renormalizes* the vertical straight line flow. This idea of replacing a long orbit segment of a dynamical system for a shorter orbit segment of a different but related dynamical system is called renormalization, and is fundamental in dynamics. For the straight line flow, it was used to prove the following.

Theorem 4.6 (Masur's criterion [Mas92]). *Suppose that $g_t(X, \omega)$ is recurrent. Then the vertical straight line flow on (X, ω) is uniquely ergodic.*

The converse is not true, however the result is extremely powerful.

Theorem 4.7 (Kerckhoff-Masur-Smillie [KMS86]). *For every (X, ω) and almost every $\theta \in [0, 2\pi)$, $g_t(r_\theta(X, \omega))$ is recurrent (as θ is fixed and $t \rightarrow \infty$). Thus, for every (X, ω) , the straight line flow is uniquely ergodic in almost every direction.*

This implies the same result for billiard flows in rational polygons.

Remark 4.8. There exist (X, ω) such that the vertical flow is minimal but not uniquely ergodic. This is a bit strange; every orbit is dense, but

most orbits are not equidistributed, and hence somehow favor (spend more time than expected in) some parts of (X, ω) .

4.3. Complete periodicity. The dynamics are much more restricted for (X, ω) that lie in a 2-dimensional affine invariant submanifold. Such (X, ω) are called *lattice surfaces*, since their stabilizer is a lattice in $SL(2, \mathbb{R})$.

Theorem 4.9 (Veech dichotomy [Vee89]). *For any lattice surface, in the direction of any saddle connection the surface is periodic. In all other directions the straight line flow is uniquely ergodic.*

This is the same dichotomy that holds for $(\mathbb{C}/\mathbb{Z}[i], dz)$, where the flow is periodic in the rational directions, and uniquely ergodic in the irrational directions.

The proof of the following lemma is easy, but it will be omitted, because it requires familiarity with the geodesic flow on a finite volume complete hyperbolic surface (every geodesic is either recurrent, or it eventually goes straight out a cusp, and the action of g_t on $SL(2, \mathbb{R})/\Gamma$ is geodesic flow on the unit tangent bundle to a hyperbolic surface).

Lemma 4.10. *For lattice surface (X, ω) , either $g_t(X, \omega)$ is recurrent, or (X, ω) is stabilized by some matrix*

$$\begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}$$

with $s \neq 0$.

Proof of Veech dichotomy. It suffices to prove the statement for the vertical direction (since the surface can be rotated to make any direction vertical).

By the lemma, either $g_t(X, \omega)$ is recurrent, or (X, ω) is stabilized by the matrix above. In the first case, Masur’s criterion gives that the flow is uniquely ergodic (and it is easy to see there can be no vertical saddle connections, or else $g_t(X, \omega)$ would diverge). In the second case, Proposition 3.4 gives that the surface is the union of vertical cylinders (and so there are vertical saddle connections, on the boundary of the cylinders). □

We will also give a more modern proof of just the first statement, together with a generalization. For this we will need

Theorem 4.11 (Minsky-Weiss, Smillie-Weiss [MW02, SW04]). *The u_t -orbit closure of any (X, ω) contains a horizontally periodic surface.*

Recall $u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$.

Lemma 4.12. *For any horizontal cylinder or saddle connection on (X, ω) , and each t , there is a corresponding horizontal cylinder or saddle connection on $u_t(X, \omega)$.*

Furthermore, there is a corresponding horizontal cylinder or union of horizontal saddle connections on each translation surface (X', ω') in the u_t -orbit closure of (X, ω) .

Proof. The first statement follows because the matrix u_t fixes the horizontal direction. Now consider $(X', \omega') = \lim_{n \rightarrow \infty} u_{t_n}(X, \omega)$. For each horizontal cylinder on (X, ω) , there is a corresponding horizontal cylinder on each $u_{t_n}(X, \omega)$, and hence there is also a horizontal cylinder in the limit.

The same argument applies equally well to horizontal saddle connections, except that possibly in the limit a zero could “land” on the interior of the saddle connection, subdividing it into several horizontal saddle connections. \square

Proposition 4.13 (One part of the Veech dichotomy). *Suppose (X, ω) is a lattice surface, and has a horizontal saddle connection. Then (X, ω) is horizontally periodic.*

Proof. Let (X', ω') be horizontally periodic and in the $u_t(X, \omega)$ -orbit closure of (X, ω) . Let T be large, so $u_T(X, \omega)$ is very close to (X', ω') . The horizontal saddle connection is present on (X', ω') as a union of horizontal saddle connections (it will turn out to be only one). There must be some matrix $g \in SL(2, \mathbb{R})$ close to the identity so $gu_T(X, \omega) = (X', \omega')$, because they are both in the same orbit. Also, g must preserve the horizontal direction, since it must preserve the horizontal saddle connections. However, that means g is a unipotent upper triangular matrix $g = u_S$, so $u_{T+S}(X, \omega)$ is horizontally periodic, so (X, ω) is horizontally periodic. \square

Definition 4.14. A *rel deformation* of translation surface is a path (X_t, ω_t) , $t \in [a, b]$, in a stratum, such that for any absolute homology class γ , $\int_\gamma \omega_t$ is constant.

Thus along rel deformations absolute periods are constant, but relative periods (i.e., the complex distance between zeros of ω) can change.

Example 4.15. If (X', ω') is a translation cover of (X, ω) , then moving the branch points gives a rel deformation of (X', ω') , consisting entirely of surfaces that cover (X, ω) . (Compare to the computations in Example 2.2.)

Lemma 4.16. *Suppose that (X, ω) and (X', ω') are nearby surfaces in some stratum, and as subspaces of absolute cohomology $H^1(X, \mathbb{C})$ we*

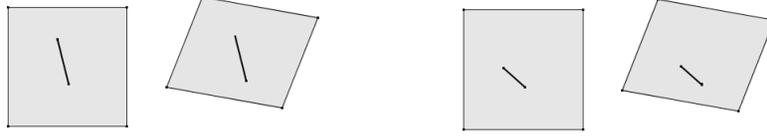


FIGURE 4.4. The translation surface on the right and the one on the left are rel deformations of each other. More generally, changing the length of the slit in the slit torus construction gives a rel deformation. It is possible to write down a basis for absolute homology consisting of cycles disjoint from the slit, which shows that the integral of any absolute homology class does not change as the complex length of the slit is changed.

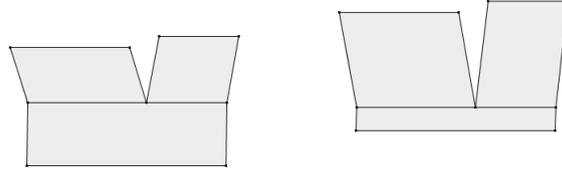


FIGURE 4.5. Opposite edges are identified (the bottom edges each consist of two saddle connections), giving two surfaces in $\mathcal{H}(1, 1)$. The translation surface on the right and the surface on the left are rel deformations of each other. The surface on the right was obtained from the one on the left by subtracting 1 from the height of the bottom cylinder, and adding 1 to the height of the two top cylinders.

have

$$\text{span}_{\mathbb{R}}(\text{Re}(\omega), \text{Im}(\omega)) = \text{span}_{\mathbb{R}}(\text{Re}(\omega'), \text{Im}(\omega')).$$

Then there is some $g \in GL^+(2, \mathbb{R})$ close to the identity such that $g(X', \omega')$ is a rel deformation of (X, ω) .

Proof. By assumption, there are constants $a, b, c, d \in \mathbb{R}$ such that, in absolute cohomology,

$$\text{Re}(\omega') = a \text{Re}(\omega) + b \text{Im}(\omega) \quad \text{and} \quad \text{Im}(\omega') = c \text{Re}(\omega) + d \text{Im}(\omega).$$

Thus if

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then $g(X, \omega)$ and (X', ω') have the same absolute periods, and hence are rel deformations of each other. In particular, the linear path in local periods joining $g(X, \omega)$ and (X', ω') is a rel deformation. \square

Definition 4.17. An affine invariant submanifold \mathcal{M} is *rank 1* if, for every $(X, \omega) \in \mathcal{M}$, there is an open neighbourhood U containing (X, ω) , such that for every $(X', \omega') \in U$ there is a $g \in GL^+(2, \mathbb{R})$ close to the identity such that there is a rel deformation in U from (X', ω') to $g(X, \omega)$.

A definition of rank will be given in the next section, so this can be considered a provisional definition of rank 1. It can be rephrased as saying that the $GL^+(2, \mathbb{R})$ directions and the directions of rel deformations in \mathcal{M} span the tangent space to \mathcal{M} at every point. By the previous lemma, it can also be rephrased as saying that $\text{span}(\text{Re}(\omega), \text{Im}(\omega))$ is locally constant on \mathcal{M} .

Proposition 4.18. *Let \mathcal{M} be a 2-dimensional affine invariant submanifold. Then \mathcal{M} is rank 1. Furthermore, let \mathcal{M}' be a connected component of the space of degree d translation covers of surfaces in \mathcal{M} branched over k points. Then \mathcal{M}' is rank 1 also.*

Proof. A 2-dimensional affine invariant submanifold \mathcal{M} is a single $GL^+(2, \mathbb{R})$ -orbit, and so is in particular rank 1 (the rel deformations aren't even required).

If $(X', \omega) \in \mathcal{M}'$ is a cover of $(X, \omega) \in \mathcal{M}$, then a neighborhood in \mathcal{M}' of (X', ω) is obtained by changing (X, ω) by a small matrix in $GL^+(2, \mathbb{R})$, and changing the location of the branch points. \square

Proposition 4.19. *The eigenform loci in genus two constructed in Section 2 are rank 1.*

Proof. These are defined by saying that the real and imaginary parts of ω should span a fixed 2-dimensional subspace of absolute homology (the \sqrt{D} -eigenspace of M). \square

Proposition 4.20 (Wright). *If $(X, \omega) \in \mathcal{M}$ and \mathcal{M} is rank 1, then (X, ω) is periodic in any direction that has a cylinder.*

That is, the proposition says that every time you find a cylinder on (X, ω) , then (X, ω) is the union of that cylinder and cylinders parallel to it. Before being established in general in [Wria] (using a different argument than the one we give here), the proposition was known in several special cases [Vee89, Cal04, LN].

Proof. Let (X', ω') be horizontally periodic and in the u_t -orbit closure of (X, ω) .

Let T be large, so $u_T(X, \omega)$ is very close to (X', ω') . The horizontal cylinder is present on (X', ω') . There must be some matrix $g \in SL(2, \mathbb{R})$ close to the identity so $gu_T(X, \omega)$ is a rel deformation of (X', ω') . Small rel deformations preserve cylinders, since the integral over their circumference curve γ must remain constant along the rel deformation. Hence g must preserve the horizontal direction, since it must preserve the horizontal cylinder which is present on both $u_T(X, \omega)$ and $gu_T(X, \omega)$.

Thus, every horizontal cylinder on (X', ω') is also horizontal on $u_T(X, \omega)$. If these cylinders do not cover $u_T(X, \omega)$, then it is possible to derive a contradiction, because then there would be more horizontal cylinders on (X', ω') . (Because every horizontal cylinder on $u_T(X, \omega)$ must also be present on every translation surface in the u_t -orbit closure, and if they do not cover $u_T(X, \omega)$ then the corresponding cylinders on (X', ω') do not cover (X', ω') , and hence there must be more cylinders on (X', ω') than on (X, ω) .) \square

Remark 4.21. In fact this proof shows that if (X, ω) has a loop γ of saddle connections in a fixed direction such that $\int_\gamma \omega \neq 0$, then (X, ω) is periodic in that direction.

We conclude by remarking that there is also a close connection between rank 1 orbit closures and real multiplication. The proof requires dynamics.

Theorem 4.22 (Filip). *If $(X, \omega) \in \mathcal{M}$ and \mathcal{M} is rank 1, then $\text{Jac}(X, \omega)$ has real multiplication.*

The conclusion is that rank 1 orbit closures are very close cousins to 2-dimensional orbit closures, which are a special case.

5. REVISITING GENUS TWO WITH NEW TOOLS

5.1. Field of definition, VHS.

Definition 5.1. The (affine) field of definition $\mathbf{k}(\mathcal{M})$ of an affine invariant submanifold \mathcal{M} is the smallest subfield of \mathbb{R} such that \mathcal{M} can be defined in local period coordinates by linear equations with coefficients in this field [Wrib]. Warning: This is not the same thing as the field of definition of \mathcal{M} viewed as a variety (where polynomial equations are allowed, and the coordinates are different).

For example, \mathcal{M} arising from branched covers over tori, (or over all surfaces in some other stratum) are defined over \mathbb{Q} . The eigenform loci are defined over $\mathbb{Q}[\sqrt{D}]$.

Let H^1 denote the flat bundle over \mathcal{M} whose fiber over $(X, \omega) \in \mathcal{M}$ is $H^1(X, \mathbb{C})$, and let H_{rel}^1 denote the flat bundle whose fiber over (X, ω) is $H^1(X, \Sigma, \mathbb{C})$, where Σ is the set of singularities of (X, ω) . Let

$$p : H_{rel}^1 \rightarrow H^1$$

denote the natural map from relative to absolute cohomology. Viewing

$$H^1(X, \Sigma, \mathbb{C}) = H_1(X, \Sigma, \mathbb{C})^* \quad \text{and} \quad H^1(X, \mathbb{C}) = H_1(X, \mathbb{C})^*,$$

the map p is just restriction of a linear functional on $H_1(X, \Sigma, \mathbb{C})$ to $H_1(X, \mathbb{C}) \subset H_1(X, \Sigma, \mathbb{C})$. The subspace $\ker(p)$ exactly corresponds to derivatives of rel deformations.

The flat connection on H^1 or H_{rel}^1 is often called the Gauss-Manin connection. From our point of view, it is an extremely simple thing. The cohomology groups $H^1(X, \mathbb{C})$ and $H^1(X, \Sigma, \mathbb{C})$ are both purely topological objects, and do not depend on the complex structure on X or the Abelian differential ω . (When the Abelian differential changes, the set Σ might move a bit by an isotopy.) Thus, varying the complex structure on X does not change these cohomology groups. In this way, if (X', ω') is nearby (X, ω) , then $H^1(X, \mathbb{C})$ is identified with $H^1(X', \omega')$ (because it is the same exact object!), and similarly for relative cohomology. This identification of nearby fibers is exactly the structure of a flat connection.

Recall that period coordinates can be considered as the map sending (X, ω) to the relative cohomology class of ω in $H^1(X, \Sigma, \mathbb{C})$. By definition, \mathcal{M} is defined in these periods by a linear subspace, which we think of as simultaneously giving \mathcal{M} in period coordinates, as well as being the tangent space to \mathcal{M} at (X, ω) . (The tangent space to a vector space, at any point, is just the vector space itself.) Thus we can consider the tangent bundle $T(\mathcal{M})$ of \mathcal{M} as a flat subbundle of H_{rel}^1 . It is flat because the subspace defining \mathcal{M} in period coordinates does not change as (X, ω) moves around in \mathcal{M} .

A *flat subbundle* E of H^1 or H_{rel}^1 is just a subbundle that is locally constant over \mathcal{M} . Associated to such a subbundle is its *monodromy representation*, which is a representation of $\pi_1(\mathcal{M})$ on the fiber of E at the chosen base point of \mathcal{M} . It is obtained by dragging cohomology classes around loops in $\pi_1(\mathcal{M})$.

A flat subbundle is called *simple* if it has no nontrivial flat subbundle, or equivalently if the monodromy representation has no nontrivial invariant subspaces. A flat subbundle is called *semisimple* if it is the direct sum of simple subbundles. Two subbundles are called Galois conjugate if their fibers are Galois conjugate; in particular, this means their monodromy representations are Galois conjugate.

Definition 5.2. The *field of definition* of a flat subbundle $E \subset H^1$ is the smallest subfield of \mathbb{C} such that locally the linear subspace E of $H^1(X, \mathbb{C})$ can be defined by linear equations (with respect to an integer basis of $H_1(X, \mathbb{Z})$) with coefficients in this field. The trace field of a flat bundle over \mathcal{M} is defined as the field generated by traces of the corresponding representation of $\pi_1(\mathcal{M})$.

Theorem 5.3 (Wright [Wrib]). *Let \mathcal{M} be an affine invariant submanifold. The field of definition of $p(T(\mathcal{M}))$ and trace field of $p(T(\mathcal{M}))$ are both equal to $\mathbf{k}(\mathcal{M})$.*

Set $\mathbb{V}_{\text{Id}} = p(\mathcal{T}(\mathcal{M}))$. There is a semisimple flat bundle \mathbb{W} , and for each field embedding $\rho : \mathbf{k}(\mathcal{M}) \rightarrow \mathbb{C}$ there is a flat simple bundle \mathbb{V}_ρ that is Galois conjugate to \mathbb{V}_{Id} , such that

$$H^1 = \left(\bigoplus_{\rho} \mathbb{V}_\rho \right) \oplus \mathbb{W}.$$

The bundle \mathbb{W} does not contain any subbundles isomorphic to any \mathbb{V}_ρ . Both \mathbb{W} and $\bigoplus \mathbb{V}_\rho$ are defined over \mathbb{Q} . All direct summands are symplectic and symplectically orthogonal.

In particular,

$$\dim_{\mathbb{C}} p(T(\mathcal{M})) \cdot \deg_{\mathbb{Q}} \mathbf{k}(\mathcal{M}) \leq 2g.$$

Corollary 5.4 (Wright). *In particular, the field of definition is a number field, and so any translation surface whose coordinates are linearly independent over $\overline{\mathbb{Q}}$ cannot be contained in a nontrivial affine invariant submanifold, and hence must have $GL^+(2, \mathbb{R})$ -orbit closure as large as possible. This provides an explicit full measure set of surfaces whose orbit closure is as large as possible.*

The direct sum decomposition of H^1 in Theorem 5.3 was previously established in the case of Teichmüller curves by Martin Möller [Möl06b], and is one of the main tools used in the study of closed $SL(2, \mathbb{R})$ -orbits.

Theorem 5.5 (Möller [Möl06b], Filip [Filb]). *The splitting above is a splitting of Variation of Hodge Structures. That is, each direct summand is equal to the direct sum of its intersection with $H^{1,0}$ and its intersection with $H^{0,1}$.*

Again the proofs use dynamics. For an elementary introduction of Variation of Hodge Structures in the context of orbit closures, see [Wri12].

When the splitting of H^1 is nontrivial ($H^1 \neq p(T(\mathcal{M}))$), then \mathcal{M} parameterizes translation surfaces whose Jacobians admit nontrivial endomorphisms.

Given that affine invariant submanifolds are varieties, parts (but not all) of the two theorems above follow from a theorem of Deligne on semi-simplicity of VHS [Del87]. However, in fact both the above theorems were established first, and used by Filip in his proof that affine invariant submanifolds are varieties. The first theorem used work of Avila-Eskin-Möller [AEM], and the following.

Theorem 5.6 (Eskin-Mirzakhani-Rafi [EMR]). *In every affine invariant submanifolds, there are lots of closed (i.e., periodic) g_t -orbits.*

This is useful because of a classical result in Teichmüller theory that says the monodromy matrix over such an orbit has simple largest and smallest eigenvalues, with eigenvectors $\operatorname{Re}(\omega)$ and $\operatorname{Im}(\omega)$. Among other things, this helps to show that there is not a second copy of $p(T(\mathcal{M}))$ in the decomposition of H^1 , since otherwise these eigenvalues would have multiplicity at least two.

The work of Avila-Eskin-Möller used in the proof also shows

Theorem 5.7 (Avila-Eskin-Möller). *$p(T(\mathcal{M}))$ is symplectic.*

Definition 5.8. The *rank* of an affine invariant submanifold \mathcal{M} is $\frac{1}{2} \dim_{\mathbb{C}} p(T(\mathcal{M}))$.

When $p(T(\mathcal{M}))$ is 2-dimensional, that means that it must be spanned by the real and imaginary parts of the absolute cohomology classes given by ω . In particular, $\operatorname{span}(\operatorname{Re}(\omega), \operatorname{Im}(\omega))$ is locally constant, since $p(T(\mathcal{M}))$ is always a flat subbundle. By Lemma 4.16, this means that a neighbourhood of any (X, ω) in \mathcal{M} can be generated using $GL^+(2, \mathbb{R})$ and rel deformations, so rank 1 according to this definition agrees with our previous definition of rank 1.

5.2. Cylinder deformations. We begin with the observation that, for each cylinder on a translation surface (X, ω) , there is a corresponding cylinder on sufficiently nearby surfaces (X', ω') . This corresponding cylinder may not have the same direction, height, circumference, or modulus, however these all change continuously with (X', ω') . The “sufficiently nearby” assumption is required since, along a path a surfaces starting at (X, ω) , the height of the cylinder might reach 0, at which point the cylinder ceases to be a cylinder.

Definition 5.9. Two cylinders C_1 and C_2 on $(X, \omega) \in \mathcal{M}$ are said to be \mathcal{M} -parallel if they are parallel, and remain parallel on all deformations

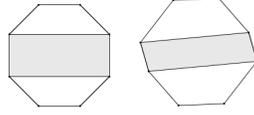


FIGURE 5.1. Consider the shaded horizontal cylinder on the regular octagon surface in $\mathcal{H}(2)$ (left). On any sufficiently small deformation of this surface, the cylinder persists (right).

of (X, ω) in \mathcal{M} . (A *deformation* is just a nearby surface, connected to (X, ω) via a path in \mathcal{M} .) The deformations are assumed to be small, so that C_1 and C_2 persist.

Example 5.10. If \mathcal{M} is 2-dimensional, it consists of a single $GL^+(2, \mathbb{R})$ -orbit. Parallel cylinders remain parallel under the action of $GL^+(2, \mathbb{R})$, so two cylinders on a translation surface in \mathcal{M} are \mathcal{M} -parallel if and only if they are parallel.

Example 5.11. On the opposite extreme, if \mathcal{M} is a connected component of a stratum, two cylinders are \mathcal{M} parallel if and only if they are parallel and their circumference curves are homologous. Indeed, suppose the core curves are γ_1 and γ_2 . Then since $\int_{\gamma_1} \omega' = \int_{\gamma_2} \omega'$ on all deformations $(X', \omega') \in \mathcal{M}$, the cylinders will always be in the same direction, which is exactly the direction in \mathbb{C} given by this integral.

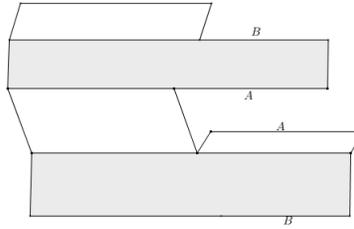


FIGURE 5.2. The two shaded cylinders are homologous on this translation surface. There is no way to deform the surface to make these two cylinders not be parallel.

Lemma 5.12. *Two cylinders C_1 and C_2 on $(X, \omega) \in \mathcal{M}$ are \mathcal{M} -parallel if and only if there is a constant $c \in \mathbb{R}$ such that on all deformations $(X', \omega') \in \mathcal{M}$,*

$$\int_{\gamma_1} \omega' = c \int_{\gamma_2} \omega',$$

where γ_i is the circumference curve of C_i .

In other words, two cylinders on $(X, \omega) \in \mathcal{M}$ are \mathcal{M} -parallel if and only if one of the linear equations defining \mathcal{M} in local period coordinates makes it so.

Proof. Both $(X', \omega') \mapsto \int_{\gamma_1} \omega'$ and $(X', \omega') \mapsto \int_{\gamma_2} \omega'$ are linear functionals on a neighbourhood of (X, ω) in \mathcal{M} , viewed in period coordinates as locally being an open set in a complex vector space. The cylinders C_1, C_2 are \mathcal{M} parallel if and only if these two linear functions are always real multiples of each other.

The only way for two linear functionals on a complex vector space always have real ratio is for one functional to be a fixed real constant c times the other. \square

The relation of being \mathcal{M} -parallel is an equivalence relation, and when we speak of an equivalence class of a cylinder, we mean the set of all cylinders \mathcal{M} -parallel to it.

Define the matrices

$$u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad a_s = \begin{pmatrix} 1 & 0 \\ 0 & e^s \end{pmatrix}, \quad r_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Let \mathcal{C} be a collection of parallel cylinders on a translation surface (X, ω) . Suppose they are all of angle $\theta \in [0, \pi)$. Define $a_s^{\mathcal{C}}(u_t^{\mathcal{C}}(X, \omega))$ to be the translation surface obtained by applying $r_{-\theta}$ to (X, ω) , then applying the matrix $a_s u_t$ to the images of the cylinders in \mathcal{C} , and then applying r_θ .

The result of $a_s^{\mathcal{C}}$ is to stretch the height of all cylinders in the collection \mathcal{C} by a factor of e^s . The result of $u_t^{\mathcal{C}}$ is to shear all the cylinders in \mathcal{C} .

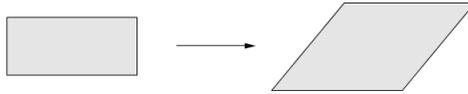


FIGURE 5.3. A horizontal cylinder (left—vertical sides are identified, but the top and bottom horizontal edges are the boundary of the cylinder.) The result of stretching and shearing this cylinder (right). Note the boundary of the cylinder stays exactly the same.

There is more than one way of thinking about these cylinder deformations. You can think of cutting out the cylinders in \mathcal{C} , and then stretching and shearing them, and then gluing them back in. Or, you can think of a polygon decomposition for (X, ω) , with one parallelogram for each cylinder in \mathcal{C} , and you can think of applying the matrix

$a_s u_t$ just to the parallelograms giving cylinders in \mathcal{C} and doing nothing to the remaining polygons.

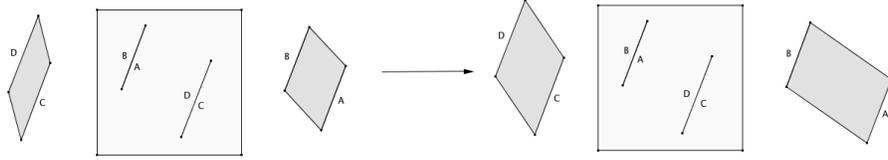


FIGURE 5.4. On the left is a translation surface (X, ω) . Let \mathcal{C} be the two shaded parallel cylinders on (X, ω) . On the right is $a_{\log(2)}^{\mathcal{C}}(X, \omega)$.

Theorem 5.13 (The Cylinder Deformation Theorem [Wria]). *Suppose that \mathcal{C} is an equivalence class of \mathcal{M} -parallel cylinders on $(X, \omega) \in \mathcal{M}$. Then for all $s, t \in \mathbb{R}$, the translation surface $a_s^{\mathcal{C}}(u_t^{\mathcal{C}}(X, \omega)) \in \mathcal{M}$.*

We call $a_s^{\mathcal{C}}$ the *cylinder stretch*, and $u_t^{\mathcal{C}}$ the *cylinder shear*.

The proof involves the dynamics of u_t , but is motivated by the then conjecture (now theorem) that affine invariant submanifolds are varieties. As $s \rightarrow \infty$, the translation surface $a_s^{\mathcal{C}}(\mathcal{C}(X, \omega))$ converges to the boundary at infinity of \mathcal{M} . The Cylinder Deformation Theorem is closely related to the boundary structure of \mathcal{M} .

More generally, the boundary of an affine invariant submanifold is related to configurations of parallel cylinders and saddle connections, since these can always be made vertical using r_θ and then shrunk using g_t .

5.3. Orbit closures in genus 2. In this subsection, we will give a qualitative description of orbit closures in genus 2. However, we will not give a classification, in that we will not discuss how many connected components the loci we discuss have, and we will not discuss how to tell if two translation surfaces have the same orbit closure. An almost complete classification in genus two was given by McMullen, the only remaining open problem being to classify orbits of square-tiled surfaces in $\mathcal{H}(1, 1)$. (Such a classification should give a finite list of invariants, such that if two square-tiled surfaces have the same invariants, then they are in the same orbit.) Thus all results in this section can be deduced as particular consequences of finer results of McMullen.

See [McM05, McM06b, McM03b, Bai07, Mukb, McM07, KM] for much finer information on orbit closures in genus 2. Note that

McMullen’s work [McM07] was done well before the work of Eskin-Mirzakhani-Mohammadi, and hence uses different techniques than what we present here.

Lemma 5.14. *In $\mathcal{H}(2)$ every $GL^+(2, \mathbb{R})$ -orbit is dense or closed. Every closed orbit is either the orbit of a square-tiled surface, or one of the eigenform loci in $\mathcal{H}(2)$ constructed in Section 2.*

Even the orbits of square-tiled surfaces can be fit into the framework of eigenforms by considering “real multiplication by $\mathbb{Q} \oplus \mathbb{Q}$ ”, but we do not pursue that perspective here.

Proof. Every orbit closure is an affine invariant submanifold. In strata with only one zero, $p : H^1(X, \Sigma, \mathbb{C}) \rightarrow H^1(X, \mathbb{C})$ is an isomorphism, and hence $T(\mathcal{M}) = p(T(\mathcal{M}))$ is symplectic. Hence because $\mathcal{M} \subset \mathcal{H}(2)$, we see that \mathcal{M} is 2-dimensional or 4-dimensional. $\mathcal{H}(2)$ is 4-dimensional and connected, so any 4-dimensional affine invariant submanifold is all of $\mathcal{H}(2)$.

Thus we may assume that \mathcal{M} is 2-dimensional, i.e., a closed orbit. If \mathcal{M} is defined over \mathbb{Q} , it contains surface (X, ω) whose periods

$$\left\{ \int_{\gamma} \omega : \gamma \in H_1(X, \mathbb{Z}) \right\} \subset \frac{1}{n} \mathbb{Z}[i]$$

for some n . This implies (X, ω) is square-tiled surface (a cover of $\mathbb{C}/(\frac{1}{n}\mathbb{Z}[i])$).

If \mathcal{M} is defined over $\mathbb{Q}[\sqrt{D}]$, then it follows from Theorem 5.3 that

$$\text{span}(\text{Re}(\omega), \text{Im}(\omega))$$

is defined over $\mathbb{Q}[\sqrt{D}]$ and is symplectically orthogonal to the Galois conjugate subspace. Hence by Section 2, (X, ω) is an eigenform for real multiplication by an order in $\mathbb{Q}[\sqrt{D}]$. \square

Similarly we obtain

Lemma 5.15. *In $\mathcal{H}(1, 1)$, every 3 dimensional rank 1 orbit closure consists either of torus covers, or of eigenforms for real multiplication by $\mathbb{Q}[\sqrt{D}]$.*

The only other dimension a rank 1 orbit closure could have in $\mathcal{H}(1, 1)$ is 2, and there are indeed closed orbits in $\mathcal{H}(1, 1)$. McMullen has shown that all but one contains a square-tiled surface [McM06b]. This one exceptional closed orbit is a two dimensional submanifold of the locus of eigenforms for real multiplication for $\mathbb{Q}[\sqrt{5}]$.

To complete the qualitative classification of orbit closures in $\mathcal{H}(1, 1)$, it remains only to show the following.

Proposition 5.16. *Any affine invariant submanifold of $\mathcal{H}(1,1)$ that is not rank 1 must be all of $\mathcal{H}(1,1)$.*

Proof. Since $p(T(\mathcal{M}))$ is symplectic, and rank is $\frac{1}{2} \dim p(T(\mathcal{M}))$, we see that $p(T(\mathcal{M})) = H^1$. (In genus 2, H^1 is 4-dimensional.) We wish to show that in fact \mathcal{M} is 5-dimensional, in which case it must be all of $\mathcal{H}(1,1)$.

Lemma 5.17. *Such \mathcal{M} must contain a horizontally periodic surface with 3 horizontal cylinders. Two of the cylinders each have a single saddle connection in their boundary, and both of these are glued to a third larger cylinder above and below, as in Figure 5.5.*

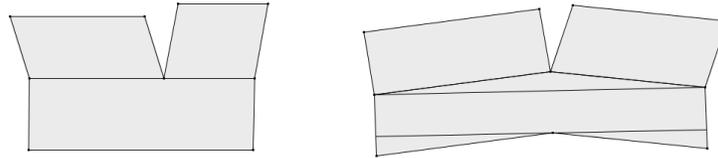


FIGURE 5.5. On the left is a translation surface in $\mathcal{H}(1,1)$. On the right is an illustration of a deformation of this surface where the three horizontal cylinders cease to be parallel. (This exact deformation is not guaranteed to be in \mathcal{M} , however one like it is.)

Given the lemma, the proposition is not so hard to prove. Indeed, it is easily verified that none of the three cylinders are \mathcal{M} parallel to each other: By Lemma 5.12, two cylinders are \mathcal{M} parallel if and only if the integrals of their circumference curves are always proportional. However, since $p(T(\mathcal{M})) = H^1(X, \mathbb{C})$, no restriction is placed on the absolute periods of a translation surface in \mathcal{M} , and we can change the circumferences in an arbitrary way. (As we do so, the relative period might be determined by the absolute ones—but this is precisely what we will show cannot happen, since \mathcal{M} will end up being $\mathcal{H}(1,1)$.)

Since none of the three cylinders are \mathcal{M} parallel to each other, we can increase the height of top two, and decrease the height of the bottom one, so as to produce a rel deformation in \mathcal{M} , as in Figure 4.5. Since $p(T(\mathcal{M})) = H^1(X, \mathcal{M})$, and $T(\mathcal{M})$ contains a vector in the 1-dimensional $\ker(p)$ (the rel deformation), we see that $T(\mathcal{M}) = H^1(X, \Sigma, \mathbb{C})$. Hence \mathcal{M} is 5-dimensional, and $\mathcal{M} = \mathcal{H}(1,1)$. \square

The proof of Lemma 5.17 is easy but will be omitted. (One can find a horizontally periodic surface in every horocycle flow orbit closure,

and work from there to get one with three horizontal cylinders. See Figure 5.6, and compare to arguments in [NW, ANW].)

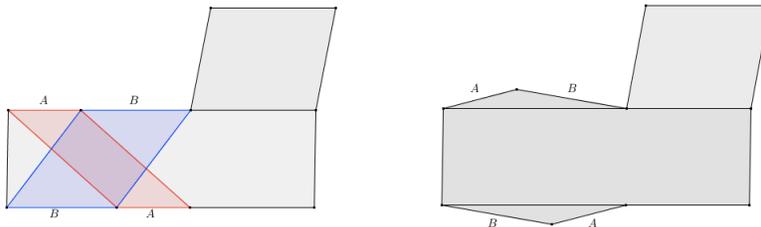


FIGURE 5.6. The proof of Lemma 5.17 is omitted, but this figure gives the idea. On the left is a translation surface with two horizontal cylinders. Shearing the two shaded non-horizontal cylinders appropriately “opposite” amounts gives the surface on the right, which now has three horizontal cylinders as desired.

5.4. Census of known orbit closures. Here we give a list of the currently known orbit closures not arising from the elementary constructions discussed in Section 2.1. We give only primitive orbit closures, i.e., those not arising from covering constructions.

Closed orbits. McMullen and Calta independently discovered infinitely many closed orbits in $\mathcal{H}(2)$ [Cal04, McM03a]. McMullen generalized his approach using real multiplication to find infinitely many primitive closed orbits in genus 3 and 4, in the Prym loci in the strata where ω has only one zero [McM06a].

There is a bi-infinite sequence of Teichmüller curves $\mathcal{T}(n, m)$ called the Veech-Ward-Bouw-Möller curves. For $n = 2$ they were discovered by Veech [Vee89]; for $n = 3$ by Ward [War98]; and in the general case by Bouw and Möller [BM10]. They have also been studied by the author and Hooper [Wri13, Hoo].

There are also two “sporadic” closed orbits known, one due to Vorobets in $H(6)$, and another due to Kenyon-Smillie in $H(1, 3)$ [HS01, KS00]. These sporadic examples correspond to billiards in the $(\pi/5, \pi/3, 7\pi/15)$ and $(2\pi/9, \pi/3, 4\pi/9)$ triangles respectively.

Rank 1 affine invariant manifolds. McMullen and Calta discovered infinitely many rank 1 affine invariant submanifolds in $\mathcal{H}(1, 1)$ [Cal04, McM03a]. McMullen generalized his construction to give infinitely many additional examples in the Prym loci in genus 3, 4, and 5 [McM06a].

A new orbit closure. In joint work in progress with Mirzakhani, the author has found the first known example of a higher rank orbit closure whose affine field of definition is not \mathbb{Q} .

REFERENCES

- [AEM] Artur Avila, Alex Eskin, and Martin Möller, *Symplectic and Isometric $SL(2, \mathbb{R})$ -invariant subbundles of the Hodge bundle*, preprint, arXiv 1209.2854 (2012).
- [ANW] David Auricino, Duc-Manh Nguyen, and Alex Wright, *Classification of higher rank orbit closures in $\mathcal{H}^{\text{odd}}(4)$* , preprint, arXiv 1308.5879 (2013), to appear in Geom. Topol.
- [Bai07] Matt Bainbridge, *Euler characteristics of Teichmüller curves in genus two*, Geom. Topol. **11** (2007), 1887–2073.
- [BM10] Irene I. Bouw and Martin Möller, *Teichmüller curves, triangle groups, and Lyapunov exponents*, Ann. of Math. (2) **172** (2010), no. 1, 139–185.
- [BM12] Matt Bainbridge and Martin Möller, *The Deligne-Mumford compactification of the real multiplication locus and Teichmüller curves in genus 3*, Acta Math. **208** (2012), no. 1, 1–92.
- [Cal04] Kariane Calta, *Veech surfaces and complete periodicity in genus two*, J. Amer. Math. Soc. **17** (2004), no. 4, 871–908.
- [Del87] P. Deligne, *Un théorème de finitude pour la monodromie*, Discrete groups in geometry and analysis (New Haven, Conn., 1984), Progr. Math., vol. 67, Birkhäuser Boston, Boston, MA, 1987, pp. 1–19.
- [DT02] Tobin A. Driscoll and Lloyd N. Trefethen, *Schwarz-Christoffel mapping*, Cambridge Monographs on Applied and Computational Mathematics, vol. 8, Cambridge University Press, Cambridge, 2002.
- [EM] Alex Eskin and Maryam Mirzakhani, *Invariant and stationary measures for the $SL(2, \mathbb{R})$ action on moduli space*, preprint, arXiv 1302.3320 (2013).
- [EMM] Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi, *Isolation theorems for $SL(2, \mathbb{R})$ -invariant submanifolds in moduli space*, preprint, arXiv:1305.3015 (2013).
- [EMR] Alex Eskin, Maryam Mirzakhani, and Kasra Rafi, *Counting closed geodesics in strata*, preprint, arXiv 1206.5574 (2012).
- [Esk06] Alex Eskin, *Counting problems in moduli space*, Handbook of dynamical systems. Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 581–595.
- [Fila] Simion Filip, *Semisimplicity and rigidity of the Kontsevich-Zorich cocycle*, preprint, arXiv:1307.7314 (2013).
- [Filb] ———, *Splitting mixed Hodge structures over affine invariant manifolds*, preprint, arXiv:1311.2350 (2013).
- [For06a] Giovanni Forni, *On the Lyapunov exponents of the Kontsevich-Zorich cocycle*, Handbook of dynamical systems. Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 549–580.
- [For06b] ———, *On the Lyapunov exponents of the Kontsevich-Zorich cocycle*, Handbook of dynamical systems. Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 549–580.

- [Hoo] W. Patrick Hooper, *Grid graphs and lattice surfaces*, preprint, arXiv 0811.0799 (2009).
- [HS01] P. Hubert and T. A. Schmidt, *Invariants of translation surfaces*, Ann. Inst. Fourier (Grenoble) **51** (2001), no. 2, 461–495.
- [HS06] Pascal Hubert and Thomas A. Schmidt, *An introduction to Veech surfaces*, Handbook of dynamical systems. Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 501–526.
- [KM] Abhinav Kumar and Ronen E. Mukamel, *Algebraic models and arithmetic geometry of Teichmüller curves in genus two*, preprint, arXiv 1406.7057 (2014).
- [KMS86] Steven Kerckhoff, Howard Masur, and John Smillie, *Ergodicity of billiard flows and quadratic differentials*, Ann. of Math. (2) **124** (1986), no. 2, 293–311.
- [KS00] Richard Kenyon and John Smillie, *Billiards on rational-angled triangles*, Comment. Math. Helv. **75** (2000), no. 1, 65–108.
- [LN] Erwan Lanneau and Duc-Mahn Nguyen, *Complete periodicity of prym eigenforms*, preprint, arXiv 1301.0783 (2013).
- [Mas82] Howard Masur, *Interval exchange transformations and measured foliations*, Ann. of Math. (2) **115** (1982), no. 1, 169–200.
- [Mas92] ———, *Hausdorff dimension of the set of nonergodic foliations of a quadratic differential*, Duke Math. J. **66** (1992), no. 3, 387–442.
- [Mas06] ———, *Ergodic theory of translation surfaces*, Handbook of dynamical systems. Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 527–547.
- [McM03a] Curtis T. McMullen, *Billiards and Teichmüller curves on Hilbert modular surfaces*, J. Amer. Math. Soc. **16** (2003), no. 4, 857–885 (electronic).
- [McM03b] ———, *Teichmüller geodesics of infinite complexity*, Acta Math. **191** (2003), no. 2, 191–223.
- [McM05] ———, *Teichmüller curves in genus two: discriminant and spin*, Math. Ann. **333** (2005), no. 1, 87–130.
- [McM06a] ———, *Prym varieties and Teichmüller curves*, Duke Math. J. **133** (2006), no. 3, 569–590.
- [McM06b] ———, *Teichmüller curves in genus two: torsion divisors and ratios of sines*, Invent. Math. **165** (2006), no. 3, 651–672.
- [McM07] ———, *Dynamics of $SL_2(\mathbb{R})$ over moduli space in genus two*, Ann. of Math. (2) **165** (2007), no. 2, 397–456.
- [McM09] ———, *Rigidity of Teichmüller curves*, Math. Res. Lett. **16** (2009), no. 4, 647–649.
- [Möl06a] Martin Möller, *Periodic points on Veech surfaces and the Mordell-Weil group over a Teichmüller curve*, Invent. Math. **165** (2006), no. 3, 633–649.
- [Möl06b] ———, *Variations of Hodge structures of a Teichmüller curve*, J. Amer. Math. Soc. **19** (2006), no. 2, 327–344 (electronic).
- [Möl08] ———, *Finiteness results for Teichmüller curves*, Ann. Inst. Fourier (Grenoble) **58** (2008), no. 1, 63–83.
- [Möl09] ———, *Affine groups of flat surfaces*, Handbook of Teichmüller theory. Vol. II, IRMA Lect. Math. Theor. Phys., vol. 13, Eur. Math. Soc., Zürich, 2009, pp. 369–387.

- [Möl13] ———, *Teichmüller curves, mainly from the viewpoint of algebraic geometry*, Moduli spaces of Riemann surfaces, IAS/Park City Math. Ser., vol. 20, Amer. Math. Soc., Providence, RI, 2013, pp. 267–318.
- [MT02] Howard Masur and Serge Tabachnikov, *Rational billiards and flat structures*, Handbook of dynamical systems, Vol. 1A, North-Holland, Amsterdam, 2002, pp. 1015–1089.
- [Muka] Ronen Mukamel, *Fundamental domains and generators for lattice Veech groups*, <http://www.stanford.edu/~ronen/papers/FundamentalDomains.pdf>, preprint.
- [Mukb] ———, *Orbifold points on Teichmüller curves and Jacobians with complex multiplication*, <http://www.stanford.edu/~ronen/papers/OrbifoldPts.pdf>, to appear in Geom. Topol.
- [MW] Carlos Matheus and Alex Wright, *Hodge-Teichmüller planes and finiteness results for Teichmüller curves*, preprint, arXiv 1308.0832 (2013).
- [MW02] Yair Minsky and Barak Weiss, *Nondivergence of horocyclic flows on moduli space*, J. Reine Angew. Math. **552** (2002), 131–177.
- [NW] Duc-Manh Nguyen and Alex Wright, *Non-Veech surfaces in $\mathcal{H}^{\text{hyp}}(4)$ are generic*, preprint, arXiv:1306.4922 (2013), to appear in Geom. Funct. Anal.
- [SW04] John Smillie and Barak Weiss, *Minimal sets for flows on moduli space*, Israel J. Math. **142** (2004), 249–260.
- [SW07] ———, *Finiteness results for flat surfaces: a survey and problem list*, Partially hyperbolic dynamics, laminations, and Teichmüller flow, Fields Inst. Commun., vol. 51, Amer. Math. Soc., Providence, RI, 2007, pp. 125–137.
- [Vee86] William A. Veech, *The Teichmüller geodesic flow*, Ann. of Math. (2) **124** (1986), no. 3, 441–530.
- [Vee89] W. A. Veech, *Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards*, Invent. Math. **97** (1989), no. 3, 553–583.
- [Vee95] William A. Veech, *Geometric realizations of hyperelliptic curves*, Algorithms, fractals, and dynamics (Okayama/Kyoto, 1992), Plenum, New York, 1995, pp. 217–226.
- [Via06] Marcelo Viana, *Ergodic theory of interval exchange maps*, Rev. Mat. Complut. **19** (2006), no. 1, 7–100.
- [War98] Clayton C. Ward, *Calculation of Fuchsian groups associated to billiards in a rational triangle*, Ergodic Theory Dynam. Systems **18** (1998), no. 4, 1019–1042.
- [Wria] Alex Wright, *Cylinder deformations in orbit closures of translation surfaces*, preprint, arXiv 1302.4108 (2013), to appear in Geom. Top.
- [Wrib] ———, *The field of definition of affine invariant submanifolds of the moduli space of abelian differentials*, preprint, arXiv 1210.4806 (2012), to appear in Geom. Top.
- [Wri12] ———, *Schwarz triangle mappings and Teichmüller curves: abelian square-tiled surfaces*, J. Mod. Dyn. **6** (2012), no. 3, 405–426.
- [Wri13] ———, *Schwarz triangle mappings and Teichmüller curves: the Veech-Ward-Bouw-Möller curves*, Geom. Funct. Anal. **23** (2013), no. 2, 776–809.

- [Yoc06] Jean-Christophe Yoccoz, *Continued fraction algorithms for interval exchange maps: an introduction*, *Frontiers in number theory, physics, and geometry. I*, Springer, Berlin, 2006, pp. 401–435.
- [Yoc09] ———, *Échanges d’intervalles et surfaces de translation*, *Astérisque* (2009), no. 326, Exp. No. 996, x, 387–409 (2010), *Séminaire Bourbaki*. Vol. 2007/2008.
- [Yoc10] ———, *Interval exchange maps and translation surfaces*, *Homogeneous flows, moduli spaces and arithmetic*, *Clay Math. Proc.*, vol. 10, Amer. Math. Soc., Providence, RI, 2010, pp. 1–69.
- [ZK75] A. N. Zemljakov and A. B. Katok, *Topological transitivity of billiards in polygons*, *Mat. Zametki* **18** (1975), no. 2, 291–300.
- [Zor06] Anton Zorich, *Flat surfaces*, *Frontiers in number theory, physics, and geometry. I*, Springer, Berlin, 2006, pp. 437–583.