The Dissertation Committee for Aaron Joshua Fenyes
certifies that this is the approved version of the following dissertation:

Warping geometric structures and abelianizing $\mathrm{SL}_2\,\mathbb{R}$ local systems

Committee:

Andrew Neitzke, Supervisor

David Ben-Zvi

Jacques Distler

Daniel Freed

Sean Keel

Warping geometric structures and abelianizing $\mathrm{SL}_2\,\mathbb{R}$ local systems

by

Aaron Joshua Fenyes, B.S.; M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2016

To: fyfraternize@googlegroups.com

(To all the friends I made along the way)

Acknowledgments

This dissertation would not have been started or finished without the tireless guidance of Andy Neitzke. I'm very grateful to Jen Berg for pointing out the argument used to prove Proposition 8.2.A, and to Sona Akopian for pointing out a key step in the proof of Theorem 9.3.K (I am, of course, responsible for any mistakes in these arguments). I've also enjoyed the benefit of conversations, some short and some long, with Jorge Acosta, David Ben-Zvi, Francis Bonahon, Luis Duque, Richard Hughes, Tim Magee, Tom Mainiero, Taylor McAdam, Amir Mohammadi, Javier Morales, and Max Riestenberg, as well as diction brainstorming with Eliana Fenyes. This research was supported in part by NSF grants 1148490 and 1160461.

Warping geometric structures and abelianizing $\mathrm{SL}_2\,\mathbb{R}$ local systems

Publication No. _____

Aaron Joshua Fenyes, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Andrew Neitzke

The *abelianization* process of Gaiotto, Hollands, Moore, and Neitzke parameterizes $\mathrm{SL}_K\,\mathbb{C}$ local systems on a punctured surface by turning them into $\mathbb{C}^\times$ local systems, which have a much simpler moduli space. When applied to an $\mathrm{SL}_2\,\mathbb{R}$ local system describing a hyperbolic structure, abelianization produces an $\mathbb{R}^\times$ local system whose holonomies encode the shear parameters of the hyperbolic structure.

This dissertation extends abelianization to $\mathrm{SL}_2\,\mathbb{R}$ local systems on a compact surface, using tools from dynamics to overcome the technical challenges that arise in the compact setting. Thurston's shear parameterization of hyperbolic structures, which has its own technical subtleties on a compact surface, once again emerges as a special case.

Table of Contents

Part I

An invitation to abelianization

Chapter 1

Introduction

1.1 Context

The space of $\mathrm{SL}_2\,\mathbb{C}$ local systems on a surface is complicated, and its geometry is an active topic of research. The space of $\mathbb{C}^\times$ local systems on a surface is simple, because a local system is determined up to isomorphism by its holonomies, and the holonomies of $\mathbb{C}^\times$ local system are just numbers. The correspondence between local systems and their holonomies identifies the space of $\mathbb{C}^\times$ local systems on a surface Σ with the space of homomorphisms $H_1(\Sigma;\mathbb{Z}) \to \mathbb{C}^\times$, which is a complex manifold isomorphic to $(\mathbb{C}^\times)^{\operatorname{rank} H_1(\Sigma;\mathbb{Z})}$.

The *abelianization* process of Gaiotto, Hollands, Moore, and Neitzke is, roughly speaking, a machine for turning $\mathrm{SL}_2\,\mathbb{C}$ local systems on a punctured surface S' into $\mathbb{C}^\times$ local systems on a branched double cover of S'. In good conditions, the process can be inverted, giving a densely defined $\mathbb{C}^\times$-valued coordinate chart on the space of $\mathrm{SL}_2\,\mathbb{C}$ local systems. When applied to local systems describing hyperbolic structures on S', abelianization unifies and extends two classical coordinate systems on the space of hyperbolic structures: W. Thurston's shear parameters and Fenchel-Nielsen coordinates [1].

1.2 Goals

This paper extends the abelianization process to a certain class of local systems on compact surfaces, confirming a conjecture of Gaiotto, Moore, and Neitzke. The conjecture is discussed further in Section 6.4. You can find a detailed statement of the main result in Section 7.1.

1.3 Invitation

To state and prove the main result, we'll need to bring in a bunch of special equipment. Though it might look arcane when it's all dumped out in a pile, much of this machinery arises naturally in the study of geometric structures on manifolds. I've therefore split the paper into two parts. The first part, which you're currently reading, is a stroll through the theory of geometric structures, meant to introduce and motivate some of the tools we'll be using later. The second part lays out all the equipment in full detail and uses it to prove the main result.

Chapter 2

Geometric structures as local systems

2.1 A general framework

2.1.1 Global to local, symmetries to sheaves

We often define symmetries as the transformations that preserve a certain geometric structure. Isometries preserve distances; conformal maps preserve angles; diffeomorphisms preserve the smoothness of functions; affine maps preserve the straightness of lines. Klein famously proposed to flip this idea on its head, defining a geometric structure as whatever is preserved by a certain group of symmetries [2]. Klein's program has been carried on by Ehresmann [3], Cartan [4], Haefliger [5], Thurston [6], Goldman [7], and many others.

Klein's vision of geometry is very broad. Thurston suggested narrowing it down by focusing on fairly rigid symmetry groups—ones where each element is determined by its restriction to any open set [6, §3.5]. This approach lends itself to a description of geometric structures using locally constant sheaves.

Pick a topological manifold X and a symmetry group $G \circlearrowright X$ which is rigid in the sense above. On any topological manifold M, we get a sheaf $\mathcal{O}(M, X)$ of local homeomorphisms from M to X. The symmetry group acts

on maps to X by postcomposition, making $\mathcal{O}(M, X)$ into a sheaf of G-sets.

There's extra structure to be found when M is X itself. The sheaf $\mathcal{O}(X, X)$ has a special subsheaf $\mathcal{G}$ consisting of maps which are locally the restrictions of elements of G. Because an element of G is determined by its restriction to any open set, $\mathcal{G}$ is locally constant, and its stalks are all G-torsors. The subsheaf $\mathcal{G}$ is therefore a G local system in the sense of Section 8.2. If we dropped the conditions on G and X that make $\mathcal{G}$ locally constant, we'd be dealing with a potentially much wilder kind of sheaf, called a *pseudogroup* [4][6, Chapter 3].

2.1.2 Geometric structures on manifolds

Klein would say the action of G on X defines a geometric structure. The local system $\mathcal{G}$ captures the local features of that structure: the features of open subsets of X that are preserved by the action of G. We can imagine reproducing the local features of $G \circlearrowright X$ on another topological manifold M by choosing a subsheaf of $\mathcal{O}(M, X)$ analogous to $\mathcal{G} \hookrightarrow \mathcal{O}(X, X)$. A subsheaf like that is called a (G, X) *structure* on M. It comes in two parts:

- A G local system $\mathcal{E}$ on M, called the *catalog*.
- A natural inclusion $\text{⫪}: \mathcal{E} \hookrightarrow \mathcal{O}(M, X)$, called the *anchor map*.

By a natural inclusion, I mean a natural transformation of sheaves of G-sets which is injective on every open set.

A (G, X) structure picks out a special class of local homeomorphisms from M to X, which we'll call *local charts*. The catalog tells you how the local charts are organized, keeping track of how they fit together and how they're shuffled around by the action of G. The anchor map links each catalog entry to an actual map from M to X. When we say that the anchor map is a natural transformation, we're saying that the catalog accurately reflects the organization of the maps it indexes.

The traditional way to specify a (G, X) structure on M is to give an atlas of charts from M to X whose transition maps come from $\mathcal{G}$ [6, Chapter 3]. That description is equivalent to the one we're using.

Proposition 2.1.A. *A catalog and anchor map contain the same information as an atlas of charts from M to X whose transition maps come from $\mathcal{G}$.*

Proof. It's not too hard to build a catalog and anchor map from an atlas of charts on M. Let $\mathcal{E}$ be the subsheaf of $\mathcal{O}(M, X)$ consisting of maps that locally restrict to charts composed with elements of G. The kind of reasoning we used earlier to show that $\mathcal{G} \hookrightarrow \mathcal{O}(M, X)$ is a G local system can also be used to show that $\mathcal{E}$ is a G local system. The catalog $\mathcal{E}$ was constructed as a subsheaf of $\mathcal{O}(M, X)$, so it comes already equipped with an anchor map.

Extracting an atlas from a catalog $\mathcal{E}$ and anchor map $\pitchfork$ takes a little more care. To construct a chart around a point $x \in M$, first recall that $\mathcal{E}$ is locally constant and M is locally connected, so x has an connected open neighborhood U on which $\mathcal{E}$ is isomorphic to a constant sheaf. In Section 8.2,

we'll introduce the term *simple* to describe a neighborhood like this. For the kind of local system we're considering here, the fact that U is simple implies that $\mathcal{E}_U$ is a G-torsor. In particular, $\mathcal{E}_U$ is nonempty; pick any $f \in \mathcal{E}_U$. Since $\pitchfork f$ is a local homeomorphism, it maps some neighborhood $V \subset U$ of x homeomorphically onto an open subset of X. The map $\pitchfork f$, restricted to V, is a chart around x.

We can use the construction above to cover M with charts to X. To verify that these charts fit together into an atlas, we just have to show that their transition maps belong to $\mathcal{G}$. Take two abstract charts $g \in \mathcal{E}_W$ and $f \in \mathcal{E}_V$ constructed in the manner above. For convenience, we'll add a $'$ to the name of an open set to denote its image under $\pitchfork f$. The transition map $(\pitchfork g)(\pitchfork f)^{-1}$ is an element of $\mathcal{O}(X, X)_{V' \cap W'}$, and we want to show that it lies in $\mathcal{G}_{V' \cap W'}$.

Consider any simple set $U \subset V \cap W$. Because $\mathcal{E}_U$ is a G-torsor, the restrictions of g and f to U are related by the action of G. To be precise,

$$\mathcal{E}_{U \subset W}\ g = \phi \cdot \mathcal{E}_{U \subset V}\ f$$

for some $\phi \in G$. We can turn this into a statement about the concrete charts $\pitchfork g$ and $\pitchfork f$ by applying the anchor map to both sides. The anchor map is a natural transformation of sheaves of G-sets, and the symmetry group acts on $\mathcal{O}(M, X)$ by postcomposition, so

$$\mathcal{O}(M, X)_{U \subset W}\ \pitchfork g = \phi \circ \mathcal{O}(M, X)_{U \subset V}\ \pitchfork f.$$

Equivalently,

$$\mathcal{O}(X,X)_{U'\subset V'\cap W'}\ (\text{⊥}g)(\text{⊥}f)^{-1} = \mathcal{O}(X,X)_{U'\subset V'\cap W'}\ \phi,$$

treating ϕ as an element of $\mathcal{G}_X \subset \mathcal{O}(X,X)_X$. This sounds much nicer when you say it in words: the restriction of $(\text{⊥}g)(\text{⊥}f)^{-1}$ to U' is the restriction of an element of G.

The argument above started from an arbitrary simple set $U \subset V \cap W$. Simple sets form a basis for M, so we can cover $V \cap W$ with them. Hence, we've shown that $(\text{⊥}g)(\text{⊥}f)^{-1}$ is given locally by restrictions of elements of G. That's exactly what it means for an element of $\mathcal{O}(X,X)_{V'\cap W'}$ to lie in $\mathcal{G}_{V'\cap W'}$. We've proven that the transition map between any two of the charts we constructed earlier belongs to $\mathcal{G}$. That means the charts fit together into an atlas, as we hoped. □

2.1.3 Analytic geometric structures

One way to come up with a rigid geometry is to make X a real- or complex-analytic manifold and pick a symmetry group G whose elements are analytic maps from X to itself. Thurston introduced (G,X) structures in this context.

When the model geometry $G\circlearrowright X$ is analytic, we can think about (G,X) structures in the analytic world, where all manifolds come with analytic structures and all maps respect them. In this world, the sheaf $\mathcal{O}(M,X)$ from which we draw our local charts is the sheaf of locally invertible analytic maps. When

we're working in the analytic world, I'll often emphasize it by talking about *analytic* (G, X) structures.

Focusing on analytic (G, X) structures can have unexpected consequences. Hyperbolic geometry, discussed in Section 2.2.2, can be naturally understood as a complex-analytic geometry, but the uniformization theorem reveals that each complex-analytic surface has exactly one analytic hyperbolic structure [8]. The resulting correspondence between complex structures and hyperbolic structures on a topological surface leads to many surprises in two-dimensional geometry.

Complex projective geometry, discussed in Section 2.2.1, is also complex-analytic, and in Sections 3.2 and 4.2 we'll learn some interesting techniques for working with analytic complex projective structures. I suspect these techniques generalize to some other analytic geometries.

The terminology we're using extendeds to other categories of manifolds and maps. If M is a smooth manifold, for instance, we can think about smooth (G, X) structures, with local charts drawn from the sheaf $\mathcal{O}(M, X)$ of local diffeomorphisms. Smooth hyperbolic structures feature prominently in Section 3.3.

2.1.4 Geometry unmoored

You can weaken the notion of a geometric structure by allowing the anchor map to produce local homeomorphisms that aren't defined everywhere, or forgetting the anchor map entirely. When a catalog is separated from its

anchor map, we'll say it's become *unmoored*. When the anchor map produces partially defined local homeomorphisms, we'll say it's *partially unmoored*. The partial anchor maps we consider will always produce local homeomorphisms defined on a fixed dense subset of the underlying manifold. We'll sometimes say a geometric structure is *anchored* to emphasize that it includes a fully defined anchor map.

If you deform a geometric structure by pushing it forward along an isotopy, the anchor map changes, but the catalog stays the same. That's because flowing a local system forward along an isotopy gives you an isomorphism between the new local system and the original. Since all manifolds have non-trivial isotopies, unmoored geometric structures inevitably drift away, never remembering their anchor maps better than up to isotopy.

For some kinds of geometric structures, like hyperbolic structures, the isomorphism class of the catalog determines the anchor map up to isotopy. In this case, specifying an unmoored geometric structure is equivalent to specifying an isotopy class of anchored ones. For other kinds of geometric structures, like complex projective structures, you can find non-isotopic geometric structures with isomorphic catalogs.[1] In this case, an unmoored geometric structure carries less information than an isotopy class of anchored ones.

[1]Many thanks to Jorge Acosta and Francis Bonahon for pointing this out to me.

2.2 Examples

2.2.1 Complex projective structures

As a first example of a geometric structure, consider the group of conformal maps from the unit sphere to itself. This group is known classically as the Möbius group. It can be identified with $\mathrm{PSL}_2\,\mathbb{C}$ by presenting the unit sphere as $\mathbb{CP}^1$. In addition to preserving angles, Möbius transformations send circles to circles. The geometric structure $\mathrm{PSL}_2\,\mathbb{C} \circlearrowright \mathbb{CP}^1$ thus provides not only a way to measure angles, but also a way to distinguish circles from other curves.

2.2.2 Hyperbolic structures

The isometry group of the hyperbolic plane can be identified with $\mathrm{PSL}_2\,\mathbb{R}$ by presenting the hyperbolic plane as the upper half-plane in $\mathbb{CP}^1$. Isometries of $\mathbb{H}^2$ preserve distances, angles, and geodesics, of course, but they also preserve more subtle things, like families of asymptotic geodesics and families of horocycles. Thus, in addition to a Riemannian metric and a way to distinguish geodesics from other curves, the geometric structure $\mathrm{PSL}_2\,\mathbb{R} \circlearrowright \mathbb{H}^2$ provides a local notion of visual boundary, and identifies certain local foliations as horocycle foliations. We'll put these more subtle features to use when we talk about deflating hyperbolic surfaces in Section 6.

2.2.3 Translation structures

The action of the translation group $\mathbb{R}^2$ on the plane $\mathbb{R}^2$ preserves lots of interesting features. Surfaces with $\mathbb{R}^2 \circlearrowright \mathbb{R}^2$ structures are called *translation surfaces*. We'll review them in depth in Section 9.2, because the whole second part of the paper revolves around their geometric and dynamical properties. We'll also see a lot of them in the first part of the paper—especially in Section 6, where we get our first good look at abelianization. I'll therefore use this section to mention some features of translation structures which will be useful to have in mind while reading Part I.

Compact translation surfaces are too constrained to be interesting, so we allow translation structures to have certain kinds of conical singularities. I don't know of a neat way to fit the singularities into the formalism of (G, X) structures, so it may be best to think of them as missing points with regularity conditions imposed on the catalog and anchor map around them. A translation structure provides a way to distinguish vertical lines from other curves, and each conical singularity has a few vertical lines diving into and shooting out of it. These *critical leaves* play an important part in the geometry and dynamics of the translation structure.

Looking at the puncture shapes in Section 9.2, you can check that the catalog of a translation structure has identity holonomy around each puncture. That means the catalog extends to a local system on the whole surface, including the singularities. The anchor map, of course, doesn't extend.

The built-in projections from $\mathbb{R}^2$ to $\mathbb{R}$ split each local chart of a translation structure into a vertical part and a horizontal part. Vertical translations only affect the vertical part of a chart, and horizontal translations only affect the horizontal part. That makes a translation structure looks a lot like a pair of independent $\mathbb{R} \circlearrowright \mathbb{R}$ structures, each with its own catalog and anchor map. The vertical and horizontal parts of a translation structure aren't true geometric structures, because their "local charts" aren't local homeomorphisms. They are, however, geometrically meaningful, as we'll see in Section 6.3.3.

Picking a complex-analytic translation structure on a Riemann surface is the same thing as picking a holomorphic 1-form [9]. The zeros of the 1-form are the conical singularities of the translation structure. If the surface has punctures, the 1-form can have poles, which appear in the translation structure as ends of certain shapes. See Appendix F for details.

2.2.4 Half-translation structures

You can generalize translation geometry by acting on the plane $\mathbb{R}^2$ with both translations and 180° rotations, which I'll call *flips*. The subgroup of $\operatorname{Isom} \mathbb{R}^2$ generated by these operations is a semidirect product $\mathbb{R}^2 \rtimes \mathbb{Z}/(2)$. This group is twice as big as the translation group, so it must preserve half as much structure. Accordingly, spaces modeled on its action are called *half-translation surfaces*.

The features of translation surfaces we discussed above all have analogues for half-translation surfaces. Like their more disciplined siblings, half-

translation surfaces are allowed to have conical singularities. The catalog doesn't always extend over the singularities, but it comes close: its holonomy around a singularity is either the identity or a flip. Picking a complex-analytic half-translation structure on a Riemann surface is the same thing as picking a holomorphic quadratic differential [9, §8.1]. This fact lends an interesting interpretation to the discussion in Section 4.2.

Chapter 3

Geometric structures as flat bundles

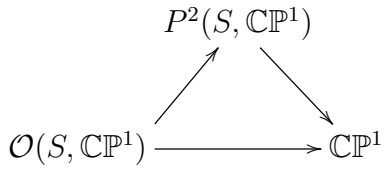
3.1 Overview

You can turn a principal G bundle with a flat connection into a G local system by taking its sheaf of flat sections. This functor from flat G bundles to G local systems is an equivalence, but not an isomorphism: you can always realize a G local system as the sheaf of flat sections of a flat G bundle, but you generally can't do it in any canonical way.

By itself, the catalog of a (G, X) structure is just a G local system. As part of a geometric structure, however, the catalog comes with extra data, linked to it by the anchor map. For some kinds of (G, X) structures, this extra data gives a natural way to turn the catalog into a flat G bundle. The analytic complex projective structures on a Riemann surface provide a striking example: they all appear naturally as flat connections on a special "carrier bundle," described in Section 3.2. A similar idea works for smooth hyperbolic structures, but with a twist, as we'll see in Section 3.3.

3.2 A carrier for complex projective structures

Let's consider analytic complex projective structures on a Riemann surface S. Their local charts are drawn from the sheaf of conformal maps $\mathcal{O}(S, \mathbb{CP}^1)$. Let $P^2(S, \mathbb{CP}^1)$ be the bundle of 2-jets of conformal maps[1] from S to $\mathbb{CP}^1$ [10, §IV.5]. At each point $x \in S$, a quotient map $\mathcal{O}(S, \mathbb{CP}^1)_x \to P^2(S, \mathbb{CP}^1)_x$ projects the germs of conformal maps down to their 2-jets. Composing with the 2-jet evaluation map, we get a factorization



of the germ evaluation map.

The action of Möbius transformations on $\mathcal{O}(S, \mathbb{CP}^1)$ descends to an action on $P^2(S, \mathbb{CP}^1)$. Every element of $P^2(\mathbb{CP}^1, \mathbb{CP}^1)$ is the 2-jet of a unique

[1] Here's a refresher on jets of locally invertible maps in the analytic world. (It works just as well in the smooth world, and probably other worlds too.) At each point x on a complex manifold M, there's a ring of germs of holomorphic functions. Within that ring, let $\mathfrak{m}$ be the ideal of germs that vanish at x. The *k-jet* of a germ in $\mathfrak{m}$ is its projection to $\mathfrak{m}/\mathfrak{m}^{k+1}$, which describes it "up to kth order." The k-jets of holomorphic functions on M fit together into a ring bundle $J^k M$, whose fiber at x is $\mathfrak{m}/\mathfrak{m}^{k+1}$.

A locally invertible analytic map $f\colon M \to N$ pulls local holomorphic functions back from N to M, inducing a bundle map $J^k M \leftarrow J^k N$. Together, the point $y = fx$ and the isomorphism $J^k M_x \leftarrow J^k N_y$ describe the action of f near x up to kth order. Playing with polynomials in local coordinates, you can see that any isomorphism $J^k M_x \leftarrow J^k N_y$ arises in this way from a locally invertible analytic map defined near x and y. The data of a point $x \in M$, a point $y \in N$, and an isomorphism $J^k M_x \leftarrow J^k N_y$ thus describe what you might call the *k-jet* of a locally invertible analytic map from M to N. These k-jets fit together into a space $P^k(M, N)$, which is naturally a bundle over both M and N. We'll always view it as a bundle over M.

Möbius transformation, so $P^2(\mathbb{CP}^1, \mathbb{CP}^1)$ is a principal $\mathrm{PSL}_2\,\mathbb{C}$ bundle. As a consequence, $P^2(S, \mathbb{CP}^1)$ is a principal $\mathrm{PSL}_2\,\mathbb{C}$ bundle too.

Take an analytic complex projective structure on S, with catalog $\mathcal{E}$ and anchor map ⚓. Composing the anchor map with the projection described above, we get a $\mathrm{PSL}_2\,\mathbb{C}$-equivariant map

$$\mathcal{E}_x \xrightarrow{\text{⚓}} \mathcal{O}(S, \mathbb{CP}^1)_x \longrightarrow P^2(S, \mathbb{CP}^1)_x$$

An equivariant map between torsors is automatically an isomorphism, so this map identifies each stalk of $\mathcal{E}$ with the corresponding fiber of $P^2(S, \mathbb{CP}^1)$.

For any open subset U of S, the map

$$\mathcal{E}_U \longrightarrow \mathcal{E}_x \xrightarrow{\text{⚓}} \mathcal{O}(S, \mathbb{CP}^1)_x \longrightarrow P^2(S, \mathbb{CP}^1)$$

varies holomorphically with respect to $x \in U$. It turns the elements of $\mathcal{E}_U$ into holomorphic sections of $P^2(S, \mathbb{CP}^1)$ over U. The tangent planes of these sections describe a flat connection A on $P^2(S, \mathbb{CP}^1)$. The sheaf of flat sections of A is identified with $\mathcal{E}$ by construction.

The flat section of $P^2(S, \mathbb{CP}^1)$ corresponding to a catalog entry $f \in \mathcal{E}_U$ tells us the 2-jet of $\text{⚓}f$ at each point in U. In particular, it tells us the value of $\text{⚓}f$ at each point in U, which is to say it tells us $\text{⚓}f$. The flat connection A thus completely describes the complex projective structure it came from, wrapping up the catalog and the anchor map in one neat geometric package.

Through the construction described here, the bundle $P^2(S, \mathbb{CP}^1)$ provides a common home for all analytic complex projective structures on S. To

honor this role, we'll call it a *carrier bundle* for analytic complex projective structures.

3.3 Carriers for hyperbolic structures

3.3.1 A canonical construction

The idea we just used to realize analytic complex projective structures as flat connections on a carrier bundle can also be applied to smooth hyperbolic structures. Its execution, however, is more subtle. Instead of building the carrier bundle directly over the surface we're studying, we'll build it over the bundle of tangent directions of the surface.

Let's consider smooth hyperbolic structures on a smooth surface S. Their local charts are drawn from the sheaf of local diffeomorphisms $\mathcal{O}(S, \mathbb{H}^2)$. Let US be the bundle of tangent directions on S, realized formally as $TS \smallsetminus 0$ modulo scaling by positive numbers. In the presence of a Riemannian metric, US earns its notation by embedding itself in TS as the unit tangent bundle. The derivative of a local diffeomorphism from S to $\mathbb{H}^2$ sends tangent directions on S to tangent directions on $\mathbb{H}^2$. This defines a map $\delta_u \colon \mathcal{O}(S, \mathbb{H}^2)_x \to U\mathbb{H}^2$ for each $u \in US_x$.

Take a smooth hyperbolic structure on S, with catalog $\mathcal{E}$ and anchor map $\pitchfork$. Bundle projections are open, so $\mathcal{E}$ pulls back straightforwardly along the projection $\pi \colon US \to S$, becoming a local system $\tilde{\mathcal{E}}$ on US. The stalk $\tilde{\mathcal{E}}_u$ over a tangent direction $u \in US$ is naturally identified with $\mathcal{E}_{\pi u}$. Composing this identification with the anchor map and the derivative map described

above, we get a $\mathrm{PSL}_2\,\mathbb{R}$-equivariant map

$$\tilde{\mathcal{E}}_u \longrightarrow \mathcal{E}_{\pi u} \xrightarrow{\perp} \mathcal{O}(S, \mathbb{H}^2)_{\pi u} \xrightarrow{\delta_u} U\mathbb{H}^2$$

An isometry of $\mathbb{H}^2$ is determined by its action on any tangent vector, so $U\mathbb{H}^2$ is a $\mathrm{PSL}_2\,\mathbb{R}$-torsor. Hence, as before, the map above is an isomorphism. We can think of it as identifying each stalk of $\tilde{\mathcal{E}}$ with the corresponding fiber of $\underline{U\mathbb{H}^2}$, the trivial $U\mathbb{H}^2$ bundle over US.

For any open subset U of US, the map

$$\tilde{\mathcal{E}}_U \longrightarrow \tilde{\mathcal{E}}_u \longrightarrow \mathcal{E}_{\pi u} \xrightarrow{\perp} \mathcal{O}(S, \mathbb{H}^2)_{\pi u} \xrightarrow{\delta_u} \underline{U\mathbb{H}^2}$$

varies smoothly with respect to $u \in U$. Thus, as before, it describes a flat connection A on $\underline{U\mathbb{H}^2}$, whose sheaf of flat sections is identified with $\tilde{\mathcal{E}}$ by construction.

By definition, $\tilde{\mathcal{E}}_U = \mathcal{E}_{\pi U}$. The flat section of $\underline{U\mathbb{H}^2}$ corresponding to a catalog entry $f \in \tilde{\mathcal{E}}_U$ tells us what the derivative of $\perp f$ does to each tangent direction in U. In particular, it tells us the value of $\perp f$ at each point in πU, which is to say it tells us $\perp f$. Thus, as before, the flat connection A completely describes the hyperbolic structure it came from. The bundle $\underline{U\mathbb{H}^2} \to US$ therefore deserves to be called a carrier bundle for smooth hyperbolic structures.

Because $\underline{U\mathbb{H}^2}$ is canonically trivial, its parallel transport maps can be seen as elements of $\mathrm{PSL}_2\,\mathbb{R}$. When you construct a flat connection on $\underline{U\mathbb{H}^2}$ from a hyperbolic structure, its parallel transport lifts each loop around a fiber of US to a loop around an elliptic subgroup of $\mathrm{PSL}_2\,\mathbb{R}$, which is not

contractible. The flat connections on $\underline{U\mathbb{H}^2}$ that describe hyperbolic structures are therefore examples of *twisted flat connections* on S [11, §10.1].

3.3.2 Construction from a reference hyperbolic structure

Building a bundle over another bundle feels a bit perverse, so you might wonder if we can find a carrier bundle for smooth hyperbolic structures on S that lives over S itself. It turns out we can, but at the cost of doing something even more perverse: fixing an arbitrary hyperbolic structure on S and describing all other hyperbolic structures in terms of it.

We start off following our construction of a carrier bundle for analytic complex projective structures. At each point $x \in S$, a quotient map $\mathcal{O}(S, \mathbb{H}^2)_x \to P^1(S, \mathbb{H}^2)_x$ projects the germs of local diffeomorphisms down to their 1-jets. You can think of a 1-jet that sends $x \in S$ to $y \in \mathbb{H}^2$ as a linear isomorphism $TS_x \to T\mathbb{H}^2_y$.

The analogy with the complex projective story immediately breaks down, because Möbius transformations don't act transitively on $P^1(S, \mathbb{H}^2)_x$, so the bundle of 1-jets isn't a principal $\mathrm{PSL}_2\,\mathbb{R}$ bundle. To patch over this difficulty, we fix a smooth hyperbolic structure on S, with catalog $\mathcal{E}^\star$ and anchor map $\mathfrak{t}^\star$. The map

$$\mathcal{E}^\star_x \xrightarrow{\mathfrak{t}^\star} \mathcal{O}(S, \mathbb{H}^2)_x \longrightarrow P^1(S, \mathbb{H}^2)_x$$

is injective, so its image is a $\mathrm{PSL}_2\,\mathbb{R}$-torsor. For an open subset U of S, the

map

$$\mathcal{E}^\star_U \longrightarrow \mathcal{E}^\star_x \xrightarrow{\mathfrak{t}^\star} \mathcal{O}(S, \mathbb{H}^2)_x \longrightarrow P^1(S, \mathbb{H}^2)_x$$

varies continuously with respect to $x \in U$, so it sweeps out a principal $\mathrm{PSL}_2\,\mathbb{R}$ bundle $Q^\star \hookrightarrow P^1(S, \mathbb{H}^2)$. The notation is meant to emphasize that $Q^\star$ depends on the reference hyperbolic structure.

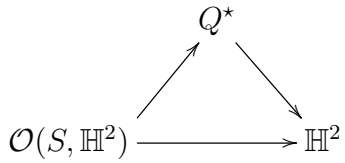
Now that we have a principal bundle over S, we can start following the complex projective story again. The next step is to find a good map $\mathcal{O}(S, \mathbb{H}^2)_x \to Q^\star_x$ at each point $x \in S$. We can't just send each germ to its 1-jet, like we did before, because the 1-jet $\Phi \in P^1(S, \mathbb{H}^2)_x$ of a germ $\phi \in \mathcal{O}(S, \mathbb{H}^2)_x$ generally won't belong to $Q^\star_x$. Instead, we'll send ϕ to the unique 1-jet in $Q^\star_x$ that lines up with Φ in a certain sense.

As I mentioned earlier, you can think of Φ as a linear isomorphism $TS_x \to T\mathbb{H}^2_{fx}$. One way to compare Φ with another 1-jet Ψ it to look at how you'd have to distort $T\mathbb{H}^2_{fx}$ to turn Φ into Ψ. That distortion is the automorphism $\Psi\Phi^{-1}$. The Riemannian metric on $\mathbb{H}^2$ makes $T\mathbb{H}^2_{fx}$ an inner product space, so we can use the polar decomposition to express $\Psi\Phi^{-1}$ uniquely as a positive-definite map followed by an orthogonal one—a stretch followed by a rotation. It seems natural to say that Ψ lines up with Φ when the rotational part of $\Psi\Phi^{-1}$ is the identity.

If Ψ is in $Q^\star_x$, we can get all the other 1-jets in $Q^\star_x$ by applying the action of $\mathrm{PSL}_2\,\mathbb{R}$. To get only the 1-jets that send x to fx, we apply only the subgroup of $\mathrm{PSL}_2\,\mathbb{R}$ that stabilizes fx, which acts faithfully on $T\mathbb{H}^2_{fx}$ by

rotations. One way to describe the polar decomposition is to say that there's a unique rotation Ω for which $\Omega\Psi\Phi^{-1}$ is positive-definite. That means $\Omega\Psi$ is the unique 1-jet in $Q^\star_x$ that lines up with Φ.

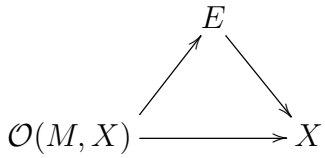
Now that we've found a bundle map $\mathcal{O}(S, \mathbb{H}^2) \to Q^\star$, we can compose it with the 1-jet evaluation map to get a factorization



of the germ evaluation map. From there, can follow the complex projective story through to the end.

3.4 A general framework

In both of the cases where we managed to build a carrier bundle for (G, X) structures directly over the manifold M we were studying, the carrier bundle E was a principal G bundle providing an equivariant factorization



of the germ evaluation map. This suggests a link between our notion of a carrier bundle and Goldman's formalism of graphs of geometric structures [7].

Goldman's approach uses an X bundle over M with structure group G, rather than a principal G bundle. To specify a (G, X) structure on M, you give the bundle two things:

- A flat connection, called the *tangent connection*.
- A global section transverse to the connection planes, called the *developing section*.

To connect Goldman's picture with ours, observe that an equivariant map from a principal G bundle to X holds the same information as a section of the associated X bundle [12, Proposition 3.11], so the X bundle associated to the carrier bundle gets a section from the factorization above. If we use that section as a developing section, a flat connection on the carrier bundle and the induced flat connection on the associated X bundle should hopefully specify the same geometric structure.

Chapter 4

Comparing geometric structures

4.1 Overview

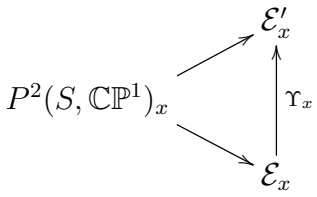
In this section, we'll look at two instruments for measuring the difference between a pair of geometric structures. The first, called the Schwarzian derivative, is classical device for comparing analytic complex projective structures. We won't need it for any technical purpose, but it illustrates the rich interaction between complex projective geometry and half-translation geometry that runs throughout this paper and its references. The second, which I'm calling the *deviation*, can be used to compare all sorts of geometric structures, and more general things too. It seems to be less widely used, but it will play a central role in this paper.

4.2 The Schwarzian derivative

4.2.1 The setup

Say we have two analytic complex projective structures on the same Riemann surface S, with catalogs $\mathcal{E}$, $\mathcal{E}'$ and anchor maps ⴑ, $\text{ⴑ}'$. I'll call them the old structure and the new structure, for short. They induce flat connections A, A' on $P^2(S, \mathbb{CP}^1)$ as described in Section 3.2, realizing $\mathcal{E}$ and

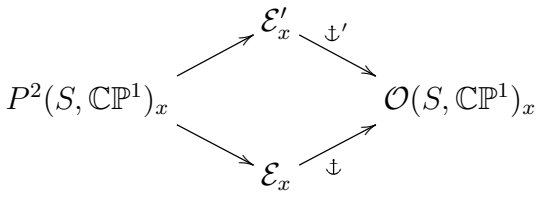
$\mathcal{E}'$ as the corresponding sheaves of flat sections. To lighten the notation, I'll implicitly identify the fibers of $P^2(S, \mathbb{CP}^1)$ with the stalks of $\mathcal{E}$ and $\mathcal{E}'$, and I won't write stalk restrictions or their inverses. The stalkwise isomorphism $\Upsilon\colon \mathcal{E} \to \mathcal{E}'$ defined by the diagram



is an essential component of the comparison we're about to discuss, though it won't be named explicitly.

4.2.2 The definition

Given an analytic complex projective structure on S, we can use the anchor map to extend any 2-jet of a map $S \to \mathbb{CP}^1$ to the germ of a local chart. In our present situation, that gives us two maps



By comparing these maps, we can see how local charts get distorted in passing from the old complex projective structure to the new one. Pursuing this idea down a path pointed out by Thurston will lead us to the Schwarzian derivative [13].

Pick a 2-jet $f \in P^2(S, \mathbb{CP}^1)_x$ that sends x to zero. We can think of the germs $z = \text{ⅅ} f$ and $z' = \text{ⅅ}' f$ as the germs of meromorphic functions. By construction, z and z' both have f as their 2-jet, so they're actually the germs of holomorphic functions that vanish at x.

Within the ring of germs of holomorphic functions, let $\mathfrak{m}$ be the ideal of germs that vanish at x. Now we can say concretely what it means for z and z' to have the same 2-jet: they represent the same element of $\mathfrak{m}/\mathfrak{m}^3$. In other words,

$$z' \in z + \mathfrak{m}^3.$$

The germ z is locally invertible, because it's the germ of a local chart. Consequently, its first derivative is nonzero—it represents a nonzero element of $\mathfrak{m}/\mathfrak{m}^2$. It follows, with a bit of thought, that z generates $\mathfrak{m}$. In particular, $\mathfrak{m}^3 = z\,\mathfrak{m}^2$, so we can rewrite the approximation above as

$$z' \in z(1 + \mathfrak{m}^2).$$

Our first chance to detect a difference between the two complex projective structures on S is to compare z and z' in $\mathfrak{m}/\mathfrak{m}^4$, at the level of 3-jets. Rewrite the approximation above as

$$z' \in z(1 + q_x)$$

for some coset $q_x \in \mathfrak{m}^2/\mathfrak{m}^3$, noting that q_x is determined uniquely by z and z'. In coordinates,

$$q_x = p_x z^2 + \mathfrak{m}^3$$

for some $p_x \in \mathbb{C}$. It will be convenient to view q_x as a bilinear form on $(1,0)$ tangent vectors, using the classical[1] isomorphism between $\mathfrak{m}^2/\mathfrak{m}^3$ and the symmetric square of $T^*_{1,0}S_x$. From this perspective, our coordinate expression for q_x should be read as

$$q_x = p_x\, dz^2.$$

We'll see later that q_x doesn't depend on which 2-jet f we started with.

We can do the calculation above at any point in S, yielding a section

$$q\colon S \to \operatorname{Sym}^2 T^*_{1,0}S$$

$$x \mapsto q_x$$

A section of $\operatorname{Sym}^2 T^*_{1,0}S$ is called a quadratic differential [9, §8.1]. Because the local charts z and z' are analytic, q_x turns out to vary holomorphically with x. The holomorphic quadratic differential q is called the *Schwarzian derivative* of the new complex projective structure with respect to the old one. As we

[1] An element of $\mathfrak{m}^2/\mathfrak{m}^3$ is the second-order part of a function that vanishes to first order. Intuitively, it's the second derivative of a function whose first derivative vanishes—a Hessian. A Hessian is typically expressed as a bilinear form on $(1,0)$ tangent vectors, classically described as a "matrix of second derivatives." This is possible because the multiplication map

$$\operatorname{Sym}^2(\mathfrak{m}/\mathfrak{m}^2) \to \mathfrak{m}^2/\mathfrak{m}^3$$
$$(g+\mathfrak{m}^2)\odot(h+\mathfrak{m}^2) \mapsto gh+\mathfrak{m}^3$$

is an isomorphism, and $\mathfrak{m}/\mathfrak{m}^2$ is the $(1,0)$ cotangent space at x.

As you might have inferred from its general terms, the discussion above makes sense on any complex manifold M. Without the references to the $(1,0)$ part of the tangent space, it makes equally good sense on any smooth manifold. Just as we define the cotangent bundle $T^*_{1,0}M$ as the bundle whose fiber at x is $\mathfrak{m}/\mathfrak{m}^2$, we could define a "Hessian bundle" $H_{1,0}M$ whose fiber at x is $\mathfrak{m}^2/\mathfrak{m}^3$. The Hessian of a function at x is an element of $H_{1,0}M_x$, just as the derivative of a function at x is an element of $T^*_{1,0}M_x$.

mentioned in Section 2.2.4, choosing a holomorphic quadratic differential on a Riemann surface is the same thing as choosing an analytic half-translation structure. The Schwarzian derivative tells us that the difference between two analytic complex projective structures is itself a geometric structure.

4.2.3 Showing the definition makes sense

When we defined the Schwarzian derivative, we got q_x from a 2-jet $f \in P^2(S, \mathbb{CP}^1)_x$ that sends x to zero. What if we'd picked a different 2-jet? Recall that $\mathrm{PSL}_2\,\mathbb{C}$ acts transitively on each fiber of $P^2(S, \mathbb{CP}^1)$, so any other 2-jet at x can be written as $M \cdot f$ for some Möbius transformation M. We want $M \cdot f$ to send x to zero, like f does, so we have to pick a Möbius transformation that fixes zero. Hence, we can use a matrix of the form

$$\begin{bmatrix} a & \cdot \\ c & 1 \end{bmatrix}$$

to represent $M \in \mathrm{PSL}_2\,\mathbb{C}$.

The anchor map for a $(\mathrm{PSL}_2\,\mathbb{C}, \mathbb{CP}^1)$ structure is $\mathrm{PSL}_2\,\mathbb{C}$-equivariant, so

$$\begin{aligned} \text{ŧ}(M \cdot f) &= M \cdot \text{ŧ} f & \qquad \text{ŧ}'(M \cdot f) &= M \cdot \text{ŧ}' f \\ &= M \cdot z & &= M \cdot z'. \end{aligned}$$

The value of $q_x \in \mathfrak{m}^2/\mathfrak{m}^3$ we get from $M \cdot f$ is thus characterized by the approximation

$$M \cdot z' \in (M \cdot z)(1 + q_x).$$

Writing M out as a rational function, we get

$$\frac{az'}{1+cz'} \in \frac{az}{1+cz}(1+q_x),$$

or equivalently

$$\begin{aligned} z' &\in \frac{1+cz'}{1+cz}\, z(1+q_x) \\ &= \left[1+\frac{c(z'-z)}{1+cz}\right] z(1+q_x). \end{aligned}$$

Recall again that z' and z have the same 2-jet, so $z'-z \in \mathfrak{m}^3$. Hence,

$$\begin{aligned} z' &\in (1+\mathfrak{m}^3)\, z(1+q_x) \\ &\subset z(1+q_x+\mathfrak{m}^3) \\ &= z(1+q_x), \end{aligned}$$

showing that the q_x we get from $M \cdot f$ is the same one we got from f.

4.2.4 Some geometric meaning

A holomorphic quadratic differential, as we remarked in Section 2.2.4, defines an analytic half-translation structure on S. This structure provides a lot of features, including a pair of singular foliations called the horizontal and vertical foliations. The horizontal and vertical foliations of the Schwarzian derivative have a cute geometric meaning.

As we mentioned in Section 2.2.1, a complex projective structure provides a way to distinguish circles from other curves. When we pass from the old complex projective structure on S to the new one, the curves that used

to be circles through x generally won't stay circular. There are a few circles through x, however, that stay circular up to third order. They're the ones tangent to the horizontal and vertical leaves of the Schwarzian derivative!

We can say this more formally using the language of germs of curves. Let $\mathcal{A}$ be the sheaf of real-analytic functions from S to $\mathbb{R}$, and let $\mathfrak{n} \subset \mathcal{A}_x$ be the ideal of germs that vanish at x. If a germ $c \in \mathfrak{n}$ has nonzero derivative, its zero set is the germ of a one-dimensional smooth submanifold of S—the germ of a curve through x. The curve depends only on the ideal c generates. We can thus identify the germs of curves through x with certain ideals in $\mathcal{A}_x$, which I'll call *curve ideals*.

Once we're viewing curves as the zero sets of functions, we can approximate curves by approximating the corresponding functions. Given a curve ideal $\mathfrak{c} \subset \mathcal{A}_x$, which describes the germ of a curve, we'll say the ideal $\mathfrak{c} + \mathfrak{n}^{k+1}$ describes the *k-jet* of that curve. Other curves with the same k-jet are described by other curve ideals contained in $\mathfrak{c} + \mathfrak{n}^{k+1}$.

We're trying to prove something about circles, so we need to characterize *circle ideals*—the curve ideals describing germs of circles. Say $z \in \mathcal{O}(S, \mathbb{CP}^1)_x$ is the germ of a local chart for the old complex projective structure. An ideal in $\mathcal{A}_x$ is a circle ideal if and only if it's generated by a germ of the form

$$\alpha z\bar{z} + \beta z + \overline{\beta z},$$

with $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{C}^\times$ [14].

Now we're ready to see which circles in the old complex projective structure stay approximately circular in the new one. Fix a circle ideal $\mathfrak{c} \subset \mathcal{A}_x$ in the old complex projective structure. Pick a 2-jet $f \in P^2(S, \mathbb{CP}^1)_x$ that sends x to zero and lays the circle tangent to the real axis. In terms of the local chart $z = \text{ŧ} f$,

$$\mathfrak{c} = \left[\alpha z\bar{z} + z + \bar{z}\right]\mathcal{A}_x$$

for some $\alpha \in \mathbb{R}$. Any circle ideal in the new complex projective structure can be exprssed in terms of the local chart $w = \text{ŧ}' f$ as

$$\mathfrak{c}' = \left[\alpha' w\bar{w} + \beta' w + \overline{\beta' w}\right]\mathcal{A}_x,$$

where $\alpha' \in \mathbb{R}$ and $\beta' \in \mathbb{C}^\times$. We want to know whether we can set α' and β' so that

$$\mathfrak{c} + \mathfrak{n}^4 = \mathfrak{c}' + \mathfrak{n}^4.$$

Assume that q_x is nonzero, so it picks out horizontal and vertical directions in the tangent space at x. Write q_x in coordinates as $p_x z^2 + \mathfrak{m}^3$. The coefficient p_x is real when $\mathfrak{c}$ describes a circle tangent to the horizontal or vertical direction of q_x. The sign of p_x determines whether the circle is horizontal or vertical; the assumption that q_x picks out horizontal and vertical directions ensures that p_x can't be zero.

From the definition of the Schwarzian derivative and the fact that z

generates $\mathfrak{m}$, we reason that

$$\begin{aligned} z(1+q_x)+\mathfrak{m}^4 &= w+\mathfrak{m}^4 \\ z+\mathfrak{m}^4 &= w(1-q_x)+\mathfrak{m}^4 \\ z &\in w(1-q_x)+\mathfrak{m}^4. \end{aligned}$$

Using this approximation, we can rewrite q_x in coordinates as $p_x w^2+\mathfrak{m}^3$. We can also deduce that

$$\begin{aligned} \mathfrak{c}+\mathfrak{n}^4 &= \left[\alpha w\bar{w}+w+\bar{w}-(wq_x+\overline{wq_x})\right]\mathcal{A}_x+\mathfrak{n}^4 \\ &= \left[\alpha w\bar{w}+w+\bar{w}-(p_x w^3+\overline{p_x w^3})\right]\mathcal{A}_x+\mathfrak{n}^4. \end{aligned}$$

By writing a general element of $\mathfrak{c}'$ out to second order, we see that we have to set $\alpha'=\alpha$ and $\beta'=1$ for $\mathfrak{c}'+\mathfrak{n}^4$ to have any chance of containing the generator

$$c=\alpha w\bar{w}+w+\bar{w}-(p_x w^3+\overline{p_x w^3}).$$

We can therefore restrict our attention to the case where $\mathfrak{c}'$ is generated by

$$c'=\alpha w\bar{w}+w+\bar{w}.$$

Writing a general multiple of c' out to third order reveals that

$$c+\mathfrak{n}^4=c'r+\mathfrak{n}^4$$

for some $r\in\mathcal{A}_x$ if and only if p_x is real. The germ r turns out to be invertible modulo $\mathfrak{n}^4$ if it exists, so we've shown that

$$\mathfrak{c}+\mathfrak{n}^4=\mathfrak{c}'+\mathfrak{n}^4$$

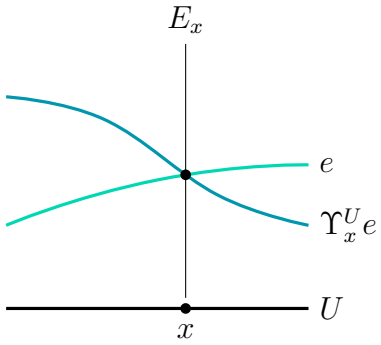
if and only if p_x is real. As we remarked earlier, p_x is real when $\mathfrak{c}$ describes a circle tangent to the horizontal or vertical direction of q_x. Hence, we've proven what we set out to prove.

4.3 Deviations of geometric structures

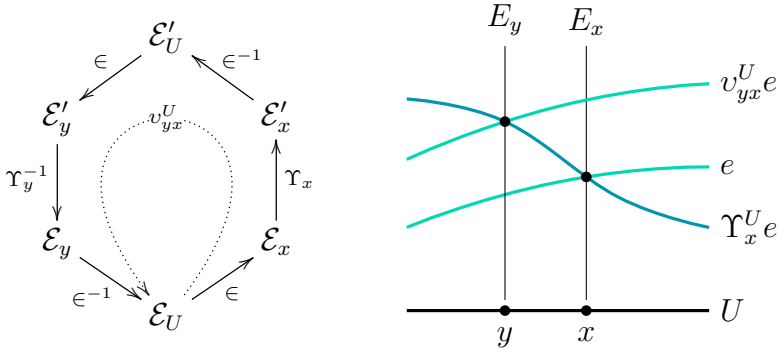
The Schwarzian derivative compares complex projective structures microscopically, using the infinitesimal details of local charts at each point. As a result, it relies on some degree of smoothness in the geometric structures and the underlying manifold. Our second instrument, the *deviation*, works on macroscopic principles, comparing local charts on open sets. Although we'll introduce it using imagery from the smooth world, it generalizes readily to geometric structures on topological manifolds. In fact, its real calling lies in the even craggier setting of Part II, where we'll use it to compare local systems on spaces quite far from being manifolds. In Section 4.4, we'll get a preview of how deviations work at that level of generality.

For now, though,[2] let's say M is a smooth manifold, and $E \to M$ is a carrier bundle for (G, X) structures on M. Suppose we have two (G, X) structures on M, represented by flat connections A and A' on E. Their catalogs $\mathcal{E}$ and $\mathcal{E}'$ are the corresponding sheaves of flat sections. The stalk $\mathcal{E}_x$ is the space of germs of flat sections of A at x. Sending each germ in $\mathcal{E}_x$ to the unique germ in $\mathcal{E}'_x$ that has the same value at x gives an isomorphism $\Upsilon_x \colon \mathcal{E}_x \to \mathcal{E}'_x$.

[2] This portion of Section 4.3 appeared previously, with minor differences, as [15, §2.3].



If you read the proof of Proposition 2.1.A, you might recall the notion of a simple set, which will be introduced formally in Section 8.2. If $U \subset M$ is simple with respect to both $\mathcal{E}$ and $\mathcal{E}'$, we can visualize Υ_x by its action $\Upsilon_x^U \colon \mathcal{E}_U \to \mathcal{E}'_U$ on flat sections over U, as shown above. Then, for any $x, y \in U$, we can define an automorphism υ_{yx}^U of $\mathcal{E}_U$ that tells us how parallel transport along A' deviates from parallel transport along A:



This automorphism is characterized by the property that $\Upsilon_y^U \upsilon_{yx}^U = \Upsilon_x^U$. If $V \subset U$ is a simple neighborhood of x and y, the automorphisms υ_{yx}^U and υ_{yx}^V commute with the restriction map $\mathcal{E}_{V \subset U}$, so all these automorphisms fit together into a natural automorphism υ_{yx} of the functor we get by restricting

$\mathcal{E}$ to the poset of simple neighborhoods of x and y. Restricting $\mathcal{E}$ further to the simple neighborhoods of three points x, y, and z, we can observe that $\upsilon_{zy}\upsilon_{yx} = \upsilon_{zx}$. Collectively, the natural automorphisms $\{\upsilon_{yx}\}_{x,y\in M}$ might be called the *deviation* of A' from A.

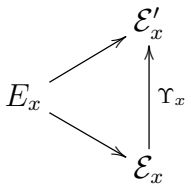
4.4 Deviations of locally constant sheaves

In Section 4.3, we used the conceptual framework of carrier bundles for geometric structures to motivate the definition of a deviation. Now, if we tear down the scaffolding, we'll see the deviation standing on its own in a much more general context. To avoid getting bogged down in details, we'll just take a quick look, enough to make sense of the way deviations are used in Section 5. A more precise treatment will appear in Section 8.3.

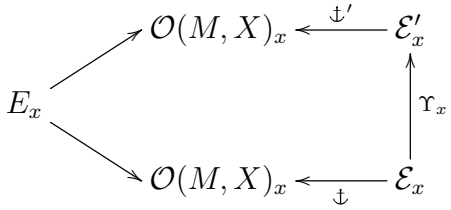
Say we have two locally constant sheaves $\mathcal{E}$ and $\mathcal{E}'$ on a topological space M. Give a stalkwise isomorphism $\Upsilon_x\colon \mathcal{E}_x \to \mathcal{E}'_x$, we can measure the deviation υ_{yx} just as we did before. The choice of stalkwise isomorphism might not be canonical anymore, so we should really talk about "the deviation of Υ" instead of "the deviation of $\mathcal{E}'$ from $\mathcal{E}$."

When M is locally connected, as we'll assume from now on, the deviation of Υ specifies Υ and $\mathcal{E}'$ uniquely up to canonical isomorphism. Because the features of a locally constant sheaf are spread out over open sets, Υ doesn't have to be defined at every point in M for this to work. It just has to be defined on a dense subset, which we'll call the *support* of Υ and its deviation.

The deviations we used earlier to compare geometric structures always had full support, because the stalkwise isomorphism Υ defined by the diagram



was defined at every point $x \in M$. In our examples, the identification of the fibers of E with the stalks of $\mathcal{E}$ and $\mathcal{E}'$ went through the anchor maps, like this:



Ultimately, Υ was defined everywhere because the anchor maps were.

Deviations with partial support appear when geometric structures come partially unmoored, with anchor maps only defined on a dense subset of M. This is a typical result of cutting and gluing constructions, which we'll explore in Section 5.

Chapter 5

Warping geometric structures

5.1 The general idea

I mentioned in Section 4.4 that a stalkwise isomorphism between locally constant sheaves is determined up to isomorphism by its deviation. We'll see in Section 8.4 that every deviation, conversely, determines a stalkwise isomorphism between locally constant sheaves. To be precise, say we're given a deviation υ from a locally constant sheaf $\mathcal{E}$, supported on $D \subset M$. We can then build a locally constant sheaf $\mathcal{E}'$ and a stalkwise isomorphism $\Upsilon\colon \mathcal{E} \to \mathcal{E}'$, supported on D, whose deviation is υ. The deviation isn't just a measuring device: it's also construction equipment.

The process of building a stalkwise isomorphism from a deviation will be called *warping*. It encompasses many classical and not-so-classical operations on geometric structures, including abelianization. It's especially well suited for cutting and gluing operations, which tend to behave well only on dense subsets of the spaces they act on. In this section, we'll get a feel for warping by using it to describe some cutting and gluing constructions related to abelianization.

5.2 Cataclysms on punctured hyperbolic surfaces

5.2.1 Setup

Take a surface S', with negative Euler characteristic, which is compact except for a finite set of punctures. Break it into triangles by drawing edges from puncture to puncture, like this:

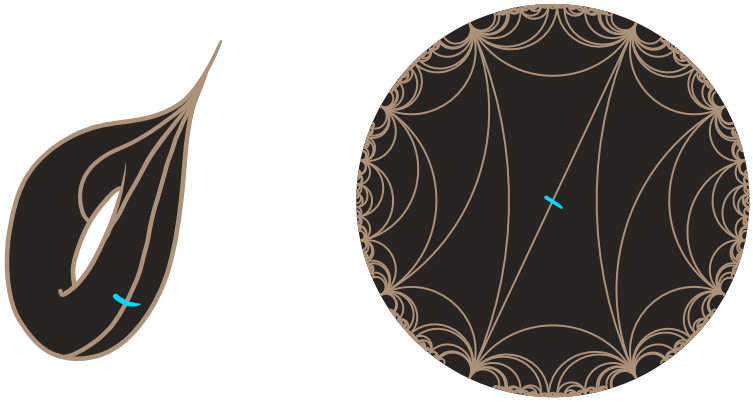
The edges of the triangles should be non-intersecting one-dimensional submanifolds of S', and each triangle should have three distinct edges. This gives an *ideal triangulation* of S' [16].

Weight each edge e of the triangulation by a real number l_e. A weighted ideal triangulation specifies an operation, called a *cataclysm*, that turns hyperbolic structures on S' into new hyperbolic surfaces [17]. Let's see how this operation works in the simplest case, where the edge weights of the triangulation sum to zero around each puncture.

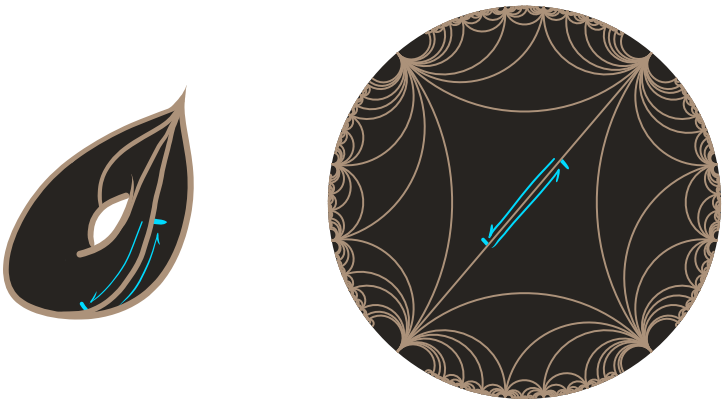
5.2.2 Cataclysms on surfaces with cusps

Give S' a hyperbolic structure, with a cusp at each puncture. Each edge of the triangulation snaps tight to a geodesic by a homotopy relative

to the punctures, making each triangle isometric to an ideal triangle in the hyperbolic plane. It's easiest to see how this works on the universal cover of S'.

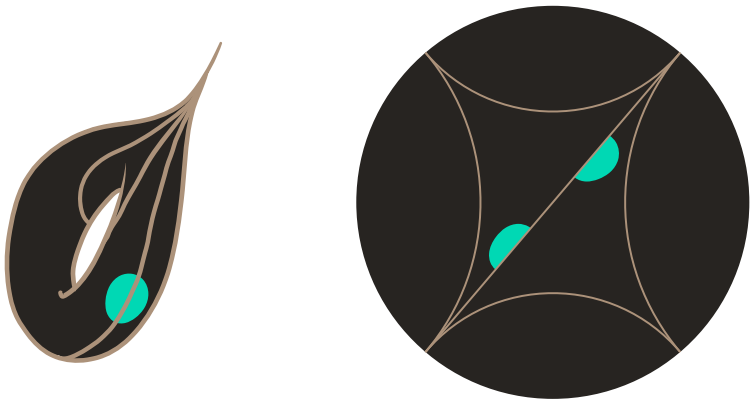


Cut along the edges of the triangulation, so S' falls apart into ideal triangles. Shift the triangles adjacent to edge e by the displacement l_e. The sign of the displacement determines the direction of the shift: when l_e is positive, someone standing on either triangle sees the other triangle move left. Glue the triangles back together, making a new surface S'' with a new hyperbolic structure. This new object is the outcome of the cataclysm.



Because we're assuming the weights of the triangulation sum to zero around each puncture, the hyperbolic structure on S'' has a cusp at each puncture, just like the structure on S' that we started with.

The new surface S'' is homeomorphic to S', but not in any canonical way. Each triangle of S'', however, is canonically isometric to the corresponding triangle of S'. This creates a new partially anchored hyperbolic structure on S', with local charts defined away from the edges of the triangulation. The new structure easier to see than it is to say, so before we go into the details of its construction, let's look at the image of one of its local charts.



The chart restricts to a chart for the original hyperbolic structure on each triangle. However, it doesn't extend continuously over edges.

Now for the details. Consider a quadrilateral $Q' \subset S'$—the union of two adjacent triangles and the edge e between them. Let Q'' be the corresponding quadrilateral in S'', and let c be the map from $Q' \smallsetminus e$ to Q'' that restricts to the canonical isometry on each triangle. Define a local chart on Q' to be an isometry $Q'' \to \mathbb{H}^2$ pulled back along c. Since local charts on Q' are maps from $Q' \smallsetminus e$ to $\mathbb{H}^2$, they can be restricted to open subsets of Q', forming a constant sheaf. The sheaves of local charts on different quadrilaterals agree on overlaps, so they glue up into a partially anchored hyperbolic structure on S'.

5.2.3 The deviation of a cataclysm

Let $\mathcal{E}'$ be the catalog of the original hyperbolic structure on S', and let $\mathcal{E}''$ be the catalog of the new hyperbolic structure produced by the cataclysm. On an open set $U \subset S'$ that doesn't cross any edges of the triangulation, the

new local charts are the same as the original ones, so $\mathcal{E}''_U$ and $\mathcal{E}'_U$ are identified. This defines a stalkwise isomorphism $K\colon \mathcal{E}' \to \mathcal{E}''$ on the complement of the edges of the triangulation. Let's measure its deviation κ.

Consider an open set $U \subset S'$ which is split in two by an edge e of the triangulation, like the one pictured a few paragraphs ago. Since U is a connected subset of a quadrilateral, it's simple with respect to $\mathcal{E}'$ and $\mathcal{E}''$, so the deviation gives an automorphism κ^U_{yx} of $\mathcal{E}'_U$ for each pair of points $y, x \in U \smallsetminus e$. If y and x lie on the same side of e, then κ^U_{yx} is the identity. If y and x lie on opposite sides of e, then κ^U_{yx} shifts each local chart $\phi\colon U \to \mathbb{H}^2$ along the geodesic containing ϕe. The displacement is l_e, as measured by someone standing at ϕx.

Notice that κ^U_{yx} doesn't act on local charts by postcomposing them with a fixed isometry of $\mathbb{H}^2$. No automorphism of $\mathcal{E}'_U$ can act that way, because an automorphism of a torsor is an equivariant map, and the group action that defines the torsor isn't equivariant unless the group is abelian.

We first defined a cataclysm as a cutting and gluing operation that turns each hyperbolic structure on S' into new hyperbolic surface S''. Then we worked our way back to S', building a new partially anchored hyperbolic structure and measuring its deviation from the original. In the end, though, our description of κ came directly from the original hyperbolic structure on S'. That means we could just as well define a cataclysm as an operation that turns each hyperbolic structure on S' into a partially supported deviation, which specifies a new partially anchored hyperbolic structure.

5.2.4 Cataclysms on surfaces with holes

So far, we've only considered weighted triangulations in which the sum of the weights around each puncture is zero. Let's relax that restriction, demanding only that the sum of the weights around each puncture is non-negative. Now a cataclysm can open cusps like camera apertures, turning them into holes with geodesic boundary. Triangulations whose edge weights are all non-negative automatically satisfy the weakened sum constraint. They describe a special class of cataclysms, called *left earthquakes* [18].

5.3 Cataclysms on compact hyperbolic surfaces

5.3.1 Generalizing weighted ideal triangulation

If you want to do a cataclysm on a punctured hyperbolic surface, your first step is to cut the surface into ideal triangles. If you want to do a cataclysm on a compact hyperbolic surface, you're stuck, because there isn't any obvious way to cut the surface into ideal triangles. There is, however, a delightfully non-obvious way to cut the surface into ideal triangles, and you can use it to do a cataclysm after all [18][17].

On a compact surface S with negative Euler characteristic, the analogue of an ideal triangulation is something called a maximal geodesic lamination. The easiest way to get your hands on one is to pick a hyperbolic structure on S. Then you can define a *geodesic lamination* on S to be a set of disjoint simple geodesics whose union is closed [19, Proposition 1]. The geodesics are called the *leaves* of the lamination. A geodesic lamination is *maximal* if no

other geodesic lamination contains it. We'll use other terms from [19] without repeating their definitions. The facts below, unless otherwise cited, can be found in or inferred from the same reference.

A maximal geodesic lamination is a very strange thing. Its complement is a finite set of disjoint ideal triangles. If it has no isolated leaves, however, its leaf set is uncountable. This is easier to picture once you learn that a transverse slice across a geodesic lamination with no isolated leaves is a Cantor set [19, Proposition 7]. For a maximal lamination, the gaps in the Cantor set are slices of the complementary triangles, which cross the transversal over and over as they wind around the surface.

A Cantor set has two kinds of points: the countably many points on the boundaries of the gaps, which I'll call the *boundary points*, and the uncountably many others, which I'll call the *bulk points*. Accordingly, a maximal geodesic lamination with no isolated leaves has two kinds of leaves: the finitely many leaves on the boundaries of the complementary triangles, which I'll call the *boundary leaves*, and the uncountably many other leaves, which I'll call the *bulk leaves*.

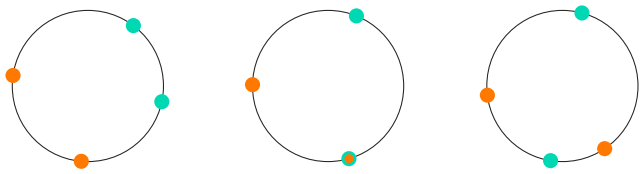
Here's a maximal geodesic lamination with no isolated leaves on a genus-2 surface.

The details of the leaves are notoriously hard to see in pictures like these, as discussed in [19, "A more explicit example"]. The complementary triangles, however, are clearly visible.

The leaves of a geodesic lamination often bunch up so closely that it doesn't make sense to assign them individual weights. Instead, more generally, we assign weights to swaths of leaves, using a gadget called a *transverse cocycle*. Transverse cocycles with non-negative weights are countably additive, so we call them *transverse measures* [20].

Geodesic laminations, in spite of the name, are purely topological objects. Here's a sketch of how to define a geodesic lamination without picking a hyperbolic structure, taken from the end of [6, §8.5]. Recall that a geodesic on S is specified by a pair of distinct points on the boundary of the universal cover $\tilde{S} \to S$. Two pairs define the same geodesic if and only if they're related by the action of $\pi_1 S$ on $\partial\tilde{S}$. A geodesic lamination is specified by a closed set λ of pairs of distinct points in $\partial\tilde{S}$, invariant under the action of $\pi^1 S$. Because

the leaves of a geodesic lamination are disjoint, no two pairs in λ are crossed in the way pictured below.



Uncrossed Uncrossed Crossed

A closed set of mutually uncrossed pairs of distinct points in $\partial\tilde{S}$ specifies a geodesic lamination. Intuitively, each pair specifies a homotopy class of curves in S relative to the boundary at infinity. These curves snap tight to geodesics when you pick a hyperbolic structure on S.

5.3.2 Performing cataclysms

Let S be a compact surface with negative Euler characteristic. Take a maximal geodesic lamination on S, with no isolated leaves, and equip it with a transverse cocycle. This is analogous to choosing a weighted ideal triangulation on a punctured surface. In the punctured case, a weighting specifies a cataclysm if the sum of the weights around each puncture is non-negative. In the compact case, a transverse cocycle specifies a cataclysm if it lies in a certain "non-negative cone" in the vector space of transverse cocycles. The cone is the one described in Theorem B of [20], but with $\geq$ in place of $>$. Transverse measures automatically satisfy the non-negativity constraint. The cataclysms they describe, like before, are called left earthquakes [18].

The cataclysm specified by a transverse cocycle is tricky to carry out. The method Bonahon uses to do it in [20, §5] is an example of warping in disguise. Bonahon works on the universal cover $\tilde{S} \to S$, where the whole space is simple with respect to the catalog $\mathcal{E}$ of the hyperbolic structure. Isometries of $\tilde{S}$ act on local charts by precomposition, which commutes with the postcomposition action of $\mathrm{PSL}_2 \mathbb{R}$, so an isometry of $\tilde{S}$ gives a torsor automorphism of $\mathcal{E}_{\tilde{S}}$. Bonahon constructs a family of isomorphisms φ_{PQ} of $\tilde{S}$, which specify a deviation from $\mathcal{E}$ supported away from the geodesic lamination. For points x, y lying in plaques $P, Q \subset \tilde{S}$, the deviation automorphism $\varphi_{xy}^{\tilde{S}}$ is given by φ_{PQ}.

Chapter 6

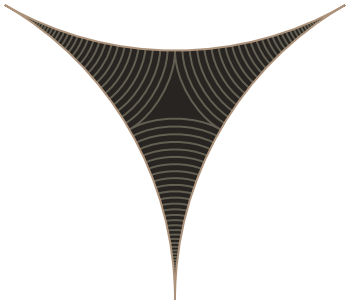
Abelianizing geometric structures

6.1 Shear parameters for cusped hyperbolic surfaces

The cataclysms we studied in Section 5 are related to abelianization through W. Thurston's *shear parameterization* of the space of hyperbolic structures on a surface. Shear parameters have many remarkable properties and applications [21][22][16], but we'll satisfy ourselves with a quick definition.

As we did in Section 5.2, take a punctured surface S' with negative Euler characteristic and equip it with an ideal triangulation. When we put a cusped hyperbolic structure on S', the edges of the triangulation snap to geodesics, and the triangles become isometric to ideal triangles in the hyperbolic plane.

The corners of an ideal triangle come intrinsically foliated by horocycles. The foliations meet around a central region I'll call the *contact triangle*, after its Euclidean analogue.



Adjacent ideal triangles are described completely, up to isometry, by the displacement between the corners of their contact triangles.

For each edge e of the triangulation of S', let $x_e \in \mathbb{R}$ be the displacement of the ideal triangles adjacent to e. Following the convention of Section 5.2, we say x_e is positive when someone standing at the corner of one contact triangle can look across e and see the other contact triangle to the left.

The numbers $X_e = \exp x_e$ are the *shear parameters* of the hyperbolic structure on S'. They tell you how to piece S' together out of ideal triangles, determining its hyperbolic structure up to isotopy. By changing the cusped hyperbolic structure, you can set the log shear parameters x_e to any values

that sum to zero around each puncture. If you allow hyperbolic structures with holes as well as cusps, you can set the x_e to any values that sum to something non-negative around each puncture.

The log shear parameters form a weighting of the ideal triangulation they come from. If you take a hyperbolic structure with log shear parameters x and do a cataclysm with weighting l, you get the hyperbolic structure with shear parameters $x + l$.

6.2 Shear parameters for compact hyperbolic surfaces

In Section 5.3, we adapted the notion of a cataclysm from punctured surfaces to compact ones by generalizing from ideal triangulations to maximal geodesic laminations. In this section, we'll do the same for the shear parameterization [22, §§3 – 4].

Take a compact surface S with negative Euler characteristic and equip it with a maximal geodesic lamination with no isolated leaves. When we put a hyperbolic structure on S, the leaves of the lamination snap to geodesics, and the complement of the lamination becomes a finite set of disjoint ideal triangles. We'd like to measure the displacements between adjacent triangles, but there aren't any: two adjacent triangles would be separated by an isolated leaf. Instead, more generally, we assign displacements to swaths of leaves, as described at the end of [22, §3.2]. Shear parameters on a compact surface form a transverse cocycle, which Bonahon calls the *shearing cocycle* [20].

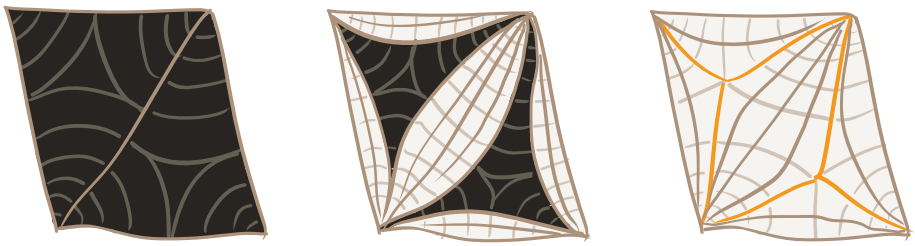
As the hyperbolic structure on S varies, the shear parameters range over a certain "positive cone" in the vector space of transverse cocycles [20, Theorem B]. If you take a hyperbolic structure with shearing cocycle x and do a cataclysm with transverse cocycle l, you get the hyperbolic structure with shearing cocycle $x + l$ [20, §0].

6.3 Abelianization on punctured surfaces

6.3.1 A geometric realization of the shear parameters

On a punctured surface with an ideal triangulation, you can compare two hyperbolic structures by taking the difference of their shear parameters. A cataclysm realizes the difference as a densely defined isometry. Another cutting and gluing process, which I'll call *deflation*, provides an embodiment of the shear parameters themselves.

Let S' be a punctured surface with a positively weighted ideal triangulation. Pick a cusped hyperbolic structure on S', snapping the edges of the triangulation to geodesics. Each edge is now isometric to a vertical line in $\mathbb{R}^2$. Expand each edge to a vertical strip whose width is the edge weight. Then, collapse each triangle to a tripod, leaving the vertical strips glued along their edges.



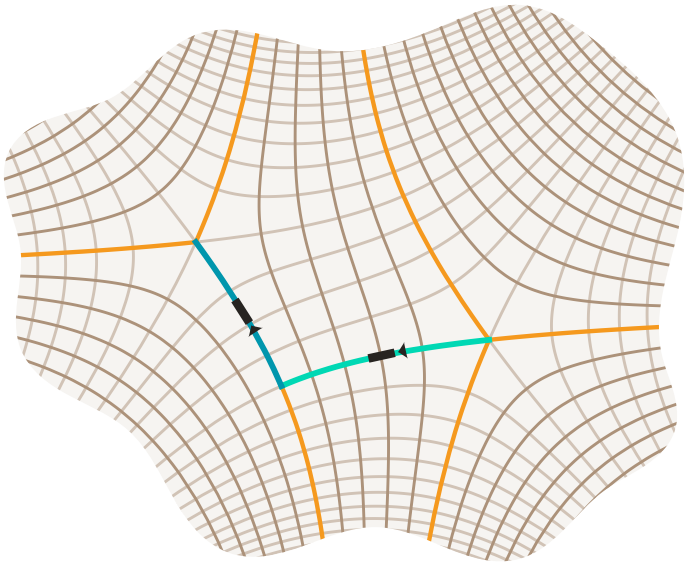
We've turned the triangulated hyperbolic surface S' into a flat surface S^{ab}, with a conical singularity at the center of each collapsed triangle. We built S^{ab} out of vertical strips in $\mathbb{R}^2$, gluing them along their edges using translations and 180° rotations. Hence, S^{ab} comes with a half-translation structure—one of the geometric structures we discussed in Section 2.2. As we mentioned there, a half-translation structure provides a way to distinguish vertical lines from other curves. The vertical lines coming out of the conical singularities are the boundaries of the strips we glued together to build S^{ab}. The strips aren't just an artifact of the construction; they're an intrinsic feature of the half-translation surface, which we'll discuss further in Section 12. Looking at S^{ab}, we can read off the log shear parameter x_e by finding the strip associated with the edge e and measuring the vertical displacement between the singularities on its edges.

6.3.2 Shear parameters as periods

It's easy enough to measure the shear parameters when you're looking at an aerial photograph of S^{ab}, as we did in the previous section. A more challenging task, and a very rewarding one, is to measure the shear parameters

from the ground.

Since the shear parameters are given by displacements on S^{ab}, our first thought might be to find them using a surveyor's wheel, an old and venerable instrument for measuring displacements on Riemannian surfaces. A surveyor's wheel is just a wheel on a stick. When it's in contact with a surface, its wide, grippy treads force it to roll in a straight line. The net angle the wheel rolls through measures displacement along the line. You can measure the vertical displacement between two singularities by rolling a surveyor's wheel along the piecewise geodesic path shown below, recording only the displacement for the vertical segment.[1]

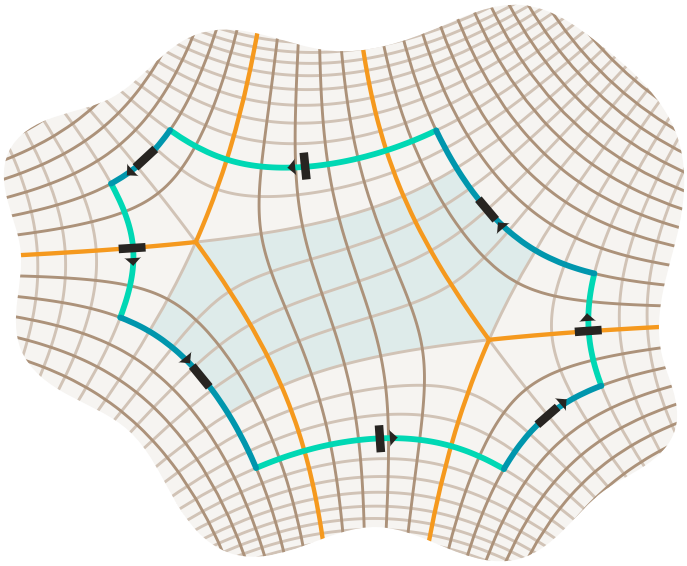


The downside of this method is that you have to start and end exactly on the singularities, inviting both error and potential injury. A safer and more

[1]Many thanks to Andy Neitzke for helping prepare the figures for this section.

forgiving method uses an improved version of the surveyor's wheel, which you might call a surveyor's omniwheel. An omniwheel is a wheel ringed with little rubber rollers, which allow it to slip sideways without turning. A surveyor's omniwheel can roll along any path, geodesic or not. It only records displacement along the direction the wheel is pointing. On a flat surface, if you fix the initial direction of the wheel, the measured displacement only depends on the homotopy class of the path relative to its endpoints.

You can measure the vertical displacement between two singularities by rolling a surveyor's omniwheel around the loop shown below, or any loop in S^{ab} minus the singularities which is homotopic to it.



On the vertical segments of the loop, you alternate between pushing and pulling the wheel, so the displacements you record are alternately positive and negative. You somehow end up pushing the wheel along both edges of

the shaded region, so the displacements you record there have the same sign. The rest of the displacements cancel out, leaving a net displacement of $-2x_e$ around the whole loop.

To understand the cancellations that make this method work, it was essential to keep track of which way the omniwheel was pointing. On a half-translation surface, there's an intrinsic notion of vertical and horizontal, and an omniwheel that starts off vertical will always stay that way. A vertical line has two orientations, though, and rolling the omniwheel around a loop can switch it from one to the other. The two vertical orientations form a double cover of S^{ab} away from the singularities. Around each singularity, the double cover looks like a punctured disk, so we can fill in the missing point to get a surface Σ^{ab} and a branched covering map $\Sigma^{\mathrm{ab}} \to S^{\mathrm{ab}}$. We'll construct Σ^{ab} more formally in Section 9.2, where we name it the *translation double cover* of the half-translation surface S^{ab}. Gaiotto, Moore, and Neitzke construct Σ^{ab} in a different way, calling it the *spectral curve*. (The two constructions don't agree in general, but they do agree for the three-pronged singularities we're considering here.) The surface Σ^{ab} comes with a 1-form that tells you how the reading on the surveyor's omniwheel will change if you roll it with a given velocity. When you roll the omniwheel around on S^{ab}, you're integrating the 1-form along your lifted path in Σ^{ab}.

By working out a good way to measure the shear parameters from the ground, we've realized those parameters as the periods of a 1-form on a branched double cover of S^{ab}. Both the cover and the 1-form are determined

by the half-translation structure of S^{ab}.

6.3.3 Shear parameters as holonomies

Introduction The deflation process described above turns a hyperbolic structure on S' into a translation surface Σ^{ab}. On the level of catalogs, it turns a $\mathrm{PSL}_2\,\mathbb{R}$ local system $\mathcal{E}$ on S' into a surface Σ^{ab} equipped with an $\mathbb{R}^2$ local system.[2] The log shear parameters x_e, as we saw, are given by certain holonomies of the vertical part of the catalog for the translation structure on Σ^{ab}. Using the exponential map $\mathbb{R} \to \mathbb{R}^\times$, we can turn that vertical part into an $\mathbb{R}^\times$ local system $\mathcal{E}^{\mathrm{ab}}$ on Σ^{ab}, whose holonomies give the shear parameters X_e.

In our construction of $\mathcal{E}^{\mathrm{ab}}$, we took advantage of the anchor map that came with the catalog $\mathcal{E}$, using local charts to talk about distances, geodesics and ideal triangles on S'. However, it's also possible to get $\mathcal{E}^{\mathrm{ab}}$ directly from the unmoored hyperbolic structure $\mathcal{E}$, using the abelianization process of Gaiotto, Hollands, Moore, and Neitzke. This version of the construction extends from the catalogs of hyperbolic structures to generic $\mathrm{PSL}_2\,\mathbb{R}$ and $\mathrm{PSL}_2\,\mathbb{C}$ local systems on S'. Just as deflation encodes the shear parameterization for hyperbolic structures, abelianization encodes an interesting coordinate system for $\mathrm{PSL}_2\,\mathbb{C}$ local systems, developed by Fock and Goncharov [23, §5].

In this section, I'll sketch out a route from the deflation process, which acts on geometric objects, to the abelianization process, which operates purely

[2] As we pointed out in Section 2.2.3, the catalog extends over the singularities, because its holonomy around each singularity is the identity.

on the level of catalogs. The description of abelianization we'll arrive at won't be the simplest or the most vivid, but its relationship with deflation should be clear. In Section 12.1, we'll review abelianization again from a more concrete point of view, taking advantage of the fact that a $\mathrm{PSL}_2\,\mathbb{C}$-torsor can be realized as the set of line decompositions of a two-dimensional vector space with a volume form. I find the concrete formulation easier to grasp, but its relationship with deflation is obscured.

Why abelianization is possible The reason we can get $\mathcal{E}^{\mathrm{ab}}$ directly from $\mathcal{E}$ is that the vertical part of each translation chart on Σ^{ab} factors through a unique hyperbolic chart on S'. To see how this works, recall that a vertical chart on a translation surface is a map to $\mathbb{R}$ that locally preserves vertical displacements. Fixing an isometry $\mathbb{R} \to \mathbb{H}^2$, we identify $\mathbb{R}$ with an oriented geodesic γ, and we can treat vertical charts as maps to γ.

Each strip $Y \subset \Sigma^{\mathrm{ab}}$ comes from an edge e of the triangulation on S'. Since Y was made by expanding e horizontally, there's a map $\eta\colon Y \to e$ that collapses it back down. Every vertical chart on Y factors through η via a unique isometry $e \to \gamma$, identifying $\mathcal{E}^{\mathrm{ab}}_Y$ with a subset of the generalized stalk $\mathcal{E}_e$.

We can find $\mathcal{E}^{\mathrm{ab}}_Y$ inside of $\mathcal{E}_e$ using nothing but information from the catalog $\mathcal{E}$. To see how, let P^+ be the subgroup that fixes the head of γ, and let P^- be the subgroup that fixes the tail. The edge e gets an orientation from the vertical orientation of Y, so we can label the punctures it connects as its

head $\mathfrak{q}$ and its tail $\mathfrak{p}$. The charts that send $\mathfrak{q}$ to the head of γ form a P^+-torsor $\mathcal{E}_e^+$ in $\mathcal{E}_e$, which is characterized by its invariance under the holonomy around $\mathfrak{q}$. Similarly, the charts that send $\mathfrak{p}$ to the tail of γ form a P^--torsor $\mathcal{E}_e^-$ in $\mathcal{E}_e$, the only one invariant under the holonomy around $\mathfrak{p}$. The intersection of $\mathcal{E}_e^+$ and $\mathcal{E}_e^-$ is $\mathcal{E}_Y^{\mathrm{ab}}$.

The reasoning above determines $\mathcal{E}^{\mathrm{ab}}$ over each strip of Σ^{ab}. Now we just have to understand how the pieces fit together across strip boundaries. Consider two adjacent strips Y' and Y in Σ^{ab} coming from edges e' and e of the triangulation on S'. Assume, for convenience, that Y' and Y are joined at the top. The edges e' and e bound a common triangle T. According to the orientations induced by Y' and Y, the edges share a head $\mathfrak{q}$, but have different tails $\mathfrak{p}'$ and $\mathfrak{p}$.

Because e' and e lie on the boundary of T, and T is simple with respect to $\mathcal{E}$, the generalized stalks $\mathcal{E}_{e'}$ and $\mathcal{E}_e$ are both canonically identified with $\mathcal{E}_T$. As subsets of $\mathcal{E}_T$, the P^+-torsors $\mathcal{E}_{e'}^+$ and $\mathcal{E}_e^+$ coincide, but the P^--torsors $\mathcal{E}_{e'}^-$ and $\mathcal{E}_e^-$ are typically different.

To make Σ^{ab}, we collapsed the triangle T along its horocycle foliation, gluing e to e' along the isometry $e \to e'$ that preserves horocycles around $\mathfrak{q}$. This isometry corresponds to an automorphism of $\mathcal{E}_T$, the one that sends $\mathcal{E}_e^-$ to $\mathcal{E}_{e'}^-$ and restricts to the identity on $\mathcal{E}_e^+$. The automorphism tells us how the pieces of $\mathcal{E}^{\mathrm{ab}}$ fit together across the boundary between Y' and Y. It can be interpreted as the transition map from $\mathcal{E}_Y^{\mathrm{ab}} \cong \mathcal{E}_e$ to $\mathcal{E}_{Y'}^{\mathrm{ab}} \cong \mathcal{E}_{e'}$ given by parallel transport across the boundary.

All the reasoning above went on under the assumption that Y' and Y are joined at the top. Similar reasoning works for strips joined at the bottom.

The abelianization process We just found a procedure for building $\mathcal{E}^{\mathrm{ab}}$ from the catalog $\mathcal{E}$. If we strip out the reasoning that justifies each step, we're left with the following instructions.

1. Construct Σ^{ab}, as a topological surface, from the combinatorics of the ideal triangulation on S'. Notice that Σ^{ab} still comes with a decomposition into strips.

2. Fix an oriented geodesic γ in $\mathbb{H}^2$. Let P^+ and P^- be the subgroups of $\mathrm{PSL}_2\,\mathbb{R}$ that fix its head and tail, respectively.

3. Over each strip $Y \subset \Sigma^{\mathrm{ab}}$, let $\mathcal{E}^{\mathrm{ab}}$ be the constant sheaf with value $\mathcal{E}_e$, where e is the edge Y came from.

4. Given two adjacent strips $Y', Y \subset \Sigma^{\mathrm{ab}}$ which are joined at the top, carry out the following steps.

    4.1. Let T be the common triangle bounded by e' and e.

    4.2. Using the fact that e' and e lie on the boundary of T, canonically identify $\mathcal{E}_{e'}$ and $\mathcal{E}_e$ with $\mathcal{E}_T$.

    4.3. Use the vertical orientations of the strips Y' and Y to orient the corresponding edges e' and e.

4.4. Let $\mathfrak{q}$ be the puncture at the shared head of e' and e, and let $\mathfrak{p}'$ and $\mathfrak{p}$ be the punctures at the tails.

4.5. Let $\mathcal{E}_e^+$ be the unique P^+-torsor in $\mathcal{E}_T$ invariant under parallel transport around $\mathfrak{q}$.

4.6. Let $\mathcal{E}_{e'}^-$ and $\mathcal{E}_e^-$ be the unique P^--torsors in $\mathcal{E}_T$ invariant under parallel transport around $\mathfrak{p}'$ and $\mathfrak{p}$, respectively.

4.7. Across the boundary between Y' and Y, set the transition map from $\mathcal{E}_{Y'}^{\mathrm{ab}} = \mathcal{E}_{e'}$ to $\mathcal{E}_Y^{\mathrm{ab}} = \mathcal{E}_e$ to be the automorphism of $\mathcal{E}_T$ that sends $\mathcal{E}_e^-$ to $\mathcal{E}_{e'}^-$ and restricts to the identity on $\mathcal{E}_e^+$.

5. Carry out the analogous steps for adjacent strips joined at the bottom.

This is the abelianization process of Gaiotto, Hollands, Moore, and Neitzke [1, §4][11, §10]. You can compare it to the more elementary formulation in Section 12.1, which replaces all the torsor gadgetry with linear algebra.

6.3.4 Ideal triangulations and half-translation structures

To make firm contact with Gaiotto, Hollands, Moore, and Neitzke's formalism for abelianization, we need to point out a subtle distinction. If you carry out the abelianization process described here, you end up with an abelianized local system $\mathcal{E}^{\mathrm{ab}}$ on a topological surface Σ^{ab}, which you pieced together according to the combinatorics of the ideal triangulation on S'. Although Σ^{ab} is homeomorphic to a branched double cover of S', it doesn't come equipped with a covering map.

Gaiotto, Hollands, Moore, and Neitzke build the abelianized local system on a translation surface Σ' that does come with a branched covering map to S'. They get it by realizing the ideal triangulation on S' as the WKB triangulation of a half-translation structure [23, §6]. Then they define Σ' as the translation double cover of S'. Although the auxiliary half-translation structure carries more information than the ideal triangulation, the extra information doesn't affect the abelianization process in any significant way.

Using the covering map $\Sigma' \to S'$, we can push $\mathcal{E}$ back to Σ' in one piece, instead of carrying it over strip by strip like we did on Σ^{ab}. The abelianization process on Σ' therefore involves cutting as well as gluing along the boundaries between the strips.

On a punctured surface, the choice between ideal triangulations and half-translation structures is just a formality. On a compact surface, the choice matters: geodesic laminations and half-translation structures feel quite different to work with, and passing between them is not straightforward. All the technical work of this paper will be done on the half-translation side.

6.4 Abelianization on compact surfaces

In the original account of abelianization, Gaiotto, Moore, and Neitzke suggested that the role of the punctures in their construction could be removed, leading to an abelianization process for compact surfaces [11, open problem 7]. In hindsight, their conjecture was already supported by at least two clues in the geometric topology literature. First, in [20], Bonahon had generalized

the shear parameterization to pleated surfaces—deformed hyperbolic surfaces whose geometries are described by a special class of $\mathrm{PSL}_2\,\mathbb{C}$ local systems. Second, Casson and Bleiler had described an operation on compact surfaces which strongly resembles our deflation process for punctured surfaces [24, proof of Lemma 6.2]. As it turns out, Casson and Bleiler's construction can indeed be adapted to produce a deflation process for compact surfaces, which takes in a hyperbolic structure and spits out a compact half-translation surface whose geometry encodes the shear parameters.[3]

The main result of this paper is to confirm that abelianization extends to a certain class of $\mathrm{SL}_2\,\mathbb{R}$ local systems on compact surfaces. Speficially, it extends to local systems with a dynamical property called *uniform hyperbolicity*, discussed in Section 11. The result and its proof should carry over to $\mathrm{SL}_2\,\mathbb{C}$ local systems, but it hasn't been fully checked in that case, as discussed in Section 7.2.

Uniform hyperbolicity generalizes a dynamical condition called the Anosov property, which plays an important role in the study of geometric structures on surfaces [25]. Because abelianization extends to local systems which are uniformly hyperbolic but not Anosov, it might be useful for studying the boundary of the region where the Anosov property holds.

[3]This construction seems to be well-known in some circles, but I haven't been able to find a printed account of its details. I hope to produce one in the course of future work.

Part II

Abelianization on compact surfaces

Chapter 7

Introduction

7.1 Results

In this part of the paper,[1] we'll learn how to abelianize uniformly hyperbolic $\mathrm{SL}_2\,\mathbb{R}$ local systems on a compact translation surface Σ, as promised in Section 6.4. The translation surface Σ is analogous to the translation double cover Σ' that appears in Section 6.3.4, although it doesn't have to be a translation double cover itself.

From a naive point of view, the abelianization process we'll use is essentially the same as the cutting and gluing construction Gaiotto, Hollands, Moore and Neitzke use on Σ', which we review in Section 12.1. The catch is that the cutting and gluing will have to be done along lines that fill Σ densely. If you've read Sections 4.3 – 4.4 and 5 of Part I, you've already seen some of the tools we'll use to make sense of this. The bulk of Part II will be spent setting up those tools, collecting more tools, and working out the conditions we need for all our equipment to work reliably.

The main result, with all the necessary tools and conditions accounted for, is summarized below. Its statement is very condensed, making it look

[1] Part II is a revised and expanded version of the preprint [15].

intimidatingly remote, but the trip from here to there will hopefully feel more like a long hike up a gentle slope than a short climb up a sheer cliff. Words and ideas that won't be introduced until later along the route are tagged with references to the relevant sections. You can get an idea of what those sections are about from the table of contents in Section 7.3.

Theorem 7.1.A (Sections 12.2 – 12.3)**.** *Let Σ be a compact translation surface (9.2.1) with generic dynamics (12.2), and let $\mathfrak{B}$ be its finite set of singularities. Let $\overleftrightarrow{\Sigma}$ be the associated divided surface (9), whose category of $\mathrm{SL}_2\mathbb{R}$ local systems is equivalent to the category of $\mathrm{SL}_2\mathbb{R}$ local systems on $\Sigma \smallsetminus \mathfrak{B}$ (Theorem 9.4.E).*

Given a uniformly hyperbolic (11) $\mathrm{SL}_2\mathbb{R}$ local system $\mathcal{E}$ on $\overleftrightarrow{\Sigma}$, we can find a new $\mathrm{SL}_2\mathbb{R}$ local system $\mathcal{F}$ and a stalkwise isomorphism $\Upsilon\colon \mathcal{E} \to \mathcal{F}$, supported on a dense subspace of $\overleftrightarrow{\Sigma}$, with the following properties:

- *The deviation (4.3 – 4.4, 8.3) of Υ behaves like the deviation of an abelianized local system from its original would behave on a punctured surface (12.1).*
- *The local system $\mathcal{F}$ splits into a direct sum of $\mathbb{R}^\times$ local systems.*

Proof. Sections 13 – 14. □

The most striking new feature of this result is the appearance of *uniform hyperbolicity*, a dynamical condition that plays no noticeable role in abelianization on punctured surfaces. The need for "generic dynamics" on Σ is also

new: the corresponding requirement in the punctured case is purely topological. As we'll see, abelianization on a compact surface involves dynamics in an essential way. This raises some interesting questions. On a punctured surface, for instance, deciding which local systems can be abelianized is pretty straightforward. The set of abelianizable local systems is open and dense, with a simple shape described by the opening sentence of Section 12.1.3. On a compact surface, deciding which local systems are uniformly hyperbolic is much trickier. The set of uniformly hyperbolic local systems is open, but not expected to be dense, and the descriptions I've seen of it suggest that its shape could be complicated. Section 11.6 and its main reference [26] say more about this.

7.2 Why not $\mathrm{SL}_2\,\mathbb{C}$?

I expect the results of this paper to generalize from $\mathrm{SL}_2\,\mathbb{R}$ to $\mathrm{SL}_2\,\mathbb{C}$ local systems. Outside of Section 11.6, where there may be some subtlety in saying what it means for an $\mathrm{SL}_2\,\mathbb{C}$ cocycle to be eventually positive, the generalization should amount to little more than declaring all the vector spaces to be complex. In our arguments about dynamical cocycles, however, we'll use results found in [26], [27], and [28], which only discuss the $\mathrm{SL}_2\,\mathbb{R}$ case. Until it can be verified that these results apply to $\mathrm{SL}_2\,\mathbb{C}$ cocycles, the results of this paper can only be stated with confidence for $\mathrm{SL}_2\,\mathbb{R}$ local systems.

7.3 Contents

Introduction

Chapter 7 Summary of results and various administrative things, including notation that will be used throughout this part of the paper.

Tools

Chapter 8 Abstracting from the idea of deforming a flat connection on a smooth bundle, we get a more general way of deforming local systems, called *warping*.

Chapter 9 After a brief introduction to translation surfaces, we describe a way to enlarge a singular translation surface by splitting its critical leaves. We show that the resulting *divided surface* has useful topological and dynamical properties, and that it resembles the original surface both dynamically and in terms of its local systems.

Chapter 10 We show how the warping process from Section 8 can be used to make sense of the idea of cutting and gluing a local sytem along the critical leaves of a divided surface, even when the critical leaves fill the surface densely.

Chapter 11 We take a well-studied condition on dynamical cocycles, called *uniform hyperbolicity*, and reinterpret it as a condition on local systems on compact translation surfaces.

Abelianization

Chapter 12 We review how abelianization works for $\mathrm{SL}_2\,\mathbb{C}$ local systems on a translation surface with punctures, point out the obstacles to carrying it out a compact translation surface, and describe how these obstacles can be overcome using the machinery from the previous sections. Section 12.3 contains the main product of the paper: instructions for abelianizing an $\mathrm{SL}_2\,\mathbb{R}$ local system on a compact translation surface, which are guaranteed to work under the conditions laid out in Section 12.2.

Chapter 13 We show that abelianization, as defined by the instructions in Section 12.3, produces a well-defined local system, assuming the conditions from Section 12.2.

Chapter 14 We show that the abelianized local system splits into a direct sum of $\mathbb{R}^\times$ local systems, assuming the conditions from Section 12.2.

Chapter 15 A non-rigorous sample computation that uses abelianization to find holomorphic coordinates on the $\mathrm{SL}_2\,\mathbb{C}$ character variety of the punctured torus.

Future directions

Chapter 16 We discuss a few of the interesting features that abelianization on a compact surface is expected to have, now that we know it can be done.

Appendices

Appendix A A pair of small technical lemmas for Section 8.

Appendix B A formalism for dynamical systems described by relations, used throughout the paper.

Appendix C Results about uniformly hyperbolic cocycles over a minimal, uniquely ergodic dynamical system, used at key points in Sections 11, 13, and 14.

Appendix D Results on infinite ordered products, used heavily in Sections 10, 13, and 14.

Appendix E Linear algebra facts about the Euclidean plane, used in Section 14.

Appendix F A list of standard puncture shapes for translation surfaces and an explanation of where they come from, included to clarify the review in Section 12.

7.4 Setup

7.4.1 Running notation

The terminology of this section hasn't been introduced yet, but will be familiar to readers familiar with translation surfaces. If you'd like to become familiar with translation surfaces, skip ahead to Section 9.2.

From now on, Σ will be a compact translation surface, $\mathfrak{B}$ its set of singularities, and $\mathfrak{W}$ the union of its critical leaves. Within $\mathfrak{W}$, let $\mathfrak{W}^+$ and

$\mathfrak{W}^-$ be the unions of the backward- and forward-critical leaves, respectively. Saying that Σ has no saddle connections is the same as saying that $\mathfrak{W}^+$ and $\mathfrak{W}^-$ are disjoint. The $\pm$ labeling is meant to evoke the fact that, in the absence of saddle connections, the vertical flow is well-defined on $\mathfrak{W}^+$ for all positive times, and on $\mathfrak{W}^-$ for all negative times.

I should stress that Σ doesn't need to be the translation double cover of a half-translation surface, and there are nice examples of abelianization where it isn't. One of these is discussed in Section 15. There are many special properties abelianization is expected to gain when Σ is a translation double cover, but our discussion of them will be limited to the speculative Section 16.

7.4.2 Index of symbols

Symbols can be hard to look up, so here's a list of unusual symbols that appear frequently in this paper, with references to the sections where they're defined. The first three are introduced in this paper, and the fourth is common, but not universal, in analysis.

- $\overleftrightarrow{\square}$ Divided interval or surface (Sections 9.3.1 and 9.4.1).
- $\langle\square\rangle$ Fractured interval or surface (same sections).
- $\square\!\!-\!\!-$ Intersection with fractured interval or surface (Section 9.4.2).
- $\square \lesssim \square$ Bounded by a constant multiple (Section 11.1).

Chapter 8

Warping local systems

8.1 Overview

In this section, we'll finish developing the tools for measuring and constructing locally constant sheaves that we introduced in Sections 4.3 – 4.4 and 5.

8.2 Conventions for local systems

8.2.1 Basics

Given a group G, we define a G local system to be a locally constant sheaf of G-sets whose stalks are all G-torsors. Until Section 11, it won't matter much what G is, so we'll often just talk about local systems in general.

The category of G-sets is a nice target category for sheaves, because it's a type of algebraic structure [29, Tag 007L]. Throughout this article, "sheaf" will mean a sheaf whose target category is a type of algebraic structure. With this said, we can define a constant sheaf to be a sheaf of locally constant functions—that is, functions constant on every connected component of their domain. In a locally connected space, every connected component of an open set is open, so the constant sheaf with value A is characterized by the property

that it sends every connected, non-empty open set to A, and every inclusion of such sets to 1_A.

If $\mathcal{F}$ is a constant sheaf on a locally connected space, the stalk restriction morphism $\mathcal{F}_{x\in X}\colon \mathcal{F}_x \leftarrow \mathcal{F}_X$ is an isomorphism for every $x \in X$. In our context, the converse is true as well:

Proposition 8.2.A. *Suppose $\mathcal{F}$ is a sheaf on a locally connected space X. If the stalk restriction $\mathcal{F}_{x\in X}$ is an isomorphism for every $x \in X$, then $\mathcal{F}$ is isomorphic to a constant sheaf.*

Proof. Let $\bar{\mathcal{F}}$ be the constant sheaf with value $\mathcal{F}_X$. For each $U \subset X$, the restriction $\mathcal{F}_{U\subset X}$ gives a morphism $\bar{\mathcal{F}}_U \to \mathcal{F}_U$, and these morphisms fit together into a natural transformation from $\bar{\mathcal{F}}$ to $\mathcal{F}$. This natural transformation induces an isomorphism on every stalk, so it's an isomorphism of the underlying sheaves of sets, and therefore an isomorphism of sheaves of algebraic structures. (Many thanks to Jen Berg for pointing out this argument.) □

To save ink, let's say a connected open subset of a space is *simple* with respect to a sheaf if the restriction of the sheaf to the subset is isomorphic to a constant sheaf. Notice that the value of a G local system on a simple set is a G-torsor.

We'll frequently and without fanfare make use of the fact that a sheaf defined on a basis for a topological space extends uniquely (up to canonical isomorphism) to a sheaf on the full poset of open sets [29, Tag 009H, Lemma 9].

8.2.2 Linear local systems

Let's say G is a linear group—a subgroup of the automorphism group of some finite-dimensional vector space R. Define a *G-structure* on a vector space V to be an isomorphism $V \to R$ modulo postcomposition by G. When a vector space is equipped with a G-structure, we'll call it a *G vector space*. Observe that G-structures pull back along isomorphisms. You can say an isomorphism between G vector spaces is *structure-preserving* if it pulls the G-structure on the target back to the G-structure on the source.

Lots of familiar structures on an n-dimensional complex vector space are examples of G-structures.

- A volume form is an $\mathrm{SL}_n \mathbb{C}$-structure.
- An inner product is a U_n-structure.
- A complete flag is a structure of the upper-triangular subgroup of $\mathrm{GL}_n \mathbb{C}$.
- A line decomposition is a structure of the diagonal subgroup of $\mathrm{GL}_n \mathbb{C}$.

In general, if G is defined as the group of automorphisms of R preserving a certain structure, you can turn around and define that structure as a G-structure.

The isomorphisms $V \to R$ that define a G-structure are interesting in their own right. We'll call them *structured frames*. The structured frames for the familiar G-structures listed above are also familiar objects.

- A unit-volume basis is an $\mathrm{SL}_n\,\mathbb{C}$-structured frame.
- An orthonormal basis is an $\mathrm{SL}_n\,\mathbb{C}$-structured frame.
- A basis subordinate to a complete flag is an upper-triangular-structured frame.
- A basis subordinate to a line decomposition is a diagonal-structured frame.

By definition, the set of structured frames for a G vector space V is the orbit of an isomorphism $V \to R$ under the action of G by postcomposition. Since G acts freely on isomorphisms $V \to R$, it acts freely and transitively on the set of structured frames. In other words, the structured frames form a G-torsor. The set of structured frames is what defines a G-structure in the first place, so there's a pithier way to say it: a G-structure is a G-torsor. That's why we're talking about G-structures in a section on G local systems.

We can define a functor from G vector spaces to G-torsors by sending each G vector space to its G-torsor of structured frames. This functor is an equivalence of categories. If we define a "compound G vector space" to be a formal direct sum of G vector spaces, the structured frames functor should extend to an equivalence between the category of compound G vector spaces and the category of G-sets generated by taking limits and colimits of G-torsors.

Let's say a *linear G local system* is a locally constant sheaf of compound G vector spaces whose stalks are all single G vector spaces. Using the equiv-

alence above, we can realize any G local system as a linear G local system. Later in the paper, when we're dealing exclusively with $\mathrm{SL}_2\,\mathbb{R}$ local systems, let's assume that all our local systems are linear. Concretely, that means we'll be working with locally constant sheaves of two-dimensional real vector spaces with volume forms.

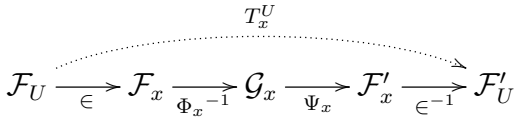
8.3 The descriptive power of deviations

In Section 4.4, we introduced the deviation as a general instrument for comparing locally constant sheaves. I claimed that the deviation of a stalkwise isomorphism Υ determines both Υ and its target up to canonical isomorphism. Now is the time to explain what that means, and to prove that it's true.

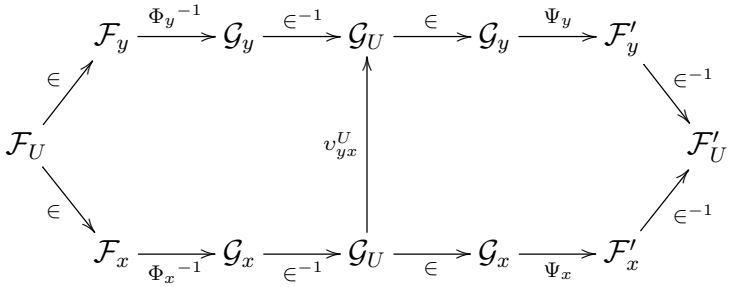
Consider three locally constant sheaves $\mathcal{G}$, $\mathcal{F}$, and $\mathcal{F}'$ on a locally connected space X. If two stalkwise isomorphisms $\Phi\colon \mathcal{G} \to \mathcal{F}$ and $\Psi\colon \mathcal{G} \to \mathcal{F}'$ have the same deviation, υ, I claim there's a unique natural isomorphism $T\colon \mathcal{F} \to \mathcal{F}'$ such that $\Psi_x = T_x\Phi_x$ for all $x \in X$. Here's why.

If we can find a natural isomorphism like this, it's clearly unique, because a natural transformation of sheaves is completely described by its action on stalks. Now, let's find one. Choose a basis $\mathcal{B}$ for the topology of X consisting of sets which are simple with respect to all three sheaves. (We can do this because the sheaves are locally constant, and X is locally conected.) Look at any basis element $U \in \mathcal{B}$. For any $x \in U$, define a morphism $T^U_x\colon \mathcal{F}_U \to \mathcal{F}'_U$

by

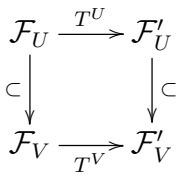


What if we had chosen another point $y \in U$ instead? Let υ be the shared deviation of Φ and Ψ, and consider the diagram



Notice that the bottom path is T_x^U, and the top path is T_y^U. The left and right chambers both commute, because υ is the deviation of both Φ and Ψ, so $T_x^U = T_y^U$ for all $x, y \in U$. Thus, we really have just one morphism $T^U \colon \mathcal{F}_U \to \mathcal{F}'_U$, which can be written in terms of any point in U.

For any other basis element $V \subset U$, writing T^V and T^U in terms of the same point $x \in V$ makes it easy to check that the square



commutes. Since we've been working with arbitrary basis elements, we now see that the morphisms $\{T^U\}_{U \in \mathcal{B}}$ fit together into a natural transformation $T \colon \mathcal{F} \to \mathcal{F}'$, and it's clear by construction that $\Psi_x = T_x \Phi_x$ for all $x \in X$.

8.4 Warping locally constant sheaves

We just saw that, under favorable conditions, a stalkwise isomorphism is determined up to canonical isomorphism by its deviation. Let's see if we can go the other way and produce a stalkwise isomorphism with a specified deviation. First, we have to say what it means to specify a deviation.

Suppose $\mathcal{F}$ is a locally constant sheaf on a locally connected space X, D is a dense subset of X, and $\mathcal{B}$ is a basis for the topology of X consisting of $\mathcal{F}$-simple sets. To specify a *deviation* from $\mathcal{F}$ with support D, defined over the basis $\mathcal{B}$, we give for each pair of points $x, y \in D$ and each neighborhood $U \in \mathcal{B}$ of x and y an automorphism υ_{yx}^U of $\mathcal{F}_U$. These automorphisms have to fit together as follows:

- If $V \subset U$ is a basis element containing x and y, the automorphisms υ_{yx}^U and υ_{yx}^V commute with the restriction morphism $\mathcal{F}_{V \subset U}$.
- For any three points $x, y, z \in D$, we have $\upsilon_{zy}^U \upsilon_{yx}^U = \upsilon_{zx}^U$.

The first condition just says that υ_{yx} is a natural automorphism of the restriction of $\mathcal{F}$ to the poset of basis elements containing both x and y.

It turns out that, given a deviation υ from $\mathcal{F}$, we can always produce a locally constant sheaf $\mathcal{F}'$ and a stalkwise isomorphism $\Upsilon \colon \mathcal{F} \to \mathcal{F}'$, supported on D, whose deviation is υ. We'll call this process *warping*. Here's how it's done. For each $U \in \mathcal{B}$, pick a point $x_U \in U \cap D$. Define $\mathcal{F}'_U$ to be the same as

$\mathcal{F}_U$, but with the warped restriction morphism

$$\mathcal{F}'_{V\subset U} = \mathcal{F}_{V\subset U}\ \upsilon^U_{x_V x_U}$$

for each basis element $V \subset U$. Because every basis element is $\mathcal{F}$-simple, $\mathcal{F}_{V\subset U}$ is an isomorphism, so $\mathcal{F}'_{V\subset U}$ is an isomorphism too. It follows that the stalk restriction $\mathcal{F}'_{x\in U}$ is an isomorphism for any $x \in U$, so every basis element is $\mathcal{F}'$-simple. The stalkwise isomorphism $\Upsilon\colon \mathcal{F} \to \mathcal{F}'$ is given by

$$\Upsilon_x = \mathcal{F}'_{x\in U}\ \upsilon^U_{x_U x}\ {\mathcal{F}_{x\in U}}^{-1}$$

for any basis element U containing $x \in D$.

There are three claims implicit in the description of $\mathcal{F} \xrightarrow{\Upsilon} \mathcal{F}'$ above:

- $\mathcal{F}'$ is a locally constant sheaf.
- The definition of Υ_x doesn't depend on our choice of neighborhood U.
- The deviation of $\mathcal{F}'$ from $\mathcal{F}$ is υ.

Let's check these claims.

$\mathcal{F}'$ is a locally constant sheaf The functoriality of $\mathcal{F}'$ follows easily from the fact that υ is a deviation. To verify that $\mathcal{F}'$ is a sheaf, pick any element U of the basis $\mathcal{B}$. Suppose that for each basis element $V \subset U$, we have an element s_V of $\mathcal{F}'_V$, and these elements commute with the restriction morphisms of $\mathcal{F}'$. We need to find an element s of $\mathcal{F}'_U$ that restricts to s_V on every basis

element $V \subset U$. Since each of the restrictions $\mathcal{F}'_{V \subset U}$ is an isomorphism, there can only be one element like this, and it will exist if and only if the elements ${\mathcal{F}'_{V \subset U}}^{-1} s_V$ match for all the basis elements $V \subset U$.

For two basis elements $W \subset V$ contained in U, we have

$$\begin{aligned} {\mathcal{F}'_{W \subset U}}^{-1} s_W &= {\mathcal{F}'_{W \subset U}}^{-1} \mathcal{F}'_{W \subset V} s_V \\ &= {\mathcal{F}'_{V \subset U}}^{-1} s_V, \end{aligned}$$

so W and V give the same element. Since any two overlapping basis elements contain another basis element in their intersection, it follows that any two overlapping basis elements give the same element as well. Because U is connected, any two basis elements can be linked by a finite sequence of overlapping basis elements (Appendix A.1). Therefore, all the basis elements contained in U give the same element. This is the element s we were looking for, completing our proof that $\mathcal{F}'$ is a sheaf.

To see that $\mathcal{F}'$ is locally constant, first recall that the elements of $\mathcal{B}$ are simple, so the restriction arrows of $\mathcal{F}$ over $\mathcal{B}$ are isomorphisms. Thus, the restriction arrows of $\mathcal{F}'$ are isomorphisms as well. Now, pick any point $x \in X$, not necessarily in D. Applying $\mathcal{F}'$ to the poset of basis elements containing x yields a downward-directed diagram whose arrows are all isomorphisms. The defining arrows from a diagram like this to its colimit are always isomorphisms (Appendix A.2). In other words, the stalk restriction $\mathcal{F}'_{x \in U}$ is an isomorphism for every $U \in \mathcal{B}$ containing x. Since x was an arbitrary point in X, it follows by Proposition 8.2.A that $\mathcal{F}'$ is locally constant.

Υ_x is well-defined To see that the definition of Υ_x doesn't depend on our choice of neighborhood, first observe that for two basis elements $V \subset U$ containing x,

$$\begin{aligned}
(\mathcal{F}'_{x\in U})\ \upsilon^U_{x_U x}\ {\mathcal{F}_{x\in U}}^{-1} &= (\mathcal{F}'_{x\in V}\ [\mathcal{F}'_{V\subset U}])\ \upsilon^U_{x_U x}\ {\mathcal{F}_{x\in U}}^{-1} \\
&= \mathcal{F}'_{x\in V}[\mathcal{F}_{V\subset U}\ (\upsilon^U_{x_V x_U}]\ \upsilon^U_{x_U x})\ {\mathcal{F}_{x\in U}}^{-1} \\
&= \mathcal{F}'_{x\in V}\ [\mathcal{F}_{V\subset U}\ (\upsilon^U_{x_V x})]\ {\mathcal{F}_{x\in U}}^{-1} \\
&= \mathcal{F}'_{x\in V}\ [\upsilon^V_{x_V x}\ (\mathcal{F}_{V\subset U}]\ {\mathcal{F}_{x\in U}}^{-1}) \\
&= \mathcal{F}'_{x\in V}\ \upsilon^V_{x_V x}\ (\mathcal{F}_{x\in V}{}^{-1}),
\end{aligned}$$

so V and U give the same isomorphism. Since any two basis elements containing x contain another basis element in their intersection, it follows that every basis element containing x gives the same isomorphism.

The deviation of $\mathcal{F}'$ from $\mathcal{F}$ is υ Finally, let δ be the deviation of Υ. It's easy to calculate δ^U_{yx} by defining Υ_x and Υ_y in terms of U:

$$\begin{aligned}
\delta^U_{yx} &= {\mathcal{F}_{y\in U}}^{-1}\ (\Upsilon_y^{-1})\ \mathcal{F}'_{y\in U}\ {\mathcal{F}'_{x\in U}}^{-1}\ (\Upsilon_x)\ \mathcal{F}_{x\in U} \\
&= {\mathcal{F}_{y\in U}}^{-1}\ (\mathcal{F}_{y\in U}\ \upsilon^U_{yx_U}\ {\mathcal{F}'_{y\in U}}^{-1})\ \mathcal{F}'_{y\in U}\ {\mathcal{F}'_{x\in U}}^{-1}\ (\mathcal{F}'_{x\in U}\ \upsilon^U_{x_U x}\ {\mathcal{F}_{x\in U}}^{-1})\ \mathcal{F}_{x\in U} \\
&= \upsilon^U_{yx_U}\ \upsilon^U_{x_U x} \\
&= \upsilon^U_{yx}.
\end{aligned}$$

Thus, $\delta = \upsilon$, as claimed.

8.5 Warping local systems

Local systems are just a special kind of locally constant sheaves, so all the constructions of the previous sections can be applied to them. In this case, the definition of a deviation can be pared down a bit, because if $\mathcal{F}$ is a G local system and U is an $\mathcal{F}$-simple open set, an automorphism of $\mathcal{F}_U$ is just an element of G.

Warping a local system always produces another local system. To see why, take a G local system $\mathcal{F}$ and warp it by some deviation. The warped sheaf $\mathcal{F}'$ is locally constant, and stalkwise isomorphic to $\mathcal{F}$ over the support of the deviation. Because the support is dense, it follows that every stalk of $\mathcal{F}'$ is isomorphic to a stalk of $\mathcal{F}$, and hence a G-torsor.

Chapter 9

Dividing translation surfaces

9.1 Overview

For working out the technical details of abelianization, it will be useful to embed the surface $\Sigma \smallsetminus \mathfrak{B}$ in a larger space $\overleftrightarrow{\Sigma}$, called the *divided surface*, whose local systems are naturally in correspondence with the local systems on $\Sigma \smallsetminus \mathfrak{B}$. On the divided surface, we can stand infinitesimally close to any critical leaf, streamlining our discussion of the abelianization process in Sections 12 and 13.

Removing the critical leaves of Σ from the divided surface yields a compact space $\overleftrightarrow{\Sigma}$, called the *fractured surface*, which can be metrized in a very natural way. Its metric properties will play a crucial role in Section 14, where we prove that abelianization does the job it's meant to do.

The divided surface is also a point of contact between the worlds of flat and hyperbolic geometry. In Section 6, we described how a hyperbolic surface with a maximal geodesic lamination can be deflated to a flat surface with a half-translation structure by collapsing the complementary triangles of the lamination. The divided version of a half-translation surface is what you might imagine a deflating hyperbolic surface looks like in the instant before it

flattens out.

9.2 A review of translation and half-translation surfaces

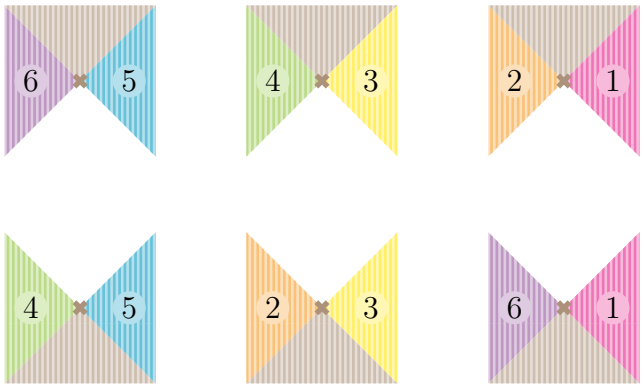
9.2.1 Translation surfaces

A non-singular *translation surface* is a manifold whose charts are open subsets of $\mathbb{R}^2$ and whose transition maps are translations. Every translation surface comes with a bunch of geometric structures induced by the translation-invariant geometric structures on $\mathbb{R}^2$, which include:

- The flat metric.
- The four cardinal directions: up, down, right, and left.
- The vertical and horizontal foliations, whose leaves are vertical and horizontal lines. Both foliations can be oriented; we'll orient them upward and rightward, respectively.
- The vertical flow, which moves points upward at unit speed. On a surface which is non-compact, as most non-singular translation surfaces are, this flow might not be defined everywhere at all times. In general, the flow at a given time will be only a *bicontinuous relation*, rather than a homeomorphism (see Appendix B for details).

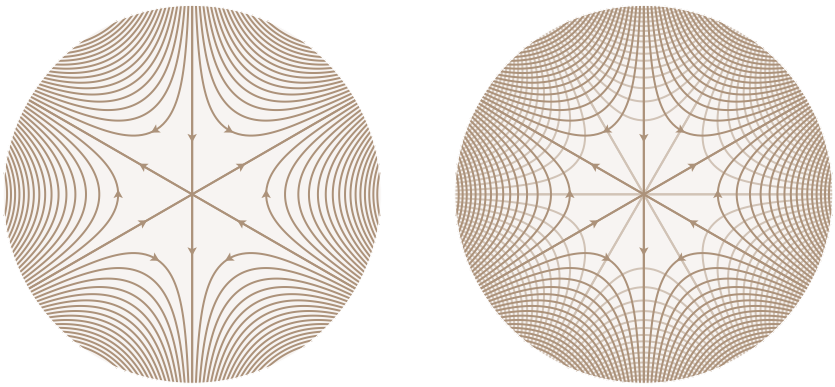
Whenever I refer to a foliation of a translation surface, I mean the vertical one, unless I say otherwise.

It's conventional, and convenient, to allow translation surfaces to have *conical singularities*, which look like this:



The like-numbered triangles are identified through translation. The triangles include the cross-marked center points, which get quotiented down to a single point by the identifications. In the case shown above, the total angle around the singularity is 6π; in general, any even multiple of π is possible. It's sometimes useful to mark a discrete set of ordinary points as "singularities" of cone angle 2π.

A compact translation surface can have only finitely many singularities. Conformally, the vertical and horizontal foliations in the neighborhood of a singularity look like this:



Vertical

Vertical and horizontal

The vertical leaves that dive into the singularity are called *forward-critical*, and the ones that shoot out of the singularity are called *backward-critical*. A leaf which is critical in both directions is called a *saddle connection*. Critical leaves are spaced evenly around the singularity at angles of π.

The vertical flow on a singular translation surface is only defined away from the singularities. It acts by bicontinuous relations, making points on the critical leaves disappear as they fall into the singularities. When restricted to the complement of the critical leaves, the vertical flow acts by homeomorphisms.

A translation surface is said to be *minimal* if all its vertical leaves are dense. We'll see in Section 9.2.2 that a non-critical leaf which is dense in one direction must be dense in both directions. Thus, on a minimal translation surface, every leaf that's not forward-critical is dense in the forward direction, and every leaf that's not backward-critical is dense in the backward direction.

A translation surface with no saddle connections is automatically minimal [30, proof of Theorem 1.8].

Away from the singularities, the topology of a translation surface has a basis consisting of *flow boxes*: open rectangles with vertical and horizontal sides. A compact translation surface can be covered by a finite collection of flow boxes and singularity charts. The special class of *well-cut* flow boxes, defined in the next section, will play an important role in this paper.

Because translations form a normal subgroup of $\mathrm{Aff}\,\mathbb{R}^2$, a translation structure can be modified by composing an element of $\mathrm{GL}_2\,\mathbb{R}$ with all its charts. In particular, a translation structure can be rotated, tilting the vertical foliation. If you rotate a translation structure through a full circle, all but countably many of the structures you pass through will have no saddle connections [30, proof of Theorem 1.8]. Moreover, for all but a measure-zero subset of the minimal structures, the vertical flow will be uniquely ergodic on the complement of the critical leaves [30, Theorem 3.5]. Hence, an arbitrarily small rotation is all it takes to turn any translation structure into a translation structure with no saddle connections and a vertical flow which is uniquely ergodic on the complement of the critical leaves.

9.2.2 First return maps

Let's say a *horizontal segment* on a translation surface is a subset that looks like an open, closed, or half-open horizontal line segment in some chart. Pick a point on a horizontal segment, and watch it as it's carried upward

by the vertical flow. On a compact translation surface, unless it falls into a singularity, the point will eventually return to the segment it started on. This is an immediate corollary of [30, Lemma 1.7], which for reference I'll restate here.

Lemma 9.2.A. *Let Z be a closed horizontal segment on a compact translation surface, and let p be one of its endpoints. Unless the vertical leaf through p is forward-critical, the vertical flow will eventually carry p back to Z.*

On any horizontal segment Z in a compact translation surface, we can define a relation α that sends each point to the place where it first returns to Z under the vertical flow. When fed a point that falls into a singularity before returning, α gives back nothing. We'll call α the *first return relation* on Z. The inverse relation α^{-1} sends each point to the place where it first returns to Z under the backward vertical flow.

When restricted to the complement of the forward-critical leaves, α becomes a function, and is called the *first return map*. Similarly, α^{-1} becomes a function when restricted to the complement of the backward-critical leaves. On the complement of all the critical leaves, α and α^{-1} are inverse functions.

Because the vertical foliation only has a few kinds of local geometry, the first return relation only has a few kinds of local behavior. Under mild conditions, it belongs to the class of transformations called *interval exchanges*, which we'll hear more about in Section 9.3.2 [31, §3]. The conditions are the price we pay for defining first return relations, horizontal segments, and inter-

val exchanges in a slightly non-standard way, whose advantages will become apparent in Section 9.5.2.

In our framework, the first return relation is only an interval exchange on a *well-cut* segment, which is a horizontal segment with the following properties:

- It looks like an open horizontal line segment bounded by critical leaves. (Both endpoints may lie on the same critical leaf.)
- The forward vertical flow drops each forward-critical boundary point into a singularity without carrying it through the segment. Similarly, the backward vertical flow drops each backward-critical boundary point into a singularity without carrying it through the segment.

When we construct the forward and backward first return relations on a horizontal segment Z, this condition prevents the vertical flow ψ^t from breaking $\psi^t Z$ across the boundaries of Z in a way inconsistent with our definition of an interval exchange. We'll call a flow box *well-cut* if every horizontal slice across it is a well-cut segment.

When critical leaves are plentiful, well-cut segments are easy to find.

Proposition 9.2.B. *If Z is an open horizontal segment bounded by critical leaves, every non-critical point on Z is contained in a well-cut subsegment of Z.*

Proof. Carry each forward-critical boundary point of Z along the forward vertical flow, marking each place it passes through Z before falling into a singularity. Do the same with each backward-critical boundary point, using the backward vertical flow. Removing the marked points breaks Z into a finite collection of open horizontal segments, which are all well-cut. □

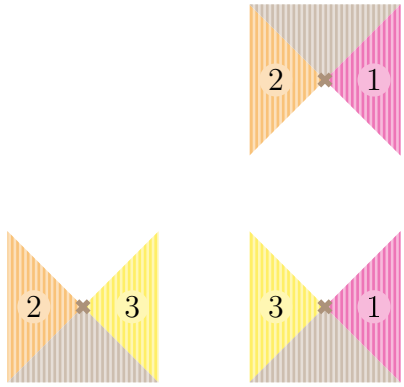
9.2.3 Half-translation surfaces

A half-translation surface is the same thing as a translation surface, but with transition maps composed of both translations and half-turn rotations. This expansion of the structure group is pretty small, so half-translation surfaces have almost as much structure as translation surfaces do:

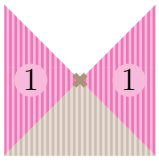
- The flat metric remains.
- The vertical and horizontal foliations remain, but they're no longer canonically oriented. In fact, it's often not possible to give either foliation a consistent global orientation.
- As a result, it's often not possible to define a global vertical flow.

The group generated by translations and half-turns has the translation group as a normal subgroup, so each of its elements is either a translation or a translation followed by a half-turn. I'll call the latter a *flip*.

In a translation surface, we saw that the angle around a conical singularity could be any even multiple of π. In a half-translation surface, any multiple of π is possible. Here's a pattern for a singularity with cone angle 3π:



The triangles labeled 3 are identified by a flip. A singularity with cone angle π can be constructed, informally, by using a flip to glue one of the notched square building blocks we've been using to itself:

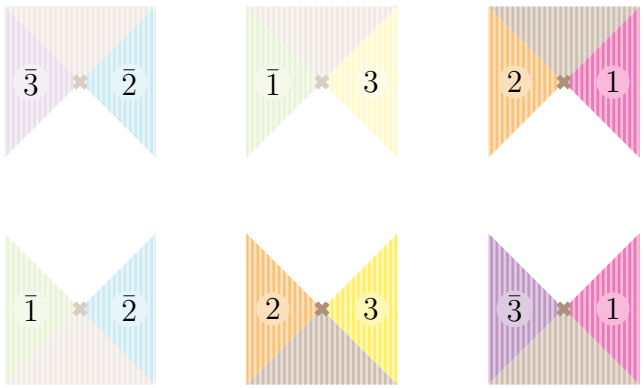


The vertical and horizontal foliations can't be oriented consistently around an odd singularity, but they can be oriented on a small enough neighborhood of any other point. In particular, on a half-translation surface, the definition of a flow box still makes sense, and the vertical and horizontal foliations can be oriented inside a flow box.

Every half-translation surface comes with a translation surface hovering over it as a branched double cover, with a projection map that preserves the half-translation structure. Away from the singularities, this *translation double*

cover is built by making two copies of each flow box on the half-translation surface, one for each possible orientation of the vertical foliation. The transition maps of the half-translation surface induce transitions between the oriented copies in a natural way.

The translation double cover can be extended over the whole surface by duplicating the cut square pieces that make up the region around each singularity. In the picture below, duplicate pieces are labeled with the same numbers. One piece from each pair is drawn in dark ink, and the other in light ink.



The new region is a double cover of the original, branched at the singularity. The branch point is a conical singularity with twice the angle of the original. The doubling process produces a branch point even when you start with a singularity of cone angle 2π, so we've found a meaning for these removable singularities: they stand for genuine singularities in the translation double cover. At the same time, the doubling process turns singularities with cone

angle π into ones with cone angle 2π, so removable singularities above mark genuine singularities below.

9.3 Dividing intervals

9.3.1 Construction of divided and fractured intervals

Dividing a translation surface is essentially a one-dimensional process, so let's start in the one-dimensional case. Let $I \subset \mathbb{R}$ be an open interval, and let W be a subset of I which is countable or smaller. To *divide* I at W, first build a new set

$$\overleftrightarrow{I} = \{\overleftarrow{w}, \hat{w}, \vec{w}\}_{w \in W} \sqcup \{s\}_{s \in I \smallsetminus W}.$$

There's an obvious map $\pi \colon \overleftrightarrow{I} \to I$ which sends $\overleftarrow{w}$, $\hat{w}$, and $\vec{w}$ to w and each point in $I \smallsetminus W$ to itself. Order $\overleftrightarrow{I}$ so that π is order-preserving and $\overleftarrow{w} < \hat{w} < \vec{w}$. Give $\overleftrightarrow{I}$ the topology generated by all the non-empty intervals (a, b) except the ones that look like $(\overleftarrow{w}, b)$ or $(a, \vec{w})$. This topology is coarser than the order topology, and non-Hausdorff: neither $\overleftarrow{w}$ nor $\vec{w}$ can be separated from $\hat{w}$ by open sets. The generating intervals described above form a basis for the topology, so I'll refer to them as "basis intervals."

As you might expect, π turns out to be a quotient map. Going the other direction, let $\iota \colon I \to \overleftrightarrow{I}$ be the map that sends $w \in W$ to $\hat{w}$ and each point in $I \smallsetminus W$ to itself. Perhaps surprisingly, ι turns out to be an embedding. (Both of these claims will be proven in Section 9.3.3.) Define the *fractured interval* $\overleftrightarrow{I}$ to be the complement of ιW in the divided interval $\overleftrightarrow{I}$.

9.3.2 Examples from dynamics

Divided intervals arise naturally in the study of some one-dimensional dynamical systems. In this setting, the fractured interval often turns out to be a familiar coding of the system. The binary shift provides an excellent example.

Divided intervals arising from interval exchange transformations will play an important role in this paper. The corresponding fractured intervals, for minimal interval exchanges, have appeared in the literature as *Cantor minimal systems* [32].

In both examples, we'll describe the dynamics using a partial map; you'll probably be able to guess from context what that means. To be precise, it means a coinjective, bicontinuous relation, in the terminology of Appendix B.

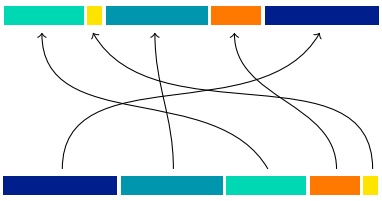
The binary shift The *binary shift* is the partial map from $(0, 1)$ to itself that sends s to $\lfloor 2s \rfloor$, returning nothing if $\lfloor 2s \rfloor = 1$. Shifting a number in $(0, 1)$ removes the first digit of its binary expansion. We can extend the binary shift to all reals by thinking of $\mathbb{R} \smallsetminus \mathbb{Z}$ as a union of copies of $(0, 1)$. Shifting a number in $\mathbb{R} \smallsetminus \frac{1}{2}\mathbb{Z}$ shifts the fractional part of its binary expansion, leaving the integer part alone. Shifting a number in $\frac{1}{2}\mathbb{Z}$ returns nothing.

Applying the shift map over and over, let $W \subset \mathbb{R}$ be the set of points that eventually "fall into a break," reaching a point where the map returns nothing. This turns out to be the set of rationals whose denominators are powers of two.

Divide $\mathbb{R}$ at W. Let $C \subset \overleftrightarrow{\mathbb{R}}$ be the interval $[\vec{0}, \overleftarrow{1}]$, noting that $\pi C = [0,1]$, and let $\ddot{C} = C \cap \overleftrightarrow{\mathbb{R}}$. We'll soon learn, from Corollary 9.3.I in Section 9.3.4, that $\ddot{C}$ is a Cantor set. In the meantime, we can see this directly by identifying $\ddot{C}$ with the space of one-sided binary sequences. A number $s \in (0,1) \smallsetminus W$ has a unique binary expansion, ιs. A number $w \in W$ has two binary expansions: $\overleftarrow{w}$, the one ending in ones, and $\vec{w}$, the one ending in zeros. The quotient map $\pi \colon \ddot{C} \to [0,1]$ interprets each binary sequence as a number.

The *shift* on a space of one-sided sequences is the map that removes the first symbol of the sequence. The shift on $\ddot{C}$ and the binary shift on $(0,1) \smallsetminus W$ commute with the embedding $\iota \colon (0,1) \smallsetminus W \to \ddot{C}$. The shift on $\ddot{C}$ extends uniquely to a relation on C that commutes with $\iota \colon (0,1) \to C$. This relation can be seen as a divided version of the binary shift. Notice that the binary shifts on C and $[0,1]$ don't quite commute with the quotient map π, because when a point $w \in W$ falls into a break, its lifts $\overleftarrow{w}$ and $\vec{w}$ keep going.

Interval exchanges An *interval exchange transformation* is a partial map from $(0,1)$ to itself that works by splitting $(0,1)$ into finitely many open subintervals and shuffling them around:



On the break points between the intervals, the map returns nothing.

An interval exchange, unlike the binary shift, is injective, so its inverse relation is also a partial map. In fact, its inverse is another interval exchange, which unshuffles the pieces of $(0, 1)$.

Just as we did with the binary shift, we can extend an interval exchange to all reals by thinking of $\mathbb{R} \setminus \mathbb{Z}$ as a union of copies of $(0, 1)$.

Pick an interval exchange, and let $W \subset \mathbb{R}$ be the set of points that eventually fall into a break under iteration of either this map or its inverse. Just as before, divide $\mathbb{R}$ at W. Let $C \subset \overleftrightarrow{\mathbb{R}}$ be the interval $[\vec{0}, \overleftarrow{1}]$, noting that $\pi C = [0, 1]$, and let $\ddot{C} = C \cap \overset{\leftrightarrow}{\mathbb{R}}$. If W is dense in $\mathbb{R}$, Corollary 9.3.I in Section 9.3.4 will tell us that $\ddot{C}$ is a Cantor set. In this case, $\ddot{C}$ can be seen as a subspace of the two-sided sequence space whose alphabet is the set of intervals shuffled by the interval exchange. The orbit of a point $s \in (0, 1) \setminus W$ under iteration of the interval exchange and its inverse is infinite in both directions, so we can get a two-sided sequence ιs by keeping track of which intervals the orbit passes through.

The orbit of a point $w \in W$ ends when it falls into a break, but it can be continued in two natural ways. One is to extend the interval exchange relation to a left-continuous map on $(0, 1]$, so each point travels with the points to the left of it, and the intervals being shuffled become closed on the right. The orbit of w becomes infinite in both directions, and keeping track of which intervals it goes through yields a two-sided sequence $\overleftarrow{w}$. The other way to continue the orbit of w is to extend the interval exchange relation to a right-continuous map on $[0, 1)$, giving a different two-sided sequence $\vec{w}$.

The *shift* on a space of two-sided sequences is the map that moves every symbol one step earlier. The shift map on $\ddot{C}$ and the interval exchange map on $(0,1) \smallsetminus W$ commute with the embedding $\iota\colon (0,1) \smallsetminus W \to \ddot{C}$. The shift on $\ddot{C}$ extends uniquely to a relation on C that commutes with $\iota\colon (0,1) \to C$. This relation can be seen as a divided version of the interval exchange. Notice that interval exchanges on C and $[0,1]$ don't quite commute with the quotient map π, because when a point $w \in W$ falls into a break, its lifts $\overleftarrow{w}$ and $\vec{w}$ keep going, following the left- and right-continuous extensions of the interval exchange.

9.3.3 Properties of divided intervals

For convenience, let $\overleftarrow{\iota}\colon I \to \overleftrightarrow{I}$ be the map that sends $w \in W$ to $\overleftarrow{w}$ and each point in $I \smallsetminus W$ to itself. Define $\vec{\iota}$ similarly.

A basis interval has a leftmost element if and only if it looks like $(\hat{w}, b) = [\vec{w}, b)$, and a rightmost element if and only if it looks like $(a, \hat{w}) = (a, \overleftarrow{w}]$. It will often be useful to *trim* a basis interval by removing its leftmost and rightmost elements, if they exist. The trimmed version of an interval $(a,b) \subset \overleftrightarrow{I}$, denoted $\operatorname{trim}(a,b)$, can be written explicitly as $(\vec{\iota}\pi a, \overleftarrow{\iota}\pi b)$. Notice that $\pi \operatorname{trim}(a,b) = \iota^{-1}(a,b) = (\pi a, \pi b)$ for any basis interval (a,b), and that trimming a basis interval does not remove any points in the image of ι. Conveniently, for any basis interval, $\pi^{-1}\iota^{-1}(a,b) = \operatorname{trim}(a,b)$.

With these tools in hand, let's prove the claims about π and ι made in the previous section.

Proof that π is a quotient map. To see that π is continuous, observe that the preimage of $(a, b) \subset I$ under π is the basis interval $(\vec{\iota}a, \overleftarrow{\iota}b)$.

To see that π is a quotient map, pick any $S \subset I$ whose preimage under π is open. We want to show S is open. For any $s \in S$, the point ιs is in $\pi^{-1}S$, so there is a basis interval $H \subset \pi^{-1}S$ containing ιs. Since $\operatorname{trim} H$ also contains ιs, and π sends trimmed basis intervals to open intervals, $\pi \operatorname{trim} H$ is an open subset of S containing s. □

Proof that ι is an embedding. To see that ι is continuous, recall that $\iota^{-1}(a, b) = (\pi a, \pi b)$ for any basis interval (a, b).

To see that ι is an embedding, observe that the image under ι of an interval $(a, b) \subset I$ is the intersection of $(\iota a, \iota b)$ with ιI. □

The continuity of ι is a way of saying that passing from I to $\overleftrightarrow{I}$ spreads out the points of W, but it doesn't spread them out too much. Here are two more reflections of this idea.

Proposition 9.3.A. *The embedding of I in $\overleftrightarrow{I}$ is dense.*

Proof. It's enough to show that ιI intersects every basis interval. Suppose the basis interval (a, b) doesn't intersect ιI, so its preimage $(\pi a, \pi b)$ under ι is empty. Since I is densely ordered, this means $\pi a = \pi b$, which is precluded by the rules defining basis intervals. □

Proposition 9.3.B. *The divided interval $\overleftrightarrow{I}$ is locally connected.*

Proof. It's enough to show that every basis interval is connected. Recall that basis intervals are non-empty by definition. Let's say the basis interval (a, b) is disconnected by two open subsets U and V. Since ιI is dense in $\overset{\leftrightarrow}{I}$, the preimages of U and V under ι are non-empty, so they disconnect the preimage $(\pi a, \pi b)$ of (a, b). □

For our purposes, the most important feature of $\overset{\leftrightarrow}{I}$ is that its local systems are naturally in correspondence with the local systems on I. This idea can be stated more precisely as follows.

Theorem 9.3.C. *For any linear group G, the direct image functors π_* and ι_* give an equivalence between the category of G local systems on $\overset{\leftrightarrow}{I}$ and the category of G local systems on I.*[1]

The reason $\overset{\leftrightarrow}{I}$ has no more local systems than I, despite having more open subsets, is that a local system on $\overset{\leftrightarrow}{I}$ is determined entirely by its values on trimmed intervals.

Lemma 9.3.D. *If $\mathcal{F}$ is a local system on $\overset{\leftrightarrow}{I}$, the restriction $\mathcal{F}_{\operatorname{trim} H \subset H}$ is an isomorphism for any basis interval H.*

Proof. If H has neither a least element nor a greatest element, $\operatorname{trim} H = H$, so there's nothing to prove. Let's assume H has a least element, but no greatest element; the remaining cases are essentially the same.

[1] Though they're stated for categories of local systems, Theorem 9.3.C and Lemma 9.3.D hold for categories of locally constant sheaves into any fixed target category. The proofs are the same, keeping in mind our convention (from Section 8.2) that the target category is a type of algebraic structure.

Since $\mathcal{F}$ is locally constant, we can pick a basis interval $A \subset H$ which contains the least element of H and is small enough that $\mathcal{F}|_A$ is constant. Since A and $\operatorname{trim} A$ are both connected, $\mathcal{F}_{\operatorname{trim} A \subset A}$ is an isomorphism. The diagram

$$\begin{array}{ccccc} & & \mathcal{F}_{\operatorname{trim} A} & \xleftarrow{\subset} & \mathcal{F}_{\operatorname{trim} H} \\ & \overset{\subset^{-1}}{\swarrow} & \downarrow{\scriptstyle =} & & \downarrow{\scriptstyle =} \\ \mathcal{F}_A & \xrightarrow[\subset]{} & \mathcal{F}_{\operatorname{trim} A} & \xleftarrow[\subset]{} & \mathcal{F}_{\operatorname{trim} H} \end{array}$$

commutes, so taking limits of its top and bottom rows gives a map $\mathcal{F}_{\operatorname{trim} H} \to \mathcal{F}_H$, which inverts $\mathcal{F}_{\operatorname{trim} H \subset H}$. □

Proof of Theorem 9.3.C. There's a canonical natural isomorphism between $\pi_*\iota_*$ and the identity functor, because $\pi\iota$ is the identity map from I to itself. Now, all we need is a natural isomorphism between $\iota_*\pi_*$ and the identity.

Pick any local system $\mathcal{F}$ on $\overset{\leftrightarrow}{I}$. For each basis interval H, recall that $\pi^{-1}\iota^{-1}H = \operatorname{trim} H$, so

$$\begin{aligned} (\iota_*\pi_*\mathcal{F})_H &= \mathcal{F}_{\pi^{-1}\iota^{-1}H} \\ &= \mathcal{F}_{\operatorname{trim} H}. \end{aligned}$$

Thus, the restriction $\mathcal{F}_{\operatorname{trim} H \subset H}$ gives a morphism from $\mathcal{F}_H$ to $(\iota_*\pi_*\mathcal{F})_H$, and Lemma 9.3.D tells us this morphism is an isomorphism. For any basis interval $H' \subset H$, the diagram

$$\begin{array}{ccc} \mathcal{F}_{H'} & \xleftarrow{\subset} & \mathcal{F}_H \\ {\scriptstyle \subset}\downarrow & & \downarrow{\scriptstyle \subset} \\ \mathcal{F}_{\operatorname{trim} H'} & \xleftarrow[\subset]{} & \mathcal{F}_{\operatorname{trim} H} \end{array}$$

commutes because all the arrows are restrictions, so we've found a natural isomorphism from $\mathcal{F}$ to $\iota_*\pi_*\mathcal{F}$. □

9.3.4 Properties of fractured intervals

One nice feature of $\overset{\leftrightarrow}{I}$ is that, with $\hat{w}$ out of the way, the points $\overleftarrow{w}$ and $\vec{w}$ can be separated by open sets. The consequence is just what you'd expect.

Proposition 9.3.E. *The fractured interval $\overset{\leftrightarrow}{I}$ is Hausdorff.*

Proof. Pick two points $s < t$ in $\overset{\leftrightarrow}{I}$. If $\pi s \neq \pi t$, we can find disjoint neighborhoods of πs and πt in I and pull them back to $\overset{\leftrightarrow}{I}$. If $\pi s = \pi t$, then $s = \overleftarrow{w}$ and $t = \vec{w}$ for some $w \in W$. Hence, $(-\infty, \overleftarrow{w}]$ and $[\vec{w}, \infty)$ are disjoint neighborhoods of s and t. □

When studying $\overset{\leftrightarrow}{I}$, we found it useful to work with basis intervals whose leftmost and rightmost elements had been removed. For studying $\overset{\leftrightarrow}{I}$, it will be useful to go the opposite direction. Let's say a basis interval is *full* if it has both a leftmost element and a rightmost element. As we saw earlier, the full intervals in $\overset{\leftrightarrow}{I}$ are the ones that look like $[\vec{a}, \overleftarrow{b}]$. The full intervals in $\overset{\leftrightarrow}{I}$ are the same.

Proposition 9.3.F. *In $\overset{\leftrightarrow}{I}$, every full interval is compact.*

Proof. Consider a full interval $[\vec{a}, \overleftarrow{b}]$. Let W' be the subset of W lying between $\vec{a}$ and $\overleftarrow{b}$. Pick a function $\kappa\colon W' \to \mathbb{R}_+$ for which the sum $K = \sum_{w \in W'} \kappa w$ is

finite, and let θ be the map from $[\vec{a}, \overleftarrow{b}] \cap \overset{\leftrightarrow}{I}$ to $\mathbb{R}$ given by the formula

$$\theta s = \pi s + \sum_{\substack{w \in W' \\ \hat{w} < s}} \kappa w.$$

It's not hard to see that θ is a homeomorphism whose image is the set

$$[a, b + K] \smallsetminus \bigcup_{w \in W'} (\theta \overleftarrow{w}, \theta \vec{w}),$$

which is closed and bounded. □

Proposition 9.3.G. *In $\overset{\leftrightarrow}{I}$, every full interval is clopen.*

Proof. In $\overset{\leftrightarrow}{I}$, the complement of a full interval $[\vec{a}, \overleftarrow{b}]$ is the union of the basis intervals $(-\infty, \overleftarrow{a}]$ and $[\vec{b}, \infty)$. □

Proposition 9.3.H. *If W is dense in I, the full intervals form a basis for $\overset{\leftrightarrow}{I}$.*

Proof. Suppose W is dense in I. Pick any point $s \in \overset{\leftrightarrow}{I}$ and any basis interval (a, b) containing it. If $\pi a \neq \pi s$, find a point of W in the interval $(\pi a, \pi s)$ and call it α. If $\pi a = \pi s$, observe that $a = \hat{\alpha}$ and $s = \vec{\alpha}$ for some $\alpha \in W$. One way or another, we've found a point $\alpha \in W$ with $a < \vec{\alpha} \leq s$. Using the same technique, we can find a point $\beta \in W$ with $s \leq \overleftarrow{\beta} < b$. The full interval $[\vec{\alpha}, \overleftarrow{\beta}]$ is a neighborhood of s contained in (a, b). □

Corollary 9.3.I. *If W is dense in I, every full interval in $\overset{\leftrightarrow}{I}$ is a Cantor set.*

Proof. Suppose W is dense in I. Because we require W to be countable or smaller, the results above imply that $\overset{\leftrightarrow}{I}$ is a Hausdorff space with a countable

basis of clopen sets. Any full interval $H \subset \overset{\diamond}{I}$ has the same properties, and in addition is compact. Therefore, H is a Cantor set as long as it has no isolated points [33, Theorem 3].

Intersecting H with a basis interval in $\overset{\diamond}{I}$ yields another basis interval. Since every basis interval contains more than one point, it follows that H has no isolated points. □

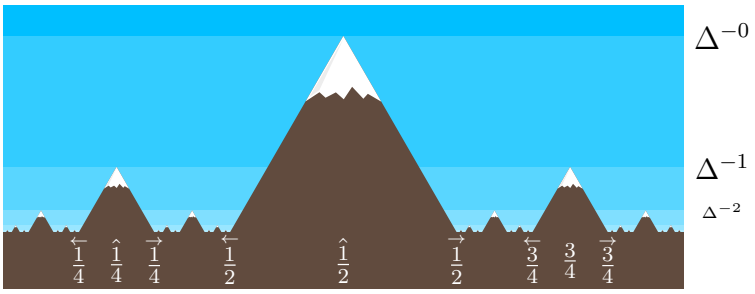
9.3.5 Metrization

The topology of I is induced by the metric that I inherits from $\mathbb{R}$. If W is dense in I, the topology of $\overset{\diamond}{I}$ can be metrized too, and there's a simple way to do it. For the examples given in Section 9.3.2, the resulting metric is dynamically meaningful.

For the rest of this section, suppose W is dense in I. Let's say we've assigned each point in W a natural number, its *grade*, and there are only finitely many points of each grade. Since W is countable or smaller, this is always possible.

For the binary shift, the points in W are the numbers whose binary expansions are eventually constant, and they're naturally graded by the number of digits before the constant tail. For an interval exchange, the points in W are the points that will eventually fall into a break under forward or backward iteration, and they're naturally graded by how long it takes for that to happen. For the sake of concreteness in our proof of Proposition 9.5.I, we'll fix a normalization by declaring the break points to have grade zero.

Pick a real number $\Delta > 1$, and define the *height* of a point in $w \in W$ to be $\Delta^{-\operatorname{grade} w}$. In $\overleftrightarrow{I}$, define the height of $\hat{w} \in \iota W$ to be the height of w, and the height of any other point to be zero. The heights of some points in the divided interval for the binary shift are illustrated below.



Let's say the distance between two points $a, b \in \overleftrightarrow{I}$ is the height of the highest point in $(a, b) \subset \overleftrightarrow{I}$. This defines a metric (in fact, an ultrametric) on $\overleftrightarrow{I}$, which I'll call the *division metric* with steepness Δ. The assumption that W is dense in I is essential here: it guarantees that distances between distinct points are positive.

Proposition 9.3.J. *The division metric induces the topology of* $\overleftrightarrow{I}$.

Proof. Let's see what the open balls of the division metric look like. Given a radius $r > 0$, let $W_{\geq r}$ be the set of points in W with heights greater than or equal to r. This set is finite, because there are only finitely many points of each grade. Listing the points in $W_{\geq r}$ from left to right as $w_1, \ldots, w_n$, we can write down all the open balls of radius r:

$$(-\infty, \hat{w}_1), (\hat{w}_1, \hat{w}_2), \ldots, (\hat{w}_{n-1}, \hat{w}_n), (\hat{w}_n, \infty).$$

From this description, it's clear that the open balls of the division metric are open subsets of $\overleftrightarrow{I}$. It's also clear that every full interval is a union of open balls, because every full interval can be written as $(\hat{a}, \hat{b})$ for $a, b \in W$. By Proposition 9.3.H, the full intervals form a basis for $\overleftrightarrow{I}$, so we're done. □

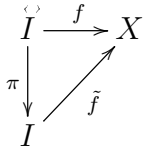
For both of the examples in Section 9.3.2, the division metric tells you how long two points in the fractured interval travel together before they end up on opposite sides of a break. Nearby points move together for a long time, while the most distant points are separated immediately. A single step of the dynamics can take a pair of points at most one step closer to being separated, increasing the distance between then by at most a factor of Δ. That means the dynamical map is Lipschitz with respect to the division metric, essentially by construction.

The division metric on $\overleftrightarrow{I}$ and the Euclidean metric on I are very different, so playing them off against one another might lead to amusing results. Let's get them talking by defining gap_r to be the minimum distance between points in $W_{\geq r}$, according to the Euclidean metric on I. We can force the two metrics to work together by putting conditions on the gap function. Now, what sort of trouble can we start?

The division metric entertains all kinds of Hölder functions, but the Euclidean metric will not allow any interesting function to have a Hölder exponent greater than one. A function f on $\overleftrightarrow{I}$ factors through π if and only if its values match at adjacent edge points, in the sense that $f\overleftarrow{w} = f\vec{w}$ for all

$w \in W$. If a Hölder function on $\overset{\frown}{I}$ were to factor through π, there might be a bit of a problem.

Theorem 9.3.K. *Suppose the gap function falls off slower than a power law, so* $\text{gap}_r \geq Mr^\alpha$ *for some positive constants* M *and* α*. Consider a function* f *from* $\overset{\frown}{I}$ *into some metric space* X *which factors through* π *as shown:*



If f *is Hölder with exponent* $\nu > 0$*, then* $\tilde{f}$ *is Hölder with exponent* ν/α*.*

Interesting, but not terribly entertaining. Let's turn up the heat.

Corollary 9.3.L. *Suppose the gap function falls off slower than* every *power law: for any* $\alpha > 0$*, no matter how small, we can find a positive constant* M *with* $\text{gap}_r \geq Mr^\alpha$*. If* f *is a Hölder function on* $\overset{\frown}{I}$ *whose values match at adjacent endpoints, then* f *is constant.*

A bit of terminology will speed our proof of Theorem 9.3.K. Let's say a point $s \in \overset{\frown}{I}$ is in the *left watershed* of $w \in W$ if $s < \hat{w}$, and every point in $(s, \hat{w}) \subset \overset{\leftrightarrow}{I}$ is lower than w. Define the *right watershed* similarly. The left and right watersheds of $\frac{1}{4}$ and $\frac{5}{8}$ for the binary shift are sketched below.

Watersheds are useful because of the following fact.

Lemma 9.3.M. *For any $s \in I$, if $\vec{\iota}s$ is in the left watershed of $w \in W$, then $\text{gap}_{d(\vec{\iota}s, \overleftarrow{w})} \leq w - s$. The mirror image statement also holds.*

Proof. Let $\hat{v} \in \iota W$ be the highest point in $(\vec{\iota}s, \hat{w}) \subset \overleftrightarrow{I}$, and let h be its height. Since $\vec{\iota}s$ is in the left watershed of w, the point w is higher than v. That means v and w are both in $W_{\geq h}$, so $w - v$ is at least gap_h. Thus, $w - s$ is at least gap_h. Since $\hat{v}$ was the highest point between $\vec{\iota}s$ and $\hat{w}$, we have $d(\vec{\iota}s, \overleftarrow{w}) = h$ by definition, so we're done. The mirror image statement is proven similarly. □

Proof of Theorem 9.3.K. Suppose f is Hölder with exponent $\nu > 0$ and scale constant $C > 0$. To see that $\tilde{f}$ is Hölder, pick any two points $a < b$ in I. Observe that

$$d(\tilde{f}a, \tilde{f}b) = d(f\vec{\iota}a, f\overleftarrow{\iota}b).$$

Let $w \in W$ be the highest point in $(\vec{\iota}a, \overleftarrow{\iota}b) \subset \overleftrightarrow{I}$. By assumption, $f\overleftarrow{w} = f\vec{w}$, so the triangle inequality and the Hölder condition on f give

$$\begin{aligned} d(\tilde{f}a, \tilde{f}b) &\leq d(f\vec{\iota}a, f\overleftarrow{w}) + d(f\vec{w}, f\overleftarrow{\iota}b) \\ &\leq Cd(\vec{\iota}a, \overleftarrow{w})^\nu + Cd(\vec{w}, \overleftarrow{\iota}b)^\nu. \end{aligned}$$

Since $\vec{\iota}a$ is in the left watershed of w, the lemma tells us that

$$\mathrm{gap}_{d(\vec{\iota}a,\overleftarrow{w})} \leq w - a \qquad\qquad \mathrm{gap}_{d(\vec{w},\overleftarrow{\iota}b)} \leq b - w.$$

Our lower bound on the gap function then ensures that

$$Md(\vec{\iota}a, \overleftarrow{w})^\alpha \leq w - a \qquad\qquad Md(\vec{w}, \overleftarrow{\iota}b)^\alpha \leq b - w,$$

so

$$d(\tilde{f}a, \tilde{f}b) \leq B[(w-a)^{\nu/\alpha} + (b-w)^{\nu/\alpha}]$$

for $B = \frac{C}{M^{\nu/\alpha}}$. Since the function $t \mapsto t^{\nu/\alpha}$ is non-decreasing, it follows[2] that

$$\begin{aligned} d(\tilde{f}a, \tilde{f}b) &\leq 2B[(w-a) + (b-w)]^{\nu/\alpha} \\ &= 2B(b-a)^{\nu/\alpha}. \end{aligned}$$

Since a and b were arbitrary, aside from the condition that $a < b$, we've shown that $\tilde{f}$ is Hölder with exponent ν/α and scale constant $\frac{2C}{M^{\nu/\alpha}}$. □

9.4 Dividing translation surfaces

9.4.1 Construction of divided and fractured surfaces

Now, let's move up to the two-dimensional case. Cover $\Sigma \smallsetminus \mathfrak{B}$ with flow boxes. Each flow box can be identified with a rectangle $I \times L \subset \mathbb{R}^2$, where I and L are open intervals in $\mathbb{R}$. The critical leaves of Σ intersect the flow box

[2] If $\rho\colon [0,\infty) \to \mathbb{R}$ is non-decreasing, and $t \leq t'$,

$$\rho t + \rho t' \leq 2\rho t' \leq 2\rho(t + t').$$

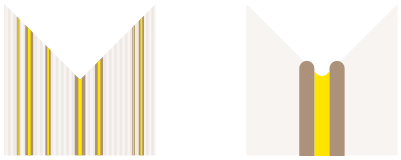
Many thanks to Sona Akopian for pointing this out.

as vertical lines $\{w\} \times L$. There are only finitely many critical leaves, and each one passes through the flow box at most countably many times. Dividing I at the positions of the critical leaves, we can produce a *divided flow box* $\overleftrightarrow{I} \times L$. In the divided flow box, each critical leaf $\{w\} \times L$ splits into a *left lane* $\{\overleftarrow{w}\} \times L$, a *median* $\{\hat{w}\} \times L$, and a *right lane* $\{\vec{w}\} \times L$.

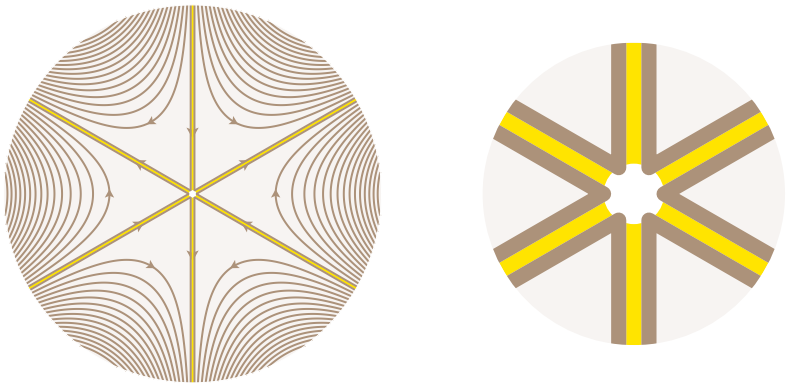
The transition maps between the flow boxes induce transitions between the divided flow boxes in a natural way. Gluing the divided flow boxes together along these transitions yields a new space—a divided version of $\Sigma \smallsetminus \mathfrak{B}$.

We still need to decide what to do with the singularities. The region around a singularity $\mathfrak{b} \in \mathfrak{B}$ can be built from cut square pieces, as described in Section 9.2.1. Here's one of them, with the critical leaves marked. (Since the critical leaves are generally dense in Σ, I've only drawn finitely many of them thick enough to see.)

The leaf through the center point is critical, of course, because the center point is $\mathfrak{b}$ itself. Here's a divided version of the same cut square:



Zooming in, you can see that I've cut away $\hat{\mathfrak{b}}$, but kept $\overleftarrow{\mathfrak{b}}$ and $\overrightarrow{\mathfrak{b}}$. When a singularity is built from pieces that look like this one (and its half-rotation), the left and right lanes of the center leaf will connect up into lines running past the singularity, but the medians will remain disconnected rays that end at the singularity. Topologically, the region around a divided singularity looks like this:



The zoomed-in picture on the right shows how the lanes of the critical leaves join up around the singularity. For clarity, only the parts of the critical leaves adjacent to the singularity are shown.

Now that we've decided what to do with the singularities, we can extend our divided version of $\Sigma \smallsetminus \mathfrak{B}$ to a full *divided surface* $\overleftrightarrow{\Sigma}$. The quotient map discussed in the one-dimensional case extrudes naturally to a quotient map $\pi\colon \overleftrightarrow{\Sigma} \to \Sigma$. Away from the singularities, the embedding from the one-dimensional case also extrudes, yielding an embedding $\Sigma \smallsetminus \mathfrak{B} \to \overleftrightarrow{\Sigma}$. This embedding can't be extended over the singularities, because a singularity has

no median point associated with it, and sending it to an associated lane point would break continuity.

We'll refer to the π-preimage of a leaf of Σ as a *road*.[3] If $\mathcal{L}$ is a critical leaf of Σ, the *critical road* $\pi^{-1}\mathcal{L}$ splits into a left lane $\{\overleftarrow{w} : w \in \mathcal{L}\}$, a median $\{\hat{w} : w \in \mathcal{L}\}$, and a right lane $\{\vec{w} : w \in \mathcal{L}\}$, as we saw locally at the beginning of the section. Each of the lane points associated with a singularity is adjacent to a lane of a forward-critical road and a lane of a backward-critical road. For convenience, we'll consider it an honorary member of both.

As in the one-dimensional case, define the *fractured surface* $\overset{\leftrightarrow}{\Sigma}$ to be the complement of the critical leaves of Σ—or, more precisely, their images under ι—in the divided surface $\overset{\leftrightarrow}{\Sigma}$. The divided flow boxes and divided singularity charts from which we constructed $\overset{\leftrightarrow}{\Sigma}$ pull back to an atlas of *fractured flow boxes* and *fractured singularity charts* on $\overset{\leftrightarrow}{\Sigma}$.

9.4.2 Properties of divided and fractured surfaces

Within each divided flow box $\overset{\leftrightarrow}{I} \times L$, we can find more flow boxes of the form $H \times L$, where $H \subset \overset{\leftrightarrow}{I}$ is a basis interval. For convenience, let's define the notation $\dot{\dot{H}} = H \cap \overset{\leftrightarrow}{I}$. A flow box $U = H \times L$ will be called *full* if H is full, and *well-cut* if H is full and $\iota^{-1}U$ is well-cut. Define $\dot{\dot{U}}$ as $U \cap \overset{\leftrightarrow}{\Sigma}$, or equivalently $\dot{\dot{H}} \times L$, and $\operatorname{trim} U$ as $(\operatorname{trim} H) \times L$. It's apparent from the analogous one-dimensional result that $\pi^{-1}\iota^{-1}U = \operatorname{trim} U$ for any flow box U.

[3] Our one-way roads are technically different from, but morally related to, the *two-way streets* of [11].

Flow boxes form a basis for $\overleftrightarrow{\Sigma}$. If the critical leaves are dense in Σ, full flow boxes form a basis for $\overleftrightarrow{\Sigma}$, as a consequence of Proposition 9.3.H. If full flow boxes form a basis for $\overleftrightarrow{\Sigma}$, well-cut flow boxes do too, as a consequence of Proposition 9.2.B.

Propositions 9.3.A, 9.3.B, and 9.3.E carry over from dimension one straightforwardly enough that I'll state them without proof.

Proposition 9.4.A. *The embedding of Σ in $\overleftrightarrow{\Sigma}$ is dense.*

Proposition 9.4.B. *The divided surface $\overleftrightarrow{\Sigma}$ is locally connected.*

Proposition 9.4.C. *The fractured surface $\overleftrightarrow{\Sigma}$ is Hausdorff.*

Because Σ is compact, Proposition 9.3.F can be extended to a global result.

Proposition 9.4.D. *The fractured surface $\overleftrightarrow{\Sigma}$ is compact.*

Proof. Because Σ compact, it can be covered by a finite collection of flow boxes and singularity charts. Therefore, $\overleftrightarrow{\Sigma}$ can be covered by a finite collection of fractured flow boxes and fractured singularity charts. Each fractured flow box $\overleftrightarrow{I} \times L$ can be covered almost out to the edges by a "full closed box" of the form $\overleftrightarrow{H} \times C$, where $H \subset \overleftrightarrow{I}$ is full interval and $C \subset L$ is a closed interval. Similarly, each fractured singularity chart can be covered almost out to the edges by a finite collection of full closed boxes, taking crucial advantage of the fact that the non-compact medians of the critical rays have been removed.

With a little care, we can now cover $\overset{\langle\rangle}{\Sigma}$ with a finite collection of full closed boxes. Each full closed box is a product of two compact spaces, so we've covered $\overset{\langle\rangle}{\Sigma}$ with a finite collection of compact sets. □

Theorem 9.3.C carries over with a catch: dividing a translation surface opens up a tiny hole at each singularity, and a local system on the divided surface can have nontrivial holonomies around those holes.

Theorem 9.4.E. *For any linear group G, the direct image functors π_* and ι_* give an equivalence between the category of G local systems on $\overset{\leftrightarrow}{\Sigma}$ and the category of G local systems on $\Sigma \smallsetminus \mathfrak{B}$.*

The reason for the theorem remains the same.

Lemma 9.4.F. *If $\mathcal{F}$ is a local system on $\overset{\leftrightarrow}{\Sigma}$, the restriction $\mathcal{F}_{\operatorname{trim} U \subset U}$ is an isomorphism for any flow box U.*

Proof. Write U as $H \times L$ for some basis interval $H \subset \overset{\leftrightarrow}{I}$ and repeat the proof of Lemma 9.3.D, replacing A with $A \times L$. □

Proof of Theorem 9.4.E. Replace

$$\begin{array}{c} \overset{\leftrightarrow}{I} \\ \pi\big(\ \big)\iota \\ I \end{array} \quad \text{with} \quad \begin{array}{c} \overset{\leftrightarrow}{\Sigma} \\ \pi\big(\ \big)\iota \\ \Sigma \smallsetminus \mathfrak{B} \end{array}$$

in the proof of Theorem 9.3.C, and change all the basis intervals to flow boxes. □

9.5 Dynamics on divided surfaces

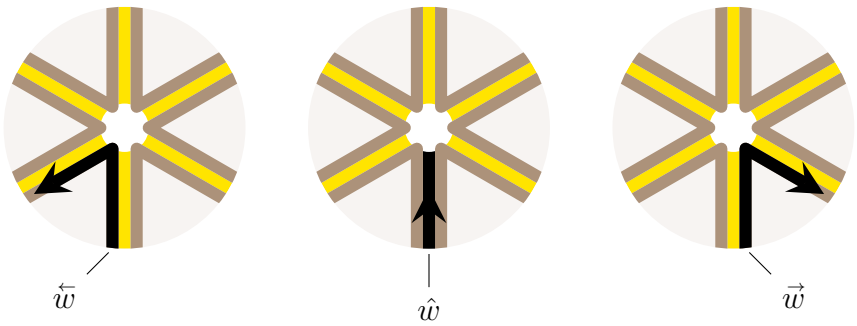
9.5.1 The vertical flow

Within a flow box $I \times L$ in Σ, the vertical flow is easy to describe:

$$\psi^t \colon (s, \zeta) \mapsto (s, \zeta + t),$$

as long as t is small enough that $\zeta + t$ is still in the interval L. This local description can be lifted directly to the divided flow box $\overleftrightarrow{I} \times L$ in $\overleftrightarrow{\Sigma}$; the only change is that s will now be a point in $\overleftrightarrow{I}$ instead of in I.

The vertical flow on a singularity chart can be lifted in the same way. It's worth thinking carefully about how the lifted flow acts on the critical roads of $\overleftrightarrow{\Sigma}$. Pick a singularity chart on Σ, and look at a point w on a forward-critical leaf that plunges into the singularity without leaving the chart. In $\overleftrightarrow{\Sigma}$, the point w splits into the three points $\overleftarrow{w}, \hat{w}, \vec{w}$ on a forward-critical road. As time runs forward, the point w falls into the singularity and disappears. Its lift $\hat{w}$, on the median of the road, does the same. The points $\overleftarrow{w}$ and $\vec{w}$, however, follow the left and right lanes past the singularity, peeling off in different directions:



On a backward-critical road, the story is the same, but told backward.

Just as the local vertical flows on flow boxes and singularity charts fit together into a global vertical flow on $\Sigma \smallsetminus \mathfrak{B}$, their lifts fit together into a global vertical flow on $\overleftrightarrow{\Sigma}$. For convenience, I'll refer to this vertical flow also as ψ. The vertical flows on $\overleftrightarrow{\Sigma}$ and $\Sigma \smallsetminus \mathfrak{B}$ commute with the embedding ι. They don't quite commute with the quotient map π, because when a point on a critical leaf of $\Sigma \smallsetminus \mathfrak{B}$ falls into a singularity, its left- and right-lane lifts keep going. The forward vertical flows on $\Sigma \smallsetminus \mathfrak{W}^-$ and its π-preimage, however, do commute with π. The same goes for the backward vertical flows on $\Sigma \smallsetminus \mathfrak{W}^+$ and its π-preimage.

The vertical flow on $\overleftrightarrow{\Sigma}$, like the one on $\Sigma \smallsetminus \mathfrak{B}$, is a flow by bicontinuous relations. This should not be taken for granted: the topology of a divided interval was engineered to make it so. The medians of the critical leaves are the only parts of $\overleftrightarrow{\Sigma}$ that vanish into the singularities under the vertical flow. Removing them leaves an invariant subspace, $\overleftrightarrow{\Sigma}$, on which the vertical flow acts by homeomorphisms. Thus, while the vertical flows on $\Sigma \smallsetminus \mathfrak{B}$ and $\overleftrightarrow{\Sigma}$ may look a bit ugly, the vertical flow on $\overleftrightarrow{\Sigma}$ is remarkably well-behaved: it's a flow by homeomorphisms on a compact Hausdorff space.

The divided and fractured surfaces are foliated by the orbits of the vertical flow—or they would be, at least, if foliations were usually defined on more general spaces than manifolds. With that in mind, we'll sometimes refer to the orbits of the vertical flows on $\overleftrightarrow{\Sigma}$ and $\overleftrightarrow{\Sigma}$ as vertical leaves.

9.5.2 First return maps

Let's say a *horizontal segment* on a divided surface is a subset that looks like a horizontal basis interval in some flow box. More precisely, it's a horizontal slice $H \times \{\zeta\}$ across a divided flow box $H \times L$. The quotient map π projects horizontal segments on $\overleftrightarrow{\Sigma}$ down to horizontal segments on Σ, as defined in Section 9.2.2. Lemma 9.2.A, which made it sensible to talk about first return maps on a translation surface, has an analogue on the divided surface.

Lemma 9.5.A. *Let Z be a horizontal segment on $\overleftrightarrow{\Sigma}$, and let p be a point in Z. Unless p is on the median of a forward-critical road, the vertical flow will eventually carry p back to Z.*

The proof is edifying, but rather long, so I've postponed it to the end of the section. The most important consequence of this lemma is that we can define a first return relation on any horizontal segment in $\overleftrightarrow{\Sigma}$, just like we did for horizontal segments in Σ.

Suppose Z is a horizontal slice across a well-cut flow box $U = H \times L$ in $\overleftrightarrow{\Sigma}$. Then $\iota^{-1}Z$ is a well-cut horizontal segment in Σ, so the first return relation on $\iota^{-1}Z$ is an interval exchange. The divided version of that interval exchange, constructed as in Section 9.3.2, is precisely the first return relation on Z.

Identifying Z and $\iota^{-1}Z$ with H and $\iota^{-1}H$ in the obvious way, we can think of the first return relations as relations on H and $\iota^{-1}H$. These relations don't depend on which slice across $H \times L$ we take, so we can talk about the

first return relations on H and $\iota^{-1}H$ defined by a well-cut flow box $H \times L$ without picking a slice at all. The first return relation becomes a function when restricted to $\dot{\dot{H}}$.

We'll be working with these first return relations a lot, so it will be useful to set down a pattern of notation for talking about them. For convenience, I'll refer to the first return relations on H and $\iota^{-1}H$ both as α. Let $W \subset \iota^{-1}H$ be the positions of the critical leaves, recalling that $\dot{\dot{H}} = H \smallsetminus \iota W$. Within W, let W^+ and W^- be the positions of the backward- and forward-critical leaves. If Σ has no saddle connections, W^+ and W^- are disjoint, and thus form a partition of W.

Let $B^+ \subset W^+$ be the places where the backward-critical leaves first pass through $\iota^{-1}U$ after shooting out of their singularities. Similarly, let $B^- \subset W^-$ be the places where the forward-critical leaves last pass through $\iota^{-1}U$ before diving into their singularities. The sets B^+ and B^- are the break points of the backward and forward first return relations α^{-1} and α on $\iota^{-1}H$, as described in Section 9.3.2.

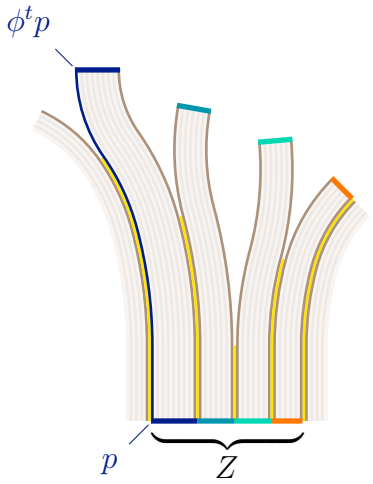
Proof of Lemma 9.5.A. If p is in $\iota\Sigma$, we can just apply Lemma 9.2.A to an appropriate closed subinterval of πZ, and we're done. The only time we have to do something less direct is when p is in the left or right lane of a critical road.

Suppose $p = \vec{w}$ for some $w \in \mathfrak{W} \smallsetminus \mathfrak{W}^-$. Because it's in the right lane of a critical road, p can't be the rightmost element of a horizontal basis interval.

We can therefore assume, without loss of generality, that Z has no rightmost element. In this case, the only way for $\pi^{-1}\pi Z$ to contain more points than Z is for Z to have a leftmost point $\vec{a}$, in which case $\pi^{-1}\pi Z = \{\overleftarrow{a}, \hat{a}\} \cup Z$.

By Lemma 9.2.A, $\psi^t w$ returns to πZ at some time $t > 0$. Away from $\mathfrak{W}^-$ and its π-preimage, the forward vertical flows on $\overleftrightarrow{\Sigma}$ and Σ commute, so $\pi\psi^t p \in \pi Z$. In other words, $\psi^t p \in \pi^{-1}\pi Z$. Points traveling along the vertical flow on $\overleftrightarrow{\Sigma}$ never change lanes, so $\psi^t p$ is in a right lane. Hence, knowing that $\psi^t p \in \pi^{-1}\pi Z$, we can conclude that $\psi^t p \in Z$, which is what we wanted to show. The same reasoning can be used when $p = \overleftarrow{w}$ for some $w \in \mathfrak{W} \setminus \mathfrak{W}^-$.

On the other hand, suppose $p = \vec{w}$ for some $w \in \mathfrak{W}^-$. In this case, we can assume without loss of generality that p is the leftmost point of Z. Follow $\psi^t Z$ upward along the vertical flow until $\psi^t p$ passes a singularity, exiting the forward-critical road it started on and merging onto the adjacent backward-critical road. By this time, a few pieces of $\psi^t Z$ may have already hit singularities and peeled off to the right, but some piece of $\psi^t Z$ is still traveling with $\psi^t p$.



The backward-critical road $\psi^t p$ is now following might also be forward-critical—a saddle connection. In that case, we can repeat the same maneuver. Because Σ is compact, it has only finitely many forward-critical roads, so $\psi^t p$ will eventually end up on either a road that isn't forward-critical or a road that it's traveled before. In the latter case, because $\psi^t p$ can only merge onto a road at its very beginning, $\psi^t p$ will eventually return to a point it's passed through before. Since ψ is a flow by bicontinuous relations, as defined in Appendix B, it follows that $\psi^t p$ is defined for all $t \in \mathbb{R}$, and periodic. As a result, $\psi^t p$ will eventually return to its starting point, and thus to Z.

This leaves us with the case where $\psi^t p$ finally merges onto a road that isn't forward critical. In other words, at some time $t > 0$, we have $\pi\psi^t p \in \mathfrak{W} \setminus \mathfrak{W}^-$. Let Y be the piece of $\psi^t Z$ that's stuck with $\psi^t p$ all this time. We showed earlier that the vertical flow eventually carries $\psi^t p$ back to Y. The vertical flow is injective, so $\psi^t p$ must have passed through $\psi^{-t} Y$ on its way

back to Y. Since $\psi^{-t}Y \subset Z$, we've proven that $\psi^t p$ eventually returns to Z. $\square$

9.5.3 Minimality

Assuming that the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal will make our discussion of abelianization and the conditions required for it go much more smoothly—so much so that I suspect it's a necessary condition for abelianization to make sense. Fortunately, this condition follows from a simple condition on Σ.

Proposition 9.5.B. *If Σ has no saddle connections, the forward and backward vertical flows on $\overset{\leftrightarrow}{\Sigma}$ are both minimal as dynamical systems.*

Proof. The forward and backward cases are mirror images of each other, so let's focus on the forward vertical flow. Suppose Σ has no saddle connections, recalling that this implies Σ is minimal. We want to prove that the forward orbit $\mathcal{P}$ of any point $p \in \overset{\leftrightarrow}{\Sigma}$ is dense.

Suppose $p = \iota q$ for some $q \in \Sigma \smallsetminus \mathfrak{W}$. Then we can just observe that the minimality of Σ implies that the forward orbit of q is dense in Σ. Since $\iota\Sigma$ is dense in $\overset{\leftrightarrow}{\Sigma}$ by Proposition 9.4.A, we're done.

Suppose p is in the left or right lane of a backward-critical road. In other words, $p \in \{\overleftarrow{w}, \vec{w}\}$ for some $w \in \mathfrak{W}^+$. Let $\mathcal{W}$ be the forward orbit of w. Because Σ has no saddle connections, the backward-critical leaf w lies on can't also be forward-critical, so the minimality of Σ implies that $\mathcal{W}$ is dense in Σ.

Because Σ is minimal, full flow boxes form a basis for $\overset{\leftrightarrow}{\Sigma}$, as pointed out in Section 9.4.2. Thus, to show that $\mathcal{P}$ is dense in $\overset{\leftrightarrow}{\Sigma}$, we just have to show that it intersects every full flow box $\dot{\dot{U}}$. Note that U is a full flow box in $\overset{\leftrightarrow}{\Sigma}$. Since ι is continuous, $\iota^{-1}U$ is an open subset of Σ, and therefore intersects the dense orbit $\mathcal{W}$. That means $\pi^{-1}\iota^{-1}U$ intersects $\iota\mathcal{W}$. Recall from Section 9.4.2 that $\pi^{-1}\iota^{-1}U = \operatorname{trim} U$. In $\overset{\leftrightarrow}{\Sigma}$, any open subset that contains a point in the median of a critical road also contains the corresponding points in the left and right lanes, so $\operatorname{trim} U$ intersects $\mathcal{P}$. Since $\operatorname{trim} U \subset U$, we've shown that U intersects $\mathcal{P}$. Since $\mathcal{P} \subset \overset{\leftrightarrow}{\Sigma}$, it follows that $\dot{\dot{U}}$ intersects $\mathcal{P}$. Since $\dot{\dot{U}}$ could have been any full flow box in $\overset{\leftrightarrow}{\Sigma}$, we've proven that $\mathcal{P}$ is dense in $\overset{\leftrightarrow}{\Sigma}$, under the assumption that p is in the left or right lane of a backward-critical road.

Finally, suppose p is in the left or right lane of a forward-critical road. The forward vertical flow will eventually carry p past a singularity, where it will leave its current road and merge onto an adjacent backward-critical road. That means the forward orbit of p contains the left or right lane of a backward-critical road. We just proved that the left and right lanes of a backward-critical road are both dense in $\overset{\leftrightarrow}{\Sigma}$, so we're done. □

On the fractured surface, a minimal vertical flow induces minimal first return maps.

Proposition 9.5.C. *Let $H \times L$ be a well-cut flow box in $\overset{\leftrightarrow}{\Sigma}$. If the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal, the first return map on $\dot{\dot{H}}$ is too.*

Proof. Suppose the first return map on $\ddot{H}$ is not minimal. Find a closed invariant subset $C \subset \ddot{H}$ which is neither empty nor all of $\ddot{H}$. Let $\mathcal{C}$ be the orbit of $C \times L$ under the vertical flow. Notice that $\mathcal{C}$ can't intersect the open set $(\ddot{H} \smallsetminus C) \times L$: if it did, the first return map on $\ddot{H}$ would send some element of C into $\ddot{H} \smallsetminus C$. Since $\mathcal{C}$ is made of vertical orbits, we've shown that not every orbit of the vertical flow is dense in $\overset{\leftrightarrow}{\Sigma}$. Hence, the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is not minimal. □

9.5.4 Ergodic theory

Proposition 9.4.A says that $\overset{\leftrightarrow}{\Sigma}$ is no bigger than Σ with respect to continuous functions, and Theorem 9.4.E says the same with respect to locally constant sheaves. It will be useful to have a similar result with respect to vertically invariant measures. Such a thing should be true, because passing from $\Sigma \smallsetminus \mathfrak{W}$ to $\overset{\leftrightarrow}{\Sigma}$ just means adding finitely many vertical leaves, which ought to have measure zero.

To turn these intuitions into something tangible, first equip Σ and $\overset{\leftrightarrow}{\Sigma}$ with their Borel σ-algebras. You can show, with a little thought, that any vertical slice across a flow box of $\overset{\leftrightarrow}{\Sigma}$ is measurable. It follows that every vertical leaf of $\overset{\leftrightarrow}{\Sigma}$ is measurable, allowing us to formulate the following proposition.

Proposition 9.5.D. *If the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal, an invariant probability measure on $\overset{\leftrightarrow}{\Sigma}$ assigns measure zero to every vertical leaf.*

Proof. Assume the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal. Let $\mathcal{L}$ be a vertical leaf.

Pick a segment $\mathcal{L}_0$ of $\mathcal{L}$ that looks like $\{s\} \times [0, \tau)$ in some divided flow box $\ddot{H} \times (-\tau, \tau)$. Observe that $\mathcal{L}_0$ is measurable. Let $\mathcal{L}_n = \psi^{n\tau}\mathcal{L}_0$ for each integer n. The segments $\mathcal{L}_n$ can't overlap locally, and a minimal flow on $\overset{\leftrightarrow}{\Sigma}$ has no periodic orbits, so in fact all the segments are disjoint.

Let μ be an invariant probability measure on $\overset{\leftrightarrow}{\Sigma}$. The segments $\mathcal{L}_n$ tile $\mathcal{L}$, so

$$\mu\mathcal{L} = \sum_{n\in\mathbb{Z}} \mu\mathcal{L}_n.$$

Because it's invariant, μ assigns the same measure to each $\mathcal{L}_n$. Hence,

$$\mu\mathcal{L} = \sum_{n\in\mathbb{Z}} \mu\mathcal{L}_0.$$

For the sum on the right-hand side to converge, its value must be zero. □

Corollary 9.5.E. *If the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal, an invariant probability measure on $\overset{\leftrightarrow}{\Sigma}$ assigns measure one to $\iota(\Sigma \smallsetminus \mathfrak{W})$.*

Proof. The only parts of $\overset{\leftrightarrow}{\Sigma}$ that don't lie in $\iota(\Sigma \smallsetminus \mathfrak{W})$ are the vertical leaves containing the left and right lanes of the critical roads. There are only finitely many of these, and each one has measure zero by the previous proposition. □

The corollary above leads to the following proposition, which links the ergodic properties of $\overset{\leftrightarrow}{\Sigma}$ and $\Sigma \smallsetminus \mathfrak{W}$.

Proposition 9.5.F. *If the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal, then it's uniquely ergodic if and only if the vertical flow on $\Sigma \smallsetminus \mathfrak{W}$ is.*

Proof. Assume the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal. The pushforward of an ergodic measure along a morphism of dynamical systems is ergodic, and the restriction of an ergodic measure to a measure-one invariant subspace is also ergodic. Therefore, $\overset{\leftrightarrow}{\Sigma}$ has an ergodic measure if and only if $\Sigma \smallsetminus \mathfrak{W}$ does.

Suppose $\Sigma \smallsetminus \mathfrak{W}$ is not uniquely ergodic. Distinct ergodic measures on $\Sigma \smallsetminus \mathfrak{W}$ push forward along ι to distinct ergodic measures on $\overset{\leftrightarrow}{\Sigma}$, so $\overset{\leftrightarrow}{\Sigma}$ isn't uniquely ergodic either.

On the other hand, suppose $\Sigma \smallsetminus \mathfrak{W}$ is uniquely ergodic. Then every ergodic measure on $\overset{\leftrightarrow}{\Sigma}$ must restrict to the same measure on $\Sigma \smallsetminus \mathfrak{W}$. All ergodic measures on $\overset{\leftrightarrow}{\Sigma}$ assign $\iota(\Sigma \smallsetminus \mathfrak{W})$ measure one, so if they agree on $\iota(\Sigma \smallsetminus \mathfrak{W})$, they agree everywhere. Hence, $\overset{\leftrightarrow}{\Sigma}$ is uniquely ergodic. □

Just like minimality, unique ergodicity of the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ can be localized to any well-cut flow box, as long as the vertical flow is minimal.

Proposition 9.5.G. *Let $H \times L$ be a well-cut flow box in $\overset{\leftrightarrow}{\Sigma}$. If the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal and uniquely ergodic, the first return map on $\dot{H}$ is too.*

The proof should bring to mind a fractured analogue of the idea that a minimal translation surface can be expressed as a suspension of the first return map on any well-cut segment [34]. We don't need the full power of this idea, however, so we won't develop it in detail.

Proof. Pick a horizontal slice $Z = H \times \{\zeta\}$ across $H \times L$. Removing the break points of the first return map from Z turns it into a disjoint union of full basis

intervals $Z_1, \ldots, Z_k$. As each segment Z_j is carried back to Z by the vertical flow, it sweeps out a "full half-open box" of the form $Y_j \cong Z_j \times [0, T_j)$. Because the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal, $\overset{\cdot\cdot}{Y}_1, \ldots, \overset{\cdot\cdot}{Y}_k$ cover $\overset{\leftrightarrow}{\Sigma}$, forming a measurable partition of it.

A measure on $\overset{\cdot\cdot}{Z}$ restricts to a measure on each segment $\overset{\cdot\cdot}{Z}_j$. The measures on $\overset{\cdot\cdot}{Z}_1, \ldots, \overset{\cdot\cdot}{Z}_k$ induce product measures on $\overset{\cdot\cdot}{Y}_1, \ldots, \overset{\cdot\cdot}{Y}_k$, which fit together into a measure on $\overset{\leftrightarrow}{\Sigma}$. In this way, a measure μ on $\overset{\cdot\cdot}{Z}$ can be extruded to a measure $\tilde{\mu}$ on $\overset{\leftrightarrow}{\Sigma}$. If μ is invariant under the first return map, $\tilde{\mu}$ is invariant under the vertical flow. If μ is a probability measure, $\tilde{\mu}\overset{\leftrightarrow}{\Sigma}$ is finite, so we can normalize $\tilde{\mu}$ to get a probability measure. Different probability measures on Z produce different probability measures on $\overset{\leftrightarrow}{\Sigma}$.

Let μ be an ergodic measure on Z, and $\tilde{\mu}$ the measure on $\overset{\leftrightarrow}{\Sigma}$ constructed from it. Consider an invariant subset $\mathcal{Q}$ of $\overset{\leftrightarrow}{\Sigma}$. Because $\mathcal{Q}$ is invariant, its intersection with the box $\overset{\cdot\cdot}{Y}_j$ looks like $Q_j \times [0, T_j)$, where $Q_j \subset Z_j$ is a measurable set. The $\tilde{\mu}$-measure of $\mathcal{Q}$ is given in terms of μ by

$$\tilde{\mu}\mathcal{Q} = T_1\, \mu Q_1 + \ldots + T_k\, \mu Q_k.$$

The disjoint union $Q = Q_1 \cup \ldots \cup Q_k$ is an invariant subset of Z, so its μ-measure is either zero or one. It follows that $\tilde{\mu}\mathcal{Q}$ is either zero or $T_1\, \mu Z_1 + \ldots + T_k\, \mu Z_k$, which is equal to $\tilde{\mu}\overset{\leftrightarrow}{\Sigma}$. We now see that $\tilde{\mu}$ assigns each invariant subset either zero measure or full measure, so its normalization is an ergodic measure on $\overset{\leftrightarrow}{\Sigma}$.

We've shown that any ergodic measure on Z can be used to produce an ergodic measure on $\overset{\leftrightarrow}{\Sigma}$, and we remarked earlier that different probability

measures on Z produce different probability measures on $\overset{\leftrightarrow}{\Sigma}$. Hence, if the first return map on Z isn't uniquely ergodic, the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ can't be either. □

When the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is uniquely ergodic, we can make simple and accurate long-term predictions about how often it will carry a point through a given well-cut segment. To see how this works, consider a horizontal slice $Z = H \times \{\zeta\}$ across a well-cut flow box $H \times L$ in $\overset{\leftrightarrow}{\Sigma}$. For each $h \in H$ and $n \in \mathbb{Z}$, let τ_h^n be the time it takes for the vertical flow to carry (h, ζ) back to Z for the nth time. Negative values of n refer to returns that happened in the past, so n and τ_h^n always have the same sign. Notice that $\psi^{\tau_h^n}(h, \zeta) = (\alpha^n h, \zeta)$.

Let W^+ and W^- be the places where the backward- and forward-critical leaves pierce Z. When $h \in W^+$, the backward vertical flow only carries (h, ζ) through Z finitely many times before it falls into a singularity, so τ_h^n is undefined for large negative n. Similarly, when $h \in W^-$, the time τ_h^n is undefined for large positive n. We should therefore think of τ as a relation, rather than a function. We'll call it the *return time relation* on Z. It restricts to an honest function when we consider forward returns to $H \smallsetminus W^-$, or backward returns to $H \smallsetminus W^+$. We can also think of τ as a relation on the well-cut segment πZ in Σ by restricting it to $\iota\pi Z$.

Proposition 9.5.H. *Consider a horizontal slice $Z = H \times \{\zeta\}$ across a well-cut flow box $H \times L$ in $\overset{\leftrightarrow}{\Sigma}$, with $W^\pm \subset H$ defined as above. Let τ be the return time relation on H. When the first return map on $\dot{H}$ is uniquely ergodic, there's an*

average return time $\bar{\tau}$ *characterized by the fact that*

$$\lim_{n\to\infty} \frac{\tau_h^{\pm n}}{n} = \pm\bar{\tau}$$

uniformly over $h \in H \smallsetminus W^{\mp}$.

This result is a divided version of the one discussed after Lemma 1 of [35], and its proof is essentially the same.

Proof of Proposition 9.5.H. The relation $h \mapsto \tau_h^1$ restricts to a continuous function on $\ddot{H}$. The first return map on $\ddot{H}$ is always continuous, and we're assuming it's uniquely ergodic as well. Recalling that $\ddot{H}$ is compact and metrizable, we can apply a version of the uniform ergodic theorem[4] to get the desired forward and backward uniform convergence over $\ddot{H}$ [37, Proposition 4.7.1].

To see how the forward convergence extends over $H \smallsetminus W^-$, consider any $\hat{w} \in W^+ \smallsetminus W^-$. The point $(\hat{w}, \zeta)$ lies on the median of a road which is backward-critical, but not forward-critical. Looking back to Section 9.5.1, we see that the three points $(\overleftarrow{w}, \zeta), (\hat{w}, \zeta), (\vec{w}, \zeta)$ travel together forever under the forward vertical flow. As a result, $\tau_{\hat{w}}^n$ is defined for all positive n, and it has the same value as $\tau_{\overleftarrow{w}}^n$ and $\tau_{\vec{w}}^n$ when it's defined. Since $\overleftarrow{w}$ and $\vec{w}$ are in $\ddot{H}$, we're back to the previously solved case. Switching directions, we can use the same argument to show that the uniform convergence of the backward limit extends over $H \smallsetminus W^+$. □

[4] As far as I can tell, it started out as a corollary to Lin's uniform ergodic theorem [36].

9.5.5 The fat gap condition

On a generic translation surface, you should expect the critical leaves to fill up each flow box more or less evenly as you follow them out from the singularities, rather than clumping together. The *fat gap condition* formalizes this property. When horizontal distances on the fractured surface are measured using the the division metric from Section 9.3.5, the fat gap condition implies the hypothesis of Corollary 9.3.L, which restricts the behavior of Hölder functions on horizontal segments. This control over horizontal Hölder functions will be the crux of our argument in Section 14 that abelianization does what it's supposed to.

For convenience, we'll keep track of first return relation break points in this section using the height function from Section 9.3.5, with steepness Δ. The definition we're about to present won't depend on the height function, but stating it in terms of the height function will allow for more consistent notation, and it will help us connect the fat gap condition with the properties of the division metric.

The local definition Consider a well-cut horizontal segment Z in Σ. As usual, let $W \subset Z$ be the positions of the critical leaves. Recall from Section 9.3.5 that $W_{\geq r}$ is the set of points in W with heights greater than or equal to r, and gap_r is the minimum distance between points in $W_{\geq r}$ according to the Euclidean metric on Z. Let's say Z satisfies the *fat gap condition* if

for all $\lambda > 0$, no matter how small,

$$\Delta^{\lambda n} \operatorname{gap}_{\Delta^{-n}}$$

is bounded below by a positive number as n varies in $\mathbb{N}$. Intuitively, the fat gap condition says that the break points of the iterated first return relations α^n and α^{-n} don't cluster too much as n grows. Specifically, the gaps between the break points shrink more slowly than any exponential. We'll say Σ satisfies the fat gap condition if all its well-cut segments do.

A global sufficient condition The fat gap condition, as stated, can be a bit of a pain to check, because it says something about the first return map on every well-cut segment. Fortunately, there's a more globally defined condition which implies the fat gap condition, and is still weak enough to hold for generic translation surfaces.

Following the terminology of [38], let's say a translation surface is φ-Diophantine if it satisfies the conclusion of part (1) of Theorem 1.1 in [39]. Here, $\varphi\colon [0,\infty) \to (0,\infty)$ is a strongly decreasing function.

Proposition 9.5.I. *Suppose Σ has no saddle connections, and the vertical flow on $\Sigma \smallsetminus \mathfrak{W}$ is uniquely ergodic. If Σ is $(\Delta^{-\lambda t})$-Diophantine for every $\lambda > 0$, then Σ satisfies the fat gap condition.*

For any $\lambda > 0$, Proposition 1.2 of [39] can be used to show that if you rotate a translation structure through a full circle, almost all the structures

you pass through will be $(\Delta^{-\lambda t})$-Diophantine. Hence, in any rotation family of translation structures, almost all will be $(\Delta^{-\lambda t})$-Diophantine for every $\lambda \in \{\frac{1}{m} : m \in \mathbb{N}_{>0}\}$, and thus for every $\lambda > 0$.

To prove Proposition 9.5.I, we'll need to introduce the notion of a *generalized saddle connection* on a translation surface. This is a geodesic, with respect to the flat metric, from one singularity to another. Equivalently, it's a saddle connection with respect to a rotated translation structure. In [39], Marchese refers to generalized saddle connections simply as saddle connections.

Every translation surface has a canonical complex-valued 1-form ω that sends horizontal unit vectors to 1 and vertical unit vectors to i. In a local translation chart $z\colon \Sigma \to \mathbb{R}^2$, thinking of $\mathbb{R}^2$ as $\mathbb{C}$, this is just the 1-form dz. For any generalized saddle connection γ on Σ, let

$$Q_\gamma = \int_\gamma \omega.$$

The real and imaginary parts of Q_γ measure the horizontal and vertical distances traveled by γ.

Proof of Proposition 9.5.I. Pick any well-cut segment Z in Σ. As usual, let W^+ and W^- be the places where the backward- and forward-critical leaves pierce Z, and let $B^\pm \subset W^\pm$ be the break points of $\alpha^{\mp 1}$. Because Σ has no saddle connections, W^+ and W^- are disjoint, so the relation $\alpha^\pm$ restricts to a function on $W^\pm$.

Let's start with some tools for keeping track of time. Let τ be the return time relation on Z. Let T be the maximum forward return time for a

point on Z, observing that $|\tau_b^n| < |n|T$ for all $b \in B^+ \cup B^-$ and $n \in \mathbb{Z}$. For each $b \in B^\pm$, let $-\rho_b$ be the time at which the vertical flow carries b into its singularity $\mathfrak{b}_b$. Notice that ρ_b is positive for $b \in B^+$ and negative for $b \in B^-$. Let P be the maximum of $|\rho_b|$ over all break points b.

For each $n \in \mathbb{N}$, find a pair of points in $W_{\geq \Delta^{-n}}$ that are as close together as possible, so the distance between then is $\text{gap}_{\Delta^{-n}}$. Because they're in W, these points can be expressed as $\alpha^k u$ and $\alpha^l v$ for some break points $u, v \in B^+ \cup B^-$. For concreteness, let's say $\alpha^k u$ is to the left of $\alpha^l v$. Because $\alpha^k u$ and $\alpha^l v$ are in $W_{\geq \Delta^{-n}}$, we know that $|k|$ and $|l|$ are at most n.

All the break points of $\alpha^{\pm(n+1)}$ are in $W_{\geq \Delta^{-n}}$, so none of them can lie between $\alpha^k u$ and $\alpha^l v$. Thus, as t varies from 0 to $-\tau_u^k$, the vertical flow ψ^t keeps the interval $(\alpha^k u, \alpha^l v)$ in one piece, eventually bringing it to the interval $(u, \alpha^{l-k} v)$. In fact, if we nudge t out to $-\tau_u^k - \rho_u$, the interval still holds together, coming to rest with its left endpoint at $\mathfrak{b}_u$. Similarly, ψ^t keeps the interval $(\alpha^k u, \alpha^l v)$ in one piece as t varies from 0 to $-\tau_v^l - \rho_v$, finally parking it with its right endpoint at $\mathfrak{b}_v$. The vertical flow from times $-(\tau_u^k + \rho_u)$ to $-(\tau_v^l + \rho_v)$ therefore sweeps out a flow box U_n with $\mathfrak{b}_u$ at one corner and $\mathfrak{b}_v$ at the opposite corner.

The diagonal of U_n from $\mathfrak{b}_u$ to $\mathfrak{b}_v$ is a generalized saddle connection, which I'll call γ_n. Recall that γ_n was constructed from a pair of points in $W_{\geq \Delta^{-n}}$ that are as close together as possible. There could be several such pairs, but it won't matter which one we picked. We can find Q_{γ_n} by observing that its real and imaginary parts are the width and height of U_n, with appropriate signs.

The real part is the distance from $\alpha^k u$ to $\alpha^l v$, which we set up to be $\operatorname{gap}_{\Delta^{-n}}$. The imaginary part is $-(\tau_v^l + \rho_v) + (\tau_u^k + \rho_u)$. Letting ℓ be the length of the segment Z, we get the bound

$$\begin{aligned} |Q_{\gamma_n}| &\leq |-(\tau_v^l + \rho_v) + (\tau_u^k + \rho_u)| + \operatorname{gap}_{\Delta^{-n}} \\ &\leq |\tau_v^l| + |\tau_u^k| + |\rho_v| + |\rho_u| + \ell. \end{aligned}$$

Recalling the constants T and P we defined earlier, we can simplify this to

$$\begin{aligned} |Q_{\gamma_n}| &\leq |l|T + |k|T + 2P + \ell \\ &\leq 2nT + 2P + \ell. \end{aligned}$$

Now, suppose Σ is $(\Delta^{-\lambda t})$-Diophantine for all $\lambda > 0$. This means that, for any $\lambda > 0$, there are only finitely many generalized saddle connections γ with

$$|\operatorname{Re} Q_\gamma| < \Delta^{-\lambda |Q_\gamma|}.$$

We want to prove Z satisfies the fat gap condition. In other words, given any $\mu > 0$, we want to show that

$$\Delta^{\mu n} \operatorname{gap}_{\Delta^{-n}}$$

is bounded below by a positive number as n varies in $\mathbb{N}$.

Pick any $\mu > 0$, and let $\lambda = \frac{\mu}{2T}$. Because Σ is minimal, $\operatorname{gap}_{\Delta^{-n}}$ goes to zero as n grows, so each generalized saddle connection has only finitely many chances to appear in the sequence γ_n. Hence, by the Diophantine condition,

$$|\operatorname{Re} Q_{\gamma_n}| < \Delta^{-\lambda |Q_{\gamma_n}|}$$

for only finitely many $n \in \mathbb{N}$. Because $\Delta^{-\lambda t}$ is decreasing, we can infer from our earlier bound on $|Q_{\gamma_n}|$ that $\Delta^{-\lambda(2nT+2P+\ell)}$ is no greater than $\Delta^{-\lambda|Q_{\gamma_n}|}$, so

$$|\operatorname{Re} Q_{\gamma_n}| < \Delta^{-\lambda(2nT+2P+\ell)}$$

for only finitely many $n \in \mathbb{N}$. Recalling that $|\operatorname{Re} Q_{\gamma_n}| = \operatorname{gap}_{\Delta^{-n}}$ and rearranging the right-hand side, we see that

$$\operatorname{gap}_{\Delta^{-n}} < \Delta^{-\lambda(2P+\ell)}\Delta^{-\mu n}$$

for only finitely many $n \in \mathbb{N}$. In other words,

$$\Delta^{\mu n} \operatorname{gap}_{\Delta^{-n}} < \Delta^{-\lambda(2P+\ell)}$$

for only finitely many $n \in \mathbb{N}$. Since the right-hand side is positive and independent of n, it follows immediately that the left-hand side is bounded below by a positive number as n varies in $\mathbb{N}$. □

Implications for the division metric Now that we have a practical way to enforce the fat gap condition, let's see how it implies the hypothesis of Corollary 9.3.L, restricting the behavior of Hölder functions on horizontal segments.

Proposition 9.5.J. *Let Z be a well-cut segment in Σ. Suppose Z satisfies the fat gap condition. The gap function on Z then has the property that for any $\lambda > 0$, we can find a positive constant M with $\operatorname{gap}_r \geq Mr^\lambda$.*

Proof. Pick any $\lambda > 0$. By the fat gap condition, there's some $\varepsilon > 0$ such that

$$\Delta^{\lambda n} \operatorname{gap}_{\Delta^{-n}} \geq \varepsilon$$

for all large enough $n \in \mathbb{N}$. Equivalently,

$$\operatorname{gap}_{\Delta^{-n}} \geq \varepsilon \Delta^{-\lambda} \Delta^{-\lambda(n-1)}$$

for all large enough $n \in \mathbb{N}$. That means

$$\operatorname{gap}_{\Delta^{-n}} \geq M(\Delta^{-n+1})^{\lambda}$$

for some $M > 0$ small enough to absorb the constant $\varepsilon\Delta^{-\lambda}$ and deal with the transient behavior of $\operatorname{gap}_{\Delta^{-n}}$ at small n. For every $r > 0$ we have $\Delta^{-n} \leq r \leq \Delta^{-n+1}$ for some n, so it follows that

$$\operatorname{gap}_r \geq Mr^{\lambda}$$

for all r. □

Chapter 10

Warping local systems on divided surfaces

10.1 Overview

Many classic geometric constructions involve cutting, shifting, and regluing a local system on a manifold along something akin to a codimension-one submanifold. For example, a Fenchel-Nielsen twist shifts the $\mathrm{PSL}_2\,\mathbb{R}$ local system encoding the hyperbolic structure of a surface along a closed geodesic. A grafting or a cataclysm shifts a $\mathrm{PSL}_2\,\mathbb{R}$ local system along a geodesic lamination on a hyperbolic surface. Higher versions of these processes act on $\mathrm{PSL}_2\,\mathbb{C}$ and $\mathrm{PSL}_n\,\mathbb{R}$ local systems [40][41].

The version of abelianization described in this paper shifts an $\mathrm{SL}_2\,\mathbb{R}$ local system along the critical leaves of a translation surface Σ. It's most conveniently carried out by pushing the local system up to the divided surface $\overset{\leftrightarrow}{\Sigma}$ and warping it along a deviation supported on $\overset{\diamond}{\Sigma}$. The special class of deviations used in this process will be the subject of this section.

Our discussion of warping only makes sense on a locally connected space, so the local connectedness of the divided surface is now playing an important role. The fact that we can warp local systems on the divided surface, like the equivalence of categories of local systems given by Theorem 9.4.E, can

be seen as a reflection of the general idea that local systems on the divided surface tend to be well-behaved.

10.2 Critical leaves in a flow box

Consider a flow box $U = H \times L$ in $\overleftrightarrow{\Sigma}$. The median of each critical leaf passes through U at most countably many times, intersecting it in a collection of vertical lines I'll call *dividers*. The dividers are naturally ordered by their positions in H. Given two points y and x in $\dot{\dot{U}}$, with y sitting to the left of x along H, let's write $(y \mid x)^U$ to denote the ordered set of dividers in U that lie between y and x.

10.3 Deviations defined by jumps, conceptually

Let $\mathcal{F}$ be a locally constant sheaf on $\overleftrightarrow{\Sigma}$. The $\mathcal{F}$-simple flow boxes form a basis for the topology of $\overleftrightarrow{\Sigma}$. To specify a *jump* in $\mathcal{F}$, we give for each divider P in a simple flow box U an automorphism j_P of $\mathcal{F}_U$. These automorphisms have to fit together as follows:

- If the basis element U contains the basis element V, and the divider P in U contains the divider Q in V, the automorphisms j_Q and j_P commute with the restriction morphism $\mathcal{F}_{V \subset U}$.

Consider a simple flow box U and a pair of points y and x in $\dot{\dot{U}}$, with y to the left of x. In the presence of a jump j, the ordered set $(y \mid x)^U$ of dividers between y and x becomes an ordered set of automorphisms of $\mathcal{F}_U$,

which are just begging to be composed. There are probably infinitely many of them, so composing them may not make sense, but let's do it anyway, and call the result

$$\delta^U_{yx} = \prod_{P \in (y|x)^U} j_P.$$

For good measure, define δ^U_{xy} to be the inverse of δ^U_{yx}, so we don't have to worry about checking that y is to the left of x in U.

Our notation makes δ look like a deviation from $\mathcal{F}$ with support $\overset{\leftrightarrow}{\Sigma}$, defined over the basis of $\mathcal{F}$-simple flow boxes. In fact, δ really will be a deviation as long as the compositions defining it make sense, and behave in the way you'd expect. We'll see this concretely in the next section, where we specialize to the case of jumps in local systems.

10.4 Deviations defined by jumps, concretely

Let's say $\mathcal{F}$ is a G local system. In this case, a jump in $\mathcal{F}$ assigns an element of G to each divider. Because G is a linear group, it comes with a topology that can be used to make sense of infinite ordered products, as described in Appendix D. Using the properties of these products, we can show that the construction in the previous section really does produce a deviation δ from a jump j, as long as all the products converge.

10.4.1 The restriction property

We want to prove that for any simple flow boxes $V \subset U$ and any two points $y, x \in V \cap \overset{\leftrightarrow}{\Sigma}$, the automorphisms δ^U_{yx} and δ^V_{yx} commute with the restriction morphism $\mathcal{F}_{V \subset U}$. We can assume, without loss of generality, that y is to the left of x. Jumps are required to commute with restriction, so

$$\prod_{P \in (y|x)^U} j_P = \prod_{P \in (y|x)^V} \mathcal{F}^{-1}_{V \subset U}\, j_P\, \mathcal{F}_{V \subset U}.$$

Recalling that conjugation is a topological group automorphism, we apply Proposition D.4.A and get

$$\prod_{P \in (y|x)^U} j_P = \mathcal{F}^{-1}_{V \subset U} \left(\prod_{P \in (y|x)^V} j_P \right) \mathcal{F}_{V \subset U},$$

which is what we wanted to show.

10.4.2 The composition property

We want to prove that $\delta^U_{zy}\delta^U_{yx} = \delta^U_{zx}$ for any simple flow box U and any three points $z, y, x \in U \cap \overset{\leftrightarrow}{\Sigma}$. If z, y, x happen to be ordered from left to right, we're trying to prove that

$$\left(\prod_{P \in (z|y)^U} j_P \right) \left(\prod_{P \in (y|x)^U} j_P \right) = \prod_{P \in (z|x)^U} j_P.$$

Since, in the notation of Section D, $(z \mid y)^U \sqcup (y \mid x)^U = (z \mid x)^U$, this follows directly from Proposition D.3.A.

Now, suppose the three points are ordered differently. If the ordering from left to right is y, z, x, we can use the reasoning above to conclude that

$\delta^U_{yz}\delta^U_{zx} = \delta^U_{yx}$, rewrite this as $(\delta^U_{zy})^{-1}\delta^U_{zx} = \delta^U_{yx}$, and rearrange to get the desired result. The other cases work similarly.

Chapter 11

Uniform hyperbolicity for $\mathrm{SL}_2\,\mathbb{R}$ dynamics

11.1 Motivation and notation

In Section 10.3, we saw a way to turn a jump into a deviation by taking infinite products of jump automorphisms. To apply that construction to a given jump in a local system $\mathcal{E}$ on $\overset{\leftrightarrow}{\Sigma}$, we need some way of showing that all those infinite products converge. A convenient approach is to show that the jump automorphisms decay rapidly as you follow them out along the critical leaves of $\overset{\leftrightarrow}{\Sigma}$.

Uniform hyperbolicity is a powerful and well-studied decay condition on the sections of a local system over a dynamical base. It will be a key player in our abelianization procedure, with a role going far beyond ensuring that jumps converge.

A bit of notation will streamline our reasoning about growth and decay. For positive functions f and g, we'll write $f \lesssim g$ to say that f is bounded by a constant multiple of g. When f and g depend on several parameters, we might say that $f \lesssim g$ over a parameter t to specify that we're thinking of f and g as functions of t, with the other parameters held fixed.

11.2 The global version

Parallel transport along the vertical flow $\psi\colon \mathbb{R} \times \overset{\leftrightarrow}{\Sigma} \to \overset{\leftrightarrow}{\Sigma}$ induces a flow Ψ on the stalks of $\mathcal{E}$. At time t, the parallel transport flow gives a morphism $\Psi^t\colon \mathcal{E}_x \to \mathcal{E}_{\psi^t x}$ for each $x \in \overset{\leftrightarrow}{\Sigma}$. Pick an inner product on the stalks of $\mathcal{E}$ over $\overset{\leftrightarrow}{\Sigma}$ which is continuous in the sense that, for any two vectors $u, v \in \mathcal{E}_U$ over an open set $U \subset \overset{\leftrightarrow}{\Sigma}$, the inner product of u and v in the stalk $\mathcal{E}_x$ varies continuously as a function of $x \in \overset{\leftrightarrow}{\Sigma}$. Because $\overset{\leftrightarrow}{\Sigma}$ is compact, it doesn't matter which inner product we pick.

Saying $\mathcal{E}$ is *globally uniformly hyperbolic* means that for every $x \in \overset{\leftrightarrow}{\Sigma}$, the dynamics of Ψ split $\mathcal{E}_x$ into a direct sum of two one-dimensional subspaces $\mathcal{E}_x^+$ and $\mathcal{E}_x^-$, called the *forward-stable* and *backward-stable* lines, respectively. These lines are characterized by the following properties:

1. The parallel transport map Ψ^t sends $\mathcal{E}_x^\pm$ to $\mathcal{E}_{\psi^t x}^\pm$.

2. There is a constant $K > 0$ such that

$$\|\Psi^{\pm t} v\| \lesssim e^{-Kt}\|v\|$$

over all $x \in \overset{\leftrightarrow}{\Sigma}$, $v \in \mathcal{E}_x^\pm$, and $t \in [0, \infty)$.

We'll call the constant K a *bounding exponent* for $\mathcal{E}$. Bounding exponents are not by any means unique: if K is a bounding exponent for a uniformly hyperbolic local system, every positive number less than K is a bounding exponent too.

The decay condition in the definition of uniform hyperbolicity only talks about forward parallel transport, but we can turn it around to get an equivalent growth condition on backward parallel transport. Our argument depends crucially on the fact that x and v are treated as variables in the decay condition, rather than parameters that can be held fixed. That detail is the "uniform" part of "uniform hyperbolicity."

Proposition 11.2.A. *For a given value of K and choice of sign, Condition 2 in the definition of global uniform hyperbolicity holds if and only if*

$$e^{Kt}\|v\| \lesssim \|\Psi^{\mp t}v\|$$

over all $x \in \overset{\leftrightarrow}{\Sigma}$, $v \in \mathcal{E}^{\pm}_x$, and $t \in [0,\infty)$.

Proof. Suppose Condition 2 holds. Then there's a constant C such that

$$\|\Psi^{\pm t}v\| \leq Ce^{-Kt}\|v\|$$

for all $x \in \overset{\leftrightarrow}{\Sigma}$, $v \in \mathcal{E}^{\pm}_x$, and $t \in [0,\infty)$. In particular,

$$\|\Psi^{\pm t}\Psi^{\mp t}v'\| \leq Ce^{-Kt}\|\Psi^{\mp t}v'\|$$

for all $x \in \overset{\leftrightarrow}{\Sigma}$, $v' \in \mathcal{E}^{\pm}_{\psi^{\pm t}x}$, and $t \in [0,\infty)$. Simplifying the inequality and rewriting the quantifier over x, we see that

$$\|v'\| \leq Ce^{-Kt}\|\Psi^{\mp t}v'\|$$

for all $x' \in \overset{\leftrightarrow}{\Sigma}$, $v' \in \mathcal{E}^{\pm}_{x'}$, and $t \in [0,\infty)$. This implies the condition in the proposition.

We've now shown that Condition 2 implies the condition in the proposition. The same kind of reasoning can be used to prove the reverse implication. □

When the dynamics of Σ are favorable, uniform hyperbolicity implies much stronger conditions than it demands. Although it doesn't restrict the way the stable lines vary from leaf to leaf, it turns out to ensure that they vary continuously. Furthermore, although it only asks for vectors on the stable lines to decay at least exponentially fast, it ends up persuading them to decay at a precise exponential rate.

Lemma 11.2.B. *Suppose that Σ is minimal, and the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is uniquely ergodic. Then the stable lines $\mathcal{E}^{\pm}_x$ of a uniformly hyperbolic local system $\mathcal{E}$ vary continuously*[1] *with respect to $x \in \overset{\leftrightarrow}{\Sigma}$.*

Lemma 11.2.C. *When Σ is minimal, and the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is uniquely ergodic, Condition 2 in the definition of uniform hyperbolicity sharpens to the following:*

2′. *There is a constant $\Lambda > 0$ such that*

$$\lim_{t\to\infty} \frac{1}{t} \log \frac{\|\Psi_x^{\pm t} v\|}{\|v\|} = -\Lambda$$

uniformly[2] *over all $x \in \overset{\leftrightarrow}{\Sigma}$ and $v \in \mathcal{E}^{\pm}_x$.*

[1] To be precise, over any simple open set U in $\overset{\leftrightarrow}{\Sigma}$, the stable lines vary continuously in $\mathcal{E}_U$. For a more global perspective, you can realize $\mathcal{E}$ as a flat vector bundle over $\overset{\leftrightarrow}{\Sigma}$ and observe that the stable lines are continuous sections of its projectivization.

[2] In case of ambiguity, see the corresponding footnote in Lemma 11.3.C.

The constant Λ*, called the* Lyapunov exponent *of* $\mathcal{E}$*, is the supremum of the set of bounding exponents for* $\mathcal{E}$*.*

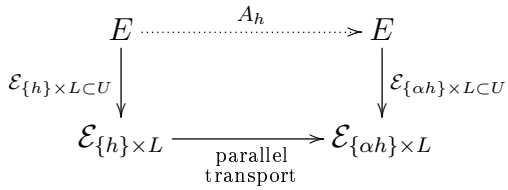
Proof of Lemma 11.2.B. Suppose $\mathcal{E}$ is globally uniformly hyperbolic. Under our assumptions, the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal and uniquely ergodic, satisfying the hypotheses of Proposition 11.4.A and Lemma 11.3.C. On each simple, well-cut flow box U in $\overset{\leftrightarrow}{\Sigma}$, the Proposition tells us that $\mathcal{E}$ is locally uniformly hyperbolic, and the Lemma goes on to imply that its stable lines vary continuously over $\ddot{U}$. Because Σ is minimal, the remarks in Section 9.4.2 ensure that simple, well-cut flow boxes cover $\overset{\leftrightarrow}{\Sigma}$. Hence, the stable lines of $\mathcal{E}$ vary continuously over $\overset{\leftrightarrow}{\Sigma}$. □

Proof of Lemma 11.2.C. Essentially the same as the proof of Proposition 11.3.C, which can be found in Appendix C. The biggest changes are in the argument that the angle between the stable lines is bounded away from zero. In this version, we use Lemma 11.2.B rather than Lemma 11.3.B to show that the stable lines are continuous, and we throw in an appeal to the continuity of the stalkwise inner product on $\mathcal{E}$. □

11.3 The local version

Consider a simple, well-cut flow box $U = H \times L$ in $\overset{\leftrightarrow}{\Sigma}$, and let $E = \mathcal{E}_U$. Let $\alpha\colon \ddot{H} \to \ddot{H}$ be the first return map discussed in Section 9.5.2. For each $h \in \ddot{H}$, parallel transport along the leaf through $\{h\} \times L$ gives an automorphism

A_h of E, defined by the commutative square



These automorphisms form a cocycle over α, which we'll call the *parallel transport cocycle*. Write A_h^n for the parallel transport along n iterations of α, starting at h. Pick an inner product on E (it doesn't matter which one).

Saying $\mathcal{E}$ is *locally uniformly hyperbolic* with respect to U means that for every $h \in \ddot{H}$, the dynamics of A split E into a direct sum of two one-dimensional subspaces E_h^+ and E_h^-, called the *forward-stable* and *backward-stable* lines, respectively. These lines are characterized by the following properties:

1. The parallel transport map A_h sends $E_h^\pm$ to $E_{\alpha h}^\pm$.

2. There is a constant $K > 0$ such that

$$\|A_h^{\pm n} v\| \lesssim e^{-Kn}\|v\|$$

   over all $h \in \ddot{H}$, $v \in E_h^\pm$, and $n \in \mathbb{N}$.

This is a standard definition of uniform hyperbolicity for dynamical cocycles [27, §2.2], specialized to the task at hand.

We can turn the decay condition around to get an equivalent growth condition on backward parallel transport, just like we did for the global version of uniform hyperbolicity.

Proposition 11.3.A. *For a given value of K and choice of sign, Condition 2 in the definition of local uniform hyperbolicity holds if and only if*

$$e^{Kn}\|v\| \lesssim \|A_h^{\mp n} v\|$$

over all $h \in \ddot{H}$, $v \in E_h^{\pm}$, and $n \in \mathbb{N}$.

Proof. Essentially the same as for Proposition 11.2.A. □

When the dynamics of the first return map are favorable, local uniform hyperbolicity implies much stronger conditions than it demands, again in analogy with the global story.

Lemma 11.3.B. *Suppose α is minimal and uniquely ergodic. Then the stable lines $E_h^{\pm}$ of a uniformly hyperbolic cocycle A vary continuously with respect to h.*

Lemma 11.3.C. *When α is minimal and uniquely ergodic, Condition 2 in the definition of uniform hyperbolicity sharpens to the following:*

2′. *There is a constant $\Lambda > 0$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\|A_h^{\pm n} v\|}{\|v\|} = -\Lambda$$

uniformly[3] *over all $h \in \ddot{H}$ and $v \in E_h^{\pm}$.*

[3]The uniform convergence of this limit means that, for any neighborhood Ω of the right-hand side, making n large enough guarantees that the left-hand side will be in Ω for all h and v.

The constant Λ*, called the* Lyapunov exponent *of* A*, is the supremum of the set of bounding exponents for* A*.*

These results hold in more generality than I've stated them in, and they may be useful outside the context of this paper. I've therefore placed their proofs in Appendix C, where they can be treated in the generality they deserve.

The sharp decay condition in Lemma 11.3.C can be turned around in the same way as the decay condition in the definition of uniformity.

Proposition 11.3.D. *For a given value of* Λ *and choice of sign, Condition* $2'$ *from the lemma holds if and only if*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\|A_h^{\mp n} v\|}{\|v\|} = \Lambda$$

uniformly over all $h \in \check{\check{H}}$ *and* $v \in E_x^{\pm}$*.*

This result, like the previous ones, is proven in more generality in Appendix C.

11.4 The two versions are usually equivalent

Suppose the vertical flow on $\overset{\leftrightarrow}{\Sigma}$ is minimal and uniquely ergodic. Then, for any simple, well-cut flow box U in $\overset{\leftrightarrow}{\Sigma}$, global uniform hyperbolicity is equivalent to local uniform hyperbolicity with respect to U. That means we can drop the distinction between them, and just talk about *uniform hyperbolicity*.

There's even a precise relationship between the local and global Lyapunov exponents, mediated by the average return time defined in Section 9.5.4.

The claims above are too much for me to prove in one go, so let's start with a fragment just strong enough to prove Lemma 11.2.C.

Proposition 11.4.A. *Suppose the vertical flow on $\overset{\circlearrowleft}{\Sigma}$ is minimal and uniquely ergodic. If the local system $\mathcal{E}$ is globally uniformly hyperbolic, then it's locally uniformly hyperbolic with respect to any simple, well-cut flow box in $\overset{\leftrightarrow}{\Sigma}$.*

Proof. Take any simple, well-cut flow box U in $\overset{\leftrightarrow}{\Sigma}$. Pick a horizontal slice $Z = H \times \{\zeta\}$ across U, and let τ be the return time relation on Z. We'll use all the notation and equipment of Sections 11.2 and 11.3. Because H is full, we can find a continuous function on $\overset{\circlearrowleft}{\Sigma}$ which is one on Z and zero outside of U. Using this function, we can pick the continuous inner product on the stalks of $\mathcal{E}$ over $\overset{\circlearrowleft}{\Sigma}$ so that it matches the inner product on E at every point in Z. Then, for any $h \in \ddot{H}$ and $v \in \mathcal{E}_{(h,\zeta)}$, we can say that $\|A_h^n v\| = \|\Psi^{\tau_h^n} v\|$, implicitly identifying E with $\mathcal{E}_{(\alpha^n h,\zeta)}$ through the stalk restriction $\mathcal{E}_{(\alpha^n h,\zeta)\in U}$.

Suppose $\mathcal{E}$ is globally uniformly hyperbolic, with bounding exponent K. Then there's a constant C such that

$$\log \frac{\|\Psi^{\pm t} v\|}{\|v\|} \leq -Kt + C$$

for all $h \in \ddot{H}$, $v \in \mathcal{E}^{\pm}_{(h,\zeta)}$, and $t \in [0, \infty)$. In particular, since $\pm\tau_h^{\pm n} \in [0, \infty)$ for all $n \in \mathbb{N}$,

$$\log \frac{\|\Psi^{\tau_h^{\pm n}} v\|}{\|v\|} \leq -K(\pm\tau_h^{\pm n}) + C$$

for all $h \in \ddot{H}$, $v \in \mathcal{E}^{\pm}_{(h,\zeta)}$, and $n \in \mathbb{N}$.

Let $E^{\pm}_h \subset E$ be the line that restricts to the stable line $\mathcal{E}^{\pm}_{(h,\zeta)} \subset \mathcal{E}_{(h,\zeta)}$. We're going to show that $E^{\pm}_h$ are stable lines for the parallel transport cocycle A. It's not hard to see that A_h sends $E^{\pm}_h$ to $E^{\pm}_{\alpha h}$, so we just need to verify the decay condition. By the inequality above,

$$\log \frac{\|A^{\pm n}_h v\|}{\|v\|} \leq -\left(\pm \frac{\tau^{\pm n}_h}{n}\right) Kn + C$$

for all $h \in \ddot{H}$, $v \in E^{\pm}_h$, and $n \in \mathbb{N}$. Pick a positive constant σ smaller than the average return time of Z. The limit defining the average return time in Proposition 9.5.H is uniform over $\ddot{H}$, so we can find a new constant $C' \geq C$ such that

$$\log \frac{\|A^{\pm n}_h v\|}{\|v\|} \leq -\sigma Kn + C'$$

for all $h \in \ddot{H}$, $v \in E^{\pm}_h$, and $n \in \mathbb{N}$. Equivalently,

$$\|A^{\pm n}_h v\| \leq e^{C'} e^{-\sigma Kn} \|v\|$$

for all $h \in \ddot{H}$, $v \in E^{\pm}_h$, and $n \in \mathbb{N}$. This verifies the decay condition, showing that $\mathcal{E}$ is locally uniformly hyperbolic with respect to U. □

With Lemma 11.2.C in hand, we can talk about Lyapunov exponents in the global setting as well as the local one, opening the way to a stronger result.

Proposition 11.4.B. *Suppose the vertical flow on $\overset{\circ}{\Sigma}$ is minimal and uniquely ergodic. Let U be a simple, well-cut flow box in $\overset{\leftrightarrow}{\Sigma}$, and let $\bar{\tau}$ be the average*

return time of a horizontal slice across U. The local system $\mathcal{E}$ is globally uniformly hyperbolic if and only if it's locally uniformly hyperbolic with respect to U. The local Lyapunov exponent is $\bar{\tau}$ times the global one.

The factor of $\bar{\tau}$ relating the local and global Lyapunov exponents reflects the fact that the local exponent is unitless, while the global one has units of inverse time.

Proof of Proposition 11.4.B. Pick a horizontal slice $Z = H \times \{\zeta\}$ across U, and let τ be the return time relation on Z. We'll keep using all the notation and equipment from the proof of Proposition 11.4.A, including the stalkwise inner product on $\mathcal{E}$ built to match the inner product on E.

In the proof of Proposition 11.4.A, we saw how a Ψ-invariant distribution of lines $\mathcal{E}_x^{\pm} \subset \mathcal{E}_x$ over $\overset{\leftrightarrow}{\Sigma}$ can be turned into an A-invariant distribution of lines $E_h^{\pm} \subset E$ parameterized by $\overset{\leftrightarrow}{H}$. An A-invariant line distribution $E_h^{\pm}$ extends uniquely to a Ψ-invariant line distribution $\mathcal{E}_x^{\pm}$ by the reverse process. We're going to show that $\mathcal{E}_x^{\pm}$ satisfies the global decay condition from Lemma 11.2.C if and only if $E_h^{\pm}$ satisfies the local decay condition from Lemma 11.3.C.

Suppose

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\|A_h^{\pm n} v\|}{\|v\|} = -\bar{\tau}\Lambda$$

uniformly over all $h \in \overset{\leftrightarrow}{H}$ and $v \in E_h^{\pm}$. From the definition in Proposition 9.5.H,

$$\lim_{n\to\infty} \frac{n}{\pm\tau_h^{\pm n}} = \frac{1}{\bar{\tau}}$$

uniformly over $h \in \ddot{H}$. It follows that

$$\lim_{n\to\infty} \frac{1}{\pm\tau_h^{\pm n}} \log \frac{\|A_h^{\pm n} v\|}{\|v\|} = -\Lambda$$

uniformly over all $h \in \ddot{H}$ and $v \in E_h^{\pm}$. Equivalently,

$$\lim_{n\to\infty} \frac{1}{\pm\tau_h^{\pm n}} \log \frac{\|\Psi^{\tau_h^{\pm n}} v\|}{\|v\|} = -\Lambda$$

uniformly over all $h \in \ddot{H}$ and $v \in \mathcal{E}^{\pm}_{(h,\zeta)}$.

Let T be the maximum forward return time for a point on Z. Because $\overset{\leftrightarrow}{\Sigma}$ is compact, and both the parallel transport flow and the stalkwise inner product on $\mathcal{E}$ are continuous, the ratio $\frac{\|\Psi^{t+\delta} u\|}{\|\Psi^t u\|}$ is bounded away from zero and infinity as u varies over the stalks of $\mathcal{E}$ and δ varies over the interval $[0, T]$. Considering the limit above in light of this fact, you can deduce that

$$\lim_{n\to\infty} \frac{1}{t} \log \frac{\|\Psi^{\pm t} v\|}{\|v\|} = -\Lambda$$

uniformly over all $h \in \ddot{H}$ and $v \in \mathcal{E}^{\pm}_{(h,\zeta)}$.

We've now shown that the local decay condition from Lemma 11.3.C, with Lyapunov exponent $\bar{\tau}\Lambda$, implies the global decay condition from Lemma 11.2.C, with Lyapunov exponent Λ. The reverse argument is similar, and slightly easier, since it avoids the subtlety of extending the limit from just the times $\pm\tau_h^{\pm n}$ to all the times $t \in [0, \infty)$. □

11.5 Extending over medians

Suppose that Σ has no saddle connections, and the vertical flow on $\Sigma \smallsetminus \mathfrak{W}$ is uniquely ergodic. Consider a point w on a critical leaf of Σ. Every

neighborhood of $\hat{w}$ contains $\overleftarrow{w}$ and $\vec{w}$, so we can take a colimit over neighborhoods U in the diagram

$$\mathcal{E}_{\overleftarrow{w}} \xleftarrow{\mathcal{E}_{\overleftarrow{w}\subset U}} \mathcal{E}_U \xrightarrow{\mathcal{E}_{\vec{w}\subset U}} \mathcal{E}_{\vec{w}}$$

to get isomorphisms

$$\mathcal{E}_{\overleftarrow{w}} \longleftarrow \mathcal{E}_{\hat{w}} \longrightarrow \mathcal{E}_{\vec{w}}$$

identifying the three stalks. I'll refer to all three as $\mathcal{E}_w$, writing $\overleftarrow{v}$, $\hat{v}$, or $\vec{v}$ when I want to think of a vector $v \in \mathcal{E}_w$ as an element of one or the other.

Let's say w is on a backward-critical leaf. The left and right lanes of this leaf are never separated by the forward vertical flow, so $\Psi^t\overleftarrow{v} = \Psi^t\vec{v}$ for all $v \in \mathcal{E}_w$ as long as $t \geq 0$. From the continuity of the inner product and the compactness of $\overleftrightarrow{\Sigma}$, you can show that the difference between $\log\|\Psi^t\overleftarrow{v}\|$ and $\log\|\Psi^t\vec{v}\|$ stays bounded as t varies over $[0,\infty)$.[4] Thus, if $\mathcal{E}$ is uniformly hyperbolic, the forward-stable lines $\mathcal{E}^+_{\overleftarrow{w}}$ and $\mathcal{E}^+_{\vec{w}}$ are the same, and we can refer to both as $\mathcal{E}^+_w$. The backward-stable lines at $\overleftarrow{w}$ and $\vec{w}$, on the other hand, are typically different. If w is on a forward-critical leaf instead of a backward-critical one, we can use the same reasoning in the other direction to define $\mathcal{E}^-_w$.

Working in a well-cut flow box $H \times L \subset \overleftrightarrow{\Sigma}$, it will be useful to describe our extension of Ψ^t in terms of the first return map. As we did in Section 9.5.2,

[4]Here's a sketch of the proof. Cover $\overleftrightarrow{\Sigma}$ with a finite collection $\mathcal{U}$ of simple open sets. For each $U \in \mathcal{U}$, the closure of $\dot{\dot{U}}$ in $\overleftrightarrow{\Sigma}$ is compact. Hence, the difference between $\log\|v\|_x$ and $\log\|v\|_y$ stays bounded as x and y vary over $\dot{\dot{U}}$ and v varies over $\mathcal{E}_U$. In particular, the difference between $\log\|v\|_{\overleftarrow{w}}$ and $\log\|v\|_{\vec{w}}$ is bounded over all $w \in \mathfrak{W} \cap \pi U$. Now, just combine the bounds over all $U \in \mathcal{U}$.

let $W \subset \pi H$ be the positions of the critical leaves, and partition it into the backward- and forward-critical sets W^+ and W^-. By our previous reasoning, $A_{\overleftarrow{w}} = A_{\overrightarrow{w}}$ for all $w \in W^+$. Defining $A_{\hat{w}}$ to be equal to both, we can extend the forward parallel transport cocycle A over the medians of all backward-critical points. If $\mathcal{E}$ is uniformly hyperbolic, the forward-stable lines $E^+_{\overleftarrow{w}}$ and $E^+_{\overrightarrow{w}}$ of $E = \mathcal{E}_U$ match, so we can define $E^+_{\hat{w}}$ to be equal to both. The backward cocycle A^{-1} extends over the medians of all forward-critical points in the same way, allowing us to define $E^-_{\hat{w}}$ for all $w \in W^-$.

11.6 Constructing uniformly hyperbolic local systems

Our abelianization process can only be carried out when $\mathcal{E}$ is uniformly hyperbolic, so it will be nice to have a way of constructing uniformly hyperbolic $\mathrm{SL}_2\,\mathbb{R}$ local systems on $\overleftrightarrow{\Sigma}$. We might as well assume that Σ is minimal and the vertical flow on $\Sigma \smallsetminus \mathfrak{W}$ is uniquely ergodic, since we'll need those conditions for abelianization anyway. The construction of uniformly hyperbolic local systems can then be done locally, in a well-cut flow box $H \times L \subset \overleftrightarrow{\Sigma}$.

Recall that $\alpha\colon \ddot{H} \to \ddot{H}$ is the fractured version of an interval exchange. The parallel transport cocycle over α is constant on each of the exchanged intervals. I'll call this kind of cocycle an *interval cocycle*. A local system on $\overleftrightarrow{\Sigma}$ is determined up to isomorphism by the parallel transport cocycle it induces over α. Conversely, any interval cocycle over α is the parallel transport cocycle of some local system on $\overleftrightarrow{\Sigma}$.

To get a sense of why the claims above are true, first recall that Theo-

rem 9.4.E lets us pass from local systems on $\overset{\leftrightarrow}{\Sigma}$ to ones on $\Sigma \smallsetminus \mathfrak{B}$. Because Σ is minimal, we can express $\Sigma \smallsetminus \mathfrak{B}$ as a suspension of the first return relation on $\iota^{-1}H$ [34]; you can see roughly what that means by looking at the proof of Proposition 9.5.G. Any local system on $\Sigma \smallsetminus \mathfrak{B}$ can be trivialized over the flow boxes that make up the suspension, with transition morphisms given by the parallel transport cocycle. Conversely, the transition morphisms given by an interval cocycle can be used to construct a local system over $\Sigma \smallsetminus \mathfrak{B}$.

Now all we need is a way of constructing uniformly hyperbolic $\mathrm{SL}_2\,\mathbb{R}$ cocycles over $\alpha\colon \ddot{H} \to \ddot{H}$. For convenience, let's have our cocycles act on $\mathbb{R}^2$ with the standard volume form. Let $\mathcal{A}$ be the intersections of the intervals exchanged by α and the ones exchanged by α^{-1}. Consider a pair of functions $u, v\colon \mathcal{A} \to \mathbb{R}^2$ that assign a basis for $\mathbb{R}^2$ to every interval in $\mathcal{A}$. We can think of u and v as continuous functions $\ddot{H} \to \mathbb{R}^2$ by composing them with the function $\ddot{H} \to \mathcal{A}$ that sends each point to its interval. An interval cocycle over α is *positive* with respect to u, v if at every $h \in \ddot{H}$ it maps u_h and v_h into the interior of the cone generated by $u_{\alpha h}$ and $v_{\alpha h}$. A cocycle is *eventually positive* if its nth iteration is positive, for some $n \in \mathbb{N}$. Theorems 3 and 4 of [26] show that an interval cocycle over α is uniformly hyperbolic if and only if it's eventually positive with respect to some basis $u, v\colon \mathcal{A} \to \mathbb{R}^2$.

Using this result, you can easily construct a lot of uniformly hyperbolic $\mathrm{SL}_2\,\mathbb{R}$ local systems on $\Sigma \smallsetminus \mathfrak{B}$: just write down interval cocycles whose matrix entries are all positive with respect to the standard basis for $\mathbb{R}^2$. You can also see that uniformly hyperbolic cocycles form an open set in the space of all

interval cocycles, which is a product of copies of $\mathrm{SL}_2\,\mathbb{R}$.

Chapter 12

Abelianization in principle

12.1 Overview

Now that we have the tools we need, we can turn again to our goal of extending abelianization to surfaces without punctures. At this point, it will be useful to take a closer look at the original description of abelianization, which is implicit in [11, §10], but first appears explicitly in [1, §4].

Our review will be simplified in two important ways. First, we'll restrict our attention to $\mathrm{SL}_2\,\mathbb{C}$ local systems, though Gaiotto, Hollands, Moore, and Neitzke show how to abelianize special linear local systems of any rank. Second, we'll only talk about abelianization using translation structures, leaving aside Gaiotto, Moore, and Neitzke's more powerful and general *spectral networks*.

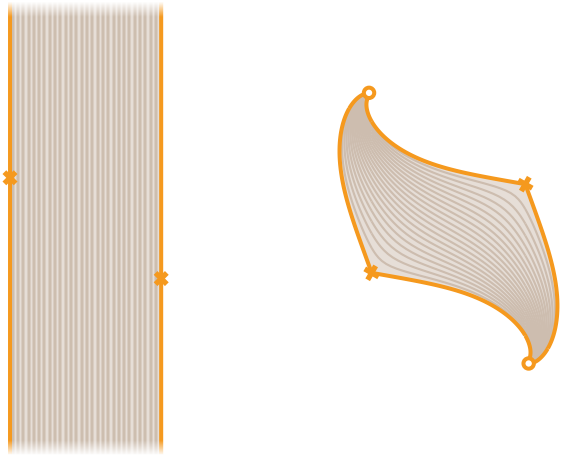
12.1.1 Setting

Our review takes place on a translation surface Σ' which is compact except for a finite set of *punctures*. A puncture is a region homeomorphic to a punctured disk, with a translation structure taken from a certain family of shapes. This definition of a puncture is analogous to our earlier definition of

a singularity. For simplicity, let's consider a translation surface whose punctures all have the simplest shape: a half-infinite cylinder. A complete list of puncture shapes, and an explanation of where they come from, can be found in Appendix F.

Let's assume Σ' has no saddle connections. In this case, if you follow a vertical leaf in some direction, your fate is easy to describe. If you're following a critical leaf in the critical direction, you will by definition fall into a singularity after a finite amount of time. Otherwise, you'll end up falling forever into a puncture; in our case, that means spiraling down the long end of a half-infinite cylinder.

Every non-critical leaf on Σ' is thus associated with two punctures, not necessarily distinct: the punctures its ends spiral into. If you remove the critical leaves $\mathfrak{W}$, the surface Σ' falls apart into infinite vertical strips of leaves that all go into the same punctures. Each strip is bounded by four critical leaves, joined at two singularities, as illustrated below. Gaiotto, Hollands, Moore, and Neitzke compactify the surface by filling in the punctures, so the closure of each strip becomes a quadrilateral with singularities as two of its vertices and punctures as the other two.



Strip Compactified strip

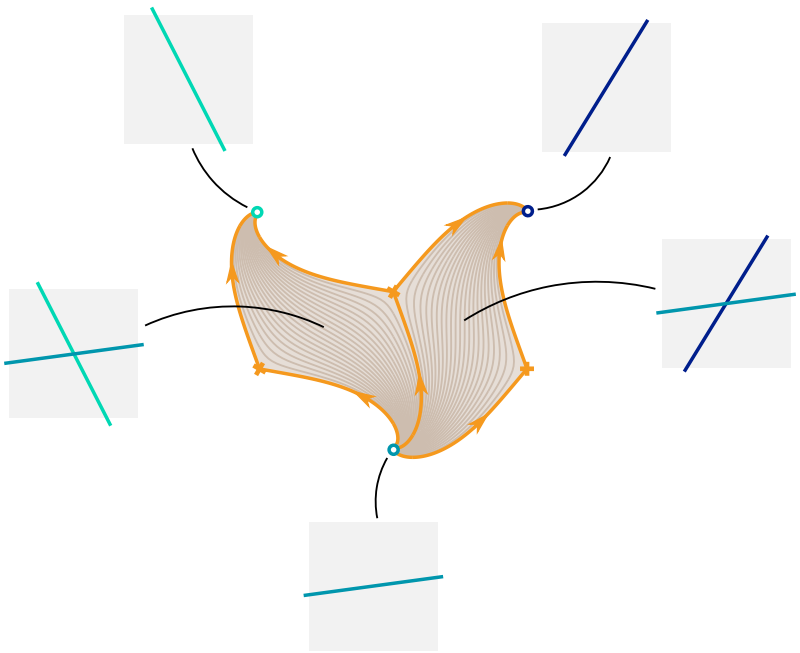
12.1.2 Framings

To abelianize an $\mathrm{SL}_2\,\mathbb{C}$ local system $\mathcal{E}$ on Σ', we first need to give it a bit of extra structure, called a *framing* (or *flag data*, in [11]). In the abstract settings of [1] and [11], a framing is specified geometrically by giving a projectively flat section of $\mathcal{E}$ on a neighborhood of each puncture (a framing can also be specified analytically, as we'll recall in Section 12.1.4). For reasons that will become apparent later, I'll refer to the sections given by the framing as *stable lines*. Framings always exist, because an operator on a finite-dimensional complex vector space always has at least one eigenvector.

The framing gives a pair of lines in every stalk of $\mathcal{E}$ over a non-critical leaf. One, which I'll call the *forward-stable* line, is gotten by following the forward vertical flow until you fall into a puncture, grabbing the stable line, and carrying it back by parallel transport. The *backward-stable* line is gotten

in the same way by following the backward vertical flow. The forward- and backward-stable lines fit together into sections of $\mathcal{E}$ over every strip of $\Sigma' \smallsetminus \mathfrak{W}$.

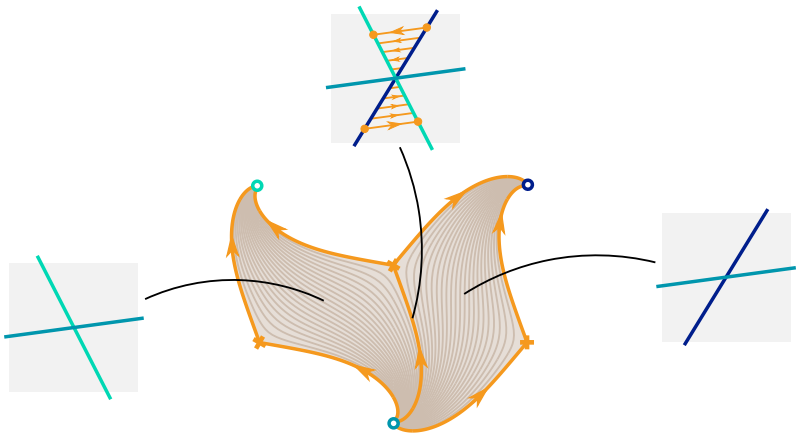
The framing also gives a line in every stalk of $\mathcal{E}$ over a critical leaf—the stable line from the puncture the critical leaf falls into. This line matches the backward- or forward-stable lines in the strips on either side, depending on whether the leaf is forward- or backward-critical. Hence, as you cross the boundary between two strips, one of the stable lines stays fixed, although the other can change.



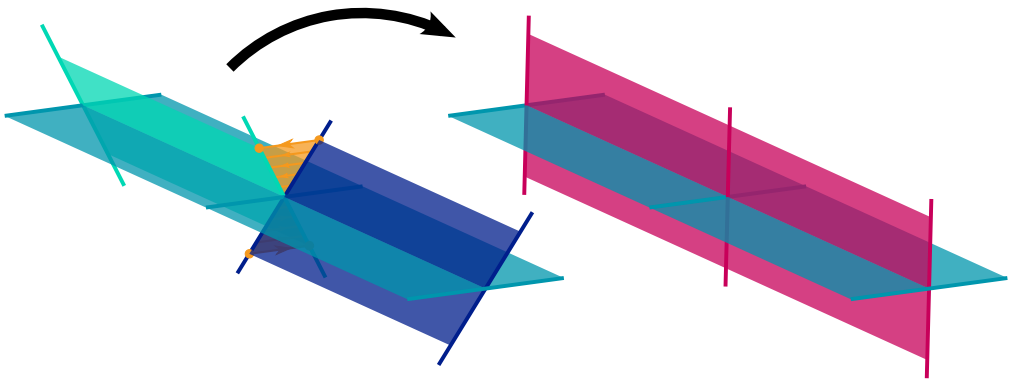
The stable lines over two adjacent strips

12.1.3 Abelianization

For a generic local system, the forward and backward-stable lines over each strip are complementary, splitting $\mathcal{E}$ into a direct sum of $\mathbb{C}^\times$ local systems over $\Sigma' \smallsetminus \mathfrak{W}$. The changes in the stable lines at the boundaries between strips generally prevent this splitting from extending over all of Σ'. At each critical leaf, however, there's a unique element of $\mathrm{SL}_2\,\mathbb{C}$ that sends the stable lines on one side to the stable lines on the other, acting by the identity on the line that stays fixed.



By cutting $\mathcal{E}$ along the critical leaves, shifting it by this automorphism, and gluing it back together, we can match up the stable lines across the boundaries of the strips, so the splitting they give becomes global.



Cutting and gluing to match up the stable lines

That's abelianization.

12.1.4 Abelianization without punctures

If we want to carry out the process above on a surface without punctures, there are two questions we have to settle. One is conceptual: what should play the role of the framing? If you've read Section 11, our suggestive terminology may have already told you the answer. When $\mathcal{E}$ is uniformly hyperbolic, it comes with complementary forward- and backward-stable lines over every non-critical leaf, which can be used as the forward- and backward-stable lines in the process above. The discussion in Section 11.5 amounts to saying that one of the stable lines stays fixed when you cross a critical leaf, so we can define the automorphisms over the critical leaves just as we did before.

This approach feels different than the one in [11], because the stable lines of a uniformly hyperbolic local system are a property of the local system,

rather than additional data. It's very reminiscent, however, of the more concrete approach taken in [23], where the framing comes from a specially chosen inner product on the stalks of the local system.[1] The stable line at a puncture consists of the sections that shrink as you fall in, just as the stable lines in our approach consist of the sections that shrink as you follow the vertical flow.

The stable lines in [11] and [23] live at the punctures in the surface, which lift to points on the boundary of the universal cover. Similarly, the stable lines in our approach can be seen as living on the boundary of the universal cover, at the points where the lifts of the vertical leaves begin and end. This point of view is central to Bonahon and Dreyer's generalization of the shear parameterization to higher-rank local systems on compact surfaces [42]. In our work, though, it won't be used.

The other question is just a technical difficulty. On a compact translation surface with no saddle connections, every vertical leaf is dense, so how do we think about shifting $\mathcal{E}$ along the automorphisms over the critical leaves? How do we know the process is well-defined? How do we know the resulting local system actually splits into a direct sum of $\mathbb{C}^\times$ local systems? The three parts of this question are answered in Sections 12.3, 13, and 14, respectively.

[1] That inner product is the celebrated *harmonic metric* from the theory of Hitchin systems. See Sections 6.5 and 13.2 of [23] for more information.

12.2 Running assumptions

The compact translation surface Σ introduced in Section 7.4.1 will, of course, stay with us. We'll discuss the abelianization of a fixed $\mathrm{SL}_2\mathbb{R}$ local system $\mathcal{E}$ on $\overleftrightarrow{\Sigma}$. Abelianization will yield a new $\mathrm{SL}_2\mathbb{R}$ local system $\mathcal{F}$ and a stalkwise isomorphism $\Upsilon\colon \mathcal{E} \to \mathcal{F}$, supported on $\overleftrightarrow{\Sigma}$. As discussed in Section 7.2, our results should generalize from $\mathrm{SL}_2\mathbb{R}$ to $\mathrm{SL}_2\mathbb{C}$ without too much trouble.

From now on, we'll assume the following things about Σ:

- The surface Σ has no saddle connections.
- The vertical flow on $\Sigma \smallsetminus \mathfrak{W}$ is uniquely ergodic.
- The surface Σ satisfies the fat gap condition of Section 9.5.5.

We'll also make one crucial assumption about $\mathcal{E}$:

- The local system $\mathcal{E}$ is uniformly hyperbolic.

If you happen to have picked a surface Σ that doesn't satisfy the required conditions, have no fear. The remarks in Sections 9.2.1 and 9.5.5 show that you can fix this problem with an arbitrarily small rotation of the translation structure on Σ. In fact, if you rotate the translation structure on Σ through a full circle, all but a measure-zero subset of the structures you pass through will satisfy all the conditions needed. Once you've fixed a good translation structure on Σ, you can use the results of Section 11.6 to find lots of uniformly hyperbolic local systems $\mathcal{E}$.

12.3 The slithering jump

12.3.1 Overview

Working on $\overleftrightarrow{\Sigma}$ gives us a convenient way to talk about the stable lines on either side of a critical leaf: using the identification in Section 11.5, we can compare the stable lines over the left and right lanes. The automorphisms that match up the stable lines across the median can be encoded as a jump in the local system $\mathcal{E}$, as defined in Section 10. This jump contains essentially the same information as *slithering maps* defined by Bonahon and Dreyer in [42], so we'll call it the *slithering jump*. We abelianize $\mathcal{E}$ by warping it along the deviation defined by the slithering jump. More explicitly, we abelianize $\mathcal{E}$ by carrying out the following steps:

1. Compute the slithering jump, using the formulas in Section 12.3.3.
2. Turn the slithering jump into a deviation, as described in Section 10.3.
3. Warp $\mathcal{E}$ along the deviation, as described in Section 8.4.

We'll prove in Sections 13 and 14 that these instructions have the desired effect, as long as the conditions in Section 12.2 are satisfied.

12.3.2 Definition

Consider a point w on a backward-critical leaf of Σ. Because $\mathcal{E}_w$ is two-dimensional, and $\mathrm{SL}_2\,\mathbb{R}$ is the group of volume-preserving linear maps, there's a unique automorphism s_w of $\mathcal{E}_w$ that sends $\mathcal{E}^-_{\overrightarrow{w}}$ to $\mathcal{E}^-_{\overleftarrow{w}}$ and is the identity on $\mathcal{E}^+_w$.

This induces an automorphism of $\mathcal{E}_U$ for any simple flow box U containing $\hat{w}$. If w is on a forward-critical leaf instead of a backward-critical one, we can define s_w in the same way, with the roles of the backward-stable and forward-stable lines reversed.

Given a divider P in a simple flow box U, it's not hard to see that s_w induces the same automorphism of $\mathcal{E}_U$ for every $w \in P$. Call this automorphism s_P. As P varies over all dividers in all simple flow boxes, the automorphisms s_P fit together into a jump s in the local system $\mathcal{E}$—the *slithering jump*.

12.3.3 Formulas

Assuming, for convenience, that w is on a backward-critical leaf, we can get an explicit expression for s_w by choosing representatives

$$u' \in \mathcal{E}^-_{\overleftarrow{w}} \qquad\qquad v \in \mathcal{E}^+_w \qquad\qquad u \in \mathcal{E}^-_{\overrightarrow{w}}.$$

Observe that $\{v, u'\}$ and $\{v, u\}$ are ordered bases for $\mathcal{E}_w$. By rescaling u' and u, we can ensure that both ordered bases have unit volume. The transformation s_w is then given by[2]

$$v \mapsto v \qquad\qquad u \mapsto u'.$$

[2]Given a pre-existing basis for $\mathcal{E}_w$, you can find s_w by solving the matrix equation

$$\left[\begin{array}{c|c} v & u' \end{array}\right] = s_w \left[\begin{array}{c|c} v & u \end{array}\right],$$

which I have found convenient for numerical work.

A quick calculation with the volume form D gives the relation

$$u' = u + D(u', u)\, v,$$

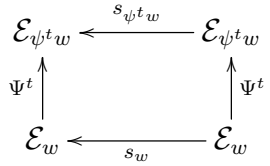
revealing that s_w is a shear transformation whose off-diagonal part $s_w - 1$ is given by

$$v \mapsto 0 \qquad\qquad u \mapsto D(u', u)v.$$

When w is on a forward-critical leaf, similar expressions can be obtained.

12.3.4 Flow invariance

Suppose w is on a backward-critical leaf. Because s_w is defined in terms of the stable lines $\mathcal{E}^{\pm}_w$ and the volume form on $\mathcal{E}_w$, which are all invariant under the vertical flow, the diagram



commutes for all positive times t. If w is on a forward-critical leaf, the same is true for all negative times.

This flow invariance property is not unique to the slithering jump. In fact, it holds for all jumps, as a direct consequence of the defining consistency condition. Its introduction has been delayed until now only for convenience.

Chapter 13

Abelianization converges

13.1 Overview

To show that the slithering jump defines a deviation σ, as discussed in Section 10, we need to verify that the infinite product defining the automorphism σ_{yx}^{U} converges for every simple flow box $U \subset \overleftrightarrow{\Sigma}$ and every pair of points $y, x \in \ddot{U}$. Because jumps commute with restrictions, it's enough to check for convergence on a set of simple flow boxes that cover $\overleftrightarrow{\Sigma}$. We'll use the simple, well-cut flow boxes for this purpose.

Consider a simple, well-cut flow box $U = H \times L$ in $\overleftrightarrow{\Sigma}$, and let $E = \mathcal{E}_U$. Pick an inner product on E, so we can use the asymptotic growth conditions given by the uniform hyperbolicity of $\mathcal{E}$. As usual, let $W \subset \pi H$ be the positions of the critical leaves, recalling that $\ddot{H} = H \smallsetminus \iota W$. Label each divider $\{\hat{w}\} \times L$ in U by the point $w \in W$ it sits above. As we did in Section 9.5.2, let $B^+ \subset W^+$ and $B^- \subset W^-$ be the break points of α^{-1} and α, respectively.

We'll keep using the shorthand $\blacksquare \lesssim \blacksquare$ to say that one positive function is bounded by a constant multiple of another, as we started doing in Section 11.1.

13.2 Bounding the jump

For any $b \in B^+$, we can choose representatives

$$u' \in E^-_{\overleftarrow{b}} \qquad\qquad v \in E^+_{\hat{b}} \qquad\qquad u \in E^-_{\overrightarrow{b}}$$

for which $D(v, u')$ and $D(v, u)$ are 1 and conclude that $s_b - 1$ is given by

$$v \mapsto 0 \qquad\qquad u \mapsto D(u', u)\, v,$$

as described in the previous section. Applying the flow invariance of jumps, we see that $s_{\alpha^n b} - 1$ is given by

$$A^n_{\hat{b}} v \mapsto 0 \qquad\qquad A^n_{\hat{b}} u \mapsto D(u', u)\, A^n_{\hat{b}} v$$

for all positive n.

We see from the formula above that $D(u', u)\, A^n_{\hat{b}} v$ spans the image of $s_{\alpha^n b} - 1$. We also learn that the shortest vector $s_{\alpha^n b} - 1$ sends to $D(u', u)\, A^n_{\hat{b}} v$ is the orthogonal projection of $A^n_{\hat{b}} u$ onto the orthogonal complement of $A^n_{\hat{b}} v$. From this, we can calculate the operator norm of $s_{\alpha^n b} - 1$:

$$\|s_{\alpha^n b} - 1\| = \frac{|D(u', u)|}{d_\angle(A^n_{\hat{b}} v, A^n_{\hat{b}} u)} \frac{\|A^n_{\hat{b}} v\|}{\|A^n_{\hat{b}} u\|},$$

where $d_\angle$ is the function that takes two nonzero vectors in E and measures the sine of the angle between the lines they span. Rearranging a bit, we get

$$\|s_{\alpha^n b} - 1\| = \frac{|D(u', u)|}{d_\angle(A^n_{\hat{b}} v, A^n_{\hat{b}} u)} \frac{\|v\|}{\|u\|} \cdot \frac{\|A^n_{\hat{b}} v\|}{\|v\|} \cdot \frac{\|u\|}{\|A^n_{\hat{b}} u\|}.$$

Because $\mathcal{E}$ is uniformly hyperbolic, the first-return cocycle A is uniformly hyperbolic too, as a consequence of Proposition 11.4.B. Pick a bounding exponent

$K > 0$ for A. The stable lines of A vary continuously (Lemma 11.3.B), and $\ddot{H}$ is compact (Proposition 9.3.F), so $d_\angle(A_b^n v, A_b^n u)$ is bounded away from zero. That and the uniform hyperbolicity of A tell us that

$$\|s_{\alpha^n b} - 1\| \lesssim e^{-2Kn}$$

over all $b \in B^+$ and $n \in \mathbb{N}$.

Applying the same reasoning in the other direction, we see more generally that

$$\|s_{\alpha^{\pm n} b} - 1\| \lesssim e^{-2Kn}$$

over all $b \in B^\pm$ and $n \in \mathbb{N}$.

13.3 Showing its product converges

Recall that σ is the deviation we're hoping will be defined by the slithering jump. Pick any two points $y, x \in \ddot{U}$. Since we're labeling the dividers in U by points of W, we can think of $(y \mid x)^U$ as a subset of W, and write

$$\sigma_{yx}^U = \prod_{w \in (y|x)^U} s_w.$$

Proposition D.6.A in Appendix D tells us that this product converges if the sum

$$C_{yx} = \sum_{w \in (y|x)^U} \|s_w - 1\|$$

does. (I've given the sum a name because its value, as a function of y and x, will be useful to us later.)

Let's say every point in $(y \mid x)^U$ takes at least n iterations of α or α^{-1} to hit a break point. Then the set

$$\bigcup_{m \geq n} \Big[\{\alpha^m b : b \in B^+\} \cup \{\alpha^{-m} b : b \in B^-\} \Big]$$

contains all the points in $(y \mid x)^U$, so

$$C_{yx} \leq \sum_{m \geq n} \left[\sum_{b \in B^+} \|s_{\alpha^m b} - 1\| + \sum_{b \in B^-} \|s_{\alpha^{-m} b} - 1\| \right].$$

Applying the bound from the previous section, we see that

$$\begin{aligned} C_{yx} &\lesssim \sum_{m \geq n} \left[\sum_{b \in B^+} e^{-2Km} + \sum_{b \in B^-} e^{-2Km} \right] \\ &\lesssim \sum_{m \geq n} e^{-2Km}. \end{aligned}$$

Hence, the sum defining C_{yx} converges.

Summing the geometric series, we learn that $C_{yx} \lesssim e^{-2Kn}$. But n is the grade of the highest point in $(y \mid x)^U$, so e^{-2Kn} is the distance between y and x in the division metric with steepness e^{2K}! The implied constant multiple in the bound above doesn't depend on y and x, so we've proven that $C_{yx} \lesssim d(y, x)$ over all $y, x \in \ddot{U}$.

Chapter 14

Abelianization delivers

14.1 Overview

Now we know the slithering jump defines a deviation σ, so we can warp $\mathcal{E}$ along this deviation to produce a new local system $\mathcal{F}$ and a stalkwise isomorphism $\Upsilon\colon \mathcal{E} \to \mathcal{F}$, supported on $\overset{\leftrightarrow}{\Sigma}$. By design, Υ matches up the stable lines of $\mathcal{E}$ across the medians of $\overset{\leftrightarrow}{\Sigma}$, sending corresponding stable lines in $\mathcal{E}_{\overleftarrow{w}}$ and $\mathcal{E}_{\overrightarrow{w}}$ to the same line in $\mathcal{F}_{\hat{w}}$ for all $w \in \mathfrak{W}$. To show that $\mathcal{F}$ splits into a direct sum of two $\mathbb{R}^\times$ bundles, we need to prove that Υ matches up the stable lines on larger scales. For any simple flow box $U \subset \overset{\leftrightarrow}{\Sigma}$, we have to show that Υ sends the corresponding stable lines in all the stalks of $\mathcal{E}$ over $\dot{\bar{U}}$ to the stalk restrictions of a single line in $\mathcal{F}_U$. Because of the way deviations restrict, it's enough to prove the desired result on a set of simple flow boxes that cover $\overset{\leftrightarrow}{\Sigma}$, just like in Section 13. Once again, we'll use the simple, well-cut flow boxes for this purpose.

We'll keep all the notation from Section 13, and add to it the shorthand $F = \mathcal{F}_U$. To make the geometry facts from Appendix E available, scale the inner product on E so that the unit square has unit volume. To make the results from Sections 13 and 9.3.5 available, pick a bounding exponent $K > 0$

for A, and give $\ddot{H}$ the division metric with steepness e^{2K}.

The argument we're about to do is somewhat technical, so let's first recall how it works over a punctured surface Σ', where it's so straightforward that we barely mentioned it earlier. Think of the stable lines of $\mathcal{E}$ as lines in E parameterized by the points of $\ddot{U}$, and think of their images in $\mathcal{F}$ under Υ as lines in F. The stable lines are constant in E away from the critical leaves of Σ', and the slithering jump only disturbs $\mathcal{E}$ at the critical leaves. Hence, the images under Υ of the stable lines are constant in F away from the critical leaves. By design, the images of the stable lines under Υ are also constant across the critical leaves, so they must be constant everywhere. In other words, Υ matches up the images of the stable lines in F all across $\ddot{U}$.

On the unpunctured surface Σ, it's hard to get away from the critical leaves, which fill the surface densely. The property of being constant away from the critical leaves thus has no clear meaning, and we'll have to replace it with something else if we want to reproduce the argument above. Corollary 9.3.L and the discussion at the end of Section 9.5.5 suggest that in our scenario, Hölder continuity might be a viable substitute. Most of the work below is concerned with proving that the stable lines and their images under Υ vary Hölder continuously over $\ddot{U}$.

14.2 The stable distributions are Hölder

Let's say the distance between two lines in E is the sine of the angle between them. This puts a metric on the projective space $\mathbf{P}E$, which I'll call

the *sine metric*. I'll write $d_\angle(u, v)$ to mean the distance in $\mathbf{P}E$ between the lines generated by $u, v \in E$.

Let's collect all the forward- and backward-stable lines over $\ddot{H}$ into a pair of functions $E^\pm \colon \ddot{H} \to \mathbf{P}E$, the *stable distributions*. Our regularity conditions on Σ and $\mathcal{E}$ guarantee that $E^\pm$ are Hölder. We can see this by applying a recent theorem of Araújo, Bufetov, and Filip to the parallel transport cocycle over $\ddot{H}$ [43, Theorem A]. Once we check that its conditions are satisfied, the theorem will tell us that $E^\pm$ are Hölder on any appropriate *regular block* in $\ddot{H}$. We'll then show that all of $\ddot{H}$ is an appropriate regular block.

14.2.1 The conditions of the theorem are satisfied

As we discussed in Section 9.5.2, the first return map on $\ddot{H}$ is the fractured version of an interval exchange on πH, with break points ιB^-. Removing the dividers over the break points from U turns it into a disjoint union of full flow boxes $H_1 \times L, \dots, H_k \times L$. Each of these flow boxes stays in one piece as the vertical flow carries it around $\overset{\leftrightarrow}{\Sigma}$ and back to U. As a result, the parallel transport cocycle $A \colon \ddot{H} \to \operatorname{Aut} E$ is constant on each of the intervals $\ddot{H}_1, \dots, \ddot{H}_k$, which form an open partition of $\ddot{H}$.

It immediately follows that $\log \|A\|$ and $\log \|A^{-1}\|$ are integrable with respect to the ergodic probability measure on $\ddot{H}$. Because the intervals $\ddot{H}_1, \dots, \ddot{H}_k$ are full, distances between points in different intervals are bounded away from zero, so the fact that A is constant on each interval also implies that A is Hölder.

We saw in Section 9.3.5 that the fractured version of an interval exchange is Lipschitz with respect to the division metric it defines. The first return map on $\ddot{H}$ is therefore Lipschitz.

14.2.2 The stable lines are Hölder on appropriate regular blocks

Lyapunov exponents play a role in the definition of a regular block, so we'll take advantage of Lemma 11.3.C, which gives us a cheap way to talk about the Lyapunov exponents of uniformly hyperbolic $\mathrm{SL}_2\,\mathbb{R}$ local systems. The limit we've been calling the "Lyapunov exponent" is really the top Lyapunov exponent of A, and we'll refer to it that way here for consistency with the language of [43].

Regular blocks are parameterized by two real numbers $\varepsilon > 0$ and $\ell > 1$. Araújo, Bufetov, and Filip take ℓ, for convenience, to be an integer, and the same could be done here as well. Their proof of Theorem A, found in [43, §3.2], shows that $E^{\pm}$ are Hölder on any regular block for which ε is small enough and ℓ is large enough. Referring back to [43, §2.2], where the relevant thresholds are defined, we see that ε must be smaller than a tenth of the minimum gap between the Lyapunov exponents of A, and ℓ must be large enough for the regular block to have positive measure.

In our case, the Lyapunov exponents of A are $\pm\Lambda$, so the minimum gap is 2Λ. In the next section, we'll choose ε to be less than $\frac{1}{5}\Lambda$, and we'll make ℓ large enough for the resulting regular block to be the whole of $\ddot{H}$, which of course has measure one.

14.2.3 The whole interval is an appropriate regular block

In our context, the *regular block* with parameters $\ell \in (1, \infty)$ and $\varepsilon \in (0, \frac{1}{5}\Lambda)$ consists of the points $h \in \ddot{H}$ at which

$$\begin{aligned} \tfrac{1}{\ell} e^{(-\Lambda-\varepsilon)n-\varepsilon|m|}\|v\| \le \|A^{\pm n}_{\alpha^m h} v\| \le \ell e^{(-\Lambda+\varepsilon)n+\varepsilon|m|}\|v\| \\ \tfrac{1}{\ell} e^{(\Lambda-\varepsilon)n-\varepsilon|m|}\|v\| \le \|A^{\mp n}_{\alpha^m h} v\| \le \ell e^{(\Lambda+\varepsilon)n+\varepsilon|m|}\|v\| \end{aligned}$$

for all $m \in \mathbb{Z}$, $v \in E^{\pm}_{\alpha^m h}$, and $n \ge 0$, and

$$\sqrt{1 - d_{\angle}(E^{+}_{\alpha^n h}, E^{-}_{\alpha^n h})^2} \le 1 - \tfrac{1}{\ell} e^{-\varepsilon|n|}$$

for all $n \in \mathbb{Z}$ [43, §2.2]. We'll see that, given any ε, we can choose ℓ large enough to make both conditions hold for all $h \in \ddot{H}$.

To see that the norm condition is satisfied everywhere for large enough ℓ, first note that it's enough to show that

$$\begin{aligned} \tfrac{1}{\ell} e^{(-\Lambda-\varepsilon)n}\|v\| \le \|A^{\pm n}_{\alpha^m h} v\| \le \ell e^{(-\Lambda+\varepsilon)n}\|v\| \\ \tfrac{1}{\ell} e^{(\Lambda-\varepsilon)n}\|v\| \le \|A^{\mp n}_{\alpha^m h} v\| \le \ell e^{(\Lambda+\varepsilon)n}\|v\| \end{aligned}$$

for all $m \in \mathbb{Z}$, $v \in E^{\pm}_{\alpha^m h}$, and $n \ge 0$. Now m plays no role other than to move our starting point, so really we only need to show that

$$\begin{aligned} \tfrac{1}{\ell} e^{(-\Lambda-\varepsilon)n}\|v\| \le \|A^{\pm n}_{h} v\| \le \ell e^{(-\Lambda+\varepsilon)n}\|v\| \\ \tfrac{1}{\ell} e^{(\Lambda-\varepsilon)n}\|v\| \le \|A^{\mp n}_{h} v\| \le \ell e^{(\Lambda+\varepsilon)n}\|v\| \end{aligned}$$

for all $v \in E^{\pm}_{h}$ and $n \ge 0$. Equivalently,

$$\begin{aligned} -\Lambda - \left(\varepsilon + \frac{\log \ell}{n}\right) \le \frac{1}{n} \log \frac{\|A^{\pm n}_h v\|}{\|v\|} \le -\Lambda + \left(\varepsilon + \frac{\log \ell}{n}\right) \\ \Lambda - \left(\varepsilon + \frac{\log \ell}{n}\right) \le \frac{1}{n} \log \frac{\|A^{\mp n}_h v\|}{\|v\|} \le \Lambda + \left(\varepsilon + \frac{\log \ell}{n}\right) \end{aligned}$$

for all $v \in E^{\pm}_h$ and $n \geq 0$. The uniform convergence in Lemma 11.3.C and Proposition 11.3.D guarantees that

$$-\Lambda - \varepsilon \leq \frac{1}{n} \log \frac{\|A_h^{\pm n} v\|}{\|v\|} \leq -\Lambda + \varepsilon$$
$$\Lambda - \varepsilon \leq \frac{1}{n} \log \frac{\|A_h^{\mp n} v\|}{\|v\|} \leq \quad \Lambda + \varepsilon$$

for large enough n, and we can always choose ℓ large enough to contain the transient behavior at small n.

To show that the angle condition is satisfied for large enough ℓ, it's enough to prove that $d_\angle(E^+_{\alpha^n h}, E^-_{\alpha^n h})$ is bounded away from zero. This holds because $\breve{H}$ is compact (Proposition 9.3.F), the stable lines vary continuously (Lemma 11.3.B), and the forward- and backward-stable lines can never coincide.

14.3 The stable distributions after abelianization

Recall that warping $\mathcal{E}$ along σ has given us a new local system $\mathcal{F}$ and a stalkwise isomorphism $\Upsilon \colon \mathcal{E} \to \mathcal{F}$, supported on $\breve{\Sigma}$. We're using the shorthand $F = \mathcal{F}_U$. Because U is simple, the stalk restrictions of $\mathcal{E}$ and $\mathcal{F}$ identify $\mathcal{E}_p$ with E and $\mathcal{F}_p$ with F for every $p \in \breve{U}$. We can thus view Υ as a map from $\breve{U}$ to $\mathrm{SL}(E, F)$. Because σ comes from a jump, Υ is constant along the vertical leaves of $\breve{\Sigma}$, so in fact we can treat Υ as a map from $\breve{H}$ to $\mathrm{SL}(E, F)$. This section takes place entirely within the flow box U, so we'll abbreviate σ^U_{yx} as σ_{yx} and $(y \mid x)^U$ as $(y \mid x)$.

Just as parallel transport in $\mathcal{E}$ along the vertical flow gave the linear cocycle $A\colon \ddot{H} \to \mathrm{SL}(E)$, parallel transport in $\mathcal{F}$ along the vertical flow gives a linear cocycle $\ddot{H} \to \mathrm{SL}(F)$. Define $F^{\pm}_h \subset F$ as the images of the lines $E^{\pm}_h$ under Υ_h. Like $E^{\pm}$, the distributions $F^{\pm}$ are invariant under the parallel transport cocycle.[1] Let's put an inner product on F by declaring Υ_a, for some arbitrary $a \in \ddot{H}$, to be an isometry. We then get a sine metric on $\mathbf{P}F$, and we can ask whether the functions $F^{\pm}$ are Hölder.[2]

Recalling that $\Upsilon_h = \Upsilon_a \sigma_{ah}$, we see that $F^{\pm}_h = \Upsilon_a \sigma_{ah} E^{\pm}_h$ for all $h \in \ddot{H}$. Since Υ_a is, by definition, an isometry,

$$d_{\angle}(F^{\pm}_y, F^{\pm}_x) = d_{\angle}(\sigma_{ay} E^{\pm}_y, \sigma_{ax} E^{\pm}_x)$$

for all $y, x \in \ddot{H}$. We might therefore be able to prove that $F^{\pm}$ are Hölder by looking at how σ^U affects distances in $\mathbf{P}E$.

14.4 The abelianized stable distributions are still Hölder

14.4.1 The deviation between nearby points is close to the identity

Remember the bound C_{yx} we used in Section 13.3? We'll soon see that $\|\sigma_{yx} - 1\| \lesssim C_{yx}$ over all $y, x \in \ddot{H}$. Combining this with the bound

[1] As a matter of fact, $\mathcal{F}$ should be uniformly hyperbolic, with $F^{\pm}$ as its stable distributions. We don't need to know that, though.

[2] In light of the previous footnote, you might hope to show that $F^{\pm}$ are Hölder the same way we showed that $E^{\pm}$ are Hölder, by applying the theorem of Araújo, Bufetov, and Filip. The difficulty in this approach is that we don't know much about the parallel transport cocycle for F, making it hard to check the conditions of the theorem.

$C_{yx} \lesssim d(y,x)$ proven at the end of Section 13.3, we'll learn that

$$\|\sigma_{yx} - 1\| \lesssim d(y,x)$$

over all $y, x \in \ddot{H}$.

Let's get down to business. Recall that σ_{yx} is the ordered product

$$\prod_{w \in (y|x)} s_w.$$

For any $w' \in W$, observe that

$$\begin{aligned}
\|\sigma_{yx} - 1\| &= \left\| \left(\prod_{w \in (y|x)} s_w \right) - 1 \right\| \\
&= \left\| \left(\prod_{w \in (y|x) \smallsetminus w'} s_w \right) + \left(\prod_{w \in (y|w')} s_w \right) (s_{w'} - 1) \left(\prod_{w \in (w'|x)} s_w \right) - 1 \right\| \\
&\leq \left\| \left(\prod_{w \in (y|x) \smallsetminus w'} s_w \right) - 1 \right\| + \|s_{w'} - 1\| \prod_{w \in (y|x) \smallsetminus w'} \|s_w\| \\
&\leq \left\| \left(\prod_{w \in (y|x) \smallsetminus w'} s_w \right) - 1 \right\| + \|s_{w'} - 1\| \prod_{w \in (y|x)} \|s_w\|.
\end{aligned}$$

Recalling the definition of C_{yx}, we see that $\prod_{w \in (y|x)} \|s_w\| \leq \exp C_{yx}$ by Proposition D.6.C. Thus,

$$\|\sigma_{yx} - 1\| \leq \left\| \left(\prod_{w \in (y|x) \smallsetminus w'} s_w \right) - 1 \right\| + \|s_{w'} - 1\| \exp C_{yx}.$$

By repeating the argument above, we see that for any finite subset S of $(y \mid x)$,

$$\|\sigma_{yx} - 1\| \;\leq\; \left\| \left(\prod_{w \in (y|x) \smallsetminus S} s_w \right) - 1 \right\| \;+\; \left(\sum_{w \in S} \|s_w - 1\| \right) \exp C_{yx}.$$

The convergence of the product $\prod_{w \in (y|x)} s_w$ tells us that as S grows, the first term of the inequality above goes to zero, leaving us with the bound

$$\begin{aligned} \|\sigma_{yx} - 1\| &\leq \left(\sum_{w \in (y|x)} \|s_w - 1\| \right) \exp C_{yx} \\ &= C_{yx} \exp C_{yx}. \end{aligned}$$

Since $C_{yx} \lesssim d(y, x)$, and distances in $\dot{\ddot{H}}$ are bounded, we can bound $\exp C_{yx}$ by a constant. Hence, $\|\sigma_{yx} - 1\| \lesssim C_{yx}$. It follows, as explained at the beginning of the section, that

$$\|\sigma_{yx} - 1\| \lesssim d(y, x)$$

over all $y, x \in \dot{\ddot{H}}$.

14.4.2 The abelianized stable distributions are Hölder

We know from Section 14.2 that $E^{\pm}$ are Hölder, say with exponent ν. Because distances in $\dot{\ddot{H}}$ are bounded, Hölder continuity with a given exponent implies Hölder continuity with all lower exponents, so we might as well assume for convenience that $\nu \leq 1$. For any $y, x \in \dot{\ddot{H}}$, as discussed in Section 14.3,

$$\begin{aligned} d_{\angle}(F_y^{\pm}, F_x^{\pm}) &= d_{\angle}(\sigma_{ay} E_y^{\pm}, \sigma_{ax} E_x^{\pm}) \\ &= d_{\angle}(\sigma_{ay} E_y^{\pm}, \sigma_{ay} \sigma_{yx} E_x^{\pm}). \end{aligned}$$

Proposition E.D gives

$$d_\angle(F_y^\pm, F_x^\pm) \leq \|\sigma_{ya}\|^2 \, d_\angle(E_y^\pm, \sigma_{yx} E_x^\pm).$$

Because distances in $\ddot{H}$ are bounded, the result of the previous section ensures that $\|\sigma_{ya}\|$ is bounded as well. Therefore,

$$\begin{aligned} d_\angle(F_y^\pm, F_x^\pm) &\lesssim d_\angle(E_y^\pm, \sigma_{yx} E_x^\pm) \\ &\leq d_\angle(E_y^\pm, E_x^\pm) + d_\angle(E_x^\pm, \sigma_{yx} E_x^\pm) \end{aligned}$$

over all $y, x \in \ddot{H}$. Proposition E.C combines with the bound from the previous section to show that

$$\begin{aligned} d_\angle(E_x^\pm, \sigma_{yx} E_x^\pm) &\leq \|\sigma_{yx} - 1\| \\ &\lesssim d(y, x). \end{aligned}$$

Using the fact that distances in $\ddot{H}$ are bounded and the assumption that $\nu \leq 1$, we conclude that

$$d_\angle(E_x^\pm, \sigma_{yx} E_x^\pm) \lesssim d(y, x)^\nu.$$

Meanwhile, the Hölder continuity of $E^\pm$ gives

$$d_\angle(E_y^\pm, E_x^\pm) \lesssim d(y, x)^\nu.$$

Therefore, altogether,

$$d_\angle(F_y^\pm, F_x^\pm) \lesssim d(y, x)^\nu$$

over all $y, x \in \ddot{H}$. In other words, the abelianized stable distributions $F^\pm$ are Hölder.

14.5 The abelianized stable distributions are constant

By construction, the values of the functions $F^{\pm}\colon \ddot{H} \to \mathbf{P}F$ match at adjacent edge points, as defined before Theorem 9.3.K. Since we just saw that $F^{\pm}$ are Hölder, and we're assuming the translation structure on Σ satisfies the fat gap condition of Section 9.5.5, Corollary 9.3.L tells us that $F^{\pm}$ are constant.

Globally, this means the stable distributions $\mathcal{F}^{\pm}$ are constant with respect to the local system $\mathcal{F}$, so they're actually $\mathbb{R}^{\times}$ local subsystems of $\mathcal{F}$. In fact, $\mathcal{F}$ is the direct sum of the $\mathbb{R}^{\times}$ local systems $\mathcal{F}^{+}$ and $\mathcal{F}^{-}$.

Chapter 15

A quick example

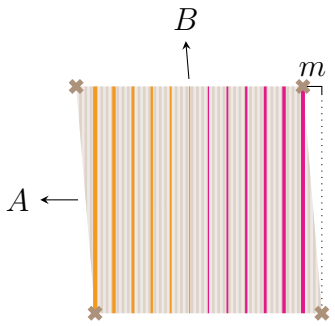
15.1 Overview

Now that we've proven abelianization works, let's see an example of what it does. The calculations in this section aren't rigorous, but I'll try to indicate what it would take to make them rigorous. Since abelianization is expected to work for $\mathrm{SL}_2\,\mathbb{C}$ cocycles, and in this case it does, we'll work over $\mathbb{C}$ rather than $\mathbb{R}$.

15.2 Setting the scene

15.2.1 A translation surface

Construct a torus with a translation structure by gluing the opposite sides of a parallelogram, inserting a singularity of cone angle 2π at the corner. For concreteness, let's fix the base of the parallelogram to be horizontal with length one, and set the height to be one as well. This leaves only one degree of freedom in the translation structure: the slope parameter m labeled in the drawing below.



The torus has one forward-critical leaf and one backward-critical leaf. The drawing follows the critical leaves a little ways out from the singularity, so you can get an idea of how they wind around the surface. We'll assume m is irrational, ensuring that neither critical leaf is a saddle connection.

The details of the computation depend on which way the parallelogram is leaning—a first hint of cluster-like behavior. To match the drawings, we'll show the work for the left-leaning case.

15.2.2 A variety of local systems

An $\mathrm{SL}_2\,\mathbb{C}$ local system on the torus minus the singularity is specified, up to isomorphism, by the group elements $A, B \in \mathrm{SL}_2\,\mathbb{C}$ that describe the parallel transport across the sides of the parallelogram, as shown in the drawing. Any values of A and B are possible.

Let's restrict ourselves to the dense open subset of the character variety in which B has distinct eigenvalues. In this region, we can hit every

isomorphism class using group elements of the form

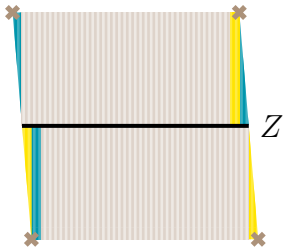
$$A = \begin{bmatrix} \mu & \rho \\ \rho & \nu \end{bmatrix} \qquad\qquad B = \begin{bmatrix} \lambda & \cdot \\ \cdot & \frac{1}{\lambda} \end{bmatrix},$$

where $|\lambda| < 1$. Restricting further to the dense open subset in which $\mu\nu \notin (-\infty, 1]$, we can make ρ a holomorphic function of μ and ν by noting that $\det A = 1$ and imposing the additional constraint $\operatorname{Re}\rho > 0$. This gives a holomorphic parameterization of a dense open subset of the character variety by the three variables μ, ν, and λ, which vary over the domain

$$\mu\nu \notin (-\infty, 1] \qquad\qquad |\lambda| < 1.$$

15.2.3 Reduction to an interval exchange

Let Z be the horizontal segment running across the middle of the parallelogram. Under the vertical flow, Z sweeps out a simple flow box that covers almost the whole torus. The vertical edge of the flow box is non-critical, so we can compute the abelianized local system just by looking at the parallel transport cocycle over Z. For this purpose, we'll mostly carry on with the notation from Section 13.



Identify Z with $(-1, 0)$. The first return relation α has a single break point, $b = -\frac{m}{2}$. Its inverse α^{-1} has break point $c = -1 + \frac{m}{2}$. Because Z

isn't well-cut, the break points aren't the only points where α and α^{-1} return nothing: $-m$ and $-1+m$ vanish under the actions of α and α^{-1} as well. We aren't calling the latter break points because they don't lie on critical leaves.

The forward parallel transport cocycle is constant on the intervals

$$\left(-1,-m\right) \qquad \left(-m,-\tfrac{m}{2}\right) \qquad \left(-\tfrac{m}{2},0\right),$$

where it has the values

$$B \qquad A^{-1}B \qquad BA^{-1},$$

respectively.

15.3 Abelianization

15.3.1 Approximation

There isn't an obvious way to compute the abelianized local system exactly, but there is a pretty obvious way to approximate it when m is tiny. As before, we'll show the work for the left-leaning case.

15.3.2 The slithering jumps at the break points

As m approaches zero, the sequence of $\mathrm{SL}_2\,\mathbb{C}$ elements generated by repeatedly applying the forward parallel transport cocycle to $\vec{b}$ approaches

$$\ldots B, B, B, B, BA^{-1},$$

in the sense that it takes more and more iterations to deviate from this sequence. Using the discussion after Lemma 2.2 of [27], you can deduce from

this that the forward-stable line $E^+_{\overrightarrow{b}}$ approaches the line spanned by

$$A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \mu \\ \rho \end{bmatrix} \sim \begin{bmatrix} 1 \\ \frac{\rho}{\mu} \end{bmatrix}$$

Similarly, applying the forward parallel transport cocycle to $\overleftarrow{b}$ yields a sequence approaching

$$\dots B, B, B, B, A^{-1}B$$

as m goes to zero, so $E^+_{\overleftarrow{b}}$ approaches the span of

$$B^{-1}A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\lambda}\mu \\ \lambda\rho \end{bmatrix} \sim \begin{bmatrix} 1 \\ \lambda^2 \frac{\rho}{\mu} \end{bmatrix}.$$

On the other hand, applying the backward parallel transport cocycle to b yields a sequence approaching

$$\dots B^{-1}, B^{-1}, B^{-1}, B^{-1}, B^{-1},$$

so E^-_b goes to the span of

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

The slithering jump s_b therefore approaches

$$\begin{bmatrix} 1 & \cdot \\ (\lambda^2 - 1)\frac{\rho}{\mu} & 1 \end{bmatrix}$$

as m goes to zero. A similar computation shows that s_c approaches

$$\begin{bmatrix} 1 & (\lambda^2 - 1)\frac{\rho}{\mu} \\ \cdot & 1 \end{bmatrix}$$

in the same limit.

15.3.3 The slithering jumps at all the critical points

The slithering jumps at all the critical points can be deduced from the ones at the break points using the flow-invariance property discussed in Section 12.3.4. For each $n \geq 0$, the jump $s_{\alpha^{-n}b}$ at the forward-critical point $\alpha^{-n}b$ goes to

$$B^{-n} s_b B^n = \begin{bmatrix} 1 & \cdot \\ \lambda^{2n}(\lambda^2 - 1)\frac{\rho}{\mu} & 1 \end{bmatrix}$$

as m goes to zero. The jump $s_{\alpha^n c}$ at the backward-critical point $\alpha^n c$ goes to

$$B^{n} s_c B^{-n} = \begin{bmatrix} 1 & \lambda^{2n}(\lambda^2 - 1)\frac{\rho}{\mu} \\ \cdot & 1 \end{bmatrix}$$

in the same limit.

15.3.4 The slithering deviation

Looking back at the drawing in Section 15.2.1, you can see that forward-critical points $\alpha^{-n}b$ march from right to left across Z as n grows, while the backward-critical points $\alpha^n c$ march from left to right. As m goes to zero, it takes longer and longer for the parades to meet at $-\frac{1}{2} \in Z$. It seems clear that the bound from Section 13.2 will hold more or less uniformly as m goes to zero, so we shouldn't have to worry too much about the later jumps. We can therefore compute as though the parades never meet.

In this approximation, the deviation to $-\frac{1}{2}$ from 0 is given by the product

$$\cdots \quad s_{\alpha^{-3}b} \quad s_{\alpha^{-2}b} \quad s_{\alpha^{-1}b} \quad s_b,$$

and the deviation to -1 from $-\frac{1}{2}$ is given by

$$s_c \quad s_{\alpha^1 c} \quad s_{\alpha^2 c} \quad s_{\alpha^3 c} \quad \cdots .$$

As m goes to zero, the jumps $s_{\alpha^{-n} b}$ go to commuting shears, so you should be able to show that their product approaches

$$\begin{bmatrix} 1 & \cdot \\ \frac{\rho}{\mu}(\lambda^2 - 1)\sum_{n=0}^{\infty} \lambda^{2n} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \cdot \\ -\frac{\rho}{\mu} & 1 \end{bmatrix}$$

Similarly, the product of the jumps $s_{\alpha^n c}$ should approach

$$\begin{bmatrix} 1 & \frac{\rho}{\mu}(\lambda^2 - 1)\sum_{n=0}^{\infty} \lambda^{2n} \\ \cdot & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{\rho}{\mu} \\ \cdot & 1 \end{bmatrix}$$

Now we know enough to approximate the holonomy A_{ab} of the abelianized local system around the loop that starts at $-\frac{1}{2}$, runs left to 0, wraps around to 1, and runs left back to $-\frac{1}{2}$. As m goes to zero, this holonomy approaches

$$\begin{bmatrix} 1 & \cdot \\ -\frac{\rho}{\mu} & 1 \end{bmatrix} A \begin{bmatrix} 1 & -\frac{\rho}{\mu} \\ \cdot & 1 \end{bmatrix} = \begin{bmatrix} \mu & \cdot \\ \cdot & \frac{1}{\mu} \end{bmatrix}.$$

As expected, the abelianized holonomy preserves the stable lines $E^+_{-1/2}$ and $E^-_{-1/2}$, which approach

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

as m goes to zero. In the limit, abelianization has no effect on the holonomy around a vertical loop, so B_{ab} goes to B as m goes to zero.

15.4 Spectral coordinates

We've learned that on the translation torus constructed from a left-leaning parallelogram with slope parameter m, the abelianization of the local

system

$$A = \begin{bmatrix} \mu & \rho \\ \rho & \nu \end{bmatrix} \qquad\qquad B = \begin{bmatrix} \lambda & \cdot \\ \cdot & \frac{1}{\lambda} \end{bmatrix}$$

approaches

$$A_{\mathrm{ab}} = \begin{bmatrix} \mu & \cdot \\ \cdot & \frac{1}{\mu} \end{bmatrix} \qquad\qquad B_{\mathrm{ab}} = \begin{bmatrix} \lambda & \cdot \\ \cdot & \frac{1}{\lambda} \end{bmatrix}$$

as m goes to zero. For a right-leaning parallelogram, the analogous calculation shows that the abelianization approaches

$$A_{\mathrm{ab}} = \begin{bmatrix} \frac{1}{\nu} & \cdot \\ \cdot & \nu \end{bmatrix} \qquad\qquad B_{\mathrm{ab}} = \begin{bmatrix} \lambda & \cdot \\ \cdot & \frac{1}{\lambda} \end{bmatrix}$$

as m goes to zero.

As expected, the abelianized local system splits into a pair of $\mathbb{C}^\times$ local systems comprising the forward- and backward-stable lines of the original. Restricing our attention to the forward-stable local systems, we get the limiting holonomies

$$A^+_{\mathrm{ab}} = \mu \qquad\qquad B^+_{\mathrm{ab}} = \lambda$$

in the left-leaning case, and

$$A^+_{\mathrm{ab}} = \tfrac{1}{\nu} \qquad\qquad B^+_{\mathrm{ab}} = \lambda$$

in the right-leaning case. Looking at both limits, we can recover the holomorphic coordinates μ, ν, and λ that we've been using to parameterize a dense open subset of the $\mathrm{SL}_2\,\mathbb{C}$ character variety.

Chapter 16

Future directions

16.1 Twisted character varieties

In this section, it will be convenient to use the language of twisted character varieties, which we encountered briefly in Section 3.3. I'll only give the barest sketch of how they look; a more detailed picture can be found in [44].

The *twisted* $\mathrm{SL}_2\,\mathbb{C}$ *character variety* of a compact Riemann surface C, which I'll denote $\mathcal{M}C$, is the space of irreducible flat $\mathrm{SL}_2\,\mathbb{C}$ vector bundles over the bundle of tangent directions UC. The topological type of a bundle in $\mathcal{M}C$ is determined by its holonomy around the fiber of UC, which turns out to always be a square root of unity times the identity [45]. Hence, classifying bundles by topological type splits the character variety into two pieces, $\mathcal{M}_1C$ and $\mathcal{M}_{-1}C$, labeled by the square roots of unity.

16.2 Abelianization should be more

The main result of this paper has been to show that, for a generic compact translation surface Σ, abelianization gives a well-defined map from an open subspace of the $\mathrm{SL}_2\,\mathbb{R}$ local systems on $\Sigma \smallsetminus \mathfrak{B}$ to the space of $\mathbb{R}^\times$ local systems on $\Sigma \smallsetminus \mathfrak{B}$. In the previously studied case of abelianization

on a punctured half-translation surface, however, the mere existence of the abelianization map is the least interesting of its properties. Many of the special features of the abelianization map are expected to persist for compact half-translation surfaces.

Let's review the setting for this version of abelianization. Since we're speculating anyway, let's assume the results of this paper can be extended to $\mathrm{SL}_2\,\mathbb{C}$ local systems, as discussed in Section 7.2. Suppose Σ is the translation double cover of a compact half-translation surface C. Pushback of twisted local systems along the covering $\Sigma \to C$ gives a map from $\mathcal{M}_{-1}C$ to $\mathcal{M}_1\Sigma$. Composing this map with the abelianization process, we get an abelianization map that sends twisted $\mathrm{SL}_2\,\mathbb{C}$ local systems on C to $\mathbb{C}^\times$ local systems on $\Sigma \smallsetminus \mathfrak{B}$. The abelianized local systems turn out to have holonomy -1 around each singularity, so abelianization actually sends twisted $\mathrm{SL}_2\,\mathbb{C}$ local systems on C to twisted $\mathbb{C}^\times$ local systems on Σ.

16.3 It should be a Darboux chart on $\mathcal{M}_{-1}C$

When the compact translation surface C is replaced with a punctured half-translation surface C', Gaiotto, Moore, and Neitzke showed that the abelianization map is a symplectomorphism onto its image [11, §§ 10.4 and 10.8]. It can therefore be seen as a Darboux coordinate system on the dense open subset of $\mathcal{M}_{-1}C'$ where it's defined. This is the *spectral coordinate system* discussed in [1, §4.4]. In the compact case, we've only managed to define abelianization for uniformly hyperbolic local systems, so its domain is no longer

expected to be dense. Nonetheless, it should still be a symplectomorphism onto its image.

16.4 It may be a generalized cluster coordinate chart

In the punctured case, Gaiotto, Moore, and Neitzke showed that the spectral coordinates are actually Fock-Goncharov coordinates, as we mentioned in Section 6.3.3. In particular, the spectral coordinate systems produced by rotating the half-translation structure of C' fit together into a cluster algebra. In the compact case, rotationally related spectral coordinate systems also appear to fit together into something bigger, but that something doesn't seem to be a cluster algebra: it has mutation-like behavior, but no readily identifiable clusters. Some preliminary investigations of this structure are briefly reported in [46].

16.5 It should be holomorphic

On a punctured half-translation surface, the source and target of the abelianization map are not only symplectic manifolds, but holomorphic symplectic manifolds, and abelianization is a holomorphic symplectomorphism [11, §10.4]. The holomorphicity of the abelianization map isn't special to half-translation surfaces, and it appears to persist in the compact case. Here's a sketch of an argument that the abelianization map for a compact translation surface Σ is holomorphic. Note that Σ doesn't have to be the translation double cover of a half-translation surface.

An easy way to see the complex structure on the moduli space of $\mathrm{SL}_2\,\mathbb{C}$ local systems on $\Sigma \smallsetminus \mathfrak{B}$ is to pick a well-cut flow box in $\overset{\leftrightarrow}{\Sigma}$. As we observed in Section 11.6, a local system on $\Sigma \smallsetminus \mathfrak{B}$ is described up to isomorphism by an interval cocycle over the flow box's first return relation α, and interval cocycles cut across all the isomorphism classes. An interval cocycle over α is just an element A of $(\mathrm{SL}_2\,\mathbb{C})^{\mathcal{A}^+}$, where $\mathcal{A}^+$ is the set of intervals exchanged by α. The complex structure of this space matches the one on the space of local systems.

Suppose A is uniformly hyperbolic. As in Section 12.3, write the slithering jump for A at a critical point w of α as $s_w \in \mathrm{SL}_2\,\mathbb{C}$. The deviation σ that abelianizes A is given by ordered products of s_w over the critical points. The Lyapunov exponent of A can be bounded when A varies over a small enough region [27, proof of Proposition 2.6], so the bound on the size of s_w in Section 13.2 should be uniform with respect to A. The products that define σ should therefore converge uniformly with respect to A. If this works, the task of showing that σ varies holomorphically with A is reduced to the task of showing that s_w does for each critical point w.

Say w is a backward-critical point. The formulas in Section 12.3 make it clear that s_w depends holomorphically on the stable lines $E^-_{\overleftarrow{w}}, E^+_w, E^-_{\overrightarrow{w}} \in \mathbf{P}\mathbb{C}^2$. The line $E^{\pm}_h$ is approximated by the line most contracted by $A^{\pm n}_h$ when n is large [27, proof of Proposition 2.1]. In fact, the convergence of the most contracted line to the stable line is uniform with respect to A [27, discussion around Equation 2.6], so we just need to show that the most contracted line depends holomorphically on A.

Using the standard inner product on $\mathbb{C}^2$, we can observe that the line most contracted by $A_h^{\pm n}$ lies in the eigenspace of $(A_h^{\pm n})^\dagger(A_h^{\pm n})$ with the smallest eigenvalue. Since A is uniformly hyperbolic, it should be safe to assume that $(A_h^{\pm n})^\dagger(A_h^{\pm n})$ has distinct eigenvalues for large enough n, so the most contracted line is just the eigenline with the smallest eigenvalue. The relationship between an operator and its eigenlines is holomorphic, and the map that sends A to $A_h^{\pm n}$ is too, so we should be done.

Appendices

Appendix A

Technical tools for warping local systems

A.1 The lily pad lemma

Lemma A.1.A. *Suppose $\mathcal{U}$ is an open cover of a connected space. For any two points a and b in the space, there is a finite sequence of elements of $\mathcal{U}$ in which the first element contains a, the last element contains b, and every element intersects the next one.*

Proof. Let's call a finite sequence of elements of $\mathcal{U}$ a *lily path* if every element intersects the next one. We'll say two points a and b can be "connected by a lily path" if there's a lily path whose first element contains a and whose last element contains b.

A lily path connecting a to b also connects a to every other point in the last element of the path, which is an open neighborhood of b. Hence, the set of points that can be connected to a by a lily path is open.

On the other hand, suppose b can't be connected to a by a lily path. Since $\mathcal{U}$ is a cover, there's some $U \in \mathcal{U}$ containing b. If a point in U could be connected to a by a lily path, adding U to the end of that path would give a lily path connecting a to b. Hence, no point in U can be connected to a by a

lily path. Therefore, the set of points that can't be connected to a by a lily path is also open. □

A.2 Collapsing downward-directed colimits

Proposition A.2.A. *Suppose Λ is a downward-directed set, C is some category, and $F\colon \Lambda \to C$ is a diagram in which all the arrows are isomorphisms. Then F has a colimit, and the defining arrows from the diagram to its colimit are isomorphisms.*

Proof. Pick any object s of Λ. For any other object t, we can get an isomorphism $f_t\colon F(t) \to F(s)$ by picking a common lower bound $\check{t}$ of s and t and taking the composition $F(\check{t} \leftarrow s)^{-1}F(\check{t} \leftarrow t)$. The isomorphism we get doesn't depend on our choice of $\check{t}$.

Now, pick any object c of C. Suppose that for each object t of Λ, we have an arrow $\phi_t\colon F(t) \to c$, and these arrows commute with the arrows of the diagram F. Observing that

$$\begin{aligned}\phi_s f_t &= \phi_s F(\check{t} \leftarrow s)^{-1} F(\check{t} \leftarrow t)\\ &= \phi_{\check{t}} F(\check{t} \leftarrow t)\\ &= \phi_t,\end{aligned}$$

we see that the object $F(s)$, equipped with the isomorphisms $f_t\colon F(t) \to F(s)$, is a colimit of the diagram F. □

Appendix B

Relational dynamics

Some dynamical systems, including vertical flows on singular translation surfaces, interval exchanges, and even the humble doubling map, are discontinuous if you insist on defining them at every point. (These particular examples are discussed in Sections 9.2.1 and 9.3.2.) You can recover a kind of continuity, however, if you describe the dynamics using relations rather than maps, allowing points to get lost as they fall into singularities, breaks between intervals, or whatever.

Consider a relation ϕ between topological spaces Y and X. Define

$$\phi A = \{y \in Y : y \,\phi\, a \text{ for some } a \in A\}$$
$$B\phi = \{x \in X : b \,\phi\, x \text{ for some } b \in B\}$$

for subsets $A \subset X$ and $B \subset Y$. For convenience, we'll relax the distinction between singletons and points, denoting $\phi\{x\}$, for example, by ϕx. If ϕ is a function $Y \leftarrow X$, then $\phi x \in Y$ is the value of ϕ at a point $x \in X$, and $B\phi \subset X$ is the preimage of a subset $B \subset Y$.

Define ϕ to be *injective* if $y\phi$ is a singleton for all $y \in Y$, *coinjective* if ϕx is a singleton for all $x \in X$, and *biinjective* if it's both injective and

coinjective. In less baroque language, a coinjective relation is just a partially defined function.

Define ϕ to be *continuous* if $V\phi$ is open whenever $V \subset Y$ is open, *cocontinuous* if ϕU is open whenever $U \subset X$ is open, and *bicontinuous* if it's both continuous and cocontinuous. If ϕ is a function, "continuous" means what it usually means, and "cocontinuous" means "open." Local systems can be pushed forward and backward along a bicontinuous relation.

A *flow by bicontinuous relations* on X can be defined as a relation ψ between X and $\mathbb{R} \times X$ with the following properties:

- As a whole, ψ is continuous and coinjective.
- For all $t \in \mathbb{R}$, the relation $\psi^t = \blacksquare\, \psi\, (t, \blacksquare)$ is bicontinuous and biinjective.
- For all $t, s \in \mathbb{R}$, we have $\psi^t \psi^s = \psi^{t+s}$.

This kind of partially defined flow acts a lot like an ordinary flow by homeomorphisms. In particular, if X carries a local system $\mathcal{E}$, it gives a parallel transport morphism $\mathcal{E}_{\psi^t U} \leftarrow \mathcal{E}_U$ for every open subset U of X and every time t.

Appendix C

Uniform hyperbolicity and Lyapunov exponents

Let $\alpha\colon X \to X$ be a homeomorphism of a compact, metrizable space. Consider a dynamical cocycle over α specified by a continuous map $A\colon X \to \mathrm{SL}(E)$, where E is the Euclidean plane as characterized in Section E. In this more general setting, the local uniform hyperbolicity condition discussed in Section 11.3 is just called *uniform hyperbolicity* [27, §2.2]. Its definition carries over word for word, substituting X for $\ddot{H}$. The terminology of stable lines and bounding exponents introduced in Section 11.3 will be used here as well.

The purpose of this appendix is to prove Lemma 11.3.B, Lemma 11.3.C, and Proposition 11.3.D from Section 11.3 in the general setting described above. Like the definition of uniform hyperbolicity, the statements of these lemmas carry over word for word, substituting X for $\ddot{H}$. Condition $2'$ from Lemma 11.3.C is stronger than Condition 2 in the definition of uniform hyperbolicity. It can therefore be substituted in to give an alternate characterization of uniform hyperbolicity for cocycles over a minimal, uniquely ergodic dynamical system.

We get the first lemma by stringing together a few results from the literature.

Proof of Lemma 11.3.B. Suppose A is uniformly hyperbolic. By Proposition 2.1 of [27], there's a constant $\Delta > 1$ such that $\Delta^n \lesssim \|A_x^n\|$ for all $x \in X$, with n varying over $\mathbb{N}$. The constant multiple implied in this bound can be made the same for all $x \in X$. Corollary 4.4 of [28] then gives us a constant $\Lambda > 0$ such that[1]

$$\lim_{n\to\infty} \frac{1}{n} \log \|A_x^n\| = \Lambda$$

uniformly over all $x \in X$. Applying Theorem 3 of [28], we learn that $E_x^{\pm}$ vary continuously with respect to x. □

The proof of the second lemma is more involved. We start by proving Proposition 11.3.D, which turns around the decay condition from Lemma 11.3.C to get an equivalent growth condition on backward parallel transport. As you might expect from the discussion before the analogous Proposition 11.2.A, our argument depends crucially on the uniform convergence of the limit in the decay condition.

Proof of Proposition 11.3.D. Suppose Condition 2′ holds. Then, given any neighborhood Ω of Λ, we can find some $N \in \mathbb{N}$ such that

$$\frac{1}{n} \log \frac{\|A_x^{\pm n} v\|}{\|v\|} \in -\Omega$$

for all $x \in X$, $v \in E_x^{\pm}$, and $n \geq N$. In particular,

$$\frac{1}{n} \log \frac{\|A_x^{\pm n} A_{\alpha^{\pm n}x}^{\mp n} v'\|}{\|A_{\alpha^{\pm n}x}^{\mp n} v'\|} \in -\Omega$$

[1] To avoid confusion when comparing with the article cited, note that uniformity and uniform hyperbolicity are distinctly different conditions, though in our current setting they are closely related.

for all $x \in X$, $v' \in E^{\pm}_{\alpha^{\pm n}x}$, and $n \geq N$. Rearranging the expression and rewriting the quantifier over x, we see that

$$\frac{1}{n} \log \frac{\|A^{\mp n}_{x'} v'\|}{\|v'\|} \in \Omega$$

for all $x' \in X$, $v' \in E^{\pm}_{x'}$, and $n \geq N$. We found the N in this expression starting from an arbitrary neighborhood Ω of Λ, so we've shown that the limit in the proposition converges uniformly.

We now see that Condition 2′ implies the condition in the proposition. The same kind of reasoning can be used to prove the reverse implication. □

Armed with the proposition above, we're ready to prove the second lemma.

Proof of Lemma 11.3.C. Suppose A is uniformly hyperbolic. We want to show that it satisfies the uniform decay condition 2′, and that its Lyapunov exponent is the supremum of its bounding exponents.

The uniform decay condition is satisfied We can find $\|A^n_x\|$ by taking the supremum of

$$\frac{\|A^n_x(u+v)\|}{\|u+v\|}$$

over all $u \in E^-_x$ and $v \in E^+_x$ with $u + v \neq 0$. Using Proposition E.E,

$$\begin{aligned} \frac{\|A^n_x(u+v)\|}{\|u+v\|} &\leq \frac{\|A^n_x u\|}{\|u+v\|} + \frac{\|A^n_x v\|}{\|u+v\|} \\ &\leq \frac{4}{d_\angle(u,v)^2} \left(\frac{\|A^n_x u\|}{\|u\|} + \frac{\|A^n_x v\|}{\|v\|} \right). \end{aligned}$$

The right-hand side is the same for all $u \in E_x^-$ and $v \in E_x^+$, so

$$\|A_x^n\| \leq \frac{4}{d_\angle(u,v)^2}\left(\frac{\|A_x^n u\|}{\|u\|} + \frac{\|A_x^n v\|}{\|v\|}\right)$$

for any $u \in E_x^-$ and $v \in E_x^+$.

We know from Lemma 11.3.B that the stable lines $E_x^\pm$ vary continuously with respect to x. Since X is compact, and the forward- and backward-stable lines can never coincide, it follows that $d_\angle(E_x^-, E_x^+)$ is bounded away from zero. We can therefore summarize the inequality above by saying that

$$\|A_x^n\| \lesssim \frac{\|A_x^n u\|}{\|u\|} + \frac{\|A_x^n v\|}{\|v\|},$$

over all $x \in X$, $u \in E_x^-$, and $v \in E_x^+$.

Rewrite the summarized inequality as

$$\|A_x^n\| \lesssim \frac{\|A_x^n u\|}{\|u\|}\left(1 + \frac{\|A_x^n v\|}{\|v\|} \cdot \frac{\|u\|}{\|A_x^n u\|}\right).$$

The uniform hyperbolicity of A, with some bounding exponent $K > 0$, implies that

$$\frac{\|A_x^n v\|}{\|v\|} \cdot \frac{\|u\|}{\|A_x^n u\|} \lesssim e^{-2Kn}$$

over all $x \in X$, $u \in E_x^-$, $v \in E_x^+$, and $n \in \mathbb{N}$. Consequently,

$$\|A_x^n\| \lesssim \frac{\|A_x^n u\|}{\|u\|}$$

over all $x \in X$, $u \in E_x^-$, and $n \in \mathbb{N}$.

At this point, it's convenient to pick a constant e^C such that

$$e^C \|A_x^n\| \leq \frac{\|A_x^n u\|}{\|u\|}$$

for all $x \in X$, $u \in E_x^-$, and $n \in \mathbb{N}$. Combining this bound with the definition of the operator norm, we see that

$$e^C \|A_x^n\| \;\leq\; \frac{\|A_x^n u\|}{\|u\|} \;\leq\; \|A_x^n\|$$

(learning, incidentally, that C must be negative). Logarithmically,

$$\frac{C}{n} + \frac{1}{n} \log \|A_x^n\| \;\leq\; \frac{1}{n} \log \frac{\|A_x^n u\|}{\|u\|} \;\leq\; \frac{1}{n} \log \|A_x^n\|.$$

The left- and right-hand sides both converge uniformly to Λ as n grows, so we've proven that

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\|A_x^n u\|}{\|u\|} = \Lambda$$

uniformly over all $x \in X$ and $u \in E_x^-$. It follows, by Proposition 11.3.D, that vectors in E^- decay as desired under backward iteration of A.

Now we just need to show that vectors in E_x^+ decay as desired under forward iteration of A. Pick any nonzero $u \in E_x^-$ and $v \in E_x^+$. Since the maps A_x are volume-preserving, Proposition E.A tells us that

$$\|A_x^n u\| \|A_x^n v\| \; d_\angle(A_x^n u, A_x^n v) = \|u\| \|v\| \; d_\angle(u, v)$$

for all $n \in \mathbb{N}$. Rearranging, we see that

$$\frac{\|A_x^n v\|}{\|v\|} = \frac{\|u\|}{\|A_x^n u\|} \; \frac{d_\angle(u, v)}{d_\angle(A_x^n u, A_x^n v)}$$

for all $n \in \mathbb{N}$. Logarithmically,

$$\frac{1}{n} \log \frac{\|A_x^n v\|}{\|v\|} = -\left(\frac{1}{n} \log \frac{\|A_x^n u\|}{\|u\|}\right) + \frac{1}{n} \log \frac{d_\angle(u, v)}{d_\angle(A_x^n u, A_x^n v)}$$

for all $n \in \mathbb{N}$. We argued earlier that the distance between the stable lines is bounded away from zero, the definition of the sine metric ensures that the distance is bounded above by one. That means the distance ratio term on the right-hand side goes uniformly to zero as n grows. The other term, which describes the growth of $u \in E_x^-$, is familiar from before: we proved that the part in brackets converges to Λ uniformly over all $x \in X$ and $u \in E_x^-$. Thus,

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\|A_x^n v\|}{\|v\|} = -\Lambda$$

uniformly over all $x \in X$ and $v \in E_x^+$.

The supremum of the bounding exponents is Λ Let's start with the quick direction: showing that any bounding exponent for A is less than or equal to Λ. If $K > 0$ is a bounding exponent,

$$e^{Kn}\|u\| \lesssim \|A_x^n u\|$$

as n varies over $\mathbb{N}$ for all $x \in X$ and $u \in E_x^-$, with the same implied constant multiple for all x. Combining this with the bound $\|A_x^n u\| \leq \|A_x^n\|\|u\|$, we learn that

$$e^C e^{Kn} \leq \|A_x^n\|$$

for some constant C. Logarithmically,

$$\frac{C}{n} + K \leq \frac{1}{n} \log \|A_x^n\|.$$

Taking limits of both sides as n grows, we see that $K \leq \Lambda$.

Now for the tedious direction: showing that any positive number less than Λ is a bounding exponent for A. Pick some $K \in (0, \Lambda)$. Since

$$\frac{1}{n} \log \|A_x^n\| = \Lambda$$

uniformly over x, we can find some $N \in \mathbb{N}$ such that

$$K \leq \frac{1}{n} \log \|A_x^n\|$$

for all $n \geq N$. In other words,

$$e^{Kn} \leq \|A_x^n\|$$

for all $n \geq N$. Since X is compact, and A_x^n is never zero, $\|A_x^n\|$ is bounded away from zero for any fixed n. That means we can find a constant multiple of e^{Kn} small enough to slip under the $\|A_x^n\|$ for all $n < N$ and $x \in X$. In other words,

$$e^{Kn} \lesssim \|A_x^n\|$$

over all $x \in X$ and $n \in \mathbb{N}$.

We saw earlier that

$$\|A_x^n\| \lesssim \frac{\|A_x^n u\|}{\|u\|}$$

over all $x \in X$, $u \in E_x^-$, and $n \in \mathbb{N}$. Combining this with the bound above, we learn that

$$e^{Kn} \lesssim \frac{\|A_x^n u\|}{\|u\|}$$

over all $x \in X$, $u \in E_x^-$, and $n \in \mathbb{N}$. This is the uniform forward growth condition we want for vectors in E_x^-.

Now we just need to prove the associated uniform forward decay condition for vectors in E_x^+. The growth condition we just proved implies that

$$\frac{\|u\|}{\|A_x^n u\|} \lesssim e^{-Kn}$$

over all $x \in X$, $u \in E_x^-$, and $n \in \mathbb{N}$. Recall that

$$\frac{\|A_x^n v\|}{\|v\|} = \frac{\|u\|}{\|A_x^n u\|} \, \frac{d_\angle(u, v)}{d_\angle(A_x^n u, A_x^n v)}$$

for all $u \in E_x^-$ and $v \in E_x^+$, and the distance ratio factor on the right-hand side is both bounded above and bounded away from zero. Together, these facts tell us that

$$\frac{\|A_x^n v\|}{\|v\|} \lesssim e^{-Kn}$$

over all $x \in X$, $v \in E_x^+$, and $n \in \mathbb{N}$. This is the forward uniform decay condition we want for vectors in E_x^+. $\square$

Appendix D

Infinite ordered products

D.1 Definition

Suppose M is a Hausdorff topological monoid, like the one formed by the endomorphisms or automorphisms of a finite-dimensional vector space, and A is a totally ordered set. Given a function $m\colon A \to M$, we'd like to make sense of the potentially infinite ordered product

$$\prod_{p \in A} m_p,$$

which I'll refer to in writing as "the product of m over A."

Recall that a function from a directed set Λ into a topological space is called a *net*. For any $s \in \Lambda$, let's call the set $\{t \in \Lambda : t \geq s\}$ the *shadow* of s. A net is said to *converge* to a point if, for every open neighborhood Ω of that point, there is some element of Λ whose shadow is sent by the net into Ω. A net into a Hausdorff space, like M, converges to at most one point.

The finite subsets of A form a directed set under inclusion, and the product of m over any finite subset is well-defined, so we can define the product of m over A to be the limit of the net that sends each finite subset of A to the product of m over that subset. I'll call this net the "product net" for short.

D.2 Calculation

Suppose Λ' and Λ are directed sets, n is a net on Λ, and $f\colon \Lambda' \to \Lambda$ is an order-preserving map whose image intersects the shadow of every element of Λ. In these circumstances, the net $n \circ f$ is called a *subnet* of n. If n converges to a certain point, every subnet of n converges to that point.

In particular, let A be a totally ordered set, and Λ its finite subsets. If $f\colon \mathbb{N} \to \Lambda$ is an increasing sequence of subsets whose union is all of A, the image of f intersects the shadow of every finite subset. Thus, if a product over A converges, we can find it by looking at partial products over any sequence of finite subsets whose union is A. Such a sequence must exist if A is countable.

D.3 Composition

Given two totally ordered sets B and A, let $B \sqcup A$ be the disjoint union of B and A with the total order that makes the inclusions order-preserving and puts every element of B to the left of every element of A.

Proposition D.3.A. *Say we have a function $m\colon B \sqcup A \to M$. If the products of m over B and A converge, then the product over $B \sqcup A$ converges too, and*

$$\prod_{p \in B \sqcup A} m_p = \left(\prod_{p \in B} m_p \right) \left(\prod_{p \in A} m_p \right).$$

Proof. Let β and α be the products of m over B and A, respectively. Given an open neighborhood Ω of $\beta\alpha$, we want to find a finite subset of $B \sqcup A$ whose shadow is sent by the product net into Ω.

Using the fact that multiplication in M is continuous, find open neighborhoods Ω_B of β and Ω_A of α with $\Omega_B\Omega_A \subset \Omega$. Then, find finite subsets S_B and S_A of B and A whose shadows are sent into Ω_B and Ω_A, respectively.

For any finite subset R of $B \sqcup A$ containing $S_B \cup S_A$, we can use the fact that $B > A$ to rewrite the product of m over R as

$$\left(\prod_{p\in R\cap B} m_p\right)\left(\prod_{p\in R\cap A} m_p\right).$$

Observing that $R \cap B$ contains S_B and $R \cap A$ contains S_A, we conclude that the product of m over R is in Ω. □

D.4 Equivariance

Proposition D.4.A. *If the product of $m\colon A \to M$ over A converges,*

$$\phi\left(\prod_{p\in A} m_p\right) = \prod_{p\in A} \phi(m_p)$$

for any continuous homomorphism $\phi\colon M \to M$.

Proof. Given an open neighborhood Ω of the left-hand side, we want to find a finite subset S of A with the property that

$$\prod_{p\in R} \phi(m_p) \in \Omega$$

for all finite subsets $R \subset A$ containing S.

Since ϕ is continuous, $\phi^{-1}(\Omega)$ is open neighborhood of the product of m over A. By the definition of convergence, we can find a finite subset S of A

with the property that

$$\prod_{p \in R} m_p \in \phi^{-1}(\Omega)$$

for all finite subsets $R \subset A$ containing S. Applying ϕ to both sides, we see that S is just what we wanted. □

D.5 Inversion

Instead of just a map into a topological monoid, suppose we have a map $g\colon A \to G$ into a topological group. Let A^{op} be A with the opposite order.

Proposition D.5.A. *If the product of g over A converges,*

$$\prod_{p \in A^{\mathrm{op}}} g_p^{-1} = \left(\prod_{p \in A} g_p \right)^{-1}.$$

Proof. Analogous to the proof of Proposition D.4.A. □

We won't use this result for anything. It's only here to reassure you that the directionality of our construction of deviations from jumps in Section 10.3 doesn't introduce any actual asymmetry.

D.6 Convergence

Any Banach algebra, like $\operatorname{End} \mathbb{R}^2$ with the operator norm, can be thought of as a topological monoid by forgetting the addition. Since all the ordered products we care about will be taken in $\mathrm{SL}_2 \mathbb{R}$, a closed submonoid of $\operatorname{End} \mathbb{R}^2$, understanding ordered products in a Banach algebra will be very helpful to us.

The following results generalize Theorem 2.3, Corollary 2.4, and a simplified version of Theorem 2.7 from [47] to products over arbitrarily ordered index sets.

Proposition D.6.A. *Suppose we have a totally ordered set A, a unital Banach algebra $\mathcal{X}$, and a function $x\colon A \to \mathcal{X}$. If the sum $\sum_{p\in A} \|x_p - 1\|$ converges, the product $\prod_{p\in A} x_p$ converges as well.*

Proposition D.6.B. *If the sum in the proposition above converges, and each factor x_p is invertible, the product is invertible.*

Our proof will give a handy bound for free.

Proposition D.6.C. *If the sum in Proposition D.6.A converges, the product $\prod_{p\in A} \|x_p\|$ converges as well, and*

$$\left\|\prod_{p\in A} x_p\right\| \quad\leq\quad \prod_{p\in A} \|x_p\| \quad\leq\quad \exp\left(\sum_{p\in A} \|x_p - 1\|\right).$$

Let's start with a less ambitious result.

Proposition D.6.D. *If the sum in Proposition D.6.A converges, then for any $C > \exp\left(\sum_{p\in A} \|x_p - 1\|\right)$, there's a finite subset S of A with the property that*

$$\prod_{p\in R} \|x_p\| < C$$

for all finite subsets $R \subset A$ containing S.

Proof. Assume $\sum_{p\in A}\|x_p-1\|$ converges, and consider any positive constant C with $\log C > \sum_{p\in A}\|x_p-1\|$. By the definition of convergence, we can find a finite subset S of A with the property that

$$\sum_{p\in R}\|x_p-1\| < \log C$$

for all finite subsets $R\subset A$ containing S. Since

$$\log\|x_p\| \le \big|\|x_p\|-1\big| \qquad \text{and} \qquad \big|\|x_p\|-1\big| \le \|x_p-1\|$$

for all $p\in A$, we know

$$\sum_{p\in R}\log\|x_p\| < \log C.$$

for all finite subsets $R\subset A$ containing S. Exponentiating both sides gives the desired result. □

Proof of Proposition D.6.A. By definition, $\mathcal{X}$ is complete, so we can prove that the product converges by showing that the product net is Cauchy. To that end, given any $\epsilon>0$, we want to find a finite subset S of A with the property that

$$\left\|\prod_{p\in R}x_p - \prod_{p\in S}x_p\right\| < \epsilon$$

for all finite subsets $R\subset A$ containing S.

Assume the sum $\sum_{p\in A}\|x_p-1\|$ converges, and pick a constant $C > \exp\left(\sum_{p\in A}\|x_p-1\|\right)$. Every convergent net is Cauchy [48, Proposition 3.2], so we can find a finite subset S' of A with the property that

$$\left|\sum_{p\in R}\|x_p-1\| - \sum_{p\in S'}\|x_p-1\|\right| < \epsilon/C$$

for all finite subsets $R \subset A$ containing S'. This inequality simplifies to

$$\sum_{p \in R \smallsetminus S'} \|x_p - 1\| < \epsilon/C.$$

By Proposition D.6.D, we can also find a finite subset S'' of A with the property that

$$\prod_{p \in R} \|x_p\| < C$$

for all finite subsets $R \subset A$ containing S''. Defining S as $S' \cup S''$, and observing that $R \smallsetminus S'$ contains $R \smallsetminus S$, we see that

$$\sum_{p \in R \smallsetminus S} \|x_p - 1\| < \epsilon/C \qquad \text{and} \qquad \prod_{p \in R} \|x_p\| < C$$

for all finite subsets $R \subset A$ containing S.

Put the elements of $R \smallsetminus S$ in some order $r_1, \ldots, r_n$. Let

$$R_0 = S,$$

$$R_1 = R_0 \cup \{r_1\},$$

$$R_2 = R_1 \cup \{r_2\},$$

and so on. Let

$$\Delta_{k+1} = \left\| \prod_{p \in R_{k+1}} x_p - \prod_{p \in R_k} x_p \right\|$$

Notice that

$$\prod_{p \in R_{k+1}} x_p - \prod_{p \in R_k} x_p = \left(\prod_{\substack{p \in R_k \\ p > r_{k+1}}} x_p \right) (x_{r_{k+1}} - 1) \left(\prod_{\substack{p \in R_k \\ r_{k+1} > p}} x_p \right),$$

yielding the bound

$$\begin{aligned}\Delta_{k+1} &\leq \|x_{r_{k+1}} - 1\| \prod_{p \in R_k} \|x_p\| \\ &\leq \|x_{r_{k+1}} - 1\| C.\end{aligned}$$

Finally, observe that

$$\begin{aligned}\left\| \prod_{p \in R} x_p - \prod_{p \in S} x_p \right\| &\leq \Delta_1 + \ldots + \Delta_n \\ &\leq \|x_{r_1} - 1\| C + \ldots + \|x_{r_n} - 1\| C \\ &= C \sum_{p \in R \smallsetminus S} \|x_p - 1\| \\ &< \epsilon.\end{aligned}$$

Since R could have been any finite subset of A containing S, and the method we used to find S works for any $\epsilon > 0$, we've proven that the product net is Cauchy. □

Proof of Proposition D.6.B. Given an ordered set I, let I^{op} be the same set in the opposite order. Assume $\sum_{p \in A} \|x_p - 1\|$ converges. Equivalently, because addition is commutative, $\sum_{p \in A^{\mathrm{op}}} \|x_p - 1\|$ converges. This is only possible if $\|x_p - 1\| \leq 6/7$ for all but finitely many $p \in A^{\mathrm{op}}$. It follows, by the calculation in the proof of [47, Lemma 2.6], that $\|x_p^{-1}\| \leq 7$ for all but finitely many $p \in A^{\mathrm{op}}$. That means $\|x_p^{-1}\|$ has a maximum over all $p \in A^{\mathrm{op}}$, which I'll call M. Just as in the proof of [47, Theorem 2.7], observe that

$$\begin{aligned}\|x_p^{-1} - 1\| &= \|x_p^{-1}(1 - x_p)\| \\ &\leq M \|x_p - 1\|\end{aligned}$$

for all $p \in A^{\text{op}}$. Then [49, Exercise 7.40.c] tells us that $\sum_{p\in A^{\text{op}}} \|x_p^{-1} - 1\|$ converges, so $\prod_{p\in A^{\text{op}}} x_p^{-1}$ converges by Proposition D.6.A.

Define $y'' = \prod_{p\in A^{\text{op}}} x_p^{-1}$ and $y' = \prod_{p\in A} x_p$. Multiplication in a Banach algebra is continuous, so for any open neighborhood Ω of $y''y'$, we can find open neighborhoods Ω'' of y'' and Ω' of y' with $\Omega''\Omega' \subset \Omega$. By convergence, we can find finite subsets S'' and S' of A such that

$$\prod_{p\in R^{\text{op}}} x_p^{-1} \in \Omega''$$

for all finite subsets $R \subset A$ containing S'', and

$$\prod_{p\in R} x_p \in \Omega'$$

for all finite subsets $R \subset A$ containing S'. Now, defining S as $S'' \cup S'$, we can observe that

$$\left(\prod_{p\in S^{\text{op}}} x_p^{-1}\right)\left(\prod_{p\in S} x_p\right) \in \Omega.$$

But the product in the expression above is clearly equal to 1! We've shown that every open neighborhood of $y''y'$ contains 1, which means $y''y'$ is equal to 1. The same argument can be used to show that $y'y''$ is 1. Therefore, $\prod_{p\in A} x_p$ is invertible, with inverse $\prod_{p\in A^{\text{op}}} x_p^{-1}$. □

Proof of Proposition D.6.C. Assume $\sum_{p\in A} \|x_p - 1\|$ converges. By Proposition D.6.A, $\prod_{p\in A} x_p$ converges too. The convergence of $\prod_{p\in A} \|x_p\|$ follows immediately from the fact that the norm is continuous. The first inequality is easy to establish.

Because $\sum_{p\in A}\|x_p-1\|$ is a sum of non-negative numbers, its convergence implies that at most countably many of its terms are nonzero. We can therefore assume, without loss of generality, that A is countable. Combining this fact with Proposition D.6.D and the discussion in Section D.2, it's not hard to show that $\prod_{p\in A}\|x_p\|$ is less than any number greater than $\exp\left(\sum_{p\in A}\|x_p-1\|\right)$. That gives the second inequality. □

Appendix E

Linear algebra on the Euclidean plane

The Euclidean plane is a two-dimensional real inner product space E with a volume form D in which the unit square has unit volume. It will be useful to collect some basic facts about geometry in such a space. Really well-known facts will be stated without proof.

The sine metric $d_\angle$ on $\mathbf{P}E$ is defined as in Section 14.2. The area of a parallelogram can be computed from the lengths of its sides and the angle between them.

Proposition E.A.

$$|D(u,v)| = \|u\|\|v\|\, d_\angle(u,v)$$

for all $u, v \in E$.

By comparing the areas of some well-chosen parallelograms, you can deduce the law of sines.

Proposition E.B (The law of sines)**.**

$$\|u\|\, d_\angle(u, u+v) = \|v\|\, d_\angle(v, u+v)$$

for all $u, v \in E$ *with* $u + v \neq 0$.

From this we see that the extent to which a linear map T can move lines is limited by the operator norm of $T - 1$.

Proposition E.C. *For any linear map $T\colon E \to E$,*

$$d_\angle(u, Tu) \leq \|T - 1\|$$

whenever $Tu \neq 0$.

Proof. Because distances in $\mathbf{P}F$ are never greater than one, it follows from the law of sines that

$$d_\angle(u, u + v) \leq \frac{\|v\|}{\|u\|}$$

In particular,

$$d_\angle(u, Tu) \leq \frac{\|(T - 1)u\|}{\|u\|}.$$

□

The extent to which a volume-preserving map can expand angles is limited by the operator norm of its inverse.

Proposition E.D. *For any $T \in \mathrm{SL}(E)$,*

$$d_\angle(Tu, Tv) \leq \|T^{-1}\|^2 \, d_\angle(u, v)$$

for all $u, v \in E$.

Proof.

$$\begin{aligned} d_\angle(Tu, Tv) &= \frac{|D(Tu, Tv)|}{\|Tu\|\|Tv\|} \\ &= \frac{\|u\|}{\|Tu\|}\,\frac{\|v\|}{\|Tv\|}\,\frac{|D(u,v)|}{\|u\|\|v\|} \\ &= \frac{\|u\|}{\|Tu\|}\,\frac{\|v\|}{\|Tv\|}\,d_\angle(u,v) \\ &\leq \|T^{-1}\|^2\, d_\angle(u,v). \end{aligned}$$

□

The angle between two vectors puts a lower bound on how completely they can cancel out when you add them together.

Proposition E.E.

$$\frac{d_\angle(u,v)^2}{4}\,\|u\| \leq \|u+v\|$$

for all $u, v \in E$.

Proof. By the Cauchy-Schwarz inequality,

$$\begin{aligned} \|u\|\|u+v\| &\geq |\langle u, u+v\rangle| \\ &= \left|\|u\|^2 + \langle u, v\rangle\right| \\ &\geq \|u\|^2 - |\langle u, v\rangle|. \end{aligned}$$

Rearranging and dividing through by $\|u\|$, we find that

$$\|u\| \leq \frac{|\langle u, v\rangle|}{\|u\|} + \|u+v\|.$$

In other words,

$$\|u\| \leq \|v\| |\cos\theta| + \|u+v\|,$$

where θ is the angle between the lines spanned by u and v. We know from the triangle inequality that $\|v\| \leq \|u\| + \|u+v\|$, so

$$\|u\| \leq (\|u\| + \|u+v\|)|\cos\theta| + \|u+v\|.$$

Rearranging again, we see that

$$\begin{aligned}
\|u\| &\leq \frac{1+|\cos\theta|}{1-|\cos\theta|}\,\|u+v\| \\
&= \frac{(1+|\cos\theta|)^2}{1-|\cos\theta|^2}\,\|u+v\| \\
&\leq \frac{4}{|\sin\theta|^2}\,\|u+v\|
\end{aligned}$$

Observing that $d_\angle(u,v) = |\sin\theta|$ completes the proof. □

Appendix F

Standard punctures for translation surfaces

F.1 Motivation

To see where the standard puncture shapes come from, we need to talk about the complex geometry of translation surfaces, whose only role until now has been a brief appearance in the proof of Proposition 9.5.I.

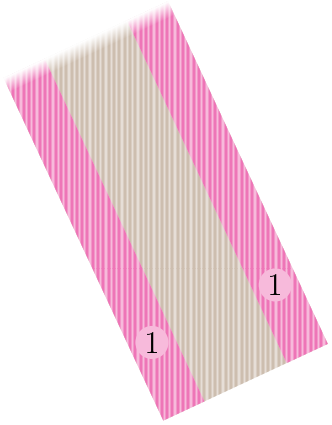
Every translation surface Σ comes with a complex structure, which we get by identifying $\mathbb{R}^2$ with $\mathbb{C}$ in the usual way, and a complex-valued 1-form ω, which sends horizontal unit vectors to 1 and vertical unit vectors to i. Observing that $\omega = dz$ for any local translation chart $z\colon \Sigma \to \mathbb{C}$, we see that ω is holomorphic. Conversely, a complex 1-manifold equipped with a holomorphic 1-form is canonically a translation surface. Where ω has a zero of order n, the translation structure has a conical singularity of angle $2(n+1)\pi$.

The complex point of view suggests a natural class of translation surfaces that are non-compact, but still well-behaved. Putting a meromorphic 1-form on a compact Riemann surface defines a translation structure on the complement of the poles. The poles correspond to the ends of the translation surface, and the Riemann surface is its end compactification [50, §1]. The poles of a 1-form have a limited variety of behaviors, so the ends of the trans-

lation surface have a limited variety of shapes, which we'll call the *standard punctures*.

F.2 First-order punctures

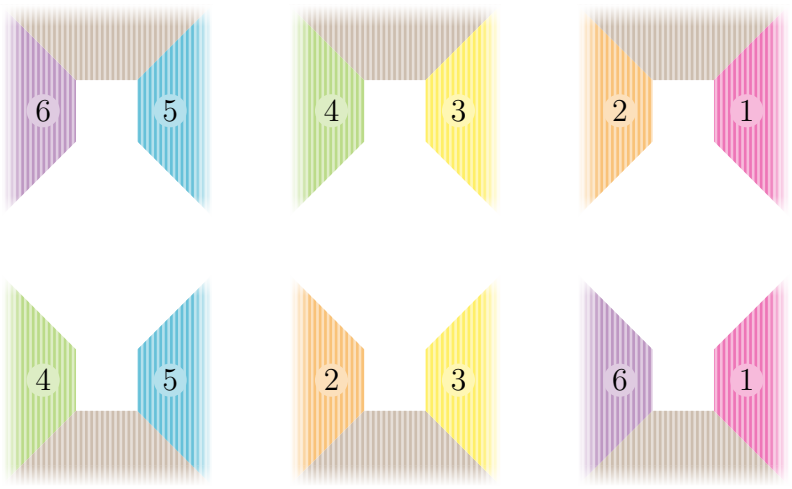
A first-order pole in ω makes a puncture shaped like a half-infinite cylinder. You can build one by rolling up a half-infinite rectangular strip and gluing its sides together:



The translation structure of the cylinder you end up with is determined by two parameters: the width of the strip and its orientation in the plane. In most orientations, the vertical leaves spiral up or down the cylinder. When the strip is horizontal, the vertical leaves close up into circles.

F.3 Higher-order punctures

The puncture created by a higher-order pole in ω can be glued together out of planes with quadrants cut away, like this:



The notches at the centers of the pieces fit together into a polygonal hole that the rest of the surface can be connected to. For the pieces to fit together, the notches all have to be the same size, which can be adjusted to accommodate the part of the surface the puncture is supposed to hook up to.

Unlike first-order punctures, which come in a $\mathbb{C}^\times$-worth of shapes parameterized by the residues at their poles, higher-order punctures depend only on the orders of their poles.

F.4 Counting ends

As I remarked earlier, a puncture is an end of the surface it lives in. Intuitively, you can think of it as a point on the boundary at infinity of the surface. The notion of an end is purely topological, however, and you might wonder if there's a notion of boundary at infinity that takes the translation structure into account. Here's one proposal, which is suitable at least for

translation surfaces with punctures.

Let's say two vertical rays on a translation surface are *translation-equivalent* if they're connected by a continuous family of vertical rays.[1] A *vertical end* is an equivalence class of vertical rays. A generic first-order puncture, whose vertical leaves spiral up or down rather than closing up into circles, is a single vertical end. A higher-order puncture, on the other hand, comprises several vertical ends, one for each building block.

[1] To be formal about it, define a vertical ray to be a local isometry of $[0, \infty)$ into a vertical leaf. We can then say a continuous family of vertical rays is a continuous map from $[0, 1] \times [0, \infty)$ into the surface which restricts to a vertical ray when the first argument is fixed.

Bibliography

[1] L. Hollands and A. Neitzke, "Spectral networks and Fenchel-Nielsen coordinates," arXiv:1312.2979 [math.GT].

[2] F. Klein, *Vergleichende Betrachtungen über neuere geometrische Forsuchungen.* Andreas Deichert, Erlangen, 1872.

[3] C. Ehresmann, "Sur les espaces localement homogènes," *L'Enseignement Mathématique* **35** (1936) .

[4] Élie Cartan, *Sur la structure des groupes de transformations finis et continus.* PhD thesis, L'École normale supérieure, 1894.

[5] A. Haefliger, "Homotopy and integrability," in *Manifolds—Amsterdam 1970*, N. H. Kuiper, ed., pp. 133 – 163. Springer, 1971.

[6] W. P. Thurston, *The Geometry and Topology of Three-Manifolds.* electronic version 1.1 ed., 2002.

[7] W. Goldman, "Geometric structures and varieties of representations," in *Geometry of Group Representations*, W. Goldman and A. R. Magid, eds., vol. 74 of *Contemporary Mathematics*, pp. 169 – 198. 1988.

[8] H. P. de Saint-Gervais, *Uniformisation des surfaces de Riemann: Retour sur un théorème centenaire.* ENS Éditions, 2010.

[9] A. Zorich, "Flat surfaces," in *Frontiers in Number Theory, Physics, and Geometry*, P. M. P. Cartier, B. Julia and P. Vanhove., eds., vol. I.

[10] S. Kobayashi, *Transformation Groups in Differential Geometry*. Springer, 1972.

[11] D. Gaiotto, G. Moore, and A. Neitzke, "Spectral networks," *Annales Henri Poincaré* **14** no. 7, (2013) 1643 – 1731.

[12] C. Kottke, "Bundles, classifying spaces, and characteristic classes." http://www.northeastern.edu/ckottke/docs/bundles.pdf.

[13] W. P. Thurston, "Zippers and univalent functions," in *The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof*, A. Baernstein, D. Drasin, P. Duren, and A. Marden, eds., no. 21 in Mathematical Surveys and Monographs, pp. 185 – 197. 1986.

[14] H. Schwerdtfeger, *Geometry of Complex Numbers*. Dover, 1962.

[15] A. Fenyes, "Abelianization of $\mathrm{SL}_2\,\mathbb{R}$ local systems," arXiv:1510.05757v1 [math.GT].

[16] S. Fomin and D. Thurston, "Cluster algebras and triangulated surfaces. Part ii: Lambda lengths," arXiv:1210.5569 [math.GT].

[17] W. P. Thurston, "Minimal stretch maps between hyperbolic surfaces," arXiv:9801039 [math.GT].

[18] W. P. Thurston, "Earthquakes in two-dimensional hyperbolic geometry," in *Low Dimensional Topology and Kleinian Groups*, D. Epstein, ed., no. 112 in London Mathematical Society Lecture Note Series.

[19] F. Bonahon, "Geodesic laminations on surfaces," in *Laminations and Foliations in Dynamics, Geometry and Topology*, vol. 269 of *Contemporary Mathematics*.

[20] F. Bonahon, "Shearing hyperbolic surfaces, bending pleated surfaces, and Thurston's symplectic form," *Annales de la Faculté des sciences de Toulouse: Mathématiques* **5** no. 2, (1996) 233 – 297.

[21] D. G. Allegretti, "Notes on quantum Teichmüller theory,".

[22] G. Théret, "Convexity of length functions and thurston's shear coordinates," arXiv:1408.5771 [math.GT].

[23] D. Gaiotto, G. Moore, and A. Neitzke, "Wall-crossing, Hitchin systems, and the WKB approximation," *Advances in Mathematics* **234** (2013) 239 – 403.

[24] A. Casson and S. Bleiler, *Automorphisms of Surfaces after Nielsen and Thurston*. No. 9 in London Mathematical Society Student Texts. 1988.

[25] S. Dowdall, "Anosov-representations: Basic definitions and properties." http://lukyanenko.net/conferences/htt2013/, 2013. 2013 Workshop on Higher Teichmüller-Thurston Theory.

[26] A. Furman, "On the multiplicative ergodic theorem for uniquely ergodic systems," *Annales de l'Institut Henri Poincaré (B)* **33** no. 6, (1997) 797 – 815.

[27] M. Viana, *Lectures on Lyapunov Exponents*, vol. 145 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 2014.

[28] D. Lenz, "Existence of non-uniform cocycles on uniquely ergodic systems," *Annales de l'Institut Henri Poincaré (B)* **40** no. 2, (2004) 197 – 206.

[29] The Stacks Project Authors, "*Stacks Project*." http://stacks.math.columbia.edu, 2013.

[30] H. Masur and S. Tabachnikov, "Rational billiards and flat structures," in *Handbook of Dynamical Systems*, B. Hasselblatt and A. Katok, eds., vol. 1A, ch. 13. 2002.

[31] J. Smillie, "The dynamics of billiard flows in rational polygons of dynamical systems," in *Dynamical Systems, Ergodic Theory and Applications*, Y. Sinai, ed., ch. 11. 2000.

[32] R. Gjerde and Ø. Johansen, "Bratteli-Vershik models for Cantor minimal systems associated to interval exchange transformations," *Mathematica Scandinavica* **90** no. 1, (2002) 87 – 100.

[33] G. Gruenhage, "MH 7750: Set theoretic topology." http://www.auburn.edu/~gruengf/fall14.html, 2014.

[34] M. Viana, "Lyapunov exponents of Teichmüller flows," in *Partially Hyperbolic Dynamics, Laminations, and Teichmüller Flow*, G. Forni, M. Lyubich, C. Pugh, and M. Shub, eds., vol. 51 of *Fields Institute Communications*. 2007.

[35] A. Zorich, "How do the leaves of a closed 1-form wind around a surface," in *Pseudoperiodic Topology*, V. Arnold, M. Kontsevich, and A. Zorich, eds., vol. 197 of *Translations of the AMS, Series 2*. 1999.

[36] M. Lin, "On the uniform ergodic theorem," *Proceedings of the American Mathematical Society* **43** no. 2, (1974) 337 – 340.

[37] M. Brin and G. Stuck, *Introduction to Dynamical Systems*. Cambridge University Press, 2002.

[38] L. Marchese, "Khinchin theorem for interval exchange transformations," arXiv:1003.5883 [math.DS].

[39] L. Marchese, "Khinchin type condition for translation surfaces and asymptotic laws for the Teichmuller flow," arXiv:1003.5887 [math.DS].

[40] D. Dumas and M. Wolf, "Projective structures, grafting, and measured laminations," arXiv:0712.0968 [math.DG].

[41] G. Dreyer, "Thurston's cataclysms for Anosov representations," arXiv:1301.6961 [math.GT].

[42] F. Bonahon and G. Dreyer, "Hitchin characters and geodesic laminations," arXiv:1410.0729v2 [math.GT].

[43] V. Araujo, A. Bufetov, and S. Filip, "Hölder-continuity of Oseledets subspaces for the Kontsevich-Zorich cocycle," arXiv:1409.8167v2 [math].

[44] M. Mulase, "Geometry of character varieties of surface groups," *Research Institute for Mathematical Sciences Kokyuroku* (2008) , arXiv:0710.5263 [math.AG].

[45] J. Martínez and V. Muñoz, "E-polynomials of SL(2, $\mathbb{C}$)-character varieties of complex curves of genus 3," arXiv:1405.7120v2 [math.AG].

[46] A. Fenyes, "Potentially cluster-like coordinates from dense spectral networks." https://www.ma.utexas.edu/users/afenyes/writing.html. Poster presented at Positive Grassmannians: Applications to integrable systems and super Yang-Mills scattering amplitudes. July 2015, CRM.

[47] S. Welstead, "Infinite products in a banach algebra," *Journal of Mathematical Analysis and Applications* **105** (1985) 523 – 532.

[48] K. D. Joshi, *Introduction to General Topology*. New Age International, 1983.

[49] E. Schechter, *Handbook of Analysis and its Foundations*. Academic Press, 1997.

[50] G. Peschke, "The theory of ends," *Nieuw Archief voor Wiskunde* **8** (1990) 1 – 12.

[51] V. Fock and A. Goncharov, "Moduli spaces of local systems and higher Teichmüller theory," arXiv:math/0311149 [math.AG].

Vita

Aaron Joshua Fenyes was one of the first third-graders to attend Birmingham Covington School, and also one of the first master's students to attend Perimeter Scholars International. He won't bother listing his high school, but he will note that a letter jacket from his high school appears in *Ferris Bueller's Day Off*. He went to college, and a lot of Sari Brown shows, at the University of Michigan. In 2010, he called the graduate admissions center at the University of Texas at Austin and found out that he was probably too far down the wait list to get in. He got in the next day. As long as this paper is formatted correctly, he should be eligible to start work as a postdoctoral researcher at the University of Toronto later this year.

Permanent address: afenyes@math.utexas.edu

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.