Involutive Translation Surfaces and Panov Planes

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

by
Charles Christopher Johnson
May 2014

Accepted by:
Dr. Martin Schmoll, Committee Chair
Dr. Jim Brown
Dr. Shuhong Gao
Dr. Hui Xue
Abstract

This dissertation is concerned with the study of Panov planes and involutive translation surfaces, motivated by questions encountered in trying to understand certain self-similar billiard trajectories in the periodic variant of the Ehrenfest wind-tree model. In particular, we outline a new approach for studying billiard trajectories in certain types of infinite billiard tables by using Panov planes. After describing how this is done in the special case of the wind-tree model, we generalize our construction to show that there are typically several Panov planes that may be associated to an involutive translation surface.

The first three chapters of the dissertation provide a brief introduction to the theory of half-translation surfaces, and are included for the convenience of the reader that may not already be familiar with this theory. The fourth chapter recalls the original example of Dmitri Panov and then generalizes this example, in particular providing criteria for the existence of a foliation of the plane with dense leaves. The fifth chapter applies Panov planes to study an infinite billiard trajectory in the Ehrenfest wind-tree model, and also explains the “self-similarity” exhibited by billiard trajectories in the eigendirection of a pseudo-Anosov map on the L-shaped surface associated to the wind-tree. The sixth chapter generalizes the relationship between Panov planes and the wind-tree model by studying involutive surfaces, particularly tori related to these surfaces by a cover-quotient relation. There is one appendix which presents two particular examples of the construction described in the sixth chapter.
Acknowledgements

I would like to sincerely thank my advisor Martin Schmoll for his help and encouragement, and for introducing me to the world of flat surfaces. I am also grateful to the members of my committee, Jim Brown, Shuhong Gao, and Hui Xue. Finally, I would like to thank my family and friends, and particularly my wife Christa, for their support.
# Table of Contents

Title Page ......................................................... i
Abstract ......................................................... ii
Acknowledgements ............................................... iii
List of Tables .................................................... vi
List of Figures ................................................... vii

I Background ....................................................... 1

1 Interval exchange transformations ............................... 2
  1.1 Definition and basic properties .............................. 2
  1.2 Rauzy-Veech induction ...................................... 6

2 Translation surfaces ............................................ 15
  2.1 Billiards .................................................... 15
  2.2 Flat surfaces .............................................. 19
  2.3 Quadratic differentials ................................... 23
  2.4 Alternative definitions of translation and half-translation surfaces . . . . . . . . . 29
  2.5 Geodesic flow .............................................. 32
  2.6 Coverings of translation surfaces .......................... 36
  2.7 Moduli space ............................................... 38

3 Affine diffeomorphisms and Veech groups ....................... 42
  3.1 Affine diffeomorphisms .................................... 42
  3.2 The action of SL(2, $\mathbb{R}$) .............................. 51

II Results ......................................................... 56

4 Panov planes ................................................... 57
  4.1 Motivating example ........................................ 57
  4.2 General Panov planes ...................................... 62
  4.3 Simple twist surfaces ..................................... 66
5 The wind-tree model ........................................... 73
  5.1 Definition and known results .......................... 73
  5.2 Connecting the wind-tree and Panov planes ........ 81
  5.3 Self-similarity in billiard trajectories ............... 88

6 Involutione surfaces ........................................... 99
  6.1 Preliminary observations .............................. 99
  6.2 Covers admitting conjugate involutions ............ 104
  6.3 Quotients of the $X_\gamma$ surfaces ................. 117

Appendices ......................................................... 123
  A Examples .................................................. 124

Bibliography ...................................................... 134
List of Tables

A.1 The covers and quotients of the Swiss cross with an invariant involution. . . 129
A.2 The covers and quotients of the regular decagon. . . . . . . . . . . . . . . 133
# List of Figures

1.1 Reducible interval exchange ........................................ 5  
1.2 Rauzy-Veech example 1 ........................................ 8  
1.3 Rauzy-Veech example 2 ........................................ 9  
1.4 Rauzy-Veech induction ........................................ 9  
1.5 Iteration of Rauzy-Veech; first case ................................ 10  
1.6 Iteration of Rauzy-Veech induction; second case ..................... 10  
1.7 Iteration of Rauzy-Veech induction ................................ 11  
1.8 Rauzy class of three letters ...................................... 13  
1.9 Hyperelliptic Rauzy class of four letters ............................... 14  
2.1 Construction of a Euclidean cone .................................. 19  
2.2 Trajectories of a quadratic differential near a pole ............... 26  
2.3 Horizontal trajectories viewed locally ............................. 27  
2.4 Trajectories of a quadratic differential near zeros of various orders 28  
2.5 Polygonal representations of translation surfaces ................... 31  
2.6 Polygonal representations of half-translation surfaces ............. 31  
2.7 Suspension of an interval exchange .................................. 36  
3.1 Dehn twist of a cylinder .......................................... 45  
3.2 Non-example of an affine diffeomorphism ............................ 51  
4.1 Panov’s folded torus ............................................ 58  
4.2 Panov’s folded plane ........................................... 60  
4.3 Geodesic in Panov’s folded plane .................................. 61  
4.4 Periodically sewn pillow cases ................................... 61  
4.5 Conventions of Lemma 4.3 ........................................ 64  
4.6 L-Shaped half-translation surfaces ................................... 67  
4.7 Orientation covers of the folded L-shaped surfaces ............... 68  
5.1 Periodic wind-tree model ........................................ 75  
5.2 Unfolded wind-tree ............................................ 77  
5.3 $\mathbb{Z}^2$-quotient of the unfolded wind-tree ..................... 78  
5.4 L-shaped surface ................................................ 78  
5.5 Escaping billiard trajectory ....................................... 79  
5.6 Pillow case appearing in Lemma 5.5 ............................... 84  
5.7 Parallelogram in a pillow case .................................... 85  
5.8 Final case in the proof of Lemma 5.5 ............................... 85
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.9</td>
<td>Recurrent $y$-values for a trajectory in the wind-tree.</td>
<td>87</td>
</tr>
<tr>
<td>5.10</td>
<td>Weierstrass points of an L-shaped surface</td>
<td>90</td>
</tr>
<tr>
<td>5.11</td>
<td>Contracting leaf used in construction of a substitution map</td>
<td>90</td>
</tr>
<tr>
<td>5.12</td>
<td>First-return IET on the L</td>
<td>91</td>
</tr>
<tr>
<td>5.13</td>
<td>Billiard trajectories emitted from nearby points</td>
<td>98</td>
</tr>
<tr>
<td>6.1</td>
<td>Slit construction of covers</td>
<td>105</td>
</tr>
<tr>
<td>6.2</td>
<td>Conjugate involution on the trivial double cover</td>
<td>106</td>
</tr>
<tr>
<td>6.3</td>
<td>Conjugate involutions on a non-trivial cover</td>
<td>112</td>
</tr>
<tr>
<td>6.4</td>
<td>Intersection of basis representatives</td>
<td>115</td>
</tr>
<tr>
<td>6.5</td>
<td>Neighborhood of a fixed point</td>
<td>116</td>
</tr>
<tr>
<td>6.6</td>
<td>Neighborhoods of fixed points</td>
<td>116</td>
</tr>
<tr>
<td>6.7</td>
<td>Bundles of quadratic differentials</td>
<td>119</td>
</tr>
<tr>
<td>A.1</td>
<td>Swiss cross with Weierstrass points</td>
<td>125</td>
</tr>
<tr>
<td>A.2</td>
<td>Canonical basis for the Swiss cross</td>
<td>126</td>
</tr>
<tr>
<td>A.3</td>
<td>Regular decagon surface</td>
<td>130</td>
</tr>
</tbody>
</table>
Part I

Background
Chapter 1

Interval exchange transformations

Interval exchange transformations are simple maps which arise as the Poincaré sections of the geodesic flow on a translation surface. Precisely, if we consider a geodesic ray emanating from a line segment on the surface, with the line segment transverse to the direction of the flow, and wait until that geodesic intersects the segment again, we obtain an interval exchange. Because of this relationship between flows on the surface and interval exchanges, dynamical properties of the flow are reflected in the dynamics of these simple maps. Studying flows can, in some special situations, be reduced to studying the interval exchanges. In this chapter we collect some of the basic facts about this important class of maps. For more information about interval exchanges, see [Via06] or [Yoc06].

1.1 Definition and basic properties

An interval exchange transformation (or IET for short) on an interval $I = [a, b)$ partitions the interval into a finite number of subintervals, and then permutes the subintervals. More precisely, an IET is a piecewise translation from a half-open interval to itself with finitely-many points of discontinuity. The IET is uniquely determined by a vector of lengths, $\lambda = (\lambda_1, ..., \lambda_n) \in \mathbb{R}_+^n$, and a permutation $\sigma \in \mathcal{S}_n$ on $n$ letters.
Given $\lambda = (\lambda_1, ..., \lambda_n) \in \mathbb{R}_{+}^n$, let $\beta_i$ denote the value

$$\beta_i = \sum_{k=1}^{i} \lambda_k$$

for $0 \leq i \leq n$, and set $X_i = [\beta_{i-1}, \beta_i)$ for $1 \leq i \leq n$. For $\sigma \in \mathfrak{S}_n$, set

$$\lambda^\sigma = (\lambda_{\sigma^{-1}(1)}, \lambda_{\sigma^{-1}(2)}, ..., \lambda_{\sigma^{-1}(n)}) .$$

Define $\beta_i^\sigma$ using $\lambda^\sigma$ as $\beta$ was defined using $\lambda$. For $x \in X_i$, set $Tx = x - \beta_{i-1} + \beta'_{\sigma(i)-1}$. We then say that $T$ is the $(\lambda, \sigma)$-IET.

Note the $\beta_i$ are the endpoints of the subintervals permuted by $T$. The vector $\lambda^\sigma$ tells us the lengths of the intervals, in the left-to-right order, after applying the map. The $\beta_i^\sigma$ are the endpoints of the images of the individual subintervals after applying the IET. The map takes a point $x \in X_i$ and translates to the left, moving the $X_i$ subinterval to the start of $I$, and then translates to the right to place the interval in the $\sigma(i)$ position.

It is clear that an IET is invertible, and that the inverse is itself an IET: just move the subintervals back into their original order. If $T$ is the $(\lambda, \sigma)$-IET, then $T^{-1}$ is the $(\lambda^\sigma, \sigma^{-1})$-IET, with $\lambda^\sigma$ defined as above.

There is a little bit of ambiguity in the literature as to how the endpoints of the subintervals are mapped. In the above we assume subintervals are of the form $[\beta_{i-1}, \beta_i)$, but other authors suppose the intervals have the form $(\beta_{i-1}, \beta_i)$ or $(\beta_{i-1}, \beta_i]$. For our purposes the inclusion or exclusion of endpoints is immaterial. Note an IET can always be rescaled, and so we assume the transformation applies to the interval $[0, 1)$ unless otherwise noted.

Given such an interval exchange transformation we get a dynamical system by iterating the map, and so naturally can ask questions about the periodicity of points, the density of orbits, ergodicity, and so on. Criteria guaranteeing minimality is given below, following [Kea75].

Let $T$ on be an IET on $[0, 1)$ which partitions $[0, 1)$ into $n \geq 2$ pieces. Let $O(x)$
denote the orbit of a point \( x \in [0,1) \) under \( T \):

\[
\mathcal{O}(x) := \bigcup_{n \in \mathbb{N}} \{T^n(x)\}.
\]

Suppose the \( \beta_i \) are the points of discontinuity in \([0,1)\). Define

\[
D^\infty := \bigcup_{i=0}^{n-1} \mathcal{O}(\beta_i) \cup \{1\}.
\]

We say that \( T \) satisfies Keane’s minimality condition if the following two conditions are met:

(i) each orbit \( \mathcal{O}(x) \) is infinite, and

(ii) if for any invariant subset \( F \subseteq [0,1) \) which is a finite union of half-open intervals whose endpoints all belong to \( D^\infty \), then either \( F = [0,1) \) or \( F = \emptyset \).

**Theorem 1.1** (p. 26 of [Kea75]). An IET \( T \) on an interval \( I \) satisfies Keane’s minimality condition if and only if each orbit \( \mathcal{O}(x) \) is dense.

Verifying that Keane’s minimality condition holds for a particular IET is generally difficult to do, but there is a sufficient condition depending only on the length and combinatorial data, \((\lambda, \sigma)\), which guarantees that the condition is met.

We say that the permutation \( \sigma \in \mathfrak{S}_n \) is irreducible if for each \( 1 \leq j \leq n-1 \),

\[
\sigma(\{1,2,\ldots,j\}) \neq \{1,2,\ldots,j\}.
\]

A permutation which is not irreducible is called reducible. If a permutation \( \sigma \) is irreducible, then any \((\lambda, \sigma)\)-IET can not be split into two IETs on disjoint subintervals. Figure 1.1 on the following page gives an example of an IET with permutation \( \sigma = (1\ 2\ 3)(4\ 5) \). Unless explicitly stated otherwise, we will always assume the permutation of an IET is irreducible.

**Lemma 1.2** (p. 27 of [Kea75]). If \( T \) is an IET with an irreducible permutation, and if
the orbits of the points of discontinuity are all infinite and distinct, then the minimality condition is satisfied.

The hypothesis of the above lemma is sometimes called (Keane’s) infinite distinct orbit condition and is abbreviated IDOC.

Keane provides a sufficient condition guaranteeing an IET satisfies the IDOC, and thus is minimal, based on how the lengths $\lambda_i$ are related to one another. We will say that an IET is irrational if the lengths of the subintervals are rationally independent.

**Lemma 1.3** (p. 27 of [Kea75]). *If $T$ is an irreducible, irrational IET, then $T$ satisfies the minimality condition.*

Given an IET $T$ on $[a, b)$, we define a $T$-connection to be a finite sequence of points in $[a, b)$, $(x_1, x_2, ..., x_n)$ where $x_{i+1} = Tx_i$ and both $x_1, x_n \in \{\beta_1, \beta_2, ..., \beta_{r-1}\}$. (Here, $T$ permutes $r$ subintervals, and the $\beta_i$ are the points of discontinuity of $T$.) We define $C(T)$ to be the union of the $T$-connections. Notice that $C(T)$ is necessarily finite: a $T$-connection is a finite sequence beginning and ending at a discontinuity, and there are only finitely many discontinuities. It is also clear that $C(T) = \emptyset$ if and only if $T$ satisfies the IDOC: if the orbits were not distinct, some piece of an orbit would form a $T$-connection; if an orbit was finite, it would be periodic, and we would have a $T$-connection starting and ending at the same discontinuity.

**Proposition 1.4** (Proposition 2.9 of [Bos88]). *Let $T$ be an IET on $[a, b)$ which does not
satisfy the IDOC, so \( C(T) \neq \emptyset \). Order the points of \( C(T) \) as \( c_1 < c_2 < \ldots < c_t \). Let \( c_0 = 0 \) and \( c_{t+1} = \beta_r \). Let \( Y_j = [c_j, c_{j+1}) \). We then have a partition \( [a, b) = \bigsqcup_j Y_j \). Let \( T_j : Y_j \to Y_j \) denote the first return map for \( Y_j \) induced by \( T \). Then \( T_j \) is either the identity or an IET satisfying the IDOC, and so is minimal.

The above proposition tells us that an IET with \( T \)-connections necessarily decomposes into subintervals where \( T \) is periodic, and subintervals where \( T \) is minimal.

### 1.2 Rauzy-Veech induction

Given an IET \( T : I \to I \), a natural question to ask is what the first-return map to a subinterval of \( I \) looks like.

**Lemma 1.5** (p. 128 of [CFS82]). Let \( T \) be an IET of \( n \) subintervals of \( I = [0,1) \). Let \( J = [a, b) \) be any subinterval of \( I \). Let \( T_J \) denote the first-return map of \( T \) on \( J \). Then \( T_J \) is itself an IET splitting \( J \) into at most \( n + 2 \) subintervals.

**Proof.** For each point of discontinuity \( \beta_i \) let \( s(\beta_i) \) be the smallest \( s \geq 0 \) such that \( T^{-s} \beta_i \in [a, b) \) – if such an \( s \) actually exists. The points for which \( s \) is defined partition \( [a, b) \) by \( T^{-s(\beta_i)} \beta_i \). Suppose there are \( \ell \) subintervals, which we will denote \( J_1, J_2, \ldots, J_\ell \), ordered from left-to-right. Notice \( 1 \leq \ell \leq n + 2 \). For each \( J_i \) let \( k_i \) denote the smallest \( k \geq 1 \) such that \( T^k J_i \cap J \neq \emptyset \). Note the existence of such a \( k \) is guaranteed by the Poincaré recurrence theorem.

For each \( 1 \leq p \leq k_i \), the transformation \( T^p \) is a continuous map on \( J_i \). We claim further that \( T^{k_i} J_i \subseteq [a, b) \). If not, then for some \( 1 \leq p \leq k_i - 1 \), \( T^p J_i \) would contain some point \( a, b, \beta_1, \ldots, \beta_n \), which we will call \( y \). (That is, either \( J_i \) gets cut into several pieces by hitting some \( \beta_m \), or is disjoint from \( J \) after iteration.) Notice we would then have \( s(y) = p \), and so \( T^{-s(y)} y = T^{-s(y)} y \in J_i \). This, however, contradicts the definition of the \( J_i \). This means the return time \( k_{J_i}(x) \) for each \( x \in J_i \) is \( k_i \). As the first-return map \( T_J \) coincides with \( T^{k_i} \) on \( J_i \), we have that the map is an IET. \( \square \)
Notice that the proof of the above theorem is non-constructive, in the sense that it does not tell us how to determine the combinatorial and length data of the first-return IET given the combinatorial and length data of the original IET. Rauzy-Veech induction provides a controlled method of constructing a subinterval $J$ of $I$ where the length and combinatorial data of the first-return IET to $J$ is easily determined.

In [Kea75], Keane first introduced IETs and conjectured that minimality implied unique ergodicity. A counterexample was constructed in [KN76], and the authors went on to ask whether or not the Lebesgue measure was the only invariant measure in the event that Lebesgue measure is ergodic. In [Kea77] a counterexample was provided by considering the map induced by an IET to a special subinterval. Keane then conjectured that almost all (in the sense of Lebesgue) interval exchanges were uniquely ergodic. This was proven, independently, by Masur and Veech. The main technical tool in Veech’s proof is a generalization of the construction of [Kea77] which was studied in [Vee78]. The reason we care about this construction is because it allows us to make statements about ergodicity of flows on translation surfaces by reducing the flow to an interval exchange where criteria for ergodicity is well understood.

The version of the induction algorithm which we will describe appears in [Via06]. To make the process simpler, we will adopt a notation for IETs which is more convenient for describing the algorithm. Suppose that $\mathcal{A}$ is some alphabet of $d$ letters, and let $\pi_0, \pi_1 : \mathcal{A} \to \{1, ..., d\}$ be bijections. These two bijections tell us the order of the labels of the subintervals for the IET. Let $\lambda \in \mathbb{R}_+^d$ and let $I_\alpha$ be the set of pairs $[0, \alpha) \times \{\alpha\}$; $\lambda^* = \sum \lambda_\alpha$; and $I = [0, \lambda^*)$. Define a map which $j_0 : I_\alpha \to I$ which places the point $(x, \alpha)$ into the interval $I$ by

$$j_0(x, \alpha) = x + \sum_{\pi_0(\beta) < \pi_0(\alpha)} \lambda_\beta.$$ 

The map $j_1 : I_\alpha \to I$ is defined similarly,

$$j_1(x, \alpha) = x + \sum_{\pi_1(\beta) < \pi_1(\alpha)} \lambda_\beta.$$
We then set $T: I \to I$ to be $T(x) = j_1 \circ j_0^{-1}$.

The change in relative position of two intervals after applying $T$ is encoded in a $d \times d$ matrix $\Omega$:

$$\Omega_{\alpha,\beta} = \begin{cases} 
1 & \text{if } \pi_0(\beta) < \pi_0(\alpha) \text{ and } \pi_1(\beta) > \pi_1(\alpha), \\
-1 & \text{if } \pi_0(\beta) > \pi_0(\alpha) \text{ and } \pi_1(\beta) < \pi_1(\alpha), \\
0 & \text{otherwise}.
\end{cases}$$

Notice that $\Omega$ is antisymmetric: $\Omega^T = -\Omega$. Calculating the translation vector of $T$, $\delta = \Omega \lambda$, allows us to express $T$ as $T(x) = x + \delta_\alpha$ for $x \in I_\alpha$.

We want to shrink the interval $I$ by a small amount, in a controlled way, that allows us to easily determine the first return map on the new subinterval. First let $\alpha_0$ denote the label $\pi_0^{-1}(d)$, and let $\alpha_1 = \pi_1^{-1}(d)$. Suppose $\lambda_{\alpha_0} \neq \lambda_{\alpha_1}$. If $\lambda_{\alpha_0} > \lambda_{\alpha_1}$, set $\varepsilon = 0$; otherwise set $\varepsilon = 1$. Then $\lambda_{\alpha_\varepsilon}$ is the length of the longer subinterval at the end of $I$, and $\lambda_{\alpha_1-\varepsilon}$ is the length of the shorter subinterval. Set $\lambda' = \lambda - \lambda_{\alpha_1-\varepsilon}$ and $I' = [0, \lambda')$. We have thus taken the interval $I$ and cut off the shorter right-most piece. See Figures 1.2 and 1.3 below for an example of each case. Notice that in each of these examples we have the same permutation of the intervals, only the length data has changed.

Consider induction in the first example. Here the interval has length 10 and the pieces $A$ through $E$ have respective lengths 2, 1, 2, 3, and 2. Because $\lambda_B < \lambda_E$, we are going to cut off the $B$ interval. This gives us a new subinterval of length 9. However there
is a slight problem in that our $E$ interval in the original configuration has shrunk, and we have a $B$ interval in the original configuration but not the second. See Figure 1.4 below.

One way to remedy this is to take the $E$ interval in the second configuration, shrink it to the appropriate size, then add on the $B$ piece. If this new exchange is to shift the shortened $E$ piece as much as it shifted the original one, we will have to place the $B$ piece at the very end, obtaining the interval exchange given in Figure 1.5. Notice that this is just the first return map of the original transformation applied to the shortened interval $[0, 9)$.

Notice that the $\pi_0$ bijection has not changed at all, and the $\pi_1$ bijection is modified only slightly. The length data has not changed any except to shorten the $E$ interval. Given this, it should not be too surprising that length and combinatorial data of the new interval exchanged can be easily (and formulaically, with our notation above) calculated. Before doing this, we consider what happens in the case where we would remove the interval from
the first configuration, as in Figure 1.3. After cutting off the $E$ interval we have situation represented in Figure 1.6.

The issue now is that the $B$ interval in the first configuration is too big, while we have an $E$ interval in the second configuration but not the first. To remedy this we will cut the $B$ interval of the first configuration down to the right size, and if we are to shift it by the same amount as in the original transformation, we will place the $E$ interval immediately after the $B$ interval in the first configuration. This gives us the transformation represented in Figure 1.7 on the following page. Notice that the inverse of this IET is the first-return map of the inverse of the original IET to the shortened interval $[0, 9)$.

Now, to determine the combinatorial and length data of this new IET, which we
Figure 1.7: The IET after one iteration of Rauzy-Veech induction.

will denote $\hat{T} : [0, \hat{\lambda}^*) \to [0, \hat{\lambda}^*)$, we set

$$\hat{\lambda}_\alpha = \begin{cases} 
\lambda_{\alpha} & : \alpha \neq \alpha_\varepsilon \\
\lambda_{\alpha} - \lambda_{\alpha_{1-\varepsilon}} & : \alpha = \alpha_\varepsilon 
\end{cases}$$

Notice that, as indicated in the examples above, one of the permutations is unchanged during the induction. In particular, if we cut the right-most interval from the first configuration, the permutation of the second configuration does not change (only the length of the last piece is shrunk). If we cut the right-most interval from the second configuration, the permutation of the first configuration does not change (again, only the last right-most subinterval is shortened); thus $\hat{\pi}_\varepsilon = \pi_\varepsilon$.

Now consider the permutation which will be changed. Cutting the right-most interval from the first configuration, we need to find a place to put $I_{\alpha_\varepsilon}$. Placing $I_{\alpha_\varepsilon}$ immediately after $I_{\alpha_{1-\varepsilon}}$, everything to the left of $I_{\alpha_{1-\varepsilon}}$ has the same position; $I_{\alpha_{1-\varepsilon}}$ also has the same position; $I_{\alpha_\varepsilon}$ comes next; and everything else is moved one space to the right.

If the right-most interval of the second configuration is cut, then again leave everything to the left $I_{\alpha_\varepsilon}$ unchanged; we leave the position of $I_{\alpha_\varepsilon}$ unchanged; put $I_{\alpha_{1-\varepsilon}}$ immediately after $I_{\alpha_\varepsilon}$; and finally slide everything else over one unit to the right. We perform the same procedure to the positions of the subintervals as we did in the first case. Thus, regardless of which configuration contains the smaller interval, we have the following:
\[ \hat{\pi}_{1-\varepsilon}(\alpha) = \begin{cases} 
\pi_{1-\varepsilon}(\alpha) & \text{if } \pi_{1-\varepsilon}(\alpha) \leq \pi_{1-\varepsilon}(\alpha\varepsilon), \\
\pi_{1-\varepsilon}(\alpha\varepsilon + 1) & \text{if } \alpha = \alpha\varepsilon, \\
\pi_{1-\varepsilon}(\alpha) + 1 & \text{if } \pi_{1-\varepsilon}(\alpha) > \pi_{1-\varepsilon}(\alpha\varepsilon) \text{ and } \alpha \neq \alpha\varepsilon. 
\end{cases} \]

This operation described above is the *Rauzy-Veech induction* on the space of interval exchanges. Denoting an IET by its length and combinatorial data, \((\lambda, \pi_0, \pi_1)\), the induction map is \(\mathcal{R}(\lambda, \pi) = (\hat{\lambda}, \hat{\pi})\), where \(\pi = (\pi_0, \pi_1)\) and \(\hat{\pi} = (\hat{\pi}_0, \hat{\pi}_1)\) in the notation above. Since this gives a new IET, the induction could be performed again. Note, however, that the induction is only defined if the right-most pieces of the configurations have different lengths: \(\lambda_{\alpha\varepsilon} \neq \lambda_{\alpha_1 - \varepsilon}\). It may be that after performing the induction, the induction cannot be performed on the new IET because the right-most intervals may have the same length. Keane’s minimality condition perfectly describes when the induction may be iterated forever without encountering this issue.

**Lemma 1.6** (Corollary 2 on p. 37 of [Yoc10]). *Let \(T : I \to I\) be an IET. We can perform Rauzy-Veech induction ad infinitum if and only if \(T\) has no connections: if \(C(T) = \emptyset\).*

Note that Rauzy-Veech induction induces two maps on the set of irreducible permutations. We call these maps \(R_0\) and \(R_1\). The \(R_0\) map takes a permutation \(\pi = (\pi_0, \pi_1)\) and replaces the \(\pi_1\) with the \(\hat{\pi}_1\) obtained after doing Rauzy-Veech induction for length data where \(\varepsilon = 1\). The \(\pi_0\) permutation is left alone. The \(R_1\) map is defined similarly: leave \(\pi_1\) alone and replace \(\pi_0\) with the \(\hat{\pi}_0\) obtained by Rauzy-Veech induction when \(\varepsilon = 0\).

For a given \(\pi = (\pi_0, \pi_1)\), we can then construct a directed graph whose vertices are irreducible permutations obtained by applying the \(R_1\) and \(R_0\) maps, and the arrows of the graph tell us how each permutation is related to the others under the \(R_1\) and \(R_0\) maps. We label the arrow 1 if \(R_1\) is applied, and 0 if \(R_0\) is applied. See Figure 1.8 for the case of irreducible permutations on three intervals.

For each \(n \geq 2\) we can construct such a graph for each irreducible permutation. The connected components of this graph are called the *Rauzy classes* of the irreducible permutations on \(n\) letters. These Rauzy classes tell us that if we perform Rauzy-Veech
induction on an interval exchange, there are only a select few permutations that can be obtained.

For each $n$, there is one particular Rauzy class we are particularly concerned with: the one containing the permutation

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ n & n-1 & n-2 & \cdots & 1 \end{pmatrix}.$$ 

The Rauzy class containing this particular permutation is called *hyperelliptic*. See Figure 1.9.

We mention the hyperelliptic Rauzy class now because later we will consider hyperelliptic surfaces and, because of some unfortunate standard terminology, there can be some confusion in how these two objects are related.
Figure 1.9: The hyperelliptic Rauzy class for four letters.
Chapter 2

Translation surfaces

In this chapter we briefly recall the main facts about translation surfaces, motivated by the problem of studying polygonal billiards. We give several equivalent definitions of translation and half-translation surfaces, each of which is helpful in certain situations. (Proofs of various theorems are usually most easily stated with respect to one definition or another, so we must necessarily understand all of the common definitions.) For more information about translation surfaces, see [MT02] or [Zor06].

2.1 Billiards

Let $P \subseteq \mathbb{R}^2$ be a polygon, by which we mean a connected, compact subset of the plane with non-empty interior and piecewise geodesic boundary. We will say the polygon is rational if the interior angle at each vertex is a rational multiple of $\pi$.

Given a polygon $P$, we consider the motion of a billiard inside the polygon. The billiard ball is an ideal point-mass that moves inside the polygon without any friction or loss of energy. The billiard moves in a straight line until it reaches the boundary of the polygon. If the billiard reaches a vertex of the polygon, its motion is undefined and so the billiard simply stops moving. If the billiard reaches a side of the polygon, it reflects off the side according to the usual rule from geometric optics: angle of incidence equals angle
of reflection. After reflection the billiard again moves in a straight line until reaching the boundary of the polygon, where it is reflected again or stops if it hits a vertex. In this way the billiard travels around the polygon indefinitely, or eventually hits a vertex. The polygon $P$ is sometimes called the billiard table, and the path inside the polygon traced out by the billiard is a billiard trajectory.

We may now ask questions about these billiard trajectories. For example, are there any periodic trajectories? Can a billiard trajectory be dense and “fill up” the polygon? If a trajectory is dense, proportionally how much time does it spend in each region of the polygon? Does it spend equal amounts of time in regions of equal size, or does the billiard prefer one region over another?

This is of course a classical problem, but despite its elementary-sounding nature surprisingly little is known about billiards in general polygons. For example, it is unknown whether or not there are necessarily any periodic billiard trajectories in the simplest of polygons: triangles. Obviously some special triangles have periodic trajectories (e.g., there is a very obvious periodic trajectory in equilateral triangles), but the question remains open for general triangles.

The rational polygons form a special class where tools from geometry and complex analysis can be used to tackle problems of polygonal billiards. These subjects are related to billiards by a process called unfolding the polygon.

### 2.1.1 Unfolding

The procedure we are about to describe first appeared in [ZK76], and has been described in various ways since then. We will elaborate on the description given in [DeM11].

Let $P$ be a rational billiard table, and suppose the sides of $P$ are labeled $S_1, S_2, ..., S_m$. Suppose the initial direction of a billiard is given by a vector $\vec{v} \in \mathbb{R}^2$. Upon hitting side $S_i$, the billiard is reflected and the new direction of the billiard is given by some vector $\vec{v}'$. Let $L_i$ denote the line in $\mathbb{R}^2$ through the origin, parallel to $S_i$; let $r_i : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection in $L_i$. Notice that $r_i$ is a linear map, and the change in direction can be calculated using
this map: \( \vec{v}' = r_i \vec{v} \).

Let \( R_P \) be the group generated by \( r_1, r_2, \ldots, r_m \). If the angles of \( P \) are all rational multiples of \( \pi \), then this group is finite. (Notice the composition of two reflections is a rotation. If both of the reflections are through lines \( y = \tan^{-1}(\theta_i) \) with \( \theta_i \in \mathbb{Q} \), then the corresponding rotation is rational. The subgroup of these rotations has index two, and since all of the rotations are rational, this group is finite.) Given any initial direction for a billiard inside the polygon, the billiard can take on only finitely-many different directions as it travels around the polygon, and these directions are parameterized by the elements of \( R_P \).

Our goal is to take one copy of \( P \) for each element of \( R_P \), and glue the copies together along their sides based on how one direction is obtained from another by reflection. For example, suppose the initial direction of the billiard is given by \( \vec{v}_0 \in S^1 \); this initial direction is associated to \( \text{id} \in R_P \). If the first side of the polygon which the billiard hits is \( S_{i_1} \), the billiard now travels in direction \( \vec{v}_1 = r_{i_1} \vec{v}_0 \), and this direction is associated to \( r_{i_1} \in R_P \). Suppose the next reflection occurs when the billiard meets side \( S_{i_2} \). The billiard then starts moving in direction

\[
\vec{v}_2 = r_{i_2} \vec{v}_1 = r_{i_2} r_{i_1} \vec{v}_0,
\]

and so we associate \( r_{i_2} r_{i_1} \) with this direction.

For each element \( r \in R_P \), let \( P_r \) be a copy of the initial billiard table. To make the gluing rule simpler, suppose \( P_r \) is the image of \( P \) under \( r \). We will glue \( P_r \) and \( P_{r'} \) together along the side \( S_i \) if \( r' = r_i r \). Points on the corresponding sides are identified by translation. Notice that gluing these sides together induces a gluing of the vertices of the copies of \( P \).

In this way we obtain a quotient

\[
X = \coprod_{r \in R_P} P_r / \sim
\]

which is equipped with a natural topology. It is clear that, away from the points obtained by gluing vertices together, the space \( X \) has the structure of a 2-manifold where chart
changes are translations. That is, if \((U, \varphi)\) and \((V, \psi)\) are overlapping coordinate charts and \(P \in U \cap V\), \(\varphi(P) = \psi(P) + c\) for some constant \(c\). In some instances the points obtained by gluing vertices together form singularities known as cone points, and the reason for this terminology will be explained shortly. The space \(\tilde{X}\) obtained by deleting the cone points of \(X\) is a smooth surface, and we can easily fill in the punctures to give \(X\) a smooth structure.

Notice that the cone points have the following interesting property: measuring the total angle around one of these points (in the natural metric inherited from the polygons), the angle may exceed \(2\pi\). By a simple geometric argument, this angle is in fact an integer multiple of \(2\pi\).

**Lemma 2.1.** The angle around a point in the unfolding of a rational billiard table obtained by gluing vertices together is an integer multiple of \(2\pi\).

**Proof.** Following [MT02], we consider a vertex of the polygon with interior angle \(\frac{m}{n}\pi\). Let \(r_1, r_2\) denote reflections in the side of the polygon incident to the vertex. The subgroup of \(\mathbb{R}P\) generated by these reflections contains \(2n\) elements, and so we glue \(2n\) copies of the polygon together at this vertex. The angle around the corresponding point on the surface is \(2n \frac{m}{n}\pi = 2m\pi\).

This means that a small neighborhood of the cone point is isometric to a Euclidean cone with the same cone angle: neighborhoods of a point with angle \(2m\pi\) are isometric to \(2m\) Euclidean half-discs glued together in the following way. Take \(m\) copies of the upper half-disc, and \(m\) copies of the lower half-disc. Glue the right-hand edge of the \(i\)-th upper half-disc to the right-hand edge of the \(i\)-th lower half-disc. The left-hand edge of the \(i\)-th lower half-disc is glued to the left-hand edge of the \((i + 1)\)-st upper half-disc. Finally, the left-hand edge of the \(n\)-th lower half-disc is glued to the left-hand edge of the first upper half-disc. See Figure 2.1.

The surface \(X\) we have obtained is thus a branched cover of the original polygon \(P\), branch points corresponding to cone points of angle greater than \(2\pi\). This surface \(X\), called the unfolding of the polygon \(P\), comes with a natural metric whose geodesics project
down to the billiard trajectories in $P$. Thus we can study billiard trajectories in polygons by looking at geodesics in the polygon’s unfolding.

(Notice that our definition of the unfolding can easily be extended to irrational polygons, but the unfolded surface will then be of infinite area and may have infinitely many cone points.)

## 2.2 Flat surfaces

Generalizing the types of surfaces obtained by unfolding a rational billiard table, we may consider flat surfaces, which are simply surfaces equipped with a metric so the surface is locally isometric to the Euclidean plane. In differential-geometric language, this means that the surface has zero Gaussian curvature. Recall the Gauss-Bonnet theorem which states that the curvature $K$ of a closed surface, $S$, satisfies the following equation:

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

This presents a problem for us in that it implies the only closed flat surface is the torus. For surfaces of higher genus, the Euler characteristic $\chi(S)$ is negative, and so a metric of constant curvature necessarily has only negative curvature. That is, most surfaces have negative total curvature: they are naturally hyperbolic and can not be given a complete flat metric. Thus if we want to study surfaces equipped with a flat metric, we will have to make some concessions.

What we will do, since we are trying to generalize the unfolded billiard tables above, is to consider surfaces which are as flat as possible: that is, away from an isolated set of
points, the surface has a flat metric. Another way to think of this is that all of the surface’s natural curvature is isolated to a few special points.

To be precise, a flat surface is a surface $X$ together with a discrete set $\Sigma \subseteq X$ such that $X \setminus \Sigma$ is given an atlas where chart changes are accomplished by Euclidean isometries. Since local coordinate changes are isometries, the Euclidean metric of $\mathbb{R}^2$ lifts to a global metric of $X \setminus \Sigma$. A natural question is what happens at the points of $\Sigma$ when we consider the metric completion of $X \setminus \Sigma$.

**Lemma 2.2.** Let $X$ be a surface, $\Sigma \subseteq X$ a discrete set, and suppose $X \setminus \Sigma$ is given a flat metric. The metric completion of $X \setminus \Sigma$ is simply $X$ where the metric at a point $p \in \Sigma$ has the form

$$ds^2 = dr^2 + \left(\frac{\alpha}{2\pi}r d\theta\right)^2$$

in polar coordinates centered at $p$, for some $\alpha > 0$.

**Proof.** Triangulate $X$ such that each point of $\Sigma$ is a vertex of the triangulation, and so that the edges of the triangles are geodesics. Consider the collection of triangles $T_1, T_2, \ldots, T_n$ which meet at $p$. In the metric of $X \setminus \Sigma$ each triangle $T_i$ becomes a triangle with a vertex removed, but we can still measure the angle between the edges which would meet at the puncture. Call this value $\alpha_i$. Thus the total angle around $p$ is $\alpha = \alpha_1 + \cdots + \alpha_n$.

This value is independent of the triangulation used for the simple reason that a refinement of the triangulation would simply break each $\alpha_i$ into pieces $\alpha_{i,1}, \ldots, \alpha_{i,m}$ with $\alpha_i = \alpha_{i,1} + \cdots + \alpha_{i,m}$. Since any pair of triangulations on a surface has a common refinement, the angle $\alpha$ is well-defined.

Finally, note that the regular Euclidean metric expressed in polar coordinates has the form

$$ds^2 = dr^2 + (r d\theta)^2.$$

This metric associates the angle $2\pi$ to a circle around the origin: the angle is measured by integrating $d\theta$ around a simple closed curve around the point.
If we wish to have a metric which is flat away from the origin, and gives the origin angle $\alpha$, we need to replace $d\theta$ above with $\frac{\alpha}{2\pi} d\theta$.

The angle around a point outside of $\Sigma$ is always $2\pi$, but the angle of a point in $\Sigma$ may be some other quantity. If $p \in \Sigma$ has angle $2\pi$, the flat metric of $X \setminus \Sigma$ may be extended to $p$, and so without loss of generality we will always assume points in $\Sigma$ have cone angle different from $2\pi$.

We can now relate the topology of the surface to the cone angles at the singularities.

**Lemma 2.3** (Combinatorial Gauss-Bonnet Theorem). Let $X$ be a closed, orientable surface of genus $g$, and $\Sigma = \{p_1, ..., p_n\}$ a finite set of points. Suppose $X \setminus \Sigma$ is given a flat metric. Let the cone angle of $p_k$ be $2\pi m_k$. Then

$$2g - 2 = \sum_{k=1}^{n} (m_k - 1).$$

**Proof.** (We simply extend the proof of [Vor96, Thm. 2.9].) Triangulate $X$ such that the points of $\Sigma$ are vertices. Let $\Sigma' = \{p_1, ..., p_n, p_{n+1}, ..., p_v\}$ denote the points of $\Sigma$ together with the non-singular vertices in the triangulation. We define $m_k = 1$ for $k \geq n + 1$. Suppose the triangulation has $t$ triangles and $e$ edges. As each edge bounds two triangles, $2e = 3t$. There are $v$ vertices, and so the Euler characteristic of the surface is $2 - 2g = v - e + t$. 

21
Some simple manipulations give us the following.

\[ 2 - 2g = v - e + t \]

\[ \implies 4 - 4g = 2v - 2e + 2t \]
\[ = 2v - 3t + 2t \]
\[ = 2v - t \]

\[ \implies 2 - 2g = v - t/2 \]

Our triangles are Euclidean, so the sum of the angles at the vertices of each triangle is \( \pi \).

Adding up the angles at every vertex in the triangulation, we thus have

\[ \sum_{i=1}^{v} 2\pi m_i = t\pi \]

\[ \implies \sum_{i=1}^{v} m_i = t/2. \]

Using this in our above formula we have

\[ 2 - 2g = v - \sum_{i=1}^{v} m_i \]
\[ = \sum_{i=1}^{v} (1 - m_i). \]

This proves the lemma.

Recall the basic fact from Riemannian geometry that, given a linear connection, we can define parallel transport of a tangent vector along a curve. If we parallel transport a tangent vector around a closed loop, the vector may “rotate” after transportation: that is, vector obtained after parallel transport around a loop may be the \( \text{SO}(2, \mathbb{R}) \)-image of the initial vector. The collection of all possible rotations forms the holonomy group of the
connection. Recall also that each Riemannian metric is equipped with the canonical Levi-Civita connection, and so once we have a metric we have a natural way to define parallel transport. See [Lee97] for details.

In the case of flat surfaces, the cone points completely determine the holonomy. In particular, parallel transport around a loop that encircles a single cone point of angle $\alpha$ results in a rotation by $\alpha$. This implies that if a surface has non-trivial holonomy, there is no globally defined notion of direction on the surface: a geodesic in some direction may loop around to come back and intersect itself moving in a new direction. If the surface did have trivial holonomy (i.e., if all of the cone points had cone angles which were even multiples of $\pi$), then we would have a globally defined notion of direction.

Our goal right now is to generalize the surfaces obtained by unfolding rational billiard tables in the hopes of having more tools at our disposal for studying billiards. Since these unfolded surfaces only have cone angles which are even multiples of $\pi$, and so trivial holonomy, a natural generalization would be to consider flat surfaces with trivial holonomy. These will turn out to be precisely the translation surfaces described below, but before discussing translation surfaces we recall one technical tool from Riemann surface theory.

### 2.3 Quadratic differentials

Recall that a meromorphic quadratic differential on a Riemann surface $X$ is an association to each chart $(U, \varphi)$ of $X$ a meromorphic function which appears in local coordinates as $f(z)$, such that if two charts $(U, \varphi)$ and $(V, \psi)$ overlap, the functions associated to these charts transform to one another in a precise way. Specifically, suppose we have $z$-coordinates from $(U, \varphi)$ and $w$-coordinates from $(V, \psi)$. Let the respective functions determined by $q$ be $f(z)$ and $g(w)$. Then for $q$ to be a quadratic differential we require

$$f(z(w)) \left( \frac{dz}{dw} \right)^2 = g(w).$$
We can write this more concisely as

\[ f(z)dz^2 = g(w)dw^2, \]

and so we may consider a quadratic differential as a tensor field which locally appears as \( q = f(z)dz^2 \) for a meromorphic function \( f \). Here, of course, \( dz^2 \) refers to the \( dz \otimes dz \) tensor. Thus a general quadratic differential is a global section of the sheaf \( (\mathcal{M}_X^{(1)}) \otimes^2 \), where \( \mathcal{M}_X^{(1)} \) is the sheaf of meromorphic 1-forms on \( X \).

Quadratic differentials are classically objects from Teichmüller theory, used in the study of extremal quasiconformal maps. (Quadratic differentials play the role of directions in which a quasiconformal map stretches as much as possible.) Our interest in quadratic differentials, though, comes from the fact that they determine a flat geometry for the surface.

Suppose that \( q \) is a quadratic differential on a Riemann surface \( X \). We will always suppose that \( q \) is non-constant and that its poles have order at most one. Suppose that the set of zeroes and poles of \( q \) is \( \Sigma \). We determine a special set of coordinates on \( X \setminus \Sigma \) by taking a point \( p \in X \setminus \Sigma \) and letting \( (U, \varphi) \) be a chart centered at \( p \). Suppose in these coordinates, which we will refer to as \( z \), \( q = f(z)dz^2 \). We then consider a chart \( (U, \varphi_f) \) by integrating:

\[ \varphi_f(x) = \int_0^{\varphi(x)} \sqrt{f(z)} \, dz. \]

Here \( \sqrt{f(z)} \) is either branch of the square root. Notice that in these new coordinates, the quadratic differential \( q \) simply appears at \( dw^2 \). The coordinates where \( q \) has this nice form are called the natural coordinates of \( q \).

Suppose that \( (U, \varphi) \), \( (V, \psi) \) are overlapping natural coordinates of \( q \). Suppose too that the center of these charts lies in the intersection (translating our coordinates as necessary, this is a harmless assumption). Say the \( (U, \varphi) \) chart is centered at \( P \) and gives us \( z \) coordinates; the \( (V, \psi) \) chart is centered at \( Q \) and gives us \( w \) coordinates. Suppose in \( z \)-coordinates, \( q = f(z) \, dz^2 \), while in \( w \)-coordinates, \( q = g(w) \, dw^2 \). Consider the simple
closed curve $\gamma$ connecting $P$, $Q$, and some chosen point $x \in U \cap V$. Notice

$$\int_{\gamma} \sqrt{f(z)} \, dz = \int_{\gamma} \sqrt{g(w)} \, dw = 0,$$

and so in $z$-coordinates,

$$\int_0^{\varphi(x)} \sqrt{f(z)} \, dz + \int_{\varphi(x)}^{\varphi(Q)} \sqrt{f(z)} \, dz + \int_0^{\varphi(Q)} \sqrt{f(z)} \, dz = 0$$

$$\Rightarrow \int_0^{\varphi(x)} \sqrt{f(x)} \, dz = \int_{\varphi(x)}^{\varphi(Q)} \sqrt{f(z)} \, dz + \int_0^{\varphi(Q)} \sqrt{f(z)} \, dz.$$

Notice $\int_{\varphi(Q)}^{\varphi(x)} \sqrt{f(z)} \, dx$ is the coordinate of $x$ in the natural coordinates coming from $w$ (or possibly its negative, depending on which square root we use), and $\int_0^{\varphi(Q)} \sqrt{f(z)} \, dz$ is a constant.

This means that in natural coordinates, chart changes have the form

$$z \mapsto \pm z + c.$$

These are Euclidean isometries, and so pulling the metric $|dz|$ back from the complex plane, we have a well-defined flat metric on $X \setminus \Sigma$. As mentioned in the previous section, the metric completion produces cone points at the points of $\Sigma$. The cone angle at these points is closely related to the orders of the points of $\Sigma$. To understand this we need to understand the trajectory structure of a quadratic differential.

At each point $p \in X \setminus \Sigma$, the quadratic differential can be applied to tangent vectors $v \in T_pX$. Suppose that in coordinates near $p$, a quadratic differential is given by $q = f(z)dz^2$. Then applying $q$ to a tangent vector $v$ we have $q(v) = f(p)v^2$, interpreting $v$ as a complex number. (Notice that since $q$ is a 2-tensor, we should really write $q(v,v)$ for this quantity. The convention, however, is that $dz^2$ here refers to squaring the complex number $v$.) We say a vector points in the horizontal direction if $q(v) > 0$, and in the vertical direction if $q(v) < 0$. (That is, we are looking at real numbers and purely imaginary numbers as our
horizontal and vertical directions. The value $f(p)$ simply rotates our directions.) Thus the quadratic differential gives us a line field on the punctured surface $X \setminus \Sigma$. (Notice this is not a vector field because we can not determine the orientation of a line. The act of squaring $v$ removes this information: $q(v) = q(-v)$. So directions on the surface are determined only up to sign; we can distinguish North from East, but we can not distinguish North from South or East from West.)

As an example, suppose $p \in \Sigma$ is a simple pole of $q$. In appropriately chosen local coordinates centered at $p$, $q$ has the form $\frac{dz^2}{z}$. The direction of a horizontal trajectory at a point $z$, in coordinates, thus corresponds to a complex number $\zeta$ such that $\frac{\zeta^2}{z} > 0$. For example, if $z \in \mathbb{R}^+$, this means $\zeta \in \mathbb{R}$. If $z = \pm i$, then $\zeta = e^{\pm \pi i / 4}$. If $z \in \mathbb{R}^-$, then $\zeta \in i\mathbb{R}$. Continuing this process for each point near $p$ allows us to construct an image of the horizontal and vertical trajectories near a simple zero, as shown in Figure 2.2.

Each of the dashed grey curves in Figure 2.2 is a horizontal line in the metric induced by the quadratic differential, while each black curve is a vertical line. In the flat metric of the surface, the collection of horizontal trajectories around $p$ appear as the horizontal lines in Figure 2.3. A small loop encircling the pole, which would appear as a half-circle if drawn in Figure 2.3. Thus the total around the pole is $\pi$, and so parallel transport around such a loop results in $180^\circ$-rotation of the tangent vector, and the holonomy group is $\mathbb{Z}/(2)$.

We can repeat this process at zeros of the quadratic differential. For a point of
order zero, we see the usual horizontal and vertical lines of the complex plane, and the cone angle at such a point is simply $2\pi$. (Note in local coordinates, such a quadratic differential appears as $dz^2$.) For a zero of order 1, we see that the horizontal trajectories partition a neighborhood of the zero into three distinct sectors, each of which contributes $\pi$ to the total angle at the singularity. In general, the cone angle of a point of order $m$ is $(2 + m)\pi$. See Figure 2.4 for examples. Combining this observation with Lemma 2.3, we have that the orders $m_1, m_2, \ldots m_n$ of zeros and poles of a quadratic differential satisfy the following equation: $4g - 4 = \sum_{i=1}^{n} m_i$.

2.3.1 Abelian differentials

Recall that an abelian differential is simply a holomorphic 1-form on a Riemann surface. Any abelian differential $\omega$ gives a quadratic differential by tensoring with itself: $q = \omega \otimes \omega$. Thus the discussion of quadratic differentials above can be adapted to the special case of abelian differentials. In particular, the existence of a global square root implies that the geometric structure determined by an abelian differential has several nice properties.

If $q = \omega \otimes \omega$, then $q$ does not have any poles, and any zeros of $q$ have even order. Supposing $P$ is a zero of $\omega$ of order $r$, then $q$ has a zero of order $2r$. The cone angle at $P$ is $(2r + 2)\pi = (r + 1)2\pi$. This means that quadratic differentials which are squares of abelian differentials have trivial holonomy. This implies the existence of a globally defined notion of direction. The horizontal and vertical trajectories of such a quadratic differential give vector fields instead of just line fields, given by $\text{Im}(\omega) = 0$ and $\text{Re}(\omega) = 0$, respectively.
Figure 2.4: Horizontal and vertical trajectories (in dashed grey and solid black, respectively) of a quadratic differentials near zeros of various orders.
Chart changes for the natural coordinates of an abelian differential are similar to those for the quadratic differentials, except there is no concern about changing signs. If, as above, $P$ and $Q$ are centers for two overlapping natural coordinate charts and $x$ is a point in the overlap, then

$$\int_{x}^{P} \omega = \int_{Q}^{x} \omega + \int_{Q}^{P} \omega.$$ 

Thus natural coordinate changes are of the form $z \mapsto z + c$.

### 2.4 Alternative definitions of translation and half-translation surfaces

Motivated by the observations in the previous section, we say that a pair $(X, q)$ with $X$ a Riemann surface and $q$ a quadratic differential with at worst simple poles is a half-translation surface. If $\omega$ is a holomorphic 1-form on $X$, then the pair $(X, \omega)$ is called a translation surface. The complex analytic definitions above are nice because they show us that complex analytic theorems and techniques can be used to study these surfaces. However there are other, equivalent, definitions which are sometimes helpful. We mention a few such definitions, each of which is useful at different times.

#### 2.4.1 Differential geometric definitions

We noted in the previous section that a half-translation surface $(X, q)$ came with an atlas of natural coordinates (away from the zeros and poles of $q$), where chart changes were of the form $z \mapsto \pm z + c$. In general, if $X$ is a surface and $\Sigma \subseteq X$ is a discrete set of points, we may define a half-translation structure on $X$ as a maximal atlas of complex charts on $X \setminus \Sigma$ where chart changes are of the form $z \mapsto \pm z + c$. These charts of course give a flat structure to $X \setminus \Sigma$, which we may then complete to define a flat geometry on $X$ with cone points. By “developing” a small closed loop around a point of $\Sigma$ into $\mathbb{R}^2$ and then measuring the angle (see [Thu97] for details), it can be shown that the cone angles we measure are integer multiples of $\pi$. 

29
Since the chart changes in a half-translation structure are holomorphic, we have a uniquely determined complex structure on the surface. The quadratic differential $dz^2$ in each chart then pulls back to a globally-defined tensor field of $X \setminus \Sigma$ without any zeros or poles. The complex structure of $X \setminus \Sigma$ easily extends to all of $X$. The quadratic differential also extends to all of $X$, but of course places zeros and poles at the punctured points, $\Sigma$. The order of these singularities is then determined by the cone angles at these points.

A translation structure is defined similarly, except that we require the chart changes to be of the form $z \mapsto z + c$. The discussion above still applies, but this restriction on coordinate changes implies the cone angles around each point of $\Sigma$ are even multiples of $\pi$, and that the 1-form $dz$ in charts pulls back to a global 1-form on the surface.

### 2.4.2 Polygonal definitions

Recall that any Riemann surface can be triangulated, and given a quadratic differential $q$ (or 1-form $\omega$), we could choose a triangulation with vertices at the poles and zeros of $q$ (zeros of $\omega$), such that the edges of the triangulation are geodesics in the associated flat metric. Cutting the surface along these edges we have a set of polygons which, when glued together, give us back the initial surface. In fact, all edge-to-edge gluings are given by translations, possibly composed with a negation.

We can actually show that all such polygonal gluings yield (half-) translation surfaces. We first describe the case of translation surfaces, as this is more straightforward.

Let $\Delta = \{D_1, D_2, D_3, \ldots\}$ be a collection of Euclidean polygons, and suppose that each edge of $D_i$ is identified by translation with a parallel edge of the same length of some $D_j$ subject to the following constraints: the inward-pointing normal vectors along the sides to be identified point in opposite directions, and each edge is glued to exactly one other, distinct edge. Assume the edges are glued together so that the surface obtained is connected. Using the polygons as chart domains, we have a translation structure for this surface, away from points obtained by gluing vertices together. Thus the description above applies, and we have a translation surface. See Figure 2.5 for examples.
A polygonal representation for a half-translation surface is similar, except that we allow edges to be glued together by translations composed with negations. As a consequence, inward-pointing normals of identified sides may both point in the same direction. Polygonal descriptions of such half-translation surfaces may thus contain “folds,” which result in points with cone angles which are odd multiples of $\pi$. See Figure 2.6.

### 2.4.3 $G$-manifolds

We very briefly note that the discussion of translation and half-translation surfaces above can be rephrased using the more general language of $G$-manifolds. We mention this because we will care about affine maps between (half-) translation surfaces, and the notion of $G$-manifolds helps to motivate why these maps are of interest. For more details about
these topics, see [Thu97] or [Rat06].

Let $G$ be a group which transitively acts by diffeomorphisms on $\mathbb{R}^n$; i.e., each element of $G$ is an $\mathbb{R}^n \to \mathbb{R}^n$ diffeomorphism. A $G$-manifold $M$ is a second-countable, Hausdorff space together with an atlas of charts $(U_i, \varphi_i)$ such that if two charts overlap, the transition function

$$\varphi_{ij} = \varphi_i \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$$

is the restriction of an element of $G$.

A $G$-diffeomorphism between $G$-manifolds is a map which locally acts like an element of $G$. If $H \leq G$, then any $H$-manifold is naturally a $G$-manifold, and so we can consider $G$-maps between $H$-manifolds.

We are interested in four different groups which act on $\mathbb{R}^2$:

- $A$ the group of affine maps,
- $F$ the group of Euclidean isometries,
- $H$ the group of half-translations ($z \mapsto \pm z + c$), and
- $T$ the group of translations.

Notice $T < H < F < A$. A translation surface is thus a $T$-manifold; a half-translation surface is an $H$-manifold; a flat surface is an $F$-manifold; and an affine surface is an $A$-manifold. In particular, there is a natural notion of $A$-maps on $T$- and $H$-manifolds: affine maps between (half-) translation surfaces. It was Veech who first noticed that the study of these affine maps yields a wealth of information about the dynamics of translation and half-translation surfaces. The next chapter mentions some important properties of affine automorphisms of (half-) translation surfaces.

### 2.5 Geodesic flow

Because (half-) translation surfaces are flat surfaces with (almost) trivial holonomy, geodesics are simply straight lines in the natural coordinates of the surface. The geodesic flow is an action of $\mathbb{R}$ on the unit tangent bundle of the surface, $UT(X)$. Given a point
\((x, \vec{v}) \in UT(X)\), the geodesic flow at time \(t\) takes \((x, \vec{v})\) and maps it to a point \(T_t(x, \vec{v}) = (x', \vec{v})\), where \(x'\) is obtained by walking in a straight line from \(x\) in direction \(\vec{v}\).

As translation surfaces come with a natural notion of vertical and horizontal, given by the vector fields \(\text{Re}(\omega) = 0\) and \(\text{Im}(\omega) = 0\), we will typically assume a flow is in the vertical direction, just to simplify language and notation. To consider flows in other directions, simply rotate the surface (i.e., replace \(\omega\) by \(e^{i\theta}\)).

2.5.1 Poincaré sections & metric cylinders

Suppose that \(I\) is a horizontal geodesic interval on a translation surface \((X, \omega)\). Denote by \(S : I \to I\) the Poincaré section over \(I\): that is, given a point \(x \in I\), define \(S(x)\) to be the first point on \(I\) obtained by flowing vertically from \(x\). Notice that there may be points where \(S(x)\) is undefined, for instance if the trajectory emitted from \(x\) hits a cone point before returning to \(I\). This is in fact the only way a geodesic can avoid coming back to \(I\).

**Theorem 2.4** (Prop. 2.4 of [Vor96]). The map \(S : I \to I\) described above is an interval exchange.

This explains why the study of interval exchange transformations and translation surfaces are intertwined. An IET is a conceptually simpler, combinatorial object and the dynamics of flow on a translation surface are reflected in the dynamics of the IET. Thus understanding the dynamics of IETs gives us a way to understand the geodesic flow on a translation surface.

Combining Theorem 2.4 with Proposition 1.4 on page 5, we have the following.

**Theorem 2.5** (Prop. 2.5 and Thm. 2.6 of [Vor96]). After deleting vertical saddle connections\(^1\), the surface partitions into invariant subsets, where the restriction of the flow is minimal, or the invariant subset is an isometrically embedded cylinder. In particular, if there are no vertical saddle connections, then the surface is either a torus or the vertical flow is minimal.

\(^1\)A saddle connection is a geodesic segment between two cone points with no cone points in its interior.
Notice that the flat structure of a (half-) translation surface allows us to consider Lebesgue measure on the surface: (half-) translations are measure-preserving, and so the Lebesgue measure on \( \mathbb{R}^2 \) pulls back to a well-defined measure on the surface. This measure is invariant under the flow: letting \( \mu \) denote the measure, if \( S \) is some measurable subset of the surface and \( T_t \) denotes the geodesic flow, then \( \mu(S) = \mu(T_t(S)) \). (This is because, locally, we simply have a straight-line flow. Thus the flow is simply a translation, and Lebesgue measure is translation-invariant.) Thus we can ask ergodic-theoretic questions about this flow.

In general, a measurable flow on a measure space is called \textit{ergodic} if its only invariant subsets are null sets, or their complement is null. Ergodicity is then a measurable notion of indecomposability: it says that a system can not be split up into smaller pieces, aside from some trivial, exceptional cases. Given any flow there is a collection of Borel invariant measures, and if the space is compact, these can be taken to be probability measures. A standard result of ergodic theory states that, provided the space is compact, the collection of flow-invariant Borel probability measures forms a convex set. The ergodic measures are precisely the extreme points of this set. See [Wal82] for details.

When there is exactly one invariant Borel probability measure, we say the measure is \textit{uniquely ergodic}. Unique ergodicity is an extremely useful notion because many classical theorems of ergodic theory, such as Birkhoff’s theorem, which hold only “almost everywhere” will hold “everywhere” if the transformation is uniquely ergodic.

In the special case of compact half-translation surfaces, the following theorem shows there is no dearth of uniquely ergodic directions.

\textbf{Theorem 2.6} ([KMS86]). \textit{The flow in almost every direction}\(^2\) on a compact half-translation surface is uniquely ergodic.

This implies, in particular, that almost every directional flow on the surface is minimal: every geodesic in that direction is dense. The existence of these minimal directions, combined with Theorem 2.4, gives us yet another way to represent translation surfaces: as

\(^2\)Here “almost every direction” considers directions as elements of \( S^1 \) with the Lebesgue measure.
suspensions of an IET. The idea here is simply that if we have a minimal direction and any orthogonal geodesic segment, a polygonal representation of the surface can be constructed so that this geodesic segment is a horizontal cross-section of the surface, and the location of discontinuities of the IET tells us where the cone points have to be (these correspond to corners of the polygonal representation). The construction we will describe appears in [Via06].

Given a minimal interval exchange \((\lambda, \pi)\) on \(I\) with \(\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)\), a suspension datum is an \(n\)-tuple of complex numbers, \((\zeta_1, \ldots, \zeta_n)\), such that \(\text{Re}(\zeta_i) = \lambda_i\),

\[
\sum_{i=1}^{k} \text{Im}(\zeta_i) > 0, \quad \text{and} \quad \sum_{i=1}^{k} \text{Im}(\zeta_{\pi^{-1}(i)}) < 0
\]

for each \(1 \leq k < n\). Given such suspension data, construct two broken geodesic lines,

\[
\Gamma_t = [0, \zeta_1, \zeta_1 + \zeta_2, \ldots, \zeta_1 + \cdots + \zeta_n] \quad \text{and} \quad \Gamma_b = [0, \zeta_{\pi^{-1}(1)}, \zeta_{\pi^{-1}(1)} + \zeta_{\pi^{-1}(2)}, \ldots, \zeta_{\pi^{-1}(1)} + \cdots + \zeta_{\pi^{-1}(n)}].
\]

Notice these two curves together form a closed curve that may or may not be simple (if the curve is not simple, it intersects itself only once). It will turn out that we can choose equivalent data which will produce a simple closed curve, but for the time being suppose the curve is simple. We then have a polygon where each side has a label \(\zeta_i\), and each \(\zeta_i\) appears exactly twice. Both sides labeled \(\zeta_i\) are parallel and of the same length, so can be identified by a translation to produce a translation surface. This surface is called a suspension of the given IET. See Figure 2.7.

Rauzy-Veech induction on the interval exchange acts on the suspension by cutting a triangular region from the right-hand side of the polygon and gluing it onto another part of the surface. Thus Rauzy-Veech induction produces another polygonal representation of the same surface, one that gets taller and skinnier as the induction is iterated. This observation, together with some dynamical properties of Rauzy-Veech induction in the space of interval
exchanges, shows that we can always choose suspension data for an IET so that the polygon constructed above does not intersect itself. See [Via06] for details.

2.6 Coverings of translation surfaces

Since translation surfaces are Riemann surfaces with some extra data (the 1-form), the most obvious maps between translation surfaces are holomorphic maps which preserve this data. That is, the 1-form of one surface pulls back to the 1-form of another. This implies that the map, in natural coordinates, appears simply as a translation. Since non-constant holomorphic maps between Riemann surfaces are always branched covers, covering maps are the natural maps to consider between translation surfaces.

Let \((X, \omega)\) and \((Y, \eta)\) be two translation surfaces, and suppose that \(\pi : X \to Y\) is a ramified covering map. Let \(\Sigma_X\) and \(\Sigma_Y\) denote the zeros of \(\omega\) and \(\eta\), respectively. Here we allow \(\omega\) and \(\eta\) to have “trivial” zeros (marked points of order 0). We say that \(\pi\) is a translation cover if:

(i) In local coordinates of \(X \setminus \Sigma_X\) and \(Y \setminus \Sigma_Y\), \(\pi\) appears as a translation; and
(ii) \( \pi(\Sigma_X) = \Sigma_Y \).

In complex-analytic terms, \( \pi^* \eta = \omega \), and we also pull back the marked points of \( Y \) to marked points of \( X \). (Notice that the addition of marked points allows us to consider maps which are ramified over removable singularities of \( \eta \).) \textit{Half-translation covers} are defined similarly. Since translation surfaces are a special type of half-translation surface, it makes sense to talk about half-translation covers from a translation surface to a half-translation surface. There is one case of particular interest.

\textbf{Theorem 2.7} (Ch. 2 of [HM79]). If \((Y, q)\) is a half-translation surface, then there exists a unique translation surface \((X, \omega)\) and a double cover \( \pi : X \to Y \), ramified at the odd-order points of \( q \), such that \( \pi^* q = \omega \otimes \omega \). That is, every half-translation surface is double-covered by a unique translation surface which doubles the angles of any cone points whose cone angle is an odd multiple of \( \pi \). We call \((X, \omega)\) the orientation cover of \((Y, q)\).

Notice geodesics on a translation surface lift to geodesics on any translation cover, and also that geodesics on a covering surface project down to geodesics. This is an important observation as it means that dynamics on one surface can be translated into dynamics on another surface by bouncing between coverings. The theory here is most easily described in terms of the Veech groups of the surfaces, which we discuss in the next chapter.

\subsection*{2.6.1 Isomorphisms and automorphisms}

Two translation surfaces, \((X_i, \omega_i)\) for \( i = 1, 2 \), are said to be \textit{isomorphic} if there is a bijection \( \varphi : X_1 \to X_2 \) which takes the set of cone points of \( X_1 \) onto the set of cone points of \( X_2 \), and appears as a translation in local coordinates. In complex-analytic terms, there is a biholomorphism \( \varphi : X_1 \to X_2 \) such that \( \omega_1 = \varphi^* \omega_2 \). An \textit{automorphism} of a translation surface \((X, \omega)\) is an isomorphism from the surface to itself, and the set of all automorphisms of \((X, \omega)\) is denoted \( \text{Aut}(X, \omega) \).

By Hurwitz’s automorphism theorem [FK92, Thm. V.I.3], the group of biholomorphisms of a genus \( g > 1 \) Riemann surface has order at most \( 84(g - 1) \), while there are
infinitely many biholomorphisms of the torus. (It is also clear in the case of the torus that infinitely many of those biholomorphisms also preserve a given 1-form. As a torus does not have any cone points, you can always shift all of the points on a flat torus a fixed amount in any given direction to obtain an automorphism.)

One general way to study a topological space is to study a group which acts on that space. Here such a group should preserve properties of the surface that we are interested in, and be large enough that there is something worthwhile to study. Since the group of automorphisms of a translation surface is typically small, we want to consider a larger group which acts on the surface. As the properties of the surface we care about concern the straight line flow of the surface, we would like to consider actions which preserve straight lines. One group to consider, then, is the group of affine diffeomorphisms of the surface, as affine maps take straight lines to straight lines. The study of this group is described in the next chapter.

2.7 Moduli space

We note in this section that half-translation surfaces naturally live in families given by the orders of the zeros and poles of the quadratic differential, and that these families have a natural, fairly nice topology. Many results about translation surfaces are obtained by studying how the surface is positioned inside of one of these families.

2.7.1 Strata of abelian and quadratic differentials

Let $X$ be a compact Riemann surface of genus $g$. Any non-zero holomorphic 1-form on $X$ corresponds to a translation structure on the surface, and any non-zero meromorphic quadratic differential with at worst simple poles corresponds to a half-translation structure. One of the most important facets of the study of translation surfaces comes from studying families of translation surfaces, because information about a particular surface can sometimes be attained by understanding how a certain collection of surfaces lives inside this
family.

Let $\mathcal{M}_g$ denote the moduli space of compact, genus $g$ Riemann surfaces. That is, $\mathcal{M}_g$ consists of equivalence classes of compact Riemann surfaces of genus $g$ where two surfaces are considered equivalent if there exists a biholomorphic map between them. The collection of all genus $g$ translation surfaces can then be identified with the vector bundle of all holomorphic 1-forms of surfaces in $\mathcal{M}_g$, denoted $\Omega\mathcal{M}_g$. Whenever we write $(X,\omega) \in \Omega\mathcal{M}_g$, we will always assume $\omega \neq 0$.

Notice that the bundle $\Omega\mathcal{M}_g$ is naturally stratified by the orders of the zeros of the 1-forms. Thus if $(X,\omega) \in \Omega\mathcal{M}_g$ and the orders of the zeros of $\omega$ are $d_1, d_2, \ldots, d_k$ (orders are repeated if several points have the same order), then we say that $(X,\omega)$ is an element of the stratum $\mathcal{H}(d_1, d_2, \ldots, d_k)$.

We can similarly consider the collection of all meromorphic quadratic differentials which live on a particular Riemann surface. This again forms a complex vector space, of dimension $3g - 3$. The collection of all quadratic differentials over genus $g$ Riemann surfaces will be denoted $\mathcal{Q}\mathcal{M}_g$. This collection is also stratified by the orders of the critical points (poles and zeros), and the collection of all quadratic differentials with fixed orders $d_1, d_2, \ldots, d_k$ is denoted $\mathcal{Q}(d_1, d_2, \ldots, d_k)$. Recall that we assume $d_i \geq -1$.

For our purposes we typically will not care about the area of a surface, and so only consider those surfaces in each stratum which are normalized to have area one. These collections of surfaces are denoted $\mathcal{H}_1(d_1, \ldots, d_k)$ and $\mathcal{Q}_1(d_1, \ldots, d_k)$.

In each stratum $\mathcal{H}(d_1, \ldots, d_k)$, we may think of a surface as being defined by a polygon with $2n$ sides identified in pairs. Parameterizing the lengths and directions of these sides by complex numbers $(\zeta_1, \ldots, \zeta_n)$, it is easy to see that small changes in the $\zeta_i$ produces a different surface within the same stratum. That is, the $\zeta_i$ may be viewed as local coordinates for the space $\mathcal{H}(d_1, \ldots, d_k)$. Alternatively, we may think of local coordinates for $(X,\omega)$ as being given by the relative periods, which assigns to $\omega$ the local coordinates of $[\omega] \in H^1(X, \{p_1, \ldots, p_k\}; \mathbb{C})$. This shows that $\mathcal{H}(d_1, \ldots, d_k)$ has dimension $2g + k - 1$. It can be shown that these coordinates give each stratum the structure of a complex orbifold.
2.7.2 Connected components of strata

Given an IET, the permutation of the IET defines some Rauzy class, \( \mathcal{R} \). By its definition, \( \mathcal{R} \) consists of all the permutations that can arise by performing Rauzy-Veech induction on the IET. As was briefly mentioned earlier, if we have a suspension over an IET, Rauzy-Veech induction cuts a triangle off of the suspension and glues it to another part of the surface. This of course gives another polygonal representation of the same surface.

Starting with a translation surface \((X, \omega) \in \mathcal{H}(d_1, \ldots, d_k)\), the first return map to a geodesic segment orthogonal to a minimal direction is an IET, and this IET has \(2g + k - 1\), \(2g + k\), or \(2g + k + 1\) intervals. (The \(k\) cone points each contribute to the number of intervals as they determine discontinuities of the IET. Additionally, the endpoints of the IET may be chosen so that the trajectory from neither endpoint hits a cone point, one endpoint hits a cone point, or both do. The “position” of these endpoints accounts for the three different possibilities for the number of exchanged intervals.)

The actual IET obtained depends of course on the particular direction chosen, the length of the interval, and where that interval is positioned on the surface. Each of the possible interval exchanges that may be obtained in this way determines a Rauzy class, and the union of these Rauzy classes is the extended Rauzy class of the surface. It was shown by Veech that the extended Rauzy class of the surface does not differ between surfaces within the same connected component of a stratum. In fact, extended Rauzy classes parameterize the connected components of the strata.

In [KZ03], it is shown that each stratum has at most three connected components. For many strata there is only one component, but certain strata have more. The connected components of these strata are distinguished by two properties: if the extended Rauzy class contains the hyperelliptic class (the hyperelliptic component), and the parity of the spin structure of the component (the even and odd components). These components will not play a role in this dissertation, except for one possibly confusing piece of terminology.

The suspension of an IET in the hyperelliptic Rauzy class has only one or two cone points, depending on whether the IET exchanges an even or odd number of intervals.
(see [Via06, Lemma 17.1]). If $X$ is a hyperelliptic Riemann surface, say $\sigma$ is the hyperelliptic involution, it is not necessarily the case that $(X, \omega)$ belongs to the hyperelliptic component of the corresponding stratum. That is, components other than the hyperelliptic component may contain hyperelliptic surfaces (the collection of hyperelliptic surfaces is called a hyperelliptic locus in the stratum).

We mention this because later we will discuss involutive translation surfaces, and in particular will care about non-hyperelliptic surfaces with an involution. At first glance, one would suspect that such surfaces necessarily belong to the hyperelliptic component of a stratum, as it would seem reasonable that any first-return IET should be in the hyperelliptic Rauzy class. The above remarks show that this is not the case, and so one needs to be careful when discussing and reasoning about involutive and hyperelliptic translation surfaces.
Chapter 3

Affine diffeomorphisms and Veech groups

As mentioned in the previous chapter, with the exception of tori, the automorphism group of a translation surface is always finite. In order to understand translation surfaces, it would be helpful if there was a larger group acting on the surface which we could study. Our main interest in these surfaces is the straight-line flow, and so we are naturally lead to the study of groups which will preserve straight lines in the sense that lines are mapped to lines, even if the directions may change.

Since the set of half-translations of the plane is a normal subgroup of the set of affine maps, every half-translation surface is naturally an affine surface. (Here, of course, we refer to the half-translation structure away from the cone points. Thus the affine structure is also defined on the complement of these cone points.) Thus we can consider affine maps between half-translation surfaces.

3.1 Affine diffeomorphisms

Let $X$ be an affine surface: that is, $X$ is equipped with an atlas where chart changes have the form of affine maps in $\mathbb{R}^2$. We will always suppose that affine maps are orientation-
preserving, unless explicitly stated otherwise. An affine map between affine manifolds is
simply a map which appears affine in the local coordinates determined by the affine struc-
ture.

Suppose that \((X, q)\) is a half-translation surface with \(\Sigma\) the set of cone points. Then
\(X \setminus \Sigma\) is naturally an affine manifold, and so we can consider the collection of affine self-
maps of \(X \setminus \Sigma\). We are particularly interested in those affine maps which preserve the
half-translation structure of \((X, q)\). These are called the affine diffeomorphisms of \((X, q)\),
and the set of such maps is denoted \(\text{Aff}^+(X, q)\). Notice that these maps naturally extend
to continuous maps on all of \(X\), and so must permute the cone points.

Requiring that an affine map preserve the half-translation structure is equivalent
to requiring that the map appears affine in natural coordinates, since it is these natural
coordinates that determine the affine structure of the surface.

**Lemma 3.1.** The derivative of an affine diffeomorphism of a half-translation surface \((X, q)\),
away from the cone points, is a constant element of \(\text{PSL}(2, \mathbb{R})\).

**Proof.** Since an affine diffeomorphism appears as an affine map in local coordinates, it is
clear the derivative is constant within each chart. In particular, if the map takes the form
\[ z \mapsto Az + b, \]
then the derivative in these coordinates is \(A\). As chart charges are given
by half-translation, we have the same \(A\) in each coordinate chart, up to sign. Thus the
derivative of an affine map, which we will denote \(D\varphi\) for \(\varphi \in \text{Aff}^+(X, q)\), is an element of
\(\text{PGL}(2, \mathbb{R})\).

Since we require that the map preserve the half-translation structure of the surface,
it must in particular preserve the volume determined by this structure, and so the derivative
has determinant one. \(\square\)

Notice that in the case of a translation surface \((X, \omega)\) – so the quadratic differential
\(q\) is the global square of an Abelian differential \(\omega\) – the derivative of an affine diffeo-
morphism is well-defined without the “up to sign” condition. Thus the derivative of an affine
diffeomorphism of a translation surface is an element of \(\text{SL}(2, \mathbb{R})\).
While automorphisms for a half-translation surface are necessarily conformal, affine
diffeomorphisms are generally only quasiconformal. (I.e., angles may be distorted under
the image of an affine map, but the distortion is bounded.) The following theorem (which
is interesting to note, though we will not need it later) tells us that in fact any two affine
diffeomorphisms act on the surface in significantly different ways.

**Theorem 3.2** (Lemma 5.7 and Corollary 5.8 of [MT02]). \( \text{Aff}^+(X,q) \) is a subgroup of the
mapping class of the surface.

As an automorphism of a translation surface is a particular type of affine diffeomor-
phism, and so there is a natural injective map \( \text{Aut}(X,\omega) \to \text{Aff}^+(X,\omega) \). Differentiation
supplies a surjection \( \text{Aff}^+(X,\omega) \to D(\text{Aff}^+(X,\omega)) \). Furthermore, automorphisms have triv-
ial derivative, and so we have a short exact sequence,

\[
0 \to \text{Aut}(X,\omega) \to \text{Aff}^+(X,\omega) \to D(\text{Aff}^+(X,\omega)) \to 0.
\]

### 3.1.1 Dehn twists and pseudo-Anosov diffeomorphisms

For our purposes we are principally concerned with two important types of affine diffeomorphisms: Dehn twists and pseudo-Anosovs. Dehn twists are simple maps on cylinders and essentially correspond to affine diffeomorphisms with parabolic derivative. Pseudo-
Anosovs, on the other hand, are more interesting maps which simultaneously push points
away from and pull points into fixed points on the surface.

In general, a Dehn twist on an oriented surface is an isotopy class of surface diffeo-
morphisms which “twists” an embedded cylinder like a screw. Letting \( \gamma \) denote a simple
closed curve on the surface, and let \( A \) be an annulus embedded which can be homotoped
down to \( \gamma \). Give the natural \( S^1 \times [0,1] \) coordinates to \( A \), and consider the map defined on
the surface by fixing those points on the surface not in the annulus, and to points \( (x,y) \in A 
\) sending

\[
(x,y) \mapsto (x + y, y).
\]
Imagine the annulus $S^1 \times [0,1]$ as a cylinder, such a map rotates points on the cylinder depending on their vertical position in the cylinder. In particular, points near the bottom of the cylinder are fixed, and points are rotated by increasing amounts as you move up the cylinder, until reaching the top of the cylinder where a full rotation is performed and so the points at the top of the cylinder are fixed.

In terms of flat geometry, a cylinder is simply a rectangle with two opposite sides identified by translation. The width to height ratio of this rectangle is called the cylinder’s modulus. Note that a Dehn twist of such a cylinder is performed by an affine map. Choosing coordinates so that the bottom, left-hand corner of a $w \times h$ cylinder is the origin, the Dehn twist is given by the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & w/h \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$  

See Figure 3.1 for an example.

The Dehn twist we have defined above is a map on a single cylinder. In some very special situations, a translation surface may be decomposed into several cylinders (this happens, for instance, when every regular geodesic in a fixed direction is periodic). Given such a cylinder decomposition, we can construct a globally defined map by twisting all of

![Figure 3.1: A Dehn twist on a cylinder. Cutting along the dashed curve and re-gluing gives back the original cylinder.](image-url)
the cylinders. Since the different cylinders are glued together at their boundary circles to form the surface, and since Dehn twists fix boundaries pointwise, it is clear that this gives us a globally defined map. However, this map may not be affine. Though the map we have defined is clearly affine on each cylinder, to be affine on the entire surface we must have constant derivative. That is, twisting each cylinder once gives us an element of $\text{Aff}^+(X, q)$ if and only if the cylinders have the same moduli: the width-to-height ratios are all the same.

More generally, we may be able to obtain an affine map by twisting some cylinders more than others in order to make the derivative of the map constant. For example, if our surface consisted of two cylinders which were $1 \times 1$ and $2 \times 1$, we can not twist each cylinder once and get a globally defined affine map. If we twist the smaller cylinder twice, and the bigger cylinder only once, then we have a map with constant derivative. In general, if we have a cylinder decomposition of a surface, we can define a global affine map on the surface if and only if the cylinders have rationally commensurate moduli.

Notice that a Dehn twist preserves the periodic direction of the cylinder. Rotating the cylinder so that this direction is horizontal, $[1, 0]^T$ is an eigenvector of a Dehn twist with eigenvalue one. This is true also of a global affine map constructed from twisting the components of a cylinder decomposition of the surface. The following theorem shows that this is the only way an affine map can have an eigenvalue of 1.

**Theorem 3.3** (Lemma 5.6 of [MT02]). Suppose $\varphi \in \text{Aff}^+(X, q)$ and $v$ is an eigenvector of $D\varphi$ with eigenvalue 1 (equivalently, $v$ points in a direction which gives a cylinder decomposition of the surface). Then there exists a $n \in \mathbb{N}$ such that $\varphi^n$ is a (possibly multiple) Dehn twist of each cylinder.

Much more interesting, at least from the point of view of dynamics on the surface, are the pseudo-Anosov maps. In general, a pseudo-Anosov map is a self-diffeomorphism $\varphi$ of a surface satisfying the following properties.

1. There exist two singular, transversal foliations on the surface, $\mathcal{F}^s$ and $\mathcal{F}^u$, called the
stable and unstable foliations, respectively, which are preserved by \( \varphi \). That is, each leaf of \( F^s \) (respectively, \( F^u \)) is sent by \( \varphi \) to another leaf of \( F^s \) (respectively, \( F^u \)). Both foliations have the same set of singularities.

2. Each foliation, \( F^s \) and \( F^u \), is equipped with a transverse measure, \( \mu^s \) and \( \mu^u \). By a transverse measure of a foliation, we simply mean a measure on the collection of transverse sections of the foliation such that for any map which preserves each leaf of the foliation, the section and its image have the same measure. (An simple example will be given below to help make this idea concrete.)

3. Finally, we require that the transverse measures \( \mu^s \) and \( \mu^u \) transform in a precise way under \( \varphi \):

\[
\varphi_*\mu^s = e^{h(\varphi)}\mu^s, \quad \text{and} \quad \varphi_*\mu^u = e^{-h(\varphi)}\mu^u
\]

where \( h(\varphi) \) is the topological entropy of \( \varphi \).

As a very simple example, consider the linear map \( \varphi \) on \( \mathbb{R}^2 \) given by

\[
\begin{pmatrix}
2 & 0 \\
0 & 1/2
\end{pmatrix}
\]

Notice that \( \varphi \) stretches by a factor of 2 in the horizontal direction, and shrinks by a factor of \( 1/2 \) in the vertical direction. Taking \( F^s \) to be the foliation of horizontal lines, and \( F^u \) is the foliation of vertical lines, we see each foliation is preserved by \( \varphi \). The transverse measures are simply \( \mu^s = |dy| \) and \( \mu^u = |dx| \). These measures simply record the total vertical change (in the case of \( |dy| \)) or total vertical change (in the case of \( |dx| \)) of a curve. For example, consider the curve \( C \) parameterized by

\[
t \mapsto (t, \sin(t)) \quad \text{for} \quad t \in [0, 2\pi].
\]
Then
\[ \mu^s(C) = \int_C |dy| = 4, \quad \text{and} \quad \mu^u(C) = \int_C |dx| = 2\pi. \]

Note how the transverse measures of the curve change under \( \varphi \) and \( \varphi^{-1} \):

\[ \begin{align*}
\mu^s(\varphi(C)) &= \int_{\varphi(C)} |dy| = 2, & \mu^u(\varphi(C)) &= \int_{\varphi(C)} |dx| = 4\pi, \\
\varphi_*\mu^s(C) &= \int_{\varphi^{-1}(C)} |dy| = 8, & \varphi_*\mu^u(C) &= \int_{\varphi^{-1}(C)} |dx| = \pi.
\end{align*} \]

The topological entropy of a linear map \( \varphi : \mathbb{R}^m \to \mathbb{R}^m \) is equal to the sum of the natural logarithms of the absolute values of the eigenvalues of \( \varphi \) whose absolute values are greater than 1 (see [War, Ch. 7].) For our linear map above this gives that the topological entropy of \( \varphi \) equals \( h(\varphi) = \log(2) \), and we see that our transverse measures do indeed satisfy the necessary conditions for \( \varphi \) to be pseudo-Anosov: for any curve \( C \),

\[ \int_{\varphi^{-1}(C)} |dy| = 2 \int_C |dy|, \quad \text{and} \quad \int_{\varphi^{-1}(C)} |dx| = \frac{1}{2} \int_C |dx|. \]

In the special case of affine diffeomorphisms on (half-) translation surfaces, whether a map is pseudo-Anosov or not is easily determined by looking at the map’s derivative.

**Lemma 3.4.** An affine diffeomorphism \( \varphi \in \text{Aff}^+(X, q) \) is pseudo-Anosov if and only if its derivative is hyperbolic. That is, its derivative is diagonalizable; it has distinct eigenvalues \( \lambda \) and \( \lambda^{-1} \); and \( |\text{tr}(D\varphi)| > 2 \).

**Proof.** Suppose \( \varphi \) is pseudo-Anosov. Since \( \varphi \) is affine, the foliations preserved by \( \varphi \) are naturally the eigenfoliations: the collection of geodesics in the directions determined by the eigenvectors of \( D\varphi \).

The simultaneous pushing and pulling of a pseudo-Anosov map is what allows for
interesting dynamics. In particular, the following lemma shows that the existence of a pseudo-Anosov affine diffeomorphism guarantees interesting phenomena for geodesic flows on a translation surface.

**Lemma 3.5.** Let \((X, \omega)\) be a translation surface and suppose \(\varphi \in \text{Aff}^+(X, \omega)\) is pseudo-Anosov. Let \(v_\pm\) denote the eigenvectors of \(D\varphi\) corresponding to eigenvalues \(\lambda^{\pm1}\) with \(\lambda > 1\). Then any geodesic on \((X, \omega)\) in direction \(v_\pm\) is dense.

**Proof.** Since \(\varphi\) is affine, it must permute the cone points of \((X, \omega)\), of which there are only finitely many. Thus there is some \(n \in \mathbb{N}\) such that \(\varphi^n\) fixes each cone point. Notice the eigenvectors of \(\varphi\) and \(\varphi^n\) point in the same direction, so to determine density of a geodesic in direction \(v_\pm\), we may suppose the cone points of \(\varphi\) are fixed.

Consider a small geodesic segment in direction \(v_+\) emitted from a (fixed) cone point. Applying \(\varphi\) leaves one endpoint of this segment (the cone point) fixed, while pushing the other endpoint further and further away. This means, in particular, that the other endpoint can not be a cone point of the surface. Thus there are no saddle connections in the direction of \(v_+\). By Theorem 2.5, this means the flow in direction \(v_+\) is minimal. Since \(\varphi^{-1}\) is also a pseudo-Anosov map with the roles of expanding and contracting exchanged, the flow in direction \(v_-\) is also minimal. \(\square\)

### 3.1.2 Construction of Affine Diffeomorphisms

Constructing an affine diffeomorphism is, sometimes, easily accomplished by looking at the polygonal representation of a surface.

**Lemma 3.6.** If a translation surface \((X, \omega)\) is represented as a polygon \(P\) with parallel sides of the same length glued together by a translation, then a map \(\varphi\) with constant derivative is an affine diffeomorphism of \((X, \omega)\) if the derivative of \(\varphi\) deforms the polygon \(P\) in such a way that the original polygon can be obtained by cutting and pasting, by translations.

**Proof.** Suppose that \(A = D\varphi\) is the derivative of \(\varphi\). Then \(A\) naturally acts on \(P\) as a subset of \(\mathbb{R}^2\). In general, it may not be possible to rearrange the pieces of \(AP\) so reconstruct \(P\)
(a simple example of this is given below), so suppose that $AP$ can be cut into pieces and rearranged to obtain $P$. Here, by “obtaining $P$” we do not simply mean that the shape of the polygon is obtained, but also the edges of the polygon are glued together in the same order as for the initial polygon.

Notice that translating pieces of the polygon around does not change the underlying translation structure as the coordinates are defined only up to translation. Thus $P$ and $AP$ determine the same surface.

In this sense affine diffeomorphisms correspond to special types of symmetries of the surface.

To help make the idea concrete, consider the following two maps of the torus. Let $(X, \omega)$ be the unit square with opposite edges glued together by translation. For convenience, imagine the lower left-hand corner of the square is at the origin. Consider the following maps:

$$\varphi_1(x + iy) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

$$\varphi_2(x + iy) = \begin{pmatrix} 1 & 1/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$}

Notice that these maps shear the square, and after shearing we can cut off the right-most “slanted” piece of the square, then glue it back onto the left-hand side. This produces two gluings of the square. In the case of $\varphi_1$ this gluing is exactly the same as our original square, and so $\varphi_1$ represents an affine diffeomorphism of the torus. The map $\varphi_2$, however, produces a different gluing pattern, and so $\varphi_2$ does not represent an affine diffeomorphism of this surface. See Figure 3.2. Note that this “half-twist” does not determine a well-defined map on the torus: the point corresponding to the four corners of the square would need to map to two distinct points on the reconstructed square.
Figure 3.2: A partial twist does not define an affine diffeomorphism on the unit square torus.

Though in this particular example a half-twist did not yield a well-defined map on the surface, there are situations when a half-twist does give an affine diffeomorphism. An example will be seen in the next chapter.

3.2 The action of $SL(2, \mathbb{R})$

The map $\varphi_2$ in the example at the end of the last section did not correspond to an affine diffeomorphism of the torus, but note that it did produce a new torus from the original torus. More precisely: the action of the matrix $D\varphi_2$ produced a new translation structure on the torus, different from the original translation structure. This is an example of a more general, and very useful, idea in the theory of translation surfaces: the group $SL(2, \mathbb{R})$ acts on the space of all translation surfaces by deforming translation structures.

The action of $SL(2, \mathbb{R})$ has already been indicated: given a polygonal representation of the surface, an element of $SL(2, \mathbb{R})$ deforms the polygon. Since pairs of parallel lines will map to pairs of parallel lines, the “deformed” polygon gives us a new translation surface. The surfaces obtained by such a deformation are clearly homeomorphic to the original surface, but the translation structure will typically be different.

By studying this $SL(2, \mathbb{R})$-action, Veech made several important contributions to the theory in his 1989 paper, [Vee89]. In honor of his achievements, many objects associated to translation surfaces are named after Veech. Particularly, the Veech group of a translation surface is the stabilizer of the surface under the action of $SL(2, \mathbb{R})$. This is group is denoted $SL(X, \omega)$, and tells us several important and interesting things about the surface.
Lemma 3.7. For any translation surface $(X,\omega)$, the Veech group of the surface is also the collection of all derivatives of affine diffeomorphisms of the surface: $\text{SL}(X,\omega) = D\text{Aff}^+(X,\omega)$.

Proof. We have already described the relation $D\text{Aff}^+(X,\omega) \subseteq \text{SL}(X,\omega)$ in the proof of Lemma 3.6. The other containment is trivial: if $A \in \text{SL}(X,\omega)$, then $z \mapsto Az$ defines an affine diffeomorphism of the surface. \square

Recall that a saddle connection is simply a geodesic segment between cone points with no cone points in the interior of the geodesic. Given a translation surface $(X,\omega)$, we let $\text{SC}(X,\omega)$ denote the developments of saddle connections in $\mathbb{R}^2$. That is, we consider line segments in $\mathbb{R}^2$ which start at the origin and point in the same direction and have the same length as saddle connections on $(X,\omega)$. It can be shown that these developments form a discrete subset of $\mathbb{R}^2$. Affine diffeomorphisms of the surface necessarily permute the set of saddle connections (since they preserve geodesics while permuting cone points), and studying this action can sometimes help us make statements about the affine diffeomorphisms of a surface.

Theorem 3.8. The Veech group is a non-uniform\(^1\) Fuchsian group.

Sketch of proof. By its definition, the Veech group is a subgroup of $\text{PSL}(2,\mathbb{R})$ and so we must show that it is discrete. The idea is to pay attention to how the Veech group acts on saddle connections. The Veech group preserves saddle connections, and so if there was a convergent sequence $(\Gamma_n)_{n \in \mathbb{N}}$ of elements in the Veech group, applying each $\Gamma_n$ to some fixed element of $\text{SC}(X,\omega)$, we would have a convergent sequence of saddle connections. Since the saddle connections are discrete, the this sequence must eventually become constant. Repeating this argument for another, non-parallel, saddle connection gives us two vectors which are preserved by all $\Gamma_n$ for all $n$ larger than some $N > 0$. These two vectors form a basis for $\mathbb{R}^2$, so the action of $\Gamma_n$ on these two vectors uniquely determines $\Gamma_n$. Since these vectors become fixed, $\Gamma_n$ must be eventually constant as well.

\(^1\)A Fuchsian group $\Gamma$ is called non-uniform if the quotient surface $\mathbb{H}^2/\Gamma$ is not compact.
We can find a sequence of elements of SL(2, ℝ) which make saddle connections of 
(X, ω) arbitrarily short. If SL(X, ω) was uniform, the corresponding sequence of SL(2, ℝ)/SL(X, ω)
would have a limit point, but then this would imply the existence of a sequence in SL(X, ω)
whose action on (X, ω) made saddle connections arbitrarily short. This is a contradiction
since SC(X, ω) is discrete and preserved under the action of SL(X, ω). See [Vor96, §3] for
a more detailed proof.

Notice that the action PSL(2, ℝ) (respectively, SL(2, ℝ)) preserves each stratum
Q(d_1, ..., d_k) (resp., H(d_1, ..., d_k)), as this action clearly does not change the number or
order of any cone points of the corresponding (half-) translation structures. Because these
maps have determinant 1, these groups in fact preserve each H_1(d_1, ..., d_k) and Q_1(d_1, ..., d_k).

There is one particular subgroup of PSL(2, ℝ) whose action is especially important
in the study of translation surfaces. The Teichmüller geodesic flow is defined as the action
of the one-parameter subgroup

\[ g_t = \begin{pmatrix} e^{t/2} & 0 \\
0 & e^{-t/2} \end{pmatrix}. \]

In terms of polygonal representations, this flow stretches a surface horizontally while shrinking it vertically. Many of the important properties of translation surfaces are shown by studying this flow, and one of the reasons for this is given by the following theorem.

**Theorem 3.9** (Masur). *The Teichmüller geodesic flow acts ergodically on each connected component of the set of area-one surfaces in each stratum.*

Properties of the Teichmüller flow form the cornerstone of many proofs in the theory, particularly the following famous theorem of Veech.

**Theorem 3.10** (The Veech dichotomy, [Vee89]). *If the Veech group SL(X, ω) of a surface is a lattice (i.e., SL(X, ω) has finite covolume in SL(2, ℝ) with respect to Haar measure – equivalently, the surface ℝ/SL(X, ω) has finite hyperbolic area), then the regular geodesics (not saddle connections) in a given direction θ ∈ S^1 satisfy the following dichotomy:*
(i) Every regular geodesic is periodic, or

(ii) Every regular geodesic is uniformly distributed; in particular, the flow in such a direction is uniquely ergodic.

Surfaces with a lattice Veech group are called Veech surfaces or lattice surfaces.

This theorem says that if a translation surface is “symmetric enough” (has a large group of affine symmetries), then the geodesic flow on the surface is dynamically optimal. This provides a generalization to the well-known theorem of Weyl that geodesics on the flat torus are uniformly distributed if they have irrational slope, and closed if their slope is rational.

**Lemma 3.11.** Any translation torus is a Veech surface.

**Proof.** It suffices to consider the torus formed by identifying opposite sides of the unit square, \((X, \omega)\), as any other translation torus is simply a \(\text{GL}(2, \mathbb{R})\)-image of this one, and so its Veech group is a \(\text{GL}(2, \mathbb{R})\)-image. It is clear that the horizontal Dehn twist and the 90°-rotation are in \(\text{SL}(X, \omega)\):

\[
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}, \begin{pmatrix}
0 & -1 \\
1 & 0
\end{pmatrix} \in \text{SL}(X, \omega).
\]

These elements generate \(\text{SL}(2, \mathbb{Z})\), so \(\text{SL}(2, \mathbb{Z}) \leq \text{SL}(X, \omega)\). As \(\text{SL}(2, \mathbb{Z})\) is a lattice (in particular, the fundamental domain of the action of \(\text{SL}(2, \mathbb{Z})\) on \(\mathbb{H}\) is a hyperbolic triangle with an ideal vertex), \(\text{SL}(X, \omega)\) must be a lattice as well.

Existence of a large group of affine symmetries has other consequences.

**Theorem 3.12.** Every Veech surface has a parabolic stabilizer.

**Proof.** This is a consequence of the more general fact (see [Kat92, Thm. 4.2.5 and Cor. 4.2.7]) that a non-uniform Fuchsian group of finite covolume always contains a parabolic element.
As remarkable as Theorem 3.10 is, it has one serious disadvantage: it is generally extremely difficult to calculate the Veech group of a surface. Hence a major avenue of research for the last two decades has been to find other ways of characterizing Veech surfaces. One possible approach is to use covers of surfaces where we are more easily able to determine if a surface is Veech or not, and then relate the Veech groups of the base and the cover.

We say that two subgroups, $G$ and $H$, of $\text{PSL}(2, \mathbb{R})$ are \textit{commensurate} if they share a common finite-index subgroup (finite index in both $G$ and $H$). We say $G$ and $H$ are \textit{commensurable} if $G$ has a finite index subgroup which is a $\text{PSL}(2, \mathbb{R})$-conjugate of a finite index subgroup of $H$. Notice that if $G$ and $H$ are commensurable, then one group is a lattice if and only if the other is as well.

**Theorem 3.13** ([GJ00]). Suppose $\pi: (X, \omega) \to (Y, \eta)$ is a translation covering. (Here, as before, we allow these surfaces to have extra marked points, but require that marked points of $Y$ pull back to marked points of $X$.) Then $\text{SL}(X, \omega)$ and $\text{SL}(Y, \eta)$ are commensurate. If $\pi$ is an affine covering instead of a translation covering, then the Veech groups are commensurable.

One particularly nice class of translation surfaces for which the above theorem instantly guarantees lattice Veech groups are the so-called arithmetic surfaces.

A translation surface $(X, \omega)$ is said to be \textit{arithmetic} if it translation covers a torus $(Y, \eta)$ with at most one ramification point. The theorem of Gutkin & Judge above then implies that arithmetic surfaces are always Veech. In fact, arithmetic surfaces also have a very nice geometric interpretation: these are precisely the surfaces which can be tiled by parallelograms. Applying an element of $\text{SL}(2, \mathbb{R})$, we can suppose that the parallelograms are squares; thus these are sometimes referred to as \textit{square-tiled surfaces} or \textit{origamis}. The Ph.D. dissertation of Gabriela Wietze-Schmithüsen, [Sch05], describes an algorithm for calculating the Veech group of such surfaces.
Part II

Results
Chapter 4

Panov planes

In this chapter we describe the first main results of the dissertation, concerning a construction the author and his adviser introduced to generalize an example of Dmitri Panov. We begin by first recalling Panov’s original example, discussing generalizations, and finally describing some interesting phenomena. In the next chapter we will show how these Panov planes may be used to study billiards in the periodic Ehrenfest wind-tree model.

4.1 Motivating example

A foliation $\mathcal{F}$ of the torus $\mathbb{T}^2$ is said to have bounded deviation if any lift of a leaf of the foliation to $\mathbb{C}$ is contained in an infinite strip. In [Wei36], André Weil showed that all oriented foliations of the torus necessarily have bounded deviation, but Dmitri Anosov constructed in [Ano89] an example of a non-orientable torus foliation which did not have bounded deviation. That is, there exist leaves of the foliation whose lift to $\mathbb{C}$ could not be contained in any infinite strip.

In a recently published paper, [Pan09], Dmitri Panov goes one step further and constructs an example of a torus foliation where every leaf lifts to a dense curve on $\mathbb{C}$. This example is constructed by using considering a particular linear foliation of a particular half-translation structure on the torus. We begin by recalling Panov’s original example.
Consider the half-translation structure on the torus indicated in Figure 4.1. Because of the identifications, a half-translation sphere is glued onto a translation torus. This half-translation sphere is constructed by folding a rectangle in half and gluing the edges together: this gives a flat sphere with four $\pi$-angle cone points. Upon gluing to the standard flat torus we obtain a torus with one $4\pi$-angle cone point and two $\pi$-angle cone points.

Note the horizontal and vertical foliations of this torus give cylinder decompositions. The horizontal foliation gives a single, long $3 \times 1$ cylinder while the vertical foliation gives two cylinders: one is $1 \times 1$ while the other is $1 \times 2$. In each case the Dehn twists of these cylinders produce a global affine map. A single twist of the horizontal cylinder is obviously a globally defined map. Twisting the $1 \times 1$ vertical cylinder once induces only a half-twist on the longer, $1 \times 2$ cylinder. (A full twist of the longer cylinder would produce a double twist of the shorter cylinder.) We refer to the twist of the horizontal cylinder as $\delta_h$ and the parabolic map which twists the shorter vertical cylinder exactly once as $\delta_v$. In local coordinates these maps appear as

$$
\delta_h = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad \delta_v = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.
$$

The induced action of these maps on the first homology group is given by

$$
\delta_{h*} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad \delta_{v*} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}
$$

Combining these parabolic maps in a particular way (Thurston's construction) pro-
duces a pseudo-Anosov affine diffeomorphism of the surface. Let \( P = \delta_v \delta_h^{-1} \delta_v \) denote this pseudo-Anosov, which appears in coordinates as

\[
P = \begin{pmatrix} -2 & -3 \\ -1 & -2 \end{pmatrix}.
\]

Notice the induced action on the first homology group is a 90°-rotation:

\[
P_* = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.
\]

Thus \( P_*^4 \) is the identity on homology, while the map \( P^4 \) acts in coordinates as

\[
P^4 = \begin{pmatrix} 97 & 168 \\ 56 & 97 \end{pmatrix}.
\]

This map has determinant 1 and has two real eigenvalues, \( \lambda_\pm = 97 \pm 56\sqrt{3} \). In the direction of the eigenvector \( v_+ \) corresponding to \( \lambda_+ \), the map acts as an expansion. The map is a contraction in the direction of eigenvector \( v_- \) corresponding to \( \lambda_- \).

As mentioned in Section 3.1.1, an affine diffeomorphism of this form (with two transverse eigendirections) admits two foliations, with the map acting as a contraction on the leaves of one foliation, and as an expansion on the leaves of the other foliation. These foliations are given by the geodesics on the surface in the direction of the eigenvectors, \( v_\pm \). Let \( \mathcal{F}^s \) denote the stable foliation, the collection of lines in the direction of the expanding eigenvector; and \( \mathcal{F}^u \) the unstable foliation, the collection of lines in the direction of the contracting eigenvector. Note that segments of leaves of the stable foliation become longer under application of \( \varphi \), but the transverse measure decreases. A comparable statement holds for segments of leaves of the unstable foliation.

**Theorem 4.1** (Theorem 1 of [Pan09]). *Any lift of a leaf of \( \mathcal{F}^s \) is dense in the universal cover of the above half-translation torus.*
We notice that the universal cover of such a half-translation torus is of course topologically just the plane, but carries some extra geometric information. Namely, the half-translation structure of the torus can be pulled back to give a half-translation structure on the plane. This half-translation structure can be visualized by taking the plane with its usual Euclidean metric, and making slits along line segments of the form \([(3m-1, n), (3m+1, n)]\), with \(m, n \in \mathbb{Z}\), and finally folding the slits together. This is indicated in Figure 4.2. In this image each horizontal line segment corresponds to a slit, and each shore of the slit is folded in half.

The endpoints of the slits are thus identified, giving a \(4\pi\)-angle cone point, and the points of the form \((3m, n)\), for \(m, n \in \mathbb{Z}\), are split into two copies each of cone angle \(\pi\). We denote these points \((3m, n)_+\) and \((3m, n)_-\).

Because of the identification, geodesics in this plane have an unusual behavior: when they enter one of the slits, they are reflected to the other side of the slit and have their orientation reversed. See Figure 4.3. It is this reflection property that allows geodesics on this plane to have radically different behavior from geodesics in the usual Euclidean plane. Since geodesics of the torus lift to geodesics of the plane, Panov’s theorem tells us there are dense geodesics on this folded plane.

The following description of the Panov plane may be more illuminating. Imagine
taking the region between the slits with center points \((3m, n)_+\) and \((3m, n+1)_-\) and pulling it away from the plane. The region between the two slits becomes a *pillow case* which is glued onto the plane. A region of this plane is shown in Figure 4.4. When a geodesic enters the region between the slits with centers \((3m, n)_+\) and \((3m, n+1)_-\), it is moving into one of these pillow cases. In a pillow case the geodesic may loop around one of the corner points of the pillow case and then move out of the pillow case in the opposite direction.

The key point in the proof of Theorem 4.1 is the 90°-rotation in homology performed by \(P_x = (\delta_v\delta_h^{-1}\delta_c)_x\). A simple geometric argument shows that the leaf through any \(\pi\)-angle cone-point on the plane will become arbitrarily close to the next \(\pi\)-angle cone point to the right, and so any such leaf will fill a *horizontal band* of cone points. The 90°-rotation, however, allows the leaf to move up to the next horizontal band. Repeating this argument shows the leaf through any \(\pi\)-angle cone point will get arbitrarily close to any other \(\pi\)-angle cone point. By density of the pseudo-Anosov on the torus, the leaf through any point on the universal cover will get arbitrarily close to some \(\pi\)-angle cone point on the cover, and hence will get close to every \(\pi\)-angle cone point. Thus the leaf must be dense on the plane.

Panov’s density result applies to a very special direction on a very special half-
translation surface, and so a natural question to ask is how this result may generalize to a wider class of directions and surfaces. This generalization will be the primary focus of the rest of this chapter.

### 4.2 General Panov planes

Let $T = (\mathbb{T}^2, q)$ be any half-translation torus, and let $C^T = (\mathbb{C}, \pi^*q)$ denote the universal cover $\pi : \mathbb{C} \to \mathbb{T}^2$ of $T$ together with the pulled-back half-translation structure. We call any such $C^T$ a Panov plane. The most obvious way to construct such a Panov plane is to start with the flat translation torus and attach to it any finite number of pillow cases, then consider the universal cover of this torus. A first generalization of Panov’s original density result is the following.

**Theorem 4.2.** Let $T = (\mathbb{T}^2, q)$ be any half-translation torus, and $C^T$ the corresponding Panov plane. If $\varphi \in \text{Aff}^+(T)$ is a pseudo-Anosov and if there exists an $n > 0$ such that $\varphi^n = 1$ in homology. Any lift of a leaf of the (un-) stable foliation of $\varphi$ to $C^T$ which passes through a $\pi$-angle cone point is dense on $C^T$.

We will obtain this theorem as a corollary to Theorem ?? below, and so forgo the proof for now.

While Theorem 4.2 is a natural generalization of Theorem 4.1, it very conspicuously does not apply in the case of a pseudo-Anosov which is an element of the Torelli group. (Recall that the Torelli group of a surface is the set of isotopy classes of surface diffeomorphisms which act trivially in the surface’s first homology group.) The lack of a homological rotation presents a difficulty in adapting Panov’s original proof, yet it is clear that Torelli pseudo-Anosovs may still give rise to foliations with dense leaves in the Panov planes. In particular, if $\varphi$ is a pseudo-Anosov satisfying the hypotheses of Theorem 4.2, then $\varphi^n$ is a Torelli pseudo-Anosov whose (un-) stable foliation must be dense on the Panov plane as well, since $\varphi$ and $\varphi^n$ carry the same transverse foliations.

To generalize this density result to Torelli pseudo-Anosovs, we note the main tech-
nical detail of the proof of Theorem 4.1 is not that we have a homological rotation, but rather that we are able to move from one cone point to any other cone point in certain lattice. Homological rotations provide a convenient method of moving between cone points, but this idea may be generalized via convex pairs.

We say that a point $P \in \mathbb{C}^T$ on a Panov plane and a direction $\theta \in \mathbb{RP}^1$ form a convex pair $(P, \theta)$ if there exists a finite set of points $P_1, P_2, ..., P_n$ such that

- each $P_i$ is a deck translate of $P$,
- each $P_i$ is in the closure of the leaf in direction $\theta$ through $P$, and
- $P$ is contained in the interior of the convex hull of $\{P_1, ..., P_n\}$.

Before going any further, we recall some simple facts about convex polygons.

**Lemma 4.3.** If $D \subseteq \mathbb{R}^2$ is a convex polygon whose vertices are elements of $\mathbb{Z}^2$, and if $0$ is in the interior of $D$, then $0$ may be written as a positive integer linear combination of some subset of the vertices.

**Proof.** We will show there exist two vectors of the same length, pointing in opposite directions, which may be written as positive rational multiples of the vertices. Adding these vectors together then shows that $0$ may be written as a positive rational multiple of the vertices. Multiplying by the denominators of these rational multiples will then give the result.

Suppose the vertices of $D$ are labeled $P_1, P_2, ..., P_n$ and $P_1 = (x_i, y_i)$. Applying an element of $\text{SL}(2, \mathbb{Z})$ and relabeling the vertices as necessary, suppose that $P_1$ is contained in the first quadrant, $P_2$ in the second quadrant, and $P_3$ is contained in the fourth quadrant.

Note that as $D$ is a convex polygon with $0$ in its interior, there exists a vertex $P_4$ contained in the lower half-plane such that $|x_4| < |x_2|$. We allow that $P_3$ may equal $P_4$; this happens in particular if $D$ is a triangle.

Note that for some positive rational value $\mu$, the vector $P_1 + \mu P_3$ is horizontal. (That $\mu$ is rational follows from the fact that each $(x_i, y_i)$ is an integer.) Repeating the
same process by adding a positive rational multiple $\eta$ to $P_3$ produces a vector $P_2 + \eta P_3$ which is horizontal but points to the left. See Figure 4.5.

We may now choose a positive rational value $\zeta$ such that $\zeta (P_1 + \eta P_3) + P_1 + \mu P_3 = 0$. Thus 0 may be written as a positive rational multiple of vertices of $D$. Multiplying by the least common multiple of the denominators of $\eta$, $\zeta$, and $\mu$, we obtain a positive integer linear combination of the vertices which equals zero. \hfill \Box

Lemma 4.3 on the previous page can now be used to generate a lattice from positive integer multiples the vertices of a convex polygon.

**Lemma 4.4.** Let $D \subseteq \mathbb{R}^2$ be a convex polygon with vertices $P_1, P_2, \ldots, P_n \in \mathbb{Z}^2$. The collection of all non-negative integer multiples of the $P_i$ forms a sublattice of $\mathbb{Z}^2$.

**Proof.** Let

$$L = \left\{ \sum_{i=1}^{n} \mu_i P_i \mid \mu_i \in \mathbb{N}_0, \text{ not all } \mu_i = 0 \right\}.$$ 

It is clear that $L$ is closed under addition, and so we simply need to show that $L$ contains inverses. Let $Q_1$ be any element of $L$. Choose elements $Q_2, Q_3, Q_4 \in L$ so that the $Q_i$ form
the vertices of a convex polygon containing 0 in its interior. By Lemma 4.3 on page 63 there exist positive integers $\alpha_i$ such that

$$\alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4 = 0.$$ 

Thus the inverse of $Q_1$ is simply

$$(\alpha_1 - 1)Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4.$$ 

As each $\alpha_i \in \mathbb{N}$, this is an element of $L$. \hfill \Box

We are now in a position to prove the following generalization of Panov’s density result.

**Theorem 4.5.** Let $T = (\mathbb{T}^2, q)$ be any half-translation torus, and $\mathbb{C}^T$ the corresponding Panov plane. Suppose that $\varphi$ is an affine pseudo-Anosov such that for some non-negative $n$, $\varphi^n_\ast = 1$ on $H_1(T)$. Suppose too that $\varphi$ lifts to a map $\Phi$ on $\mathbb{C}^T$ which fixes a $\pi$-angle cone point $P$, and the expanding direction of $\varphi$ forms a convex pair with $P$.

**Proof.** Let $D$ denote a convex polygon with $P$ in its interior whose vertices are deck translates of $P$. By Lemma 4.4 on the previous page, $D$ generates a subgroup of the deck transformation group of $\mathbb{C}^T$; call this group $L$. Let $X = \mathbb{C}^T/L$, and note this is a finite cover of $T$. Every point of $L$ is identified in $X$ to a point we will call $Q$. Note that $\varphi^n$ lifts to a pseudo-Anosov $\tilde{\varphi}$ on $X$.

Let $\ell$ denote the leaf of the foliation on $\mathbb{C}^T$ in the direction which gives the convex pair with $P$ in the interior of the polygon. As $\ell$ projects to the expanding leaf of $\tilde{\varphi}$ through $Q$, every point of $L$ is in the closure of the leaf $\ell$. Furthermore, as this corresponds to the expanding leaf of a pseudo-Anosov on $X$, the expanding leaf through any point $x \in \mathbb{C}^T$ must get arbitrarily close to some subset of $L$. Hence the leaf $\ell$ gets arbitrarily close to every point of $\mathbb{C}^T$. \hfill \Box
Though Theorem 4.5 greatly extends the original result of Panov, it can only be applied if you already know you have a pseudo-Anosov and a convex pair. We now give one simple criterion which guarantees the existence of a convex pair.

**Lemma 4.6.** If $T$ is a half-translation torus and $\varphi$ is a pseudo-Anosov such that $\varphi^n = 1$ for some $n \geq 1$, then there exists a convex pair on the associated Panov plane.

**Proof.** Note that $\varphi_*$ is an elliptic element of $\text{SL}(2, \mathbb{Z})$. Thus $\varphi_*$ is conjugate to a rotation by $30^\circ$, $60^\circ$, or $90^\circ$. Let $P$ be a $\pi$-angle cone point on $\mathbb{C}^T$ fixed by a lift $\Phi$ of $\varphi$ to $\mathbb{C}^T$. Now consider a deck translate of $P$ obtained in the following way. On the torus $T$, consider the curve obtained by moving in the direction expanding direction of $\varphi$ away from $P$ until some intersection with the the leaf in the contracting direction of $\varphi$ through $P$. Following this contracting leaf back to $P$ produces a loop on $T$, and for an appropriately chosen point of intersection between the expanding and contracting leaves, this loop is homotopically non-trivial.

On the Panov plane $\mathbb{C}^T$, the loop lifts to a curve which connects $P$ to one of its deck translates, say $P_1$. Furthermore, the expanding leaf through $P$ becomes arbitrarily close to $P_1$ since this curve intersects the contracting leaf through $P_1$. Applying $\varphi$ a finite number of times to $P_1$ produces a collection of points $P_1, P_2, ..., P_n$ which serve as the vertices of a convex polygon with $P$ in its interior. (This is because $\varphi_*$ is a rotation in homology. The points $P_2, ..., P_n$ are obtained by rotation $P_1$ around $P$.) Hence we have a convex pair. 

4.3 **Simple twist surfaces**

We end this chapter by exploring one special family of examples in depth. The folded $L$ surfaces we will discuss shortly will play a special role in the next chapter when we describe a relationship between the Ehrenfest wind-tree model and Panov planes.

Consider an L-shaped polygon which is obtained by taking a $1 \times 1$ square and attaching to its right-hand side a $\alpha \times 1$ rectangle, and to its top a $1 \times \beta$ rectangle. Let $L^1_{\alpha, \beta}$ denote the translation surface obtained by identifying opposite sides of this polygon. (We
will consider L-shaped surfaces obtained in different ways, and for the sake of sanity use the superscript 1 in $L^1_{\alpha,\beta}$ to denote an L described as above.) We will call the initial $1 \times 1$ square the core of the L. The attached $\alpha \times 1$ and $1 \times \beta$ rectangles are the horizontal and vertical legs, respectively.

In order to combine the results of Panov with the well-understood L-shaped surfaces, we look at the three most “obvious” half-translation L’s, presented in Figure 4.6. Notice that we take an L-shaped polygon, fold at least one of the legs of the L to produce identifications for one pair of sides, and then identify each other side of the L with its opposite. We shall denote these surfaces by $L^1_{\alpha,b}$, $L^1_{a,\beta}$, and $L^1_{\alpha,\beta}$. The hat indicates which of the legs is folded. Each of these surfaces is clearly a half-translation torus since a the surface is obtained by attaching half-translation spheres to a flat, rectangular torus. Counting the angles of the cone points shows the surfaces $L^1_{\alpha,\beta}$ and $L^1_{a,\beta}$ are members of $\mathcal{Q}(-1,-1,1,1)$, while $L^1_{\alpha,\beta}$ is a member of $\mathcal{Q}(-1,-1,-1,4)$. We will denote the Panov planes associated to these tori as $C^1_{\alpha,\beta}$, $C^1_{a,\beta}$ and $C^1_{\alpha,\beta}$, respectively.

Canonically associated to any half-translation surface $X$ is a translation surface $\text{Ori}(X)$, called the orientation cover of $X$, which is a double cover of $X$ branched at the cone points whose cone angles are odd multiples of $\pi$. In the case of the folded L’s above, the orientation cover is obtained by taking two copies of the original L-shaped polygon and identifying any opposite edges that are identified in the folded L. If an edge is folded, however, we identify each half of the fold with the opposite side on the other copy of the L.
Intuitively, the line foliations of a half-translation surface has non-orientable leaves because we can traverse a leaf in either direction. The orientation cover is obtained by taking two copies of the surface and segregating trajectories which travel in different directions into different copies. See Figure 4.7.

In the special case of our folded L’s, the orientation cover is also a double cover of the translation L of the same dimensions. This is easy to see by simply translating one copy of the L onto the other copy. This is an extremely important observation for what will come later, because this allows us to take results about the translation L’s and convert them into results about folded L’s via this covering.

**Proposition 4.7.** If one of $L_{\alpha,\beta}^1$, $L_{\hat{\alpha},\hat{\beta}}^1$, $L_{\alpha,\hat{\beta}}^1$, or $L_{\hat{\alpha},\hat{\beta}}^1$ is Veech, then all of the surfaces are Veech. In particular, if one of the surfaces has horizontal and vertical parabolic stabilizers (i.e., globally defined horizontal and vertical Dehn twists), then each surface has horizontal and vertical parabolic stabilizers.

**Proof.** Recall that if one translation surfaces covers another, then the surfaces have commensurate Veech groups. Each of the folded tori has an orientation cover which also double
covers the usual translation $L$. If the $L$ is Veech (which, by the Calta-McMullen classification, [Cal04] and [McM03], happens if and only if the surface has horizontal and vertical parabolic stabilizers), then the orientation covers are also Veech, and so the folded $L$’s are Veech.

Note that if any $L$ has parabolic horizontal and vertical stabilizers, the other $L$’s must have such stabilizers as well. The reason for this is simply that a horizontal/vertical twist of a cylinder with folds, gives rise a double twist on the corresponding cylinder without folds.

Given any folded $L$, the vertical and horizontal directions each give a two-cylinder decomposition of the surface, and to each cylinder we may associate an affine Dehn twist. To mimic the construction of Panov, we want to consider surfaces which contain two globally defined maps (one for the horizontal direction, and one for the vertical direction), each of which twists the longer cylinder in the corresponding decomposition exactly once. If a folded $L$ admits two such maps, we will call the surface a simple twist surface. First we classify the parameters $\alpha, \beta$ which yield simple twist surfaces $L^{1}_{\alpha, \beta}$.

Notice that the map twisting the long horizontal cylinder once is given by

$$\begin{pmatrix} 1 & 1 + \alpha \\ 0 & 1 \end{pmatrix},$$

while the map twisting the shorter horizontal cylinder once is

$$\begin{pmatrix} 1 & 4/\beta \\ 0 & 1 \end{pmatrix}.$$

(Because of the folds, the $1 \times \beta$ rectangle becomes a $2 \times \beta/2$ cylinder. The modulus of this cylinder is thus $2/(\beta/2) = 4/\beta$.)

In order to have a globally defined map we only need to consider half-twists of the
shorter cylinder, which has the form

\[
\begin{pmatrix}
1 & 2/\beta \\
0 & 1
\end{pmatrix}.
\]

Thus to have a globally defined map twisting the long horizontal cylinder once we require that for some \(k_h \in \mathbb{N}\),

\[
\begin{pmatrix}
1 & 2/\beta \\
0 & 1
\end{pmatrix}^{k_h} = \begin{pmatrix}
1 & 1 + \alpha \\
0 & 1
\end{pmatrix},
\]

which of course means

\[
\frac{2k_h}{\beta} = 1 + \alpha. \tag{4.1}
\]

Similarly, the vertical foliation gives us a long \(1 \times (1 + \beta)\) cylinder and a shorter \(\alpha/2 \times 2\) cylinder. To have a globally defined affine map twisting the \(1 \times (1 + \beta)\) cylinder exactly once, there must be a \(k_v \in \mathbb{N}\) such that

\[
\frac{2k_v}{\alpha} = 1 + \beta. \tag{4.2}
\]

Solving the Equation 4.1 for \(\beta\) and substituting into Equation 4.2 gives us a quadratic in terms of \(\alpha\) with solutions

\[
\alpha = \frac{-(1 + 2k_h - 2k_v) \pm \sqrt{(1 + 2k_h - 2k_v)^2 + 4k_v}}{2}.
\]

Let \(D\) denote the discriminant of the quadratic and notice that by simple algebra, \(D \equiv 1 \mod 8\). Combining this with the observation that \(\alpha - \beta = 2(k_h - k_v)\), we have that \(\alpha\) and \(\beta\) are of the form

\[
\alpha = \frac{\pm \sqrt{D} - 1}{2} + k \tag{4.3}
\]
\[ \beta = \pm \sqrt{\frac{D-1}{2}} - k \]  

(4.4)

such that both of these quantities are positive. Notice \( \alpha, \beta \in \mathbb{Q}[\sqrt{D}] \). In general, a pair \((\alpha, \beta) \in \mathbb{Q}_{\geq 0}[\sqrt{D}]^2 \) of the same form as above is called a simple twist pair.

**Proposition 4.8.** Each parameter, \((\alpha, \beta)\), in a simple twist pair can be chosen to be arbitrarily close to a fixed value.

**Proof.** If \( n \in \mathbb{N} \) and \( n^2 \equiv 1 \mod 8 \), then \( n \) must be odd. (If \( n \) were even, then \( n^2 \) would be even, but number congruent to 1 mod 8 is necessarily odd.) The converse is also true: if \( n \in \mathbb{N} \) is odd, then \( n^2 \equiv 1 \mod 8 \). Let \( n = 2p + 1 \), so \( n^2 = 4k^2 + 4k + 1 = 4k(k+1) + 1 \). If \( m \in \mathbb{N} \) were square free with \( m \equiv 1 \mod 8 \), then for any odd \( n \), \( n^2m = 8\ell + 1 \) for some \( \ell \).

Writing \( D = n^2m = (2p + 1)^2m \) we have

\[
\alpha = \frac{\sqrt{(2p+1)^2m-1}}{2} + k
\]
\[
= \frac{(2p+1)\sqrt{m}-1}{2} + k
\]
\[
= n\frac{\sqrt{m}}{2} + \frac{\sqrt{m} - 1}{2} + k
\]

Consider the function \( \gamma(n) = n\sqrt{m} + \frac{\sqrt{m}-1}{2} \) modulo 1. Since \( \sqrt{m} \) is irrational, the image \( \gamma(N) \) is dense in \([0, 1)\). Thus simple twist parameters have the form \((\alpha, \beta) = (\gamma(n)+k, \gamma(n)-k)\).

We can now describe an explicit class of surfaces satisfying the hypothesis of Proposition 4.2 using the observations about simple twist surfaces above.

**Proposition 4.9.** Any simple twist surface has an affine pseudo-Anosov map whose induced action on homology is elliptic of order 4, and so the foliation of its universal cover by geodesics in an eigendirection of this map has dense leaves.
Proof. We simply construct an affine pseudo-Anosov and apply Proposition 4.1. We construct this map following the original example of Panov. Let $P_h$ denote the map twisting the large, horizontal cylinder exactly once; let $P_v$ denote the map twisting the large, vertical cylinder exactly once. Consider the map

$$P_v P_h^{-1} P_v = \begin{pmatrix} 1 & 0 \\ 1 + \beta & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 - \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 + \beta & 1 \end{pmatrix}$$

$$= - \begin{pmatrix} \alpha + \beta + \alpha \beta & 1 + \alpha \\ \alpha \beta^2 + 2\alpha \beta + \beta^2 + \alpha - 1 & \alpha + \beta + \alpha \beta \end{pmatrix}.$$ 

The eigenvalues of this map are

$$\lambda_{\pm} = -(\alpha + \beta + \alpha \beta) \pm \sqrt{\alpha^2 \beta^2 + 2\alpha^2 \beta + 2\alpha \beta^2 + \alpha^2 + 2\alpha \beta - 1}.$$ 

Recalling that $\alpha$ and $\beta$, for a simple twist surface, satisfy (4.3) and (4.4), we have that the discriminant of the above equation is positive. Thus we have two distinct, real eigenvalues, so the map is pseudo-Anosov.

To check that the action of this map is elliptic in homology we simply note, as in the original example of Panov, that

$$P_h = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{ and } \quad P_v = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

and so the map $(P_v P_h^{-1} P_v)_* \text{ acts on homology by a } 90^\circ \text{-rotation. Applying Theorem 4.5 gives the result.}$
Chapter 5

The wind-tree model

In this chapter we describe the periodic Ehrenfest wind-tree model and discuss its relation to Panov planes. This relation motivates the construction described in the next chapter.

5.1 Definition and known results

5.1.1 Historical motivation and summary of recent results

The Ehrenfest wind-tree model was introduced in [EE90] by Paul and Tatyana Ehrenfest to study a statistical interpretation of the second law of thermodynamics. This wind-tree model is a special case of a Lorentz gas, as introduced by Hendrik Lorentz in [Lor05], which is a simple model for the motion of free electrons in a metal. In Lorentz’s original paper, he makes the simplifying assumption that free electrons in a metal interact only with fixed, immovable atoms by performing elastic reflections off the atoms. In the Ehrenfest’s wind-tree model, we assume the immovable atoms are randomly arranged diamonds in the plane and the electrons are point-masses moving with unit speed.

Though the wind-tree model has provided statistical physicists with intuition since its inception, there were very few rigorous results until [HW80]. In this paper Hardy & Weber study recurrence properties of a periodic variant of the wind-tree model where
identical rectangular obstacles are placed periodically at the integer lattice points in the plane. In particular, [HW80] shows the billiard flow in the periodic wind-tree is recurrent for a special set of directions and parameters.

In recent years the wind-tree model has gained attention as an example of a billiard whose unfolding is an infinite translation surface where rigorous results about the dynamics of the flow are obtained. In [HLT11], Hubert, Lelièvre, and Troubetzkoy show that for a certain special set of parameters, the billiard flow is recurrent in almost every direction. This was later extended by Avila and Hubert, [AH12], to remove the restriction on the parameters. That is, recurrence is the generic case for billiards in the wind-tree. Delecroix, Hubert, and Lelièvre then showed in [DHL] that the diffusion rate of the billiard is $2/3$. Delecroix, [Del13], then studied the divergent trajectories and has shown that for wind-tree’s whose parameters satisfy certain algebraic conditions (coming from the Calta-McMullen classification of lattice surfaces in genus two, [Cal04] and [McM03]), the set of divergent directions has Hausdorff dimension at least $1/2$. Recently, Fraczek and Ulcigrai, in [FU], have used the methods of symbolic dynamics to study density of orbits in $\mathbb{Z}^2$-covers of translation surfaces, and have shown that ergodicity in these covers is very rare. In particular, [FU] shows that for almost every choice of direction and almost every choice of parameters, the billiard flow in the wind-tree is not ergodic. This is in stark contrast to the case of billiards in finite polygons where [KMS86] shows that almost every direction is uniquely ergodic.

5.1.2 The periodic wind-tree

By the wind-tree model we will always mean the periodic, infinite billiard table described below. At each point of the integer lattice $\mathbb{Z} + i\mathbb{Z}$ in $\mathbb{C}$ we place an $a \times b$ rectangular obstacle where $(a, b) \in (0, 1)$. The billiard table obtained by removing the interiors of these rectangles is denoted $T_{a,b}$:

$$
T_{a,b} := \mathbb{C} \setminus \left( \bigcup_{m, n \in \mathbb{Z}} (m - a/2, m + a/2) + i (n - b/2, n + b/2) \right).
$$
We then consider a billiard which moves in $\mathcal{T}_{a,b}$ according to the usual rules of polygonal billiards: an ideal point-mass traveling in a straight line at unit speed until reaching the boundary of the table (i.e., an edge of one of the removed rectangles), at which point it performs an elastic reflection. See Figure 5.1 for an example of such a table, together with a small piece of a billiard trajectory.

In [HLT11], Hubert, Lelièvre, and Troubetzkoy study recurrence properties of the wind-tree billiards and show that for a dense set of parameters $(a, b)$, there exists a dense set of directions such that every regular billiard trajectory whose initial direction is in this dense set is periodic; and for almost every direction, the billiard flow is recurrent. (All measurable dynamics here are stated with respect to the Lebesgue measure in the phase space of the billiard.)

These results are obtained by studying the geodesic flow on the translation surface obtained by unfolding. In particular, this unfolded surface is a branched cover of the well-understood L-shaped surfaces studied by Calta and McMullen, and so results about these surfaces play an important role. We mention here the basic objects of study and recall the important, relevant results.
5.1.3 The wind-tree’s unfolding

Notice that once a billiard in $T_{a,b}$ starts moving, it can travel only in one of four directions: the group generated by linear reflections parallel to the sides of the obstacles only has four elements. Specifically, if $\tau_h$ denotes reflection along the horizontal line $y = 0$, and $\tau_v$ is reflection in the vertical line $x = 0$, then the group of reflections is

$$R = \langle \tau_h, \tau_v | \tau_h^2 = \tau_v^2 = \tau_h \tau_v \tau_h^{-1} \tau_v^{-1} = 1 \rangle \cong \mathbb{Z}/(2) \oplus \mathbb{Z}/(2).$$

This means that the translation surface obtained by unfolding the table is made up of four copies of $T_{a,b}$.

Let us refer to these copies as NE, NW, SE, and SW. Each edge in a copy of the billiard table can thus be referred to as the 4-tuple, $(m, n, S, D)$ where $m, n \in \mathbb{Z}$ denote the integer lattice point at the center of the obstacle; $S \in \{N, S, E, W\}$ refers to the side of the obstacle; and finally $D \in \{NE, NW, SE, SW\}$ refers to the copy of the table which the edge belongs to.

The gluings used to obtain the unfolded billiard table, which we will refer to as $U_{a,b}$ are thus

1. $(m, n, N, NE) \sim (m, -n, S, SE)$,
2. $(m, n, S, NE) \sim (m, -n, N, SE)$,
3. $(m, n, E, NE) \sim (-m, n, W, NW)$,
4. $(m, n, W, NE) \sim (-m, n, E, NW)$,
5. $(m, n, S, NW) \sim (m, -n, S, SW)$,
6. $(m, n, N, NW) \sim (m, -n, S, SW)$,
7. $(m, n, S, NW) \sim (m, -n, N, SW)$,
8. $(m, n, E, SW) \sim (-m, n, W, SE)$, and
9. \((m, n, W, SW) \sim (-m, n, E, SE)\).

This surface is indicated in Figure 5.2. In this figure, only some of the edge identifications are indicated in order to keep the diagram from being too complicated.

![Diagram of the wind-tree billiard](image)

Figure 5.2: The unfolding of the wind-tree billiard, \(U_{a,b}\)

Notice that the wind-tree billiard table is naturally tiled by L-shaped polygons. The four copies of the table used in the construction of the unfolded surface \(U_{a,b}\) are each tiled by L-shaped polygons, and the gluings used in constructing the surface give us a surface which is built from four of these L-shaped polygons and is \(\mathbb{Z}^2\)-covered by \(U_{a,b}\). This surface is shown in Figure 5.3 and will be referred to as \(R_{a,b}\). This is a genus five surface, but is naturally a four-fold cover of a single, genus two, L-shaped surface which we denote \(L_{a,b}\) as shown in Figure 5.4.

Using these coverings we can study the complicated, infinite genus \(U_{a,b}\) by looking at the much simpler, and much better understood, L-shaped surface. In particular, by the results of Calta and McMullen, we know precisely when such a surface is Veech.

Using these constructions, Hubert, Lelièvre, and Troubetzkoy show that for a certain
collection of rational parameters,

\[ \mathcal{E} = \{ (a, b) = (p/q, r/s) \in \mathbb{Q}^2 \mid p, r \text{ odd}, q, s \text{ even} \}, \]

with \( p/q \) and \( r/s \) in lowest terms, almost every geodesic in \( \mathcal{U}_{a,b} \) – and hence billiard in \( \mathcal{T}_{a,b} \) – is recurrent: they return arbitrarily close to their starting point infinitely often.

**Theorem 5.1** (Thm. 4 of [HLT11]). *If \((a, b) \in \mathcal{E}, then there is a dense set of strongly parabolic\(^1\) rational directions; and Lebesgue almost every direction is recurrent.*

Since the parameters in \( \mathcal{E} \) are rational, the L-shaped surface covered tiled by \( \mathcal{U}_{a,b} \) is arithmetic (i.e., square-tiled), and rational directions on such surfaces give a cylinder decomposition where the heights and lengths of the cylinders are bounded by certain simple

---

\(^1\)A direction is *strongly parabolic* if every regular trajectory in that direction is periodic and each trajectory has the same period.
functions of the denominator of the slope of the direction. These rational directions are then used to approximate a set of irrational directions of full measure which have a good Diophantine approximation.\(^2\)

Using Sage to perform simulations of a billiard in the wind-tree, we observed the opposite phenomenon: for a dense set of directions, billiard trajectories for wind-tree tables with parameters in \(E\) had escaping orbits. By *escaping* we mean the orbit eventually leaves every compact set. In fact, something much more interesting than merely escaping was observed: the trajectories that we will describe exhibit a sort of self-similarity. See Figure 5.5 which is the trajectory of a billiard in the wind-tree model with \(\frac{1}{2} \times \frac{1}{2}\) obstacles, the billiard emitted from the midpoint of a top edge of one of the obstacles with slope \(\sqrt{2}\). One goal is to explain precisely why this, and similar trajectories, escape and are self-similar.

The directions where we observed this phenomenon arose in trying to mimic the

\(^2\)An irrational number \(x\) has a *good* rational approximation if for a chosen \(\varepsilon > 0\), there is a sequence of rational numbers \(\left(\frac{p_n}{q_n}\right)_{n \in \mathbb{N}}\) approaching \(x\) such that \(\left| x - \frac{p_n}{q_n} \right| < \frac{\varepsilon}{q_n^2} \).
construction of dense geodesics by Panov, in the case of a surface related to the wind-tree’s unfolding. This construction will be described in detail in the next section.

After observing this phenomenon, similar results obtained by Vincent Delecroix were brought to our attention.

5.1.4 Delecroix’s results

In [Del13], Delecroix shows the existence of a dense set of directions of positive Hausdorff dimension such that regular billiard trajectories in the wind-tree model in those directions must escape. Delecroix’s paper describes a fairly complicated relationship between the sizes of obstacles and these escaping directions which allows for the explicit construction of such directions by a continued fraction.

These results are summarized below by combining the main theorem of [Del13] with some technical propositions.

**Theorem 5.2 (V. Delecroix).** Consider the wind-tree billiard with obstacles of size $a \times b$, and suppose that $\theta$ is the initial direction of a billiard trajectory in $T_{a,b}$. Consider sequences $x_k, y_k, m_k, z_k$ defined as follows:

\[
\begin{align*}
x_1 &= (1 - b) \cos \theta & x_{k+2} &= y_k \\
x_2 &= b \cos \theta & x_{k+3} &= y_{k+1} \\
y_1 &= (1 - a) \sin \theta & y_{k+2} &= x_k - m_k(y_k + y_{k+1}) \\
y_2 &= a \sin \theta & y_{k+3} &= x_{k+1} - n_k \\
m_k &= \left\lfloor \frac{x_k}{y_k + y_{k+1}} \right\rfloor & n_k &= \left\lfloor \frac{x_{k+1}}{y_{k+1}} \right\rfloor
\end{align*}
\]

If $x_k + x_{k+1} > y_k > x_{k+1}$ and $y_k + y_{k+1} > x_k > y_{k+2}$ for all $k$ and if $n_k$ is an even integer for all odd $k$, then the billiard trajectory in direction $\theta$ is escaping. Furthermore, if $(a,b) \in \mathcal{E}$, then the set of $\theta$ satisfying the above is dense in $S^1$ and has Hausdorff dimension at least
The main idea in the proof of this theorem is to explicitly construct a set of directions which are poorly approximated by rational directions on arithmetic L-shaped surfaces, in contrast to the directions of [HLT11] which are well-approximated by rational directions on arithmetic surfaces.

This is a fairly technical proof relying heavily on the machinery of Ferenczi-Zamboni induction, which is a specialized variant of the standard Rauzy-Veech induction. The results described in this dissertation, while similar in consequence to the results of [Del13], are obtained through conceptually simpler, topological and geometric arguments.

5.2 Connecting the wind-tree and Panov planes

We now describe how Panov planes may be used to study billiard trajectories in the wind-tree model. To do this we will consider L-shaped surfaces (folded or unfolded) which are constructed by removing a rectangle from the upper right-hand corner of a $1 \times 1$ square to obtain a polygon. We let $L_{a,b}$, $L_{\hat{a},b}$, $L_{a,\hat{b}}$, and $L_{\hat{a},\hat{b}}$ be the surfaces obtained by cutting a $a \times b$ rectangle from the corner, folding edges corresponding whose parameter is has a hat over it, and identifying any opposite edges which are unfolded. We denote the universal covers of the folded L’s by $\hat{C}_{\hat{a},b}$, $\hat{C}_{a,\hat{b}}$, and $\hat{C}_{\hat{a},\hat{b}}$.

Notice that each of our folded L’s can be rescaled to one of the $L_{1,\alpha,\beta}^1$, $L_{1,\hat{\alpha},\beta}^1$, $L_{1,\alpha,\hat{\beta}}^1$, $L_{1,\hat{\alpha},\hat{\beta}}^1$ surfaces by taking

$$\alpha = \frac{a}{1-a},$$
$$\beta = \frac{b}{1-b}.$$ 

Any $(\alpha, \beta)$ that appears below is obtained from $(a, b)$ in this way.

Consider the wind-tree billiard table $T_{a,b}$ together with the Panov planes $\hat{C}_{\hat{a},b}$ and $\hat{C}_{a,\hat{b}}$. We now make the following simple observation:
**Lemma 5.3.** Let \((x(t), y(t))\) be any billiard trajectory in \(\mathcal{T}_{a,b}\) and suppose \((x(0), y(0)) = (x_0, y_0)\) and the initial direction of the billiard trajectory is \(\theta\). Let \((x_1(t), y_1(t))\) and \((x_2(t), y_2(t))\) be the geodesic trajectories in the planes \(\mathbb{C}_{a,b}\) and \(\mathbb{C}_{a,\hat{b}}\), respectively, such that \((x_1(0), y_1(0)) = (x_0, y_0)\) and the geodesic has initial direction \(\theta\). Then for all time \(t\), \(y_1(t) = y(t)\) and \(x_2(t) = x(t)\).

**Proof.** As the geodesic in the Panov plane and the billiard in the wind-tree model both start from the same position in the same direction, \(y_1(t) = y(t)\) and \(x_2(t) = x(t)\) for some short period of time, until the billiard first hits an obstacle and is reflected. Suppose the billiard first hits a horizontal side of an obstacle. The geodesic in \(\mathbb{C}_{a,\hat{b}}\) then has its \(y\)-coordinate changed (so \(y(t) \neq y_2(t)\)), but the \(x\)-coordinates still agree (the billiard’s vertical velocity has changed, but the horizontal velocity remains the same). Thus \(x(t) = x_1(t)\) again. In fact, reflecting the billiard table through the horizontal line cutting the obstacle hit by the billiard in half, we see that the \(x\)-coordinates, \(x(t)\) and \(x_2(t)\) will be equal at least until the billiard intersects the next obstacle.

The argument in case the billiard first intersects a vertical side is identical, except that \(x(t)\) and \(y(t)\) are exchanged; \(x_2(t)\) is replaced by \(y_1(t)\); \(x_1(t)\) is replaced by \(y_2(t)\); and the words “horizontal” and “vertical” are exchanged.

This shows the claim is true at least until the billiard first hits an obstacle. Repeating the argument, after reflecting the billiard table through the horizontal or vertical line which cuts the intersected obstacle in half, gives the result for the next piece of the trajectory, until the billiard hits a second obstacle. Thus repeating the argument gives that the values \(x_2(t)\) and \(x(t)\) are equal for all \(t\); \(y_1(t)\) and \(y(t)\) are equal for all \(t\).

The observation that a billiard trajectory in the wind-tree is completely described by geodesics in a pair of Panov planes opens the door for using Panov planes to model infinite translation surfaces. Note that topologically Panov planes are very simple objects, homeomorphic to the complex plane, and since they are universal covers of tori, any map on the torus will always lift to the Panov plane. Neither of these statements are true for
general translation surfaces. The unfolding of the wind-tree, for example, has infinite genus; and maps on the genus-two L-shaped surface covered by the unfolding do not necessarily lift.

While these ideas are still in their infancy, they may provide alternative methods for studying infinite translation surfaces and polygonal billiards. We end this section by mentioning one such application.

5.2.1 Vertical recurrence of wind-tree trajectories

Using the observation above that a pair of Panov planes captures billiard dynamics in the wind-tree, we show how Panov planes and half-translation tori may be used to study billiard trajectories in the wind-tree. In particular, we will show how local observations about Panov planes may be used to make statements about the global wind-tree trajectory. We begin with one simple lemma about geodesics on Panov planes in the eigendirection of a pseudo-Anosov.

**Lemma 5.4.** Let $T = (\mathbb{T}^2, q)$ be a half-translation torus and $\varphi \in \text{Aff}^+(T)$ a pseudo-Anosov. Suppose the expanding and contracting directions of $\varphi$ are $\theta^\pm$. Let $C_T^T$ be the associated Panov plane. Suppose $P \in C_T^T$ is a $\pi$-angle point fixed by a lift $\tilde{\varphi}$ of $\varphi$, and let $\ell^\pm$ be the geodesics in directions $\theta^\pm$ through $P$. If $\ell^+ \cap \ell^- \neq \emptyset$, then $P$ is in the closure of $\ell^+$.

**Proof.** Applying $\tilde{\varphi}$ preserves the geodesics $\ell^\pm$ and any point on $\ell^-$ is moved closer to $P$; the distance between $P$ and a given point on $\ell^-$ shrinks exponentially with each application of $\tilde{\varphi}$. Thus there are points on the intersection $\ell^+ \cap \ell^-$ which are arbitrarily close to $P$. 

Now we give a condition for folded L-shaped surfaces $L_{\tilde{a}, \tilde{b}}$ which guarantees the intersection of the expanding and contracting leaves of a pseudo-Anosov.

**Lemma 5.5.** Let $\varphi$ be a pseudo-Anosov on a folded L-shaped surface, $L_{\tilde{a}, \tilde{b}}$, $C_T^T$ the associated Panov plane, and let $\tilde{\varphi}$ be a lift of $\varphi$ to $C_T^T$ which fixes a cone point $\pi$-angle cone point $P$. Let $\ell^\pm$ denote the expanding and contracting leaves through $P$. Let $h$ denote the horizontal saddle connection containing $P$, and $v$ the vertical saddle connection containing $P$. If $v$
separates the initial segments of \( \ell^+ \) and \( \ell^- \) (i.e., if for some \( \varepsilon > 0 \), the pieces of \( \ell^+ \) and \( \ell^- \) of length \( \varepsilon \) through \( P \) are on different sides of \( v \)) and if \( \ell^+ \) intersects \( h \) at a point distinct from \( P \), then \( \ell^+ \cap \ell^- \neq \emptyset \).

**Proof.** Let \( C \) denote the pillow case on \( \mathbb{C}^T \) which contains the point \( P \) and the initial segments of \( \ell^\pm \). Let \( \ell^+_0 \) denote the segment of \( \ell^+ \) emanated from \( P \) which stays inside of \( C \). Suppose that \( \ell^+ \) leaves \( P \) pointing to the right, and \( \ell^- \) leaves \( P \) pointing to the left.

Suppose \( \ell^+ \) intersects \( h \) at a point \( P_1 \). Let \( \ell^+_1 \) denote the segment of \( \ell^+ \) which is in \( C \), emanated from \( P_1 \), and is to the left of \( P \). See Figure 5.6.

Figure 5.6: The pillow case \( C \) in Lemma 5.5.

If \( \ell^+_1 \) intersects the vertical saddle connection \( v \), then \( \ell^- \) obviously must intersect \( \ell^+_1 \).

So suppose \( \ell^+_1 \) does not intersect \( v \). Then \( \ell^+_1 \) must intersect the horizontal saddle connection above \( P \) at the top of the pillow case \( C \). This means \( \ell^+_1 \) must also intersect the horizontal saddle connection at the top of \( C \), since the length from \( P_1 \) to \( P \) is less than the length of \( h \).

Consider the parallelogram whose left- and right-hand sides are given by the segments \( \ell^+_0 \) and \( \ell^+_1 \), and whose top and bottom sides are portions of the horizontal saddle connections at the top and bottom of \( C \). Note this parallelogram is contained within the pillow case \( C \).

Note the leaf \( \ell^- \) must eventually leave the pillow case \( C \). We will show that before leaving the pillow case, \( \ell^- \) must cross some segment of \( \ell^+ \). As \( \ell^- \) points to the left, if it does not cross the segment \( \ell^+_1 \), then it must intersect the top of the parallelogram.

Let \( d_- \) denote the distance from the midpoint of the top fold of \( C \) to the intersection
of the leaf $\ell^-$ with top fold. Let $d_+$ denote the distance from the midpoint of the top fold of $C$ to the intersection of the leaf $\ell^+_0$ with top fold. If $d_+ = d_-$, then $\ell^+_0$ and $\ell^-$ intersect at the top of the parallelogram.

If $d_+ < d_-$, as illustrated in Figure 5.7, then $\ell^-$ exits the top of the fold, now moving to the right, and must intersect the right-hand edge given by $\ell^+_0$.

Finally, suppose $d_+ > d_-$. Let $d$ denote the distance in the top fold of the pillow case from the center of the fold to the intersection with $\ell^+_1$. As this distance is less than the distance between $P$ and $P_1$, the leaf $\ell^+_1$ exits the right-hand side of the top fold and then intersects the bottom fold to the right of $P$. Call this segment of the leaf $\ell^+_2$. Now consider the parallelogram whose left- and right-hand sides are $\ell^+_1$ and $\ell^+_2$, as seen in Figure 5.8.

Note the reflection of $\ell^-$ to the right-hand side of the top fold is to the left of $\ell^+_2$. This new segment of $\ell^-$ either intersects $\ell^+_2$, or the bottom fold. If the bottom fold is intersected, we obtain a new segment of $\ell^-$ which is to the left of the initial segment. This new segment either intersects $\ell^+_1$, or the top fold. Each time the leaf intersects the top or bottom fold, it is moved closer one of the segments $\ell^+_1$ or $\ell^+_2$ on the left- and right-hand
sides of the parallelogram. Since the leaf must eventually exit the parallelogram it will, after a finite number of intersections with the top and bottom folds, intersect one of the sides \( \ell_1^+ \) or \( \ell_2^+ \).

Combining Lemma 5.5 with Lemma 5.4, we have a fairly simple condition to check that guarantees certain directions in a Panov plane contain leaves which repeatedly return to a given \( y \)-value. As billiard trajectories in the wind-tree may be modeled with a pair of Panov planes, this condition carries over to wind-tree trajectories.

**Corollary 5.6.** Let \( \mathcal{T}_{a,b} \) be a wind-tree billiard table such that \( L_{\hat{a},\hat{b}} \) admits a pseudo-Anosov \( \varphi \) satisfying the conditions of Lemma 5.5. A billiard in \( \mathcal{T}_{a,b} \) emitted from the midpoint of the top side of an obstacle in the expanding direction of the pseudo-Anosov \( \varphi \) passes through points with the same \( y \)-value as the initial point of emission infinitely often.

We give one particular example of a billiard satisfying the conditions of the above lemma. Let \( a = b = 1/2 \), and on the surface \( L_{\hat{a},\hat{b}} \) consider the affine diffeomorphisms \( P_h \), which twists the long horizontal cylinder once, and \( P_v \), which twists the long vertical cylinder once. The map \( \varphi = P_h P_v^2 \) is a pseudo-Anosov with derivative

\[
D\varphi = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}^2 = \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}.
\]

The eigenvectors of \( D\varphi \) are \([\frac{1}{2} (2 + \sqrt{6}), 1]^T\) and \([\frac{1}{2} (2 - \sqrt{6}), 1]^T\). Lifting \( \varphi \) to map on the Panov plane which fixes a \( \pi \)-angle cone point, we see the expanding and contracting directions of the map are separated by the vertical saddle connection. Now we simply need to show the leaf in the expanding direction will return to the horizontal saddle connection containing the fixed \( \pi \)-angle point. This can be done “experimentally” by following the leaf until we see a return to the saddle connection, or this can be reasoned as follows.
Let $C$ be the pillow case in the Panov plane which contains the $\pi$-angle cone point on its bottom edge, and let $\ell^+_0$ denote the first segment of the expanding leaf inside the pillow case. This segment connects the cone point to the right-hand side of the pillow case. Note that because of the fold, this pillow case has one vertical cylinder with height (which in this case refers to the horizontal length of the cylinder) $1/2$ and width (circumference) $1$. Applying $P^2_v$ performs a complete twist of this cylinder, and so produces a new intersection point of $P^2_v(\ell^+_0)$ with the horizontal saddle connection. Applying $P_h$ now preserves this horizontal saddle connection, and converts $P^2_v(\ell^+_0)$ into a longer portion of the expanding leaf which we now see has an intersection with the horizontal saddle connection distinct from the fixed point $P$.

By Corollary 5.6, a billiard in the wind-tree with obstacle sizes $1/2 \times 1/2$ emitted from the midpoint of the top of an obstacle in the direction $\left[\frac{1}{2} \left(2 + \sqrt{6}\right), 1\right]^T$ will return to the same $y$-level as the original point infinitely often. This billiard trajectory is shown in Figure 5.9.
5.3 Self-similarity in billiard trajectories

We end this chapter by explaining the “self-similarity” which appears in certain billiard trajectories in the wind-tree, such as in Figure 5.5. These self-similar trajectories occur when we consider the billiard in the eigendirection of a pseudo-Anosov diffeomorphism of the underlying L-shaped surface. Affine pseudo-Anosovs correspond to closed geodesics in the moduli space of translation surfaces, and this in turn implies certain interval exchange transformations on the surface are “self-similar” in the sense that the first return map to an appropriately chosen subinterval is simply a rescaling of the original interval exchange.

We will explain the self-similarity of billiards by using these facts about pseudo-Anosovs and interval exchanges to construct a symbolic substitution map, and show how to associate pieces of the billiard trajectory to the characters that appear in words obtained by iterating the substitution map.

5.3.1 The substitution map

To describe the construction of a substitution map, we begin with a general observation about appropriately chosen interval exchanges on translation surfaces which admit pseudo-Anosovs.

Lemma 5.7. Suppose that $(X, \omega)$ is a translation surface, and suppose $\varphi \in \text{Aff}^+(X, \omega)$ is pseudo-Anosov. Let $\theta^\pm$ denote the expanding and contracting directions of $\varphi$. Let $P$ be any fixed point of $\varphi$, and let $\ell^\pm$ be two geodesic rays emanated from $P$ in the directions of $\theta^\pm$. Note that $\ell^\pm$ have a dense set of intersections. Let $P_0$ be any one of these intersections, and let $I_0$ denote the segment of $\ell^-$ connecting $P_0$ to $P$. Let $T_0 : I_0 \to I_0$ denote the interval exchange obtained as the first-return map to $I_0$ when flowing from $I_0$ in the direction $\theta^+$. Then $I_0$ is self-similar in the sense that the first-return to a certain $I_1 \subseteq I_0$ is the same as $T_0$, but with the lengths of subintervals scaled by some factor $\mu \in (0, 1)$.

Proof. Deforming $(X, \omega)$ by applying elements of $\text{SL}(2, \mathbb{R})$, we see that $(X, \omega)$ is conjugate to a translation surface $(Y, \eta)$ where the pseudo-Anosov on $(Y, \eta)$ corresponding to $\varphi$ has
expanding horizontal direction and contracting vertical direction. Let $J_0$ denote the subinterval of $(Y, \eta)$ which corresponds to $I_0$ on $(X, \omega)$. Note $J_0$ is a vertical line segment. Let $S_0 : J_0 \to J_0$ denote the interval exchange obtained by flowing from $J_0$ in the horizontal direction until intersecting $J_0$. Note the derivative of the pseudo-Anosov on $(Y, \eta)$ has the form

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

for some $\mu \in (0, 1)$.

Applying the Teichmüller geodesic flow $g_t$ to $(Y, \eta)$ for time $t = \ln(\mu)$ brings the surface $(Y, \eta)$ back to itself: that is, there is a closed Teichmüller geodesic in moduli space through the point $(Y, \eta)$. Note that while applying $g_t$, the interval $J_0$ is continuously shrunk, while the horizontal leaves through $J_0$ are continuously stretched. The ratios of sizes between two subintervals of $J_0$ remain the same however (they are both shrunk by a factor of $e^{-t}$), and the permutation of the intervals does not change.

After flowing for time $\ln(\mu)$ we obtain a subinterval $J_1 \subseteq J_0$, and the first-return to $J_1$ is simply a rescaling of the first-return to $J_0$ by the above. The lemma thus holds on the surface $(Y, \eta)$ and so holds on all of its $\text{SL}(2, \mathbb{R})$-conjugates, notably our original surface $(X, \omega)$.

To connect the above lemma to the billiard trajectory in the wind-tree model, recall the wind-tree with obstacles of size $a \times b$ unfolds to a translation surface $U_{a,b}$ which is a cover of the L-shaped surface $L_{a,b}$. Suppose that $L_{a,b}$ is a Veech surface and so admits an affine pseudo-Anosov $\varphi$. Note that $L_{a,b}$ is genus two and so hyperelliptic. An affine diffeomorphism of $L_{a,b}$ must preserve the set of Weierstrass points of the surface, which appear in Figure 5.10. Replacing $\varphi$ with a power if necessary, we may suppose that $\varphi$ fixes each Weierstrass point and also fixes the expanding and contracting leaves through the $6\pi$-angle cone point. (Note there are three distinct expanding leaves and three distinct contracting leaves through the cone point. Any affine diffeomorphism on the surface must fix the cone point, since there is only one, but the leaves may be permuted.) Let $P$ be any
Let $\theta^+$ and $\theta^-$ denote the expanding and contracting directions of $\varphi$, respectively, and consider the billiard in $T_{a,b}$ starting from a preimage of $P$ in the direction of $\theta^+$. Suppose $\ell^\pm$ is the geodesic on $L_{a,b}$ through the Weierstrass point $P$ in the direction $\theta^\pm$. Note $\ell^+$ and $\ell^-$ intersect in a dense set of points. Let $P_0$ denote any one of these points of intersection, and let $I_0$ denote the segment of $\ell^-$ from $P_0$ to $P$. See for example Figure 5.11 where the red segment is a portion of the ray $\ell^+$, and the blue segment is $I_0$.

![Figure 5.11: The piece of a contracting leaf used in constructing a substitution map.](image)

Let $T_0 : I_0 \to I_0$ denote the interval exchange on $I_0$ obtained by flowing in the direction $\theta^+$. This map $T_0$ is necessarily an interval exchange on five subintervals: three points of discontinuity inside the interval come from the three leaves in direction $\theta^+$ emanated from
the 6π-angle cone point of the surface, and the fourth discontinuity comes from following the leaf \( \ell^+ \) from \( P \) in the opposite direction. See Figure 5.12 for an example.

Suppose the subintervals of \( I_0 \) are given labels from the set \( A_0 = \{ A_0, B_0, C_0, D_0, E_0 \} \) in left-to-right order. Let \( \sigma \) denote the permutation on five letters associated to \( T_0 \). Suppose the larger eigenvalue of \( D\varphi \) is \( \mu \), so \( \varphi \) stretches leaves in the direction of \( \theta^\pm \) by the factor \( \mu^\pm 1 \).

We obtain a subinterval \( I_1 \subseteq I_0 \) by applying \( \varphi \); i.e., \( I_1 = \varphi(I_0) \). As \( P \) is fixed by \( \varphi \), \( P \) is one endpoint of this interval; let \( P_1 \) denote the other endpoint and notice \( P_1 = \varphi(P_0) \).

Let \( T_1 : I_1 \to I_1 \) denote the first-return IET to \( I_1 \) obtained by flowing in direction \( \theta^+ \). We may iterate this procedure constructing a sequence of points \((P_n)_{n \in \mathbb{N}_0}\), intervals \((I_n)_{n \in \mathbb{N}_0}\), and interval exchanges \((T_n : I_n \to I_n)_{n \in \mathbb{N}_0}\). By Lemma 5.7, each \( T_n \) is simply \( T_0 \) rescaled by a factor of \( \mu^{-n} \), and so is an IET on five subintervals. Suppose the subintervals of \( T_n \) are given labels from \( A_n = \{ A_n, B_n, C_n, D_n, E_n \} \) such that \( A_n = \varphi(A_{n-1}) \), and likewise for the other subintervals. We will sometimes refer to elements of \( A_n \) as symbols, and sometimes as subintervals of \( I_n \). The meaning will always be clear from context.

For each \( n \geq 1 \), consider the symbolic map

\[
\zeta_n : A_n \to A_{n-1}^*,
\]

where \( A_{n-1}^* \) denotes the free monoid on the set of symbols \( A_{n-1} \), obtained in the following way. To determine \( \zeta_n(A_n) \), let \( x \) be any point in the interval labeled \( A_n \) and consider the
geodesic ray emanated from \( x \) in the direction \( \theta^+ \). This ray will first intersect \( I_n \) at the point \( T_n(x) \), but before this may intersect the segment \( I_{n-1} \). Each time the ray intersects \( I_{n-1} \), record the label from \( A_{n-1}^* \) corresponding the intersected subinterval. If the ray does not intersect \( I_{n-1} \) before returning to \( I_n \), then record the empty word \( \varepsilon \). Note this sequence of labels in \( A_{n-1}^* \) is independent of the chosen \( x \in A_n \), as no element of \( A_n \) will intersect a cone point before returning to \( I_n \). Repeating this process for each subinterval of \( I_n \), we obtain the map \( \zeta_n \).

Let \( A = \{A, B, C, D, E\} \) and for each \( n \geq 0 \) define a map \( \rho_n : A_n \to A \) by forgetting the subscript \( n \) on each letter. E.g., \( \rho_3(B_3) = B \).

**Lemma 5.8.** For every \( m, n \geq 0 \), \( \rho_n \circ \zeta_n = \rho_m \circ \zeta_m \).

**Proof.** It suffices to show that \( \rho_n \circ \zeta_n = \rho_{n+1} \circ \zeta_{n+1} \), but this follows trivially from the self-similarity described by Lemma 5.7.

Lemma 5.8 implies that each \( \zeta_m \) and \( \zeta_n \) are identical, aside from the chosen subscripts. By forgetting all subscripts, we obtain a map \( \zeta : A \to A^* \). We extend \( \zeta \) to a map \( A^* \to A^* \), also denoted \( \zeta \), by applying the map \( A \to A^* \) component-wise and concatenating the results. We refer to this \( \zeta \) as the substitution map. We will now refer to elements of \( A \) both as symbols and as subintervals of \( I_0 \) with \( A = A_0, B = B_0 \), and so on.

### 5.3.2 Symbolic orbits

Let \( \alpha_n : I_n \to A_n \) denote the map which returns the label of a subinterval a given point is contained in. That is, \( \alpha_n(x) = A_n \) if \( x \in A_n \); \( \alpha_n(x) = B_n \) if \( x \in B_n \), and so on. Let \( \alpha = \rho_0 \circ \alpha_0 \). By the symbolic orbit of a point \( x \) under an IET \( T \), we mean the infinite collection of symbols obtained by recording which subinterval \( T^n x \) is in for each \( n \). That is, the symbolic orbit of \( x \in I_n \) under \( T_n \) is the infinite word

\[
\prod_{k=0}^{\infty} \alpha_n \left( T_n^k x \right).
\]
In the special case of self-similar interval exchanges, such as the $T_n$ above, iterating the substitution map on a certain finite word produces the symbolic orbit.

**Lemma 5.9.** There exists a finite word $w \in \mathcal{A}^*$ and a map $\eta: \mathcal{A}^* \to \mathcal{A}^*$ such that iteration of $w$ under $\eta$ gives the symbolic orbit of the point $P_0$ under the map $T_0$:

$$\prod_{k=0}^{\infty} \alpha^k(T_0^k P_0) = \lim_{k \to \infty} \eta^k(w).$$

**Proof.** Note that each $n \geq 0$, there is some finite number $j$ such that $T_n^j(P_n) = P_{n+1}$. Let $w_n \in \mathcal{A}_n^*$ be the word of intermediate symbols that occur in successively applying $T_n$ to $P_n$ until reaching $P_{n+1}$:

$$w_n = \prod_{k=0}^{j} \alpha_n^k(T_n^k(P_n)).$$

Note that we may continue iterating $T_n$ until reaching $P_{n+2}$. This produces a new word in $\mathcal{A}_n^*$ which we may write as $w_n \cdot \zeta_{n+1}(w_{n+1})$. To see this, note $w_{n+1} \in \mathcal{A}_{n+1}^*$ is the word of intermediate symbols obtained by iterating $T_{n+1}$ until reaching $P_{n+2}$ from $P_{n+1}$. As $I_{n+1} \subseteq I_n$, we can convert the word $w_{n+1}$ into a word in $\mathcal{A}_n^*$ by applying $\zeta_{n+1}$, and this new word, $\zeta_{n+1}(w_{n+1})$, represents the subintervals of $I_n$ obtained by iterating $T_n$ from $P_{n+1}$ to $P_{n+2}$.

Similarly, the set of symbols in $\mathcal{A}_n$ obtained by iterating $T_n$ until reaching $P_{n+3}$ from $P_n$ is

$$w_n \cdot \zeta_{n+1}(w_{n+1}) \cdot \zeta_{n+1}(w_{n+1} \cdot \zeta_{n+2}(w_{n+2})).$$

Applying the $\rho_n$ maps to obtain symbols in $\mathcal{A}$, we have

$$\rho_0(w_0) = \rho_1(w_1) = \cdots = \rho_n(w_n) = \cdots$$

Combining this with the observation from before that $\zeta = \rho_n \circ \zeta_n$ for each $n$, we have that the symbolic orbit of $P_0$ from until $P_1$ is $w$; the orbit from $P_1$ to $P_2$ is $\zeta(w)$; the orbit from $P_2$ to $P_3 = \zeta(w \cdot \zeta(w))$; and so on. Thus the orbit from $P_0$ to $P_2$ is $w \cdot \zeta(w)$; the orbit from
$P_0$ to $P_3$ is $w \cdot \zeta(w) \cdot \zeta(w \cdot \zeta(w))$ and so on.

Consider the map $\eta : \mathcal{A}^* \to \mathcal{A}^*$ given by taking any string $x \in \mathcal{A}^*$ and concatenating $\zeta(x)$ to it: $x \mapsto x \zeta(x)$. Iterating $\eta$ precisely adds the “next piece” of the orbit to the string $w$, and so the orbit from $P_0$ to $P_n$ is given by $\eta^{n-1}(w)$. Notice that by its definition, $\eta$ adds characters to the end of the word. Hence the limit of $\eta^k(x)$ as $k$ goes to infinity makes sense for any word $x$ and is simply an infinite word in $\mathcal{A}^\mathbb{N}$.

As $\zeta$ performs a character-by-character substitution, $\zeta$ distributes across concatenation. In particular, $\zeta(w \cdot \zeta(w)) = \zeta(w) \cdot \zeta^2(w)$. This observation gives us a convenient way to write out the iterates of $\eta$, and hence the symbolic orbit of $P_0$ under the IET $T_0$. Before mentioning this, however, we recall a special integer sequence.

The *1's-counting sequence* (A000120 in the On-Line Encyclopedia of Integer sequences, [Slo]) is a sequence of non-negative integers, $\{a_n\}_{n \geq 0}$ where $a_n$ equals the number of 1’s in the binary representation of $n$. For example, $a_0 = 0$ as $0 = 0_2$, $a_1 = 1$ as $1 = 1_2$, $a_2 = 1$ as $2 = 10_2$, but $a_7 = 3$ since $7 = 111_2$.

As mentioned in [Sch91, p. 383], there is a *fast algorithm* (in the sense that the $k$-th iterate produces $2^{k-1}$ terms of the sequence) for generating the elements of $a_n$. If the terms $a_0$, $a_1$, ..., $a_{2^k-1}$ are given, the next $2^k$ terms are obtained by adding 1 to each of the previous terms. This is simplest to illustrate by considering a table whose $k$-th row contains the first $2^k$ elements of the sequence:

\[
\begin{array}{c}
0 \\
0, 1 \\
0, 1, 1, 2 \\
0, 1, 1, 2, 1, 2, 2, 3 \\
0, 1, 1, 2, 1, 2, 2, 3, 1, 2, 3, 2, 3, 3, 4 \\
\vdots
\end{array}
\]
One important aspect of this sequence mentioned in [Sch91] is its self-similarity: if every other entry of the sequence is deleted, we obtain the original sequence. This will help to explain the self-similarity of the billiard above after noting the following lemma.

**Lemma 5.10.** The k-th iterate of \( \eta \) applied to \( w \) may be written as

\[
\eta^k(w) = \zeta^{a_0}(w) \cdot \zeta^{a_1}(w) \cdots \zeta^{a_{2^k-1}}(w)
\]

where \( \{a_n\}_{n \geq 0} \) is the 1's-counting sequence.

**Proof.** Note \( \eta^0(w) = w = \zeta^0(w) \) and \( a_0 = 0 \). Supposing the result holds for \( k - 1 \geq 0 \) we have

\[
\eta^k(w) = \eta \left( \eta^{k-1}w \right) \\
= \eta \left( \zeta^{a_0}(w) \cdot \zeta^{a_1}(w) \cdots \zeta^{a_{2^{k-1}-1}}(w) \right) \\
= \zeta^{a_0}(w) \cdot \zeta^{a_1}(w) \cdots \zeta^{a_{2^{k-1}-1}}(w) \cdot \zeta \left( \zeta^{a_0}(w) \cdot \zeta^{a_1}(w) \cdots \zeta^{a_{2^{k-1}-1}}(w) \right) \\
= \zeta^{a_0}(w) \cdot \zeta^{a_1}(w) \cdots \zeta^{a_{2^{k-1}-1}}(w) \cdot \zeta^{a_0+1}(w) \cdot \zeta^{a_1+1}(w) \cdots \zeta^{a_{2^{k-1}-1}+1}(w).
\]

However, by the above fast algorithm for generating the sequence 1's-counting sequence, the powers of \( \zeta \) appearing in the above expression of \( \eta^k \) are precisely the elements of the 1's-counting sequence. \( \square \)

We now have a symbolic representation of the original billiard trajectory in the wind-tree where by applying the substitution \( \zeta \) to a particular word \( w \) in a certain way, determined by the elements of the 1's-counting sequence, we can construct long pieces of the symbolic trajectory. We would like to recover the original billiard trajectory from this symbolic representation, but first note the self-similarity of the 1's-counting sequence is manifested in the self-similarity of certain curves.
5.3.3 Building self-similar curves

Suppose for each letter $\alpha \in \mathcal{A}$, $\gamma_\alpha$ is a curve in $\mathbb{R}^2$ subject to the following constraints:

1. $\gamma_\alpha$ is oriented and has finite arclength,

2. $\gamma_\alpha$ is piecewise smooth, and

3. the tangent vectors at the ends of each $\gamma_\alpha$ and $\gamma_\beta$ are the same for every $\alpha, \beta \in \mathcal{A}$.

We may then associate an oriented curve of finite length $\gamma_x$ to each word $x \in \mathcal{A}^*$ by concatenating the curves corresponding to the letters of $x$. For instance, if $x = AABE$, then $\gamma_x$ would be the curve obtained by concatenating the curves $\gamma_A, \gamma_A, \gamma_B$, and $\gamma_E$.

Note the substitution $\zeta$ induces a transformation on these curves, taking $\gamma_x$ to $\gamma_{\zeta(x)}$. Similarly, $\eta$ induces a transformation which produces new curves from old curves by concatenating $\gamma_{\zeta(x)}$ to $\gamma_x$. For a given word $x$, let $\Gamma_n(x) = \gamma_{\eta^n(x)}$, and let $\Gamma(x)$ be the curve obtained by taking the limit as $n$ goes to infinity in $\Gamma_n(x)$. Letting $\delta_n(x)$ denote the curve $\gamma_{\zeta^n(x)}$, for notational convenience, the curve $\Gamma(x)$ may then be written as

$$\Gamma(x) = \delta_0(x) \delta_1(x) \delta_1(x) \delta_2(x) \delta_1(x) \delta_2(x) \delta_2(x) \delta_3(x) \cdots$$

with the subscripts being determined by the 1’s-counting sequence. Here, juxtaposition means two curves are concatenated at their endpoints in a way which preserves the orientation of the factors. The self-similarity of the 1’s-counting sequence carries over to the curve: deleting every other $\delta_k(x)$ produces the same curve,

$$\Gamma(x) = \delta_0(x) \delta_1(x) \delta_1(x) \delta_2(x) \delta_1(x) \delta_2(x) \delta_2(x) \delta_3(x) \cdots$$

$$= \delta_0(x) \delta_1(x) \delta_1(x) \delta_2(x) \delta_1(x) \delta_2(x) \delta_2(x) \delta_2(x) \cdots.$$
5.3.4 Recovering the billiard trajectory

Finally, to recover the wind-tree trajectory from the symbolic orbit we must associate curves in $\mathbb{R}^2$ to each letter $\alpha \in A$. There is one “hiccup” here, however, in that our letters can not determine a unique curve, but instead determine a family of related curves.

Each subinterval of $I_0$ determines a one-parameter family of pieces of billiard trajectories in the following way. Recall that the unfolding of the wind-tree model is made of four copies of the wind-tree billiard table and forms a cover of an L-shaped surface. The deck transformation group of this cover is $\mathbb{Z}^2 \rtimes (\mathbb{Z}/(2) \oplus \mathbb{Z}/(2))$, where $\mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$ acts on $\mathbb{Z}^2$ by negation in those components which contain 1. For example, $(1,0) \cdot (m,n) = (-m,n)$.

Let $\pi_L$ denote the covering map, and let $\tilde{I}_0$ be a connected component of $\pi_L^{-1}(I_0)$. Let $\tilde{I}$ be the collection of deck translates of $\tilde{I}_0$ which are contained in one of the four copies of the billiard table:

$$\tilde{I} = \bigcup_{(m,n) \in \mathbb{Z}^2} ((m,n),(0,0)) \cdot \tilde{I}_0.$$ 

Now let $\tilde{A}$ be the subset of $\tilde{I}_0$ corresponding to the subinterval $A$ of $I_0$. For each point $x \in \tilde{A}$, let $\tilde{\gamma}_A(x)$ denote the geodesic on $U_{a,b}$ in the direction $\theta^+$ from $x$ until the next intersection with $\tilde{I}$. Let $\gamma_A(x)$ denote the projection of this geodesic to the billiard table. Note that for each $x \in A$, $\gamma_A(x)$ is a broken line in the plane. For each $x \in A$, each curve $\gamma_A(x)$ has the same arclength, but the “breaks” in the curve occur at different points which vary continuously with $x$. We similarly have families of broken lines $\gamma_B, \gamma_C, \gamma_D,$ and $\gamma_E$ for each subinterval of $I_0$.

To see this, consider two billiards in the wind-tree emitted in the same direction from slightly different points, $P_1$ and $P_2$, such as in Figure 5.13. The billiards follow parallel trajectories for some amount of time, but the of the trajectories from the starting point until the first reflection (and subsequent reflections) will be slightly different. This difference can be made arbitrarily small by choosing points $P_1$ and $P_2$ to be very close.

Given a word $w \in A^*$, we obtain a family of curves by considering the set of all possible curves obtained by concatenating representatives from each family of curves for
each letter in $w$. For example, just to illustrate the idea, consider the word $AAB$. Choose an element of the family $\gamma_A$ (corresponding to the first $A$), and concatenate to the end of this curve another element of the family $\gamma_A$ (corresponding to the second $A$). Finally choose an element of $\gamma_B$ (corresponding to the $B$) and then concatenate to the curve obtained thus far. The collection of all possible curves obtained in this way is the family of curves associated with the word $AAB$.

The symbolic orbit of the billiard trajectory thus determines a family of curves, and one of these curves represents the true billiard trajectory. However, each curve in the family has the same “shape.” Suppose $\xi$ and $\kappa$ are two curves in this family. Each of these curves is a broken line, so suppose the line segments between the breaks were ordered $\xi_1, \xi_2, \xi_3, \ldots$ for $\xi$; and $\kappa_1, \kappa_2, \kappa_3, \ldots$ for $\kappa$. Each $\xi_i$ and $\kappa_i$ will be parallel, and the pieces of the curves will join together in the same way. That is, if the curves $\kappa$ and $\xi$ were given orientations such that their initial segments, $\kappa_1$ and $\xi_1$ point in the same direction, then $\kappa_i$ and $\xi_i$ will point in the same direction for all $i$.

Though the symbolic orbit does not uniquely determine the billiard trajectory, it does tell us information about the “shape” of the trajectory, and in particular explains the self-similarity observed in many experiments. Because of the variations that may occur within the families of curves, however, we should perhaps instead say the billiard trajectory is approximately self-similar.
Chapter 6

Involutive surfaces

Recall that if \( \pi : X \to Y \) is a surjection and if \( T : X \to X \) and \( S : Y \to Y \) are maps on \( X \) and \( Y \), we say that \( T \) and \( S \) are \textit{conjugate}\(^1\) if \( \pi \circ T = S \circ \pi \). In this chapter we describe some results about involutive surfaces, in particular counting the number of double covers of such a surface which admit a conjugate involution, and the number of quotients of that surface which have a given genus. This can be thought of as generalizing the relationship between Panov planes and the wind-tree model which was explained in the last chapter, in the sense that it gives a way to associate half-translation tori, and hence Panov planes, to an involutive surface.

6.1 Preliminary observations

Suppose that \((X, \omega)\) is a translation surface and that \( \sigma \in \text{Aff}^+(X, \omega) \) is an involution. This means \( D\sigma \) is an involution and as \( \sigma \) is assumed to be orientation preserving, there are only two possibilities: either \( D\sigma = I \) or \( D\sigma = -I \).

In the first case we have an order 2 automorphism of the translation structure. By an \textit{affine involution} we will always mean the second type of map: an affine diffeomorphism which is an involution, but not a translation. We will refer to a translation surface with

\(^1\)Some authors use the term \textit{semiconjugate}, reserving \textit{conjugate} for the case that \( \pi \) is a bijection.
an involution as an involutive surface, denoted by a triple \((X, \omega, \sigma)\). Our basic question, motivated by the construction in the previous chapter, is concerned with the existence of an involutive surface \((Y, \eta, \tau)\) together with a translation covering \(\pi : Y \to X\) such that \(\pi \circ \tau = \sigma \circ \pi\), and how the quotient surfaces \(X/\sigma\) and \(Y/\tau\) are related. Given that such a covering exists, how many non-isomorphic such coverings are there?

Suppose that \((X, \omega, \sigma)\) is an involutive surface. Note the derivative \(D\sigma\) described above has the natural interpretation as the Jacobian matrix of \(\sigma\), and that this matrix satisfies the Cauchy-Riemann equations. Hence an affine involution on a translation surface naturally descends to a conformal automorphism of the underlying complex structure. Typical affine diffeomorphisms are only quasiconformal and so do not descend to a biholomorphism of the Riemann surface; thus affine involutions are indeed very special maps. This observation means that affine involutions have a nice complex analytic interpretation.

**Lemma 6.1.** Let \((X, \omega)\) be a translation surface and \(\sigma : X \to X\) an involution. Then \(\sigma\) is an affine diffeomorphism of \((X, \omega)\) if and only if \(\sigma\) is a conformal map and \(\sigma^*\omega = -\omega\).

**Proof.** Suppose \(\sigma\) is an affine involution. As noted above, \(D\sigma = -I\). In a polygonal representation of the surface, \(\sigma\) acts by negation. In particular, any relative holonomy vector representing a relative period of \(\omega\) is mapped to its negative, and so \(\sigma^*\omega = -\omega\).

Suppose \(\sigma\) is a conformal involution and \(\sigma^*\omega = -\omega\). We may then interpret \(\sigma\) as a translation covering from \((X, -\omega)\) to \((X, \omega)\). In natural coordinates, the map \((X, -\omega) \xrightarrow{\sigma} (X, \omega)\) appears as a translation. This implies the map \(\sigma\) from \((X, \omega)\) to itself appears in natural coordinates as a composition of negation and translation: \(z \mapsto -z + b\). Thus \(\sigma\) is an affine diffeomorphism with derivative \(-I\).

In principle a translation surface \((X, \omega)\) may admit multiple affine involutions which \textit{a priori} need not be related. However, the complex analytic interpretation of an affine involution easily shows us that the difference between any two involutions is simply a translation.

**Lemma 6.2.** Any two affine involutions, \(\sigma_1\) and \(\sigma_2\), of a translation surface \((X, \omega)\) differ by an automorphism.
Proof. Notice that $\sigma_1^*\sigma_2^*\omega = \sigma_1^*(-\omega) = \omega$, meaning $(\sigma_2 \circ \sigma_1)^*$ is the identity on $\Omega^1 X$ and thus $\sigma_2 \circ \sigma_1 \in \text{Aut}(X, \omega)$.

Our goal will be to count the number of double covers of a surface which admit a conjugate involution, and also determine which of these covers have quotients of a certain genus. The main technical tool for doing this will be understanding the action of an involution in homology, and for this it will be helpful to recall the primary result of [Gil73] which may be stated as follows.

**Theorem 6.3** (Prop. 1 of [Gil73]). Let $X$ be a Riemann surface of genus $g$ and $\sigma$ a conformal involution with $k \geq 4$ fixed points. Suppose the quotient surface $X/\sigma$ has genus $\tilde{g}$. Then there exists a collection of simple closed curves on $X,

a_1, ..., a_{\tilde{g}}, b_1, ..., b_{\tilde{g}}, c_1, ..., c_{k/2-1}, d_1, ..., d_{k/2-1}

such that the homology classes of the curves

$a_1, ..., a_{\tilde{g}}, \sigma(a_1), ..., \sigma(a_{\tilde{g}}), c_1, ..., c_{k/2-1}$

$b_1, ..., b_{\tilde{g}}, \sigma(b_1), ..., \sigma(b_{\tilde{g}}), d_1, ..., d_{k/2-1}$

form a canonical basis in homology with $\sigma_*[c_i] = -[c_i]$ and $\sigma_*[d_i] = -[d_i]$. Here, canonical basis means the intersection form $i : H_1(X; \mathbb{Z}) \times H_1(X; \mathbb{Z}) \rightarrow \mathbb{Z}$ is easily computed in terms of this basis:

$i(a_i, b_j) = i(\sigma(a_i), \sigma(b_j)) = i(c_i, d_j) = \delta_{ij}$

and all other intersections of the basis elements are zero.

Furthermore, each $c_i$ and $d_i$ passes through exactly two fixed points, with no two curves passing through the same pair of fixed points. In fact, $c_i$ and $d_j$ share one common fixed point if $i = j$ and otherwise share no fixed points. No two $c_i$ share a fixed point, and
no two $d_j$ share a fixed point.

A basis of the form described in the theorem above is called a *canonical basis adapted to the involution* $\sigma$, or simply *adapted basis* if the involution is clear from context. For our purposes it will be convenient to relabel the elements of the adapted basis as follows. Let $m = \hat{g}$ and write $e_1^+, \cdots, e_{2m}^+$ for $a_1, \ldots, a_m, b_1, \ldots, b_m$. Let $e_i^- = \sigma(e_i^+)$. Let $n = k - 2$ and write $\nu_1, \ldots, \nu_n$ for $c_1, \ldots, c_{k/2-1}, d_1, \ldots, d_{k/2-1}$. Let $E^+$ denote the group generated by the $e_i^+$, $E^-$ the group generated by the $e_i^-$, and $N$ the group generated by the $\nu_i$.

Note we may now consider $H_1(X; \mathbb{Z})$ as the direct sum $E^+ \oplus E^- \oplus N$. We further have $\sigma_* e_i^+ = e_i^+$ and $\sigma_* \nu_i = -\nu_i$. For our purposes it will be important to consider elements of $H_1(X; \mathbb{Z})$ which have representatives that are invariant under $\sigma$. It is clear that all elements of $N$ have such a representative. No element of $E^+$ or $E^-$ has an invariant representative, since the construction of $E^\pm$ gives $\sigma(E^\pm) = E^{\mp}$. However, we easily construct $\sigma$-invariant classes within $E^+ \oplus E^-$ by adding a class and its involute. We denote the collection of all such invariant classes in $E^+ \oplus E^-$ by $I$:

$$I = \{ e + \sigma_*(e) \mid e \in E^+ \}.$$  

We denote the collection of all classes in $H_1(X; \mathbb{Z})$ which have a $\sigma$-invariant representative by $H_1^\sigma(X; \mathbb{Z})$. Letting $I_2$ and $N_2$ denote the mod 2 reductions of $I$ and $N$, respectively, we may write $H_1^\sigma(X; \mathbb{Z}/2) = I_2 \oplus N_2$.

If the space $N$ appearing above has rank $n$, we will say that $\sigma$ is an $n$-involution. Note an $n$-involution $n + 2$ fixed points if $n > 0$. If $n = 0$, then there are two possibilities: the involution may have two fixed points, or may be fixed-point-free. The second case occurs only if the involution is a translation automorphism, and this is a situation which we explicitly ignore.

**Proposition 6.4.** If $\sigma$ is a fixed-point-free affine involution on a closed translation surface $(X, \omega)$, then $\sigma$ is a translation.

**Proof.** Consider the map $f : X \to \mathbb{R}$ which measures the distance between a point and
its involute, \( f(p) = d(p, \sigma(p)) \). The function \( f \) is clearly continuous, and must attain a minimum value since \( X \) is compact. Suppose that minimum is not zero and is attained at \( p \). Consider a geodesic \( \gamma \) of length \( f(p) \) from \( p \) to \( \sigma(p) \).

If \( \sigma(\gamma) = \gamma \), then any point on \( \gamma \) other than the endpoints is closer to its involute than \( p \), contradicting that \( p \) is a minimum. In particular, the midpoint of this geodesic is fixed by \( \sigma \). If \( \sigma(\gamma) \neq \gamma \), then \( \gamma \cdot \sigma(\gamma) \) is a closed geodesic loop. Thus the flow in the direction of \( \gamma \cdot \sigma(\gamma) \) has a periodic component. Translating the loop orthogonally (i.e., left and right if we suppose the direction is vertical) we have an open, cylindrical neighborhood of the loop where \( \sigma \) acts by a half-rotation of this cylinder. Thus, in this cylinder, \( \sigma \) moves every point a fixed amount in a fixed direction: \( D\sigma = I \), and so \( \sigma \) is a translation. \( \square \)

**Corollary 6.5.** If \( \sigma \) is an affine involution of the translation surface \( (X, \omega) \) and is not a translation, then \( \sigma \) has at least two fixed points.

Finally, given any adapted basis for a conformal involution \( \sigma \), there exists one special, distinguished fixed point we will call the **ignored point** of the basis. To see this, note that the space \( N \) has rank \( n \) though there are \( n + 2 \) fixed points. Each generator \( \nu_i \) of \( N \) passes through exactly two fixed points, and each generator shares one fixed point with another generator. This observation allows us to represent elements of \( N_2 \) by binary strings whose bits determine the fixed points on an invariant representative of each class.

That is, suppose the fixed points of \( \sigma \) are labeled \( F_1, F_2, ..., F_{n+2} \) so that \( \nu_1 \) passes through \( F_1 \) and \( F_2 \); \( \nu_2 \) passes through \( F_2 \) and \( F_3 \); and so on. We may then represent elements of \( N_2 \) as elements of \( (\mathbb{Z}/2)^{n+2} \) where

\[
\nu_1 = 11000 \cdots 0000 \\
\nu_2 = 01100 \cdots 0000 \\
\vdots \\
\nu_n = 00000 \cdots 0110,
\]
and addition is performed bit-wise modulo 2. Each of these strings necessarily contains an even number of 1’s, and the last bit is always zero. The fixed point corresponding to this last bit is the ignored point.

The binary string representation of elements of $\mathcal{N}_2$ is dependent on several choices, such as our ordering of the $\nu_i$, and which fixed point plays the role of the ignored point will change if we make different choices. For the counting problem we are interested in, however, these choices will not change the count.

With these preliminary remarks out of the way, we now turn to the problem of describing which translation covers of an involutive surface admit conjugate affine involutions.

### 6.2 Covers admitting conjugate involutions

Let $(X, \omega, \sigma)$ be an involutive surface. We will construct, and classify, translation covers of $X$ which admit an involution conjugate to $\sigma$. After the author started this work, he was alerted to similar results in the Ph.D. thesis of Sergey Vasilyev, [Vas05]. Let us mention that the results described here differ from those of Vasilyev in two important ways: [Vas05] considers only hyperelliptic surfaces, and explicitly only considers quadratic differentials which are not squares of Abelian differentials.

Recall that isomorphism classes of connected, unbranched covers of a nice\(^2\) topological space $X$ are in one-to-one correspondence with conjugacy classes of subgroups of $\pi_1(X)$. A cover $\pi : X \to Y$ is called regular if the deck transformation group acts transitively on each of the fibers $\pi^{-1}(y)$. This occurs if and only if the cover is associated to a normal subgroup of $\pi_1(X)$. (These are also sometimes called normal or Galois covers.) Notice the degree of the cover is given by the index $[\pi_1(Y) : \pi_1(X)]$, and in particular degree-two covers are always regular as index-two subgroups are always normal.

Surfaces admit a symplectic intersection form which gives a signed “count” of the number of times two curves intersect. This count depends only on the homology of the

\(^2\)“Nice” here means the space has a universal cover; this occurs precisely when the space is connected, locally path-connected, and semi-locally simply connected. These conditions are automatically met by any connected manifold.
figures, and as such determines a homomorphism $i : H_1(X; \mathbb{Z}) \times H_1(X; \mathbb{Z}) \to \mathbb{Z}$. By using homology with $\mathbb{Z}/2$-coefficients, we may consider the $\mathbb{Z}/2$-intersection form, also denoted by $i : H_1(X; \mathbb{Z}/2) \times H_1(X; \mathbb{Z}/2) \to \mathbb{Z}/2$ which counts intersections modulo 2.

Fixing a class $[\gamma] \in H_1(X; \mathbb{Z}/2)$, we may consider the map $i_{[\gamma]} : H_1(X; \mathbb{Z}/2) \to \mathbb{Z}/2$ given by $[\delta] \mapsto i([\gamma],[\delta])$. The kernel of this map is a normal subgroup of $H_1(X; \mathbb{Z}/2)$, and since $H_1(X; \mathbb{Z}/2)$ is the abelianization of $\pi_1(X)$, there is a corresponding normal subgroup of $\pi_1(X)$. In this way the elements of $H_1(X; \mathbb{Z}/2)$ parameterize all of the connected double covers of $X$. We will denote the cover corresponding to $\ker(i_{[\gamma]})$ by $\pi_{[\gamma]} : X_{[\gamma]} \to X$ or simply $\pi_\gamma : X_\gamma \to X$.

Geometrically, these covers are built using the so-called slit construction. This means we pick a non-separating closed curve $\gamma$ on $X$, and consider the preimages of $\gamma$ on the trivial, disconnected double cover $\pi_0 : X_0 \to X$. We slit each component of $X_0$ along its preimage of $\gamma$, and then identify the two components across the slits. See Figure 6.1.

Our goal is to construct covers $\tilde{X}$ of $(X, \omega, \sigma)$ such that $\sigma$ lifts to a conjugate involution $\tilde{\sigma}$ of $\tilde{X}$ (i.e., $\sigma \circ \pi = \pi \circ \tilde{\sigma}$). To begin, suppose that $\pi_0 : X_0 \to X$ is the trivial, disconnected double cover. Notice there are two natural involutions on $X_0$: the deck transformation exchanging the two connected components (call this map $\delta$), and the map which performs $\sigma$ on each component (call this map $\sigma_0^+$). The composition of these two
involutions yields a third involution, $\sigma_0^- := \delta \circ \sigma_0^+$. The diagram in Figure 6.2 commutes.

We wish to define comparable involutions on each cover $X_\gamma$.

As a first step we need to understand the quotient surface $X/\sigma$.

**Lemma 6.6.** If $\sigma$ is a conformal involution with $k = 2j$ fixed points on a compact Riemann surface of genus $g$, then the quotient surface $X/\sigma$ has genus $\hat{g} = \frac{1}{2}(g - j + 1)$.

**Proof.** Note $X/\sigma$ is a smooth manifold and $\pi : X \to X/\sigma$ is a ramified double covering map. There is a unique complex structure on $X/\sigma$ so that $\pi$ is a holomorphic map. By the Riemann-Hurwitz formula,

$$
\chi(X) = 2\chi(X/\sigma) - 2j
$$

$$
\implies 2 - 2g = 4 - 4\hat{g} - 2j
$$

$$
\implies \hat{g} = \frac{g - j + 1}{2}.
$$

\hfill \square

**Lemma 6.7.** Suppose $(X, \omega, \sigma)$ is an involutive surface of genus $g$, $\sigma \notin \text{Aut}(X, \omega)$, and

$$(X, \omega) \in \mathcal{H}(f_1, \ldots, f_m, d_1, \ldots, d_n, d_{n+1}, \ldots, d_{2n})$$

where the points corresponding to zeros of $\omega$ of order $f_i$ are fixed by $\sigma$, and the points corresponding to zeros of order $d_i$ and $d_{i+n}$ are exchanged. The quotient surface $X/\sigma$ is a
half-translation surface and belongs to the stratum

\[ Q(-1, \ldots, -1, f_1 - 1, \ldots, f_m - 1, 2d_1, \ldots, 2d_n). \]

**Proof.** Consider \((X, \omega)\) as the half-translation surface \((X, \omega^\otimes 2)\) in \(Q(2f_1, \ldots, 2f_m, 2d_1, \ldots, 2d_n)\). The quotient map \(\pi : X \to X/\sigma\) is branched at the fixed points of \(\sigma\). Away from the ramification points, we place a quadratic differential \(q\) on \(X/\sigma\) as follows. Let \(P \in X/\sigma\) be an unramified point. Suppose that \(U \subseteq X/\sigma\) is an evenly covered neighborhood of \(P\), and that \(V_{\pm 1}\) are the preimages of \(U\) in \(X\). Choosing \(U\) small enough, we may suppose that \(\omega\) appears in the \(V_i\) as \(f_i(z)dz\). Thus \(\omega^\otimes 2\) appears as \(f_i(z)^2dz^2\). Notice that \(V_{\pm 1}\) are exchanged by \(\sigma\). Furthermore, as \(\sigma^*\omega = -\omega\), \(f_i(z) = -f_{-i}(z)\) and since squaring removes the sign, \(f_i^2 = f_{-i}^2\). Thus to define a quadratic differential on \(X/\sigma\), inside of the chart \(U\) we assign the function \(f_i^2\). Because the complex structure of \(X/\sigma\) is determined by \(X\) and coordinate changes in \(X\) satisfy the transformation rule for the quadratic differential \(\omega^\otimes 2\), the quadratic differential defined by \(f_i^2\) in local coordinates for \(X/\sigma\) also transforms in the required way. The orders of zeros, away from ramification points, of \(q\) agree with the orders of zeros of \(\omega^\otimes 2\), which are twice the orders of \(\omega\). (Though there are half as many such zeros.)

We can extend \(q\) to the ramification points of \(\pi\) on \(X/\sigma\), but we still need to calculate the order of \(q\) at each of these points, and this is a slightly more delicate procedure. In the following we will locally represent \(q\) as the square of some 1-form \(\eta\) in a neighborhood of the ramification point. (This, of course, does not extend in general to all of \(X/\sigma\) and only applies locally.) Suppose \(R\) is a ramification point of \(\pi\) with preimage \(P\). Note \(P\) is a fixed point of \(\sigma\), so the order of \(\omega\) at \(P\) is either 0 or some \(f_i\); call this value \(r\). In a small coordinate neighborhood \(V\) centered at \(P\), \(\omega\) is given by some \(f(z)dz\) with \(f(P) = 0\). Define a local 1-form \(\eta\) in the image of \(V\) as the push-forward of \(\omega\) by \(\pi\): locally, \(\omega = \pi^*\eta\). The only critical point of \(\eta\) in \(V\) is at \(R\), and to calculate the order of this point we consider the canonical divisors \(\text{div}(\omega)\) and \(\text{div}(\eta)\).
The canonical divisor of a pullback form $\omega = \pi^* \eta$ is related to the pullback of the canonical divisor $\pi^*(\text{div}(\eta))$ by the following formula (see [Mir95, p. 135]):

$$\text{div}(\omega) = \text{div}(\pi^* \eta) = \pi^*(\text{div}(\eta)) + R_\pi$$

(6.1)

where $R_\pi$ is the ramification divisor of $\pi$. In our local situation this divisor is simply $R_\pi = 1 \cdot P$, so applying equation (6.1), at the point $P$ we have

$$\text{div}(\omega)(P) = \pi^*(\text{div}(\eta))(P) + 1$$

$$\implies \ord_P(\omega) = 2\ord_R(\eta) + 1$$

$$\implies 2\ord_R(\eta) = r - 1.$$ 

Since (locally) $q = \eta^2$, the order of $q$ at $R$ is $2\ord_R(\eta) = r - 1$. 

Abusing language and notation slightly, we will refer to the quadratic differential $q$ in the proof of the previous lemma as the pushforward of $\omega$ by $\sigma$, and denote this differential $\sigma_* q$. Context will always make it clear whether $\sigma_*$ represents this pushforward, or the induced action of $\sigma$ in homology.

In terms of the flat geometry involved, the above theorem says that when we consider the quotient surface $X/\sigma$, the collection of points which are not fixed by $\sigma$ is cut in half: every non-fixed point is identified with another point of the same cone angle. The fixed points, on the other hand, have their angle halved. If $\sigma$ has a fixed point $P$ which is a zero of $\omega$ of order $r$, the cone angle at $P$ is $2\pi(r + 1)$. For the corresponding point on the quotient surface the angle is instead $\pi(r + 1)$, which means this point corresponds to a critical point of order $r - 1$ of the quadratic differential.

**Lemma 6.8.** Let $(X,\omega,\sigma)$ be an involutive surface with at least four fixed points, and suppose that $\gamma$ is a $\sigma$-invariant simple closed curve through two fixed points. Then $\gamma$ is not null-homologous.

**Proof.** Suppose $\gamma$ is null-homologous. It is a basic fact of surface topology that null-
homologous simple closed curves are necessarily separating (see [FM12, section 1.3]). That is, \( X \setminus \gamma \) would be disconnected with two connected components. We may compactify these components by adding a copy of \( \gamma \) to each as the boundary. Call these two surfaces with boundary \( X_{\pm} \), and let \( X' = X_+ \cup X_- \). Note that \( \sigma \) acts on \( X' \), and may leave the components invariant or exchange them. If \( \sigma \) has more than two fixed points, then the \( X_{\pm} \) must be preserved as at least one of \( X_{\pm} \) contains another fixed point.

Suppose that each \( X_{\pm} \) is \( \sigma \)-invariant. The boundaries \( \partial X_{\pm} \) inherit an orientation from \( X_{\pm} \), and since \( \sigma \) is orientation preserving, applying \( \sigma \) to each \( X_{\pm} \) preserves the orientation of the boundaries. However because we have \( \sigma \)-fixed points on the boundary, \( \sigma \) must reverse the orientation of the boundary. This is a contradiction.

**Lemma 6.9.** If \((X,\omega,\sigma)\) is an involutive surface and \( F_{\pm} \) are two fixed points on \( X \), then there exists an invariant simple closed curve \( \gamma \) through \( F_{\pm} \) which is not null-homologous.

**Proof.** We have already seen that this is true if there are more than two fixed points, so suppose that \( F_{\pm} \) are the only fixed points. Note that since we are considering an affine involution on a translation, the surface must have positive genus (there are no translation structures on the sphere) and the involution is not an automorphism of the surface.

Note that for a surface to admit an involution with only two fixed points, the surface must have even genus. This is easily seen by considering the Riemann-Hurwitz formula. Also by the Riemann-Hurwitz theorem, the quotient surface \( \tilde{X} = X/\sigma \) has genus \( g/2 \), and so \( \tilde{X} \) has positive genus. The fixed points \( F_{\pm} \) map to distinct points, denoted \( G_{\pm} \).

Let \( \delta \) be a non-null-homologous curve on \( \tilde{X} \) through \( G_{\pm} \). Suppose the \( \delta \) may be written as the concatenation of two curves, \( \delta = \zeta \cdot \eta \) where \( \zeta \) connects \( G_+ \) to \( G_- \), and \( \eta \) connects \( G_- \) to \( G_+ \). Lift the curves \( \zeta, \eta \) of \( \tilde{X} \) to curves \( Z \) and \( H \), respectively, of \( X \), connecting \( F_{\pm} \). We then concatenate the involutes of \( Z \) and \( H \) to obtain two invariant cycles, \( z = Z \cdot \sigma(Z) \) and \( h = H \cdot \sigma(H) \). We now claim at least one of \( z \) or \( h \) is non-null-homologous.

Letting \( \pi : X \rightarrow \tilde{X} \) denote the quotient map, we have \( \pi(z) = \zeta \) and \( \pi(h) = \eta \). If \( z \) and \( h \) were both null-homologous, then \( \pi(z + h) \) would also be null-homologous. However,
\[\pi(z + h)\] is homologous to our original curve \(\delta = \zeta \cdot \eta\), which by assumption was non-null-homologous. Hence one of \(z\) or \(h\) is non-null-homologous.

Let \(\gamma \in H_1^\sigma(X; \mathbb{Z}/2)\) be any non-zero class, and consider the associated cover \(\pi_\gamma : X_\gamma \to X\). Let \(\delta_\gamma\) denote the non-trivial deck transformation of \(X_\gamma\). Notice that the involution \(\sigma\) of \(X\) lifts to \(X_\gamma\) precisely because \(\gamma\) is \(\sigma\)-invariant.

**Lemma 6.10.** For \(\gamma \in H_1(X; \mathbb{Z}/2)\), the affine involution \(\sigma\) on \((X, \omega)\) lifts to an affine involution \(\sigma_\gamma\) of \((X_\gamma, \pi_\gamma^* \omega)\) if and only if \(\gamma \in H_1^\sigma(X; \mathbb{Z}/2)\).

*Proof.* Consider the map \(\tau : X_\gamma \to X\) given by \(\tau = \sigma \circ \pi_\gamma\). By standard topology, the map \(\tau\) will lift to a map \(\tilde{\tau} : X_\gamma \to X_\gamma\) if and only if \(\tau_\sharp(\pi_1(X_\gamma)) \subseteq \pi_\gamma^\sharp(\pi_1(X_\gamma))\), where \(\tau_\sharp\) denotes the homomorphism between fundamental groups induced by a continuous map \(f\). Note \(\pi_\gamma^\sharp\) is injective, and \(\sigma_\sharp\) is an isomorphism. As \(\tau_\sharp = \sigma_\sharp \pi_\gamma^\sharp\), we have \(\tau_\sharp(\pi_1(X_\gamma)) \subseteq \pi_\gamma^\sharp(\pi_1(X_\gamma))\) if and only if \(\sigma_\sharp(\pi_1(X_\gamma)) = \pi_1(X_\gamma)\), thinking of \(\pi_1(X_\gamma)\) as a subgroup of \(\pi_1(X)\). By the construction of a subgroup of \(\pi_1(X)\) as the kernel of the map \(\delta \mapsto i(\gamma, \delta)\) as outlined earlier, the fundamental group of \(X_\gamma\) consists precisely of those loops of \(\pi_1(X)\) which intersect \(\gamma\) an even number of times. That is,

\[\pi_1(X_\gamma) = \{\delta \in \pi_1(X) : i(\gamma, \delta) = 0\},\]

where \(i\) denotes the \(\mathbb{Z}/2\)-intersection form. We want to show that \(\sigma_\sharp \pi_1(X_\gamma) = \pi_1(X_\gamma)\) if and only if \(\gamma\) is \(\sigma\)-invariant. Note

\[\sigma_\sharp \pi_1(X_\gamma) = \{\delta \in \pi_1(X) : i(\gamma, \sigma_\sharp \delta) = 0\}.\]

Note too that \(\sigma_\sharp \pi_1(X_\gamma) = \pi_1(X_\gamma)\) if and only if for every \(\delta \in \pi_1(X)\), \(i(\gamma, \delta) = i(\gamma, \sigma_\sharp \delta)\). As the intersection form depends only on the homology class of the curves involved, so we may replace \(i(\gamma, \sigma_\sharp \delta) = 0\) with \(i(\gamma, \sigma_\delta) = 0\).

Since the intersection form is \(\sigma_\delta\)-invariant pairing (i.e., \(i(\alpha, \beta) = i(\sigma_\delta \alpha, \sigma_\delta \beta)\) for all \(\alpha, \beta \in H_1(X; \mathbb{Z}/2)\)), we have \(i(\gamma, \delta) = i(\gamma, \sigma_\delta) = 0\) if and only if \(i(\sigma_\delta \gamma, \delta) = i(\sigma_\delta \gamma, \sigma_\delta \delta) = 0\).
Putting these together, we have $i(\gamma, \delta) = 0$ if and only if $i(\sigma_* \gamma, \delta) = 0$. This means the maps $i(\gamma, \cdot)$ and $i(\sigma_* \gamma, \cdot)$ have the same kernel; and this happens if and only if $\gamma = \sigma_* \gamma$. □

If $\gamma \in H_1^q(X; \mathbb{Z}/2)$, the affine involution $\sigma : X \to X$ lifts to an affine involution of $X_\gamma$. There are in fact two such involutions, since we can take one involution and compose it with the deck-change transformation $\delta_\gamma$. In general the composition of two involutions does not necessarily produce a third involution (e.g., the composition of two reflections in the plane produces a rotation), but in our situation the involution was produced specifically to commute with the covering projection which in turn commutes with the deck change transformation. That is, if $\bar{\sigma} : X_\gamma \to X_\gamma$ denotes the involution described above, then we know

$$\pi_\gamma \circ \bar{\sigma} = \sigma \circ \pi_\gamma.$$

Letting $\delta_\gamma$ be the (non-trivial) deck transformation, we of course have $\pi_\gamma \delta_\gamma = \pi_\gamma$. Thus plugging this into the above,

$$\pi_\gamma \circ \delta_\gamma \circ \bar{\sigma} = \sigma \circ \pi_\gamma.$$

Intuitively, one of these involutions preserves decks of the cover while the other exchanges them. The construction in the above proof does not distinguish between these two involutions, so we need to make this precise.

Let $\bar{\tau}$ be the involution from the proof of Lemma 6.10, and choose a basis of $H_1(X_\gamma; \mathbb{Z})$ adapted to $\bar{\tau}$. Recall that once an adapted basis is chosen, we have a notion of an ignored fixed point. We will say that $\bar{\tau}$ is deck preserving if each of the preimages of the ignored fixed point are fixed by $\bar{\tau}$. If instead the two preimages are exchanged, we will say $\bar{\tau}$ is deck reversing. This notion of deck preserving or reversing of course depends on our choice of adapted basis.

Given the deck preserving (respectively, deck reversing) involution, we can recover the deck reversing (resp., deck preserving) involution by composing with the non-trivial deck transformation. In general, we will let $\sigma^+_\gamma$ denote the deck preserving involution, and $\sigma^-_\gamma$ the deck reversing involution. Note in Figure 6.3 we have the commutative diagram
Figure 6.3: Two conjugate involutions on the $X_\gamma$ cover.

analogous to Figure 6.2.

**Lemma 6.11.** The involution $\delta_\gamma$ is an automorphism of $(X_\gamma, \pi_\gamma^* \omega)$, while the $\sigma_\gamma^\pm$ are affine involutions.

**Proof.** Note that $\delta_\gamma$ is a conformal mapping of $X_\gamma$, and $\pi_\gamma \circ \delta_\gamma = \pi_\gamma$. Consider the pullback of $\pi_\gamma^* \omega$ by $\delta_\gamma$:

$$\delta_\gamma^* \pi_\gamma^* \omega = (\pi_\gamma \circ \delta_\gamma)^* \omega = \pi_\gamma^* \omega.$$ 

Thus $\delta_\gamma \in \text{Aut}(X, \omega)$.

Note the $\sigma_\gamma^\pm$ are also conformal maps. Consider the pullbacks:

$$(\sigma_\gamma^\pm)^* \pi_\gamma^* \omega = (\pi_\gamma \circ \sigma_\gamma^\pm)^* \omega = (\sigma \circ \pi_\gamma)^* \omega = \pi_\gamma^* \sigma^* \omega = \pi_\gamma^* (-\omega) = - \pi_\gamma^* \omega.$$ 

Thus the $\sigma_\gamma^\pm$ are affine involutions. \qed

**Lemma 6.12.** For each $\gamma \in H_\sigma^1(X; \mathbb{Z}/2)$, the sets of fixed points of $\sigma_\gamma^\pm$, denoted $\text{Fix}(\sigma_\gamma^\pm)$, partition the preimages $\pi^-_\gamma(\text{Fix}(\sigma))$. Furthermore, $\delta_\gamma(\text{Fix}(\sigma_\gamma^\pm)) = \text{Fix}(\sigma_\gamma^\mp)$.

**Proof.** If a point $p \in X_\gamma$ is fixed by $\sigma_\gamma^\pm$, then $p$ must be the preimage of a fixed point of $\sigma$. 112
This follows from the commutativity of Figure 6.3:

\[
\begin{align*}
\sigma^\pm_\gamma(p) &= p \\
\Rightarrow \pi_\gamma \sigma^\pm_\gamma(p) &= \pi_\gamma(p) \\
\Rightarrow \sigma \pi_\gamma(p) &= \pi_\gamma(p).
\end{align*}
\]

As \(\sigma^\pm_\gamma = \delta_\gamma \sigma^\mp_\gamma\) and \(\delta_\gamma\) has no fixed points, \(\sigma^\pm_\gamma\) and \(\sigma^\mp_\gamma\) can not share any fixed points. Furthermore, the fixed point set of one involution is obtained by applying \(\delta_\gamma\) to the fixed point set of the other involution.

Recall that if a canonical basis of \(H_1(X;\mathbb{Z})\) adapted to \(\sigma\) is given, we may write \(H_1(X;\mathbb{Z}) = \mathcal{E}^+ \oplus \mathcal{E}^- \oplus \mathcal{N}\) where \(\mathcal{N}\) is generated by classes \(\nu_1, ..., \nu_n\) with \(\sigma_\ast \nu_i = -\nu_i\). By construction of the \(\nu_i\) (rather, construction of the curves \(c_1, ..., c_{k/2-1}\) and \(d_1, ..., d_{k/2-1}\) in [Gil73, Prop. 1]), each \(\nu_i\) is represented by a simple closed curve passing through precisely two fixed points. We may order the \(\nu_i\) so that \(\nu_1\) and \(\nu_2\) share a single fixed point, \(\nu_2\) and \(\nu_3\) share a single fixed point, and so on. Labeling the fixed points \(F_1, F_2, ..., F_{n+2}\) we have that \(\nu_i\) passes through the fixed points \(F_i\) and \(F_{i+1}\). We can now describe the fixed points of \(\sigma^\pm_\gamma\) by considering which elements of \(\mathcal{N}\) appear in the homology class of a given curve. The remarks above prove the following lemma.

**Lemma 6.13.** If \(\gamma\) is an invariant simple closed curve on an involutive surface \((X,\omega,\sigma)\) representing an invariant class \([\gamma] \in H_1^\sigma(X;\mathbb{Z}/2) = \mathcal{I}_2 \oplus \mathcal{N}_2\) and if \([\gamma] = \iota + \nu_j\) for \(\iota \in \mathcal{I}_2\) and \(\nu_j\) a generator of \(\mathcal{N}\) as described above, then \(\gamma\) passes through the fixed points \(F_j\) and \(F_{j+1}\), and these are the only two fixed points on \(\gamma\).

The fixed points which live on an invariant curve \(\gamma\) determine which points of the cover \(X_\gamma\) are fixed by \(\sigma^\pm_\gamma\).

**Lemma 6.14.** If \(\gamma\) is an invariant curve which appears in homology as \(\iota + \nu_j\) for some \(\iota \in \mathcal{I}_2\) and some generator \(\nu_j\) of \(\mathcal{N}_2\), then the preimages of the points \(F_j\) and \(F_{j+1}\) are the
fixed points of $\sigma^-_\gamma$ on $X_\gamma$.

**Proof.** We prove the equivalent statement that every preimage of the $\sigma$-fixed points, except for $F_j$ and $F_{j+1}$, is fixed by $\sigma^+_\gamma$. Let $F$ be any fixed point other than $F_j$ or $F_{j+1}$, and let $R$ be the ignored fixed point. Let $\varepsilon$ be a curve from $R$ to $F$ which does not cross $\gamma$. Then the lifts of this curve to $X_\gamma$ are two curves from a preimage of $R$ to a preimage of $F$. As $R$ is fixed by $\sigma^+_\gamma$ and $\varepsilon$ does not cross the slit, $F$ is also fixed. Thus every preimage of a fixed point other than $F_j$ and $F_{j+1}$ is fixed by $\sigma^+_\gamma$.

To see that the preimages of $F_j$ and $F_{j+1}$ are not fixed by $\sigma^+_\gamma$, and so must be fixed by $\sigma^-_\gamma$, note that if $\sigma^+_\gamma$ fixed $F_j$ and $F_{j+1}$, then $\sigma^-_\gamma$ would be fixed-point-free, and so would be a translation. By Lemma 6.11, $\sigma^-_\gamma$ is not a translation, and so $F_j$, $F_{j+1}$ must be fixed by $\sigma^-_\gamma$. \(\square\)

We turn our attention to the fixed points of $\sigma^\pm_\gamma$ when $\gamma$ is a slightly more complicated curve, but first make one very simple observation.

**Lemma 6.15.** The number of fixed points of $\sigma^\pm_\gamma$ is a multiple of 4.

**Proof.** If $\gamma \in H^q_\eta(X;\mathbb{Z}/2)$, then $\sigma = \iota + \nu$ for some $\iota \in \mathcal{I}_2$ and $\nu \in \mathcal{N}_2$. The fixed points of $\sigma^\pm_\gamma$ are determined by the fixed points on a representative of $\nu$ (recall no invariant representative of $\iota \in \mathcal{I}$ may pass through any fixed points).

Suppose $F$ is a fixed point of $\sigma$ and the cover $X_\gamma$ is obtained by slitting curves which pass through $F$. If an even number number of slits are made through $F$, then $F$ will be fixed by $\sigma^+_\gamma$. To see this, note $\sigma^+_\gamma$ applies $\sigma$ to each deck in the cover, where we determine if a point remains in the same deck or not by considering whether the preimages of the ignored point are fixed or not. If an even number of slits are performed, the identifications from the slits force the point to map to itself. Equivalently, if an even number of slits through $F$ are made, preimages of $F$ are not fixed by $\sigma^-_\gamma$. Similarly, if an odd number of slits are made through $F$, then the preimages of $F$ will be fixed by $\sigma^-_\gamma$, but not $\sigma^+_\gamma$.

As we consider homology classes with $\mathbb{Z}/2$ coefficients, we may think of an even number of slits as being the same as no slits. This means every element of $\mathcal{N}_2$ (and hence
every element of $H^1_\sigma(X; \mathbb{Z}/2)$, as the elements of $\mathcal{I}_2$ pass through no fixed points) is associated with an even number of fixed points, and on the cover this number of fixed points is doubled.

Lemma 6.16. Let $\gamma$ be an invariant simple closed curve on an involutive surface $(X, \omega, \sigma)$ representing an invariant class $[\gamma] \in H^1_\sigma(X; \mathbb{Z}/2) = \mathcal{I}_2 \oplus \mathcal{N}_2$, and suppose $[\gamma] = \iota + \nu_j + \nu_{j+1}$ for $\iota \in \mathcal{I}$, and $\nu_j, \nu_{j+1}$ generators of $\mathcal{N}_2$ as described above. The fixed points of $\sigma^{-\gamma}$ are the preimages of $F_j$ and $F_{j+2}$.

Proof. Proceeding as in the proof of the previous lemma, we can show that the preimages of $F_1, ..., F_{j-1}, F_{j+3}, ..., F_{n+2}$ are $\sigma^{-\gamma}$-fixed. This leaves six points unaccounted for: the two preimages of each fixed point $F_j, F_{j+1},$ and $F_{j+2}$. As $\sigma^{-\gamma}$ fixes $4m$ points, exactly one pair of preimages gives $\sigma^{-\gamma}_+$-fixed points, and the other two points are fixed by $\sigma^{-\gamma}$.

Suppose that $F_{j+1}$ is a zero of order $2k$ of the quadratic differential of $X$. If $F_{j+1}$ was fixed by $\sigma^{-\gamma}$, then the quadratic differential of the quotient surface $X_\gamma / \sigma^{-\gamma}$ would have a zero of order $k - 1$ at the corresponding point. It is easiest to show that this is not the case by using the flat geometry of the surfaces involved.

The curves $\nu_j$ and $\nu_{j+1}$ intersect at $F_{j+1}$ as indicated in Figure 6.4. Performing the slit construction, we then see the curves of Figure 6.5 on the cover $X_\gamma$.

The darker sectors form a neighborhood of a point on the cover, and the lighter sectors form a disjoint neighborhood. Both of these neighborhoods are around points whose cone angle is the same as the angle of $F_{j+1}$ on the original surface. (Though for simplicity
we draw the picture as if this angle were $2\pi$.)

Notice $\sigma^-\gamma$ takes dark sectors on the left-hand side of the diagram to lighter sectors on the right-hand side, and vice versa. Thus in the quotient $X_{\gamma}/\sigma^-\gamma$ no two points in the darker sectors are identified with points in the darker sectors, and likewise for the lighter sectors. This means the corresponding point on the quotient has the same cone angle as the original point. Hence the lifts of $F_{j+1}$ are not fixed by $\sigma^-\gamma$.

With one last minor lemma, we will have all of the information we need to completely determine fixed points of the involutions $\sigma^\pm\gamma$.

Lemma 6.17. Let $\gamma = \iota + \nu_j + \nu_k \in H^1(X; \mathbb{Z}/2)$ with $\iota \in \mathcal{I}_2$ and $\nu_j, \nu_k \in \mathcal{N}$ distinct generators. The fixed points of $\sigma^-\gamma$ are $F_j, F_{j+1}, F_k, F_{k+1}$.

Proof. Proceeding as in the last proof, we look to see if the corresponding points of $X_{\gamma}/\sigma^-\gamma$ have half the cone angle as the original points or not. For each $\sigma$-fixed point on $\gamma$, the slit construction produces neighborhoods like those in Figure 6.6 on $X_{\gamma}$.

Applying $\sigma^-\gamma$ takes the dark region on the left to the dark region on the right, and similarly for the light regions. This means the cone angle of the corresponding points on the quotient $X_{\gamma}/\sigma^-\gamma$ are cut in half, and the points $F_j, F_{j+1}, F_k, F_{k+1}$ are fixed by $\sigma^-\gamma$. \qed
Combining the lemmas above, we have a complete classification of $\sigma_\gamma^\pm$, based on which fixed points live on the cutting curve $\gamma$.

**Lemma 6.18.** A point of $X_\gamma$ is fixed by $\sigma^-_\gamma$ if and only if it is the preimage of a fixed point on the cutting curve $\gamma$, and is not the intersection of two distinct $\nu_j$ and $\nu_k$ in $N_2$.

We can represent this algebraically as follows. Consider a homomorphism $c : H^1_\sigma(X; \mathbb{Z}/2) \to (\mathbb{Z}/2)^{n+1}$ given by mapping $c(\nu_j)$ to the tuple whose $j$-th and $(j+1)$-st entries are 1’s, all the other entries are 0’s. The map is then extended linearly. The number of fixed points of $\sigma^-_\gamma$ corresponds to twice the number 1’s in $c(\gamma)$; the number of $\sigma^+_\gamma$-fixed points is twice the number of 0’s of $c(\gamma)$. We will denote these quantities $f_+(\gamma)$ and $f_-(\gamma)$, respectively, and note $f_-(\gamma) = 2n + 4 - f_+(\gamma)$.

### 6.3 Quotients of the $X_\gamma$ surfaces

As noted in the previous section, if $(X, \omega, \sigma)$ is an involutive surface and $\gamma \in H^0_\sigma(X; \mathbb{Z}/2)$, then the cover $(X_\gamma, \pi^*_\gamma \omega)$ comes with three involutions: the deck change $\delta_\gamma$, and the deck preserving and reversing involutions, $\sigma^\pm_\gamma$. We now seek to understand these involutions and the corresponding quotient surfaces.

In order to simplify notation in what is to follow, let $(X, \omega, \sigma)$ be an involutive surface of genus $g$; suppose $\sigma$ has $n + 2 = 2j$ fixed points; suppose the quotient $X/\sigma$ has genus $\bar{g} = \frac{1}{2}(g - j + 1)$; for a fixed $\gamma \in H^0_1(X; \mathbb{Z}/2)$, suppose $X_\gamma$ has genus $\tilde{g} = 2g - 1$; $\sigma^-_\gamma$ has $2k$ fixed points, and $\sigma^+_\gamma$ has $4j - 2k = 2(2j - k)$ fixed points.

All lemmas here are stated with the above conventions.

**Lemma 6.19.** The quotient surface $\tilde{X}_\gamma^- := X_\gamma/\sigma^-_\gamma$ has genus

$$\tilde{g}_- = \frac{1}{2}(\bar{g} - k + 1) = \frac{1}{2}(2g - k).$$
The quotient surface \( \hat{X}_\gamma^+ = X_\gamma / \sigma_\gamma^+ \) has genus

\[
\hat{g}_+ = \frac{1}{2}(\hat{g} - (2j - k) + 1) = \frac{1}{2}(2g - 2j + k)
\]

**Proof.** Apply Lemma 6.6. \( \square \)

**Corollary 6.20.** Using the notation above,

\[
\tilde{g} + 2\tilde{g} = g + \hat{g}_+ + \hat{g}_-. 
\]

Note that each quotient \( \hat{X}_\gamma^\pm \) is itself an involutive surface, and each of the surfaces \( X, \hat{X}_\gamma^\pm \) all share the same quotient, \( \bar{X} \).

**Lemma 6.21.** Each of the quotient surfaces \( \hat{X}_\gamma^\pm = X_\gamma / \sigma_\gamma^\pm \) is itself an involutive surface.

**Proof.** We already know, by Lemma 6.7, that each of these surfaces is a half-translation surface. The deck transformation \( \delta_\gamma \) then descends to each quotient surface. To see this suppose \( P \in X_\gamma^\pm \) and let \( \tilde{P}_{1,2} \in X_\gamma \) be the two preimages of \( P \) in \( X_\gamma^\pm \). Note these points are exchanged by \( \sigma_\gamma^\pm \). (There might be only one such preimage point; in this case set \( \tilde{P}_1 = \tilde{P}_2 \).)

To show \( \delta \) descends to a well-defined map on \( \hat{X}_\gamma^\pm \), we simply need to show that \( \delta_\gamma(\tilde{P}_{1,2}) \) are exchanged by \( \sigma_\gamma^\pm \), but this follows from the commutativity of \( \sigma_\gamma^\pm \) and \( \delta_\gamma \):

\[
\sigma_\gamma^\pm \delta_\gamma(\tilde{P}_1) = \delta_\gamma(\sigma_\gamma^\pm(\tilde{P}_1)) = \delta_\gamma(\tilde{P}_2).
\]

Now we need to check that this map, which we will denote \( \tau_\gamma^\pm : \hat{X}_\gamma^\pm \to \hat{X}_\gamma^\pm \), is affine (in fact, it will be an automorphism). Let \( q \) denote the pushforward half-translation structure on \( X_\gamma^\pm \), so \( (\pi_\gamma^\pm)^* q = \pi_\gamma^\pm^* \omega \). Note that in Figure 6.7 we have a commutative diagram on the spaces of quadratic differentials on the surfaces involved. Here \( \delta_\gamma^* = \text{id} \), and by commutativity this forces \( (\tau_\gamma^\pm)^* = \text{id} \). \( \square \)

**Lemma 6.22.** The quotients \( X / \sigma \) and \( \hat{X}_\gamma^\pm / \tau_\gamma^\pm \) are isomorphic.
Proof. We first show there exists a biholomorphism between the quotients and then show this biholomorphism respects the quadratic differentials involved.

Let $P \in \hat{X}_\gamma^\pm/\tau_\gamma^\pm$, and suppose the preimages of $P$ are $P_1, P_{-1} \in \hat{X}_\gamma^\pm$. (We allow $P_1 = P_{-1}$ if $P$ is a ramification point.) Now consider the lifts of $P_1, P_{-1}$ to $X_\gamma$, which we will denote $P_{1,1}, P_{1,-1}, P_{-1,1}, P_{-1,-1}$. We list how all of these points are related to one another:

$$
\begin{align*}
\tau_\gamma^\pm(P_1) &= P_{-1}, \\
\tau_\gamma^\pm(P_{-1}) &= P_1, \\
\sigma_\gamma^\pm(P_{1,1}) &= P_{1,-1}, \\
\sigma_\gamma^\pm(P_{1,-1}) &= P_{1,1}, \\
\sigma_\gamma^\pm(P_{-1,1}) &= P_{-1,-1}, \text{ and} \\
\sigma_\gamma^\pm(P_{-1,-1}) &= P_{-1,1}.
\end{align*}
$$

Let $Q_{x,y} = \pi_\gamma(P_{x,y})$. Because of the commutativity $\pi_\gamma \sigma_\gamma^\pm = \sigma \pi_\gamma$, we have that $\pi_\gamma(P_{x,\pm 1}) = Q_{x,\mp 1}$. Further, since $\tau_\gamma^\pm$ was defined in terms of the deck transformation $\delta_\gamma$ on $X_\gamma$, $\pi_\gamma(P_{\pm 1,y}) = Q_{\pm 1,y}$. Finally, taking the quotient from $X$ to $X/\sigma$ we identify the points $Q_{\pm 1,\pm 1}$ and have a single point of $X/\sigma$. Thus we have a well-defined map from $\hat{X}_\gamma^\pm/\tau_\gamma^\pm$ to $X/\sigma$. Since this same process could be repeated by starting with $X/\sigma$, we can construct
an inverse map. Thus the map described above is bijective, and since all maps involved are holomorphic, we have a biholomorphism.

Finally, because all of the maps involved preserve the quadratic differentials in question under pullback and pushforward, our given map is an isomorphism of half-translation surfaces.

**Theorem 6.23.** If \((X, \omega, \sigma)\) is an involutive surface of genus \(g\) with \(n + 2\) points fixed by \(\sigma\), then there are \(2^{g-n/2} \cdot \binom{n+1}{2(g-\hat{g})}\) classes in \(H^1(X; \mathbb{Z}/2)\) so that the quotient \(\hat{X}_\gamma\) has genus \(\hat{g}\).

**Proof.** Note the genus of the quotient surface \(X_\gamma\) is determined completely by the genus of \(X\) and the number of fixed points of \(\sigma_\gamma\). If \(X\) has genus \(g\) and \(X_\gamma\) is an unbranched cover, then \(X_\gamma\) must have genus \(\tilde{g} = 2g - 1\). The number of fixed points of \(\sigma_\gamma\) is determined by the class \(\gamma = \iota + \nu \in H^1(X; \mathbb{Z}/2)\) where \(\iota \in \mathcal{I}_2\) and \(\nu \in \mathcal{N}_2\). More precisely, the number of fixed points is completely determined by the class \(\nu\).

To prove the theorem we first count the number of classes \(\gamma \in \mathcal{N}_2\) which give the quotient \(X_\gamma\) of genus \(\hat{g}\), and then multiply this value by \(2^m = 2^{g-n/2}\) (where \(n\) is the rank of \(\mathcal{N}_2\), and \(m\) is the rank of \(\mathcal{I}_2\)) to determine the number of classes in \(H^1(X; \mathbb{Z}/2)\) which give a quotient of the desired genus.

Suppose \(\gamma \in \mathcal{N}_2\). Suppose \(\sigma_\gamma\) has \(f_-(\gamma)\) fixed points. As this must be a multiple of 4, write \(f_-(\gamma) = 4k\). The quotient surface \(\hat{X}_\gamma\) has genus \(\tilde{g} = \frac{1}{2}(\tilde{g} - 2k + 1) = g - k\). The number of fixed points of \(\sigma_\gamma\), however, is twice the number of 1’s that appear in the binary representation of \(\gamma\) which was described earlier. The number of 1’s appearing in the binary representation of \(\gamma\) is \(2k = 2(g - \tilde{g})\). Thus to count the number of classes \(\gamma \in H^1(X; \mathbb{Z}/2)\) giving a surface \(\hat{X}_\gamma\) of genus \(\tilde{g}\), we should count the number of binary representations with \(2k = 2(g - \tilde{g})\) 1’s. As there are \(n + 1\) bits in the string representation of \(\gamma\), there are \(\binom{n+1}{2g-2\tilde{g}}\) strings.
Now taking the $2^{g-n/2}$ elements of $I_2$ into consideration, we have

$$2^{g-n/2} \cdot \binom{n+1}{2g-2\hat{g}}$$

elements of $H_1^\sigma(X;\mathbb{Z}/2)$ which give quotient surfaces $X_\gamma^+$ of genus $\hat{g}$.

**Corollary 6.24.** If $(X,\omega,\sigma)$ is an involutive surface of genus $g$ with $n+2 = 2j$ points fixed by $\sigma$, then there are $2^{g-n/2} \cdot \binom{n+1}{2(\hat{g}+j-g)}$ classes in $H_1^\sigma(X;\mathbb{Z}/2)$ so that the quotient $\hat{X}_\gamma^+$ has genus $\hat{g}$.

**Proof.** By Corollary 6.20, if $\hat{X}_\gamma^+$ has genus $\hat{g}$, then $\hat{X}_\gamma^−$ has genus $2g-j-\hat{g}$. So to count the number of ways $\hat{X}_\gamma^+$ has genus $\hat{g}$, we count the number of ways $\hat{X}_\gamma^−$ has genus $2g-j-\hat{g}$. By Theorem 6.23, there are

$$2^{g-n/2} \cdot \binom{n+1}{2g-2(2g-j-\hat{g})} = 2^{g-n/2} \cdot \binom{n+1}{2(\hat{g}+j-g)}$$

classes in $H_1^\sigma(X;\mathbb{Z}/2)$ giving $X_\gamma^+$ of genus $\hat{g}$. \qed

Adding the two results together we would have the total number of surfaces classes which produce quotients $\hat{X}_\gamma^\pm$ of a given genus $\hat{g}$, provided there was no double counting.

**Lemma 6.25.** There does not exist a class $\gamma \in H_1^\sigma(X;\mathbb{Z}/2)$ such that $X_\gamma^+$ and $X_\gamma^−$ have the same genus.

**Proof.** By the proofs of Theorem 6.23 and Corollary 6.24, $X_\gamma^+$ and $X_\gamma^−$ will have genus $\hat{g}$ if there are $2(\hat{g}+j-g)$ 1’s and $2(g-\hat{g})$ 0’s in the binary representation of $\gamma$. However, this binary representation consists of $n+1$ bits with $n$ an even integer: we would require an even number of 1’s and an even number 0’s in a binary string with an odd number of bits. \qed

**Lemma 6.26.** If $(X,\omega,\sigma)$ is an involutive surface and if $\sigma$ has $n+2 = 2j$ fixed points, then there are

$$2^{g-n/2} \cdot \left( \binom{n+1}{2g-2\hat{g}} + \binom{n+1}{2(\hat{g}+j-g)} \right)$$

121
quotient surfaces $X^\pm_{\tilde{g}}$ of genus $\tilde{g}$.

Finally, note that in the special case where $\sigma$ is a hyperelliptic involution, $n = 2g$. Hence the number of quotient surfaces of genus $\tilde{g}$ is

$$\binom{2g + 1}{2g - 2\tilde{g}} + \binom{2g + 1}{2\tilde{g} + 2}.$$ 

If we special are concerned with counting those quotient surfaces which are tori, so $\tilde{g} = 1$, this becomes

$$\binom{2g + 1}{2g - 2} + \binom{2g + 1}{4} = \binom{2g + 1}{3} + \binom{2g + 1}{4} = \binom{2g + 2}{4}.$$ 

In the appendices we explicitly calculate a few examples.

The results of this chapter show that if a translation surface admits an affine involution whose quotient is a sphere or a torus, then there is a covering construction which associates tori to this surface. These tori will typically be half-translation surfaces, and so this allows for the possibility of applying the Panov planes (universal covers of the half-translation tori).
Appendices
Appendix A

Examples

In this appendix we present examples of the covers and quotients described in Chapter 6. We give examples in each stratum of genus two translation surfaces where the involution considered is hyperelliptic. The first example of a hyperelliptic surface we will present in detail, but will omit the details in the second example as they are essentially the same as the first example.

A.1 Quotients of the Swiss cross

The Swiss cross is a standard example of a genus two translation surface which is contained in the stratum $\mathcal{H}(2)$; that is, the surface contains one cone point with cone angle $6\pi$ corresponding to a zero of order 2 of the abelian differential. A polygonal representation of the Swiss cross is shown in Figure A.1. In polygonal representation of Figure A.1, the hyperelliptic involution acts by $180^\circ$ rotation around the midpoint of the cross, and the six Weierstrass points (fixed points of the hyperelliptic involution) are marked.

As the hyperelliptic involution acts on the first homology group by negation, all double covers of the Swiss cross admit an involution conjugate to the hyperelliptic involution on the Swiss cross. That is, $H_1^\sigma(X;\mathbb{Z}/2) = H_1(X;\mathbb{Z})$. There are $2^{2g} = 2^4 = 16$ double covers, 15 of which are connected. (The trivial disconnected double cover corresponds to
the zero class in $H_1(X;\mathbb{Z}/2)$. By the remarks at the end of Chapter 6, there are

$$\binom{2g + 2}{4} = \binom{6}{4} = 15$$

tori associated to the Swiss cross by the construction of Chapter 6.

In Table A.1 below we list all of the connected double covers $X_\gamma$ and all of the
quotient surfaces $\hat{X}_\gamma^\pm$. In each case one of $\hat{X}_\gamma^\pm$ is a torus and the other surface is a sphere.
It is easy to determine which surfaces are spheres or tori by counting the number of 1’s
that occur in the binary representation of the homology class. To obtain such a binary
representation we first must choose a canonical basis adapted to the involution. In this
particular case the involution is hyperelliptic and so we may use the usual canonical basis.
In terms of the polygonal representation of the Swiss cross, the simple closed curves giving
the basis $\{a_1, b_1, a_2, b_2\}$ appear as geodesics on the boundary of the polygon, as shown in
Figure A.2.
Letting

\[ \nu_1 = b_2 \]
\[ \nu_2 = a_1 \]
\[ \nu_3 = b_1 \]
\[ \nu_4 = a_2 \]

We have that \( \nu_i \) passes through the fixed points \( F_i \) and \( F_{i+1} \) in Figure A.1. The ignored fixed point with respect to this basis is \( F_6 \), the center point of the cross. We may represent homology classes \( \gamma \in H_1^*(X; \mathbb{Z}/2) \) as binary strings with five bits where the \( i \)-th bit corresponds to the fixed point \( F_i \). If the \( i \)-th bit is a 1, then on the cover \( X_\gamma \) the preimages of \( F_i \) are fixed points of \( \sigma_\gamma^- \); if the \( i \)-th bit is 0, then the preimages of \( F_i \) are instead fixed by \( \sigma_\gamma^+ \). The preimages of the ignored fixed point, \( F_6 \), are always fixed by \( \sigma_\gamma^+ \) and never fixed by \( \sigma_\gamma^- \) for this particular choice of basis.

In Table A.1 below we list all of the connected double covers \( X_\gamma \) of the Swiss cross, as well as the quotients \( \hat{X}_\gamma^\pm \) of the cover. In the left-hand column we give the quotient \( \hat{X}_\gamma^+ \); the middle column contains the cover; and the right-hand column contains the quotient \( \hat{X}_\gamma^- \).
The binary string representing the homology class $\gamma$ is listed beneath the cover. When the binary representation has two 1’s, the surface $\hat{X}_\gamma^+$ is a sphere and $\hat{X}_\gamma^-$ is a torus. Otherwise $\hat{X}_\gamma^+$ is a torus and $\hat{X}_\gamma^-$ is a sphere.
Table A.1: The covers and quotients of the Swiss cross with an invariant involution.

A.2 Quotients of the decagon surface

In the last section we listed the covers and quotients of a genus two surface in the stratum $\mathcal{H}(2)$. Now we list covers and quotients for a surface in the other stratum of genus two surfaces, $\mathcal{H}(1, 1)$. The involution we consider is again the hyperelliptic involution, acting on a polygonal representation of the surface by $180^\circ$-rotation. The surface we will consider
is a regular decagon with opposite sides identified. See Figure A.3.

Notice that this surface has two cone points, each of angle $4\pi$, which alternate as the vertices of the decagon are traversed. In this case the cone points are not fixed, but are exchanged by the involution. The six fixed points of the involution are the midpoints of edges, and the center point.

Figure A.3: The decagon with opposite sides identified is a surface in $\mathcal{H}(1,1)$.

Let $\nu_i$ denote the curve consisting of the two edges which contain the points $F_i$ and $F_{i+1}$. The collection of connected covers and associated quotient surfaces appears in Table A.2.
Table A.2: The covers and quotients of the regular decagon.
Bibliography


