Dynamics of Interval Exchange Transformations and Teichmüller Flows

Marcelo Viana

July 6, 2008

Contents

Chapter 0

Overview

This a working version of notes for a course I taught at IMPA in 2005 and 2007. Chapters 0, 1, 2, 3, 4, 7, and Appendix A are close to final form. Chapters 5 and 6 are under substantial revision. Chapter 8 is not yet available.

The topics covered in this book lie at the interface of several mathematical areas, from complex analysis and topology to number theory and geometry, differential or algebraic. The main questions can often be formulated in terms of underlying dynamical systems, such as Teichmüller flows and certain renormalization operators, and ideas from dynamics and ergodic theory have indeed been most effective in providing very complete answers. That point of view permeates the whole text.

The unifying thread is the study of the geodesic flows on translation surfaces. By definition, a *translation surface* is equipped with an atlas whose coordinate changes are all translations of the plane, outside a finite number of conical singularities. One way to construct such a surface is by taking a planar polygon with an even number of edges, distributed in pairs such that edges in the same pair are parallel and have the same length, and identifying (by translation) the two edges in each of the pairs. The singularities arise from the vertices of the polygon. See Figure 1.

A translation surface comes with a flat Riemannian metric, transported from the plane through local charts, and the geodesic flow is meant with respect to this metric. In addition, a translation surface admits a non-vanishing parallel vector field, for instance the one corresponding to the vertical vector field $(0,1)$ in local coordinates. For example, if the surface is described by a planar polygon as in Figure 1, the geodesic flow corresponds to displacement with constant speed along straight lines, taking into account the identifications of the edges. Throughout, the geodesics keep a constant angle to the vertical direction.

Yet another equivalent way to describe the translation structure is through

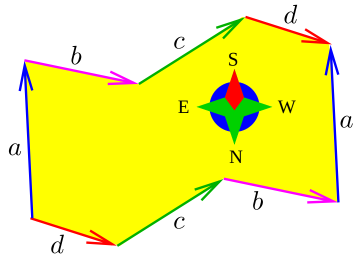


Figure 1:

the *Abelian differential* (or complex 1-form) α_z obtained by transporting the standard 1-form dz from the complex plane to surface through local charts. The geodesics $\gamma(t)$ are the curves characterized by

$$\alpha_{\gamma(t)}(\dot{\gamma}(t)) = \text{const}$$

and the argument of this constant determines the angle to the vertical vector field. The conical singularities of the Riemannian metric correspond to the zeros of the Abelian differential. In particular, their "horizon" angles are integer multiples $2\pi(m_i+1)$ of a full turn, where m_i is the multiplicity of the corresponding zero.

While the local behavior of geodesics on a translation surface is quite simple, the presence of singularities renders the global behavior very rich. A first result is that the geodesic flow in a fixed direction is *minimal*, meaning that all geodesics are dense on the surface, for all but countably many exceptional directions. This is easily deduced from a classical theorem of Maier [38], as follows. The family of geodesics in a fixed direction is a special case of a *measured foliation*, that is, a foliation whose leaves are tangent to the kernel of some closed real 1-form on the surface. Integrating the 1-form over cross-sections to the foliation one obtains a transverse arc-length measure which is invariant under all holonomy maps, and that is the reason for the denomination "measured foliation".

Maier's theorem describes the global structure of any measured foliation. In the case when there are no *saddle-connection*, that is, no leafs connecting two singularities, it implies that all leaves are dense in the surface. Hence, it suffices to observe that on any translation surface there are only countably many saddle-connections. A kind of converse is also true, by results of Calabi [9], Katok [24], and Hubbard, Masur [22]: any measured foliation without saddle-connections can be realized as the family of vertical geodesics in some translation surface.

The behavior of a measured foliation, including the special case of geodesics in a fixed direction on a translation surface, may be analyzed through its return map to some convenient cross-section. Let the cross-section be parametrized according to the arc-length measure induced on it by the closed 1-form. The return map is piecewise smooth and preserves this arc-length. Thus, it is an *interval exchange transformation*: there is a finite partition of the domain into

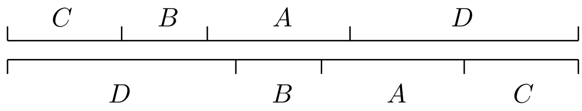


Figure 2:

subintervals such that the return map is a translation restricted to each element of that partition. See Figure 2, where we mean that each subinterval on the upper part of the figure is mapped to the subinterval with the same label on the lower part of the figure, by translation. Conversely, every interval exchange transformation may be realized as the first return map to some cross-section of the vertical geodesic flow on some translation surface (even a whole family of surfaces).

Keane [26] proved that, for all but a countable set of choices of the subintervals, the interval exchange transformation is *minimal*: every orbit is dense in the whole domain. From this one can recover the minimality of the geodesic flow discussed previously. He also conjectured a much stronger property, for Lebesgue almost all choices of the subintervals, namely that the transformation is *uniquely ergodic*: the multiples of Lebesgue measure are the only invariant measures.

The proof of the Keane conjecture, by Masur [41] and Veech [54], was a major development in this field. As a direct consequence, the vertical geodesic flow of almost every Abelian differential on a surface is uniquely ergodic. Here 'almost every' is with respect to a natural volume measure in the space of Abelian differentials, that we shall discuss in a while. Sometime afterwards, Kerckhoff, Masur, Smillie [30] refined the conclusion: for *every* Abelian differential and almost every direction, the geodesic flow in that direction is uniquely ergodic.

Recently, Avila, Forni [3] proved that every transitive interval exchange transformation f which is not a rotation is *weak mixing*, meaning that the constant functions are the only measurable eigenfunctions of the operator $\varphi \mapsto \varphi \circ f$, thus answering an old question of Veech [55]. They also proved that, on surfaces of genus $g > 1$, the geodesic flow of almost every Abelian differential in almost every direction is weak mixing. Weak *topological* mixing, where one considers continuous eigenfunctions only, had been proved by Nogueira, Rudolph [45], for almost all interval exchange transformations which are not rotations. An older result of Katok [24] asserts that interval exchange transformations are never mixing.

The Masur-Veech proof of the unique ergodicity conjecture was also the first important manifestation of a fruitful general principle: *properties of individual translation surfaces are often encoded in their orbits under the Teichmüller flow acting in the moduli space of all translation surfaces.* Let us explain this.

Let $\mathcal{A}_g$ denote the moduli space of Abelian differentials, that is the space of conformal equivalence classes of Abelian differentials, on Riemann surfaces of genus $g \geq 1$. $\mathcal{A}_g$ is a complex algebraic variety of dimension $d = 4g - 3$. It is naturally stratified into the subsets $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ of Abelian differentials whose zeros have multiplicities $m_1, \ldots, m_\kappa$. Here $\kappa \geq 0$ is the number of zeros, and the multiplicities must satisfy the compatibility relation

$$m_1 + \cdots + m_\kappa = 2g - 2.$$

Each stratum $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ is an algebraic subvariety of the moduli space, of dimension $2g + \kappa - 1$. It admits a natural holomorphic affine structure, that is, an atlas whose coordinate changes are holomorphic affine maps, as well as a natural volume measure induced by this affine structure.

The *Teichmüller flow* $\mathcal{T}^t$, $t \in \mathbb{R}$ is the action of the diagonal group

$$A^t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \quad t \in \mathbb{R}$$

on $\mathcal{A}_g$, by post-composition. That is, $\mathcal{T}^t$ acts on each Abelian differential by

$$\mathcal{T}^t(\alpha) = e^t \Re(\alpha) + i e^{-t} \Im(\alpha).$$

In geometric terms, given a translation surface represented by a polygon P as in Figure 1, its trajectory under the Teichmüller flow consists of the translation surfaces represented by the polygons $A^t(P)$, $t \in \mathbb{R}$. The flow $\mathcal{T}^t$ preserves the area of the surface and commutes with every dilation $\alpha \mapsto c\alpha$, $c > 0$. Hence, not much is lost by restricting the Teichmüller flow to the hypersurface of Abelian differentials with unit area. It is also clear that $\mathcal{T}^t$ preserves each stratum, since the number and multiplicities of the zeros remain the same along the orbit.

Masur [41] and Veech [54] proved that, for every stratum, the natural volume on the this hypersurface is finite, and it is invariant under the Teichmüller flow. In addition, and most importantly, its ergodic components coincide with the connected components of the stratum. This implies that, for almost every translation surface, its orbit under the flow returns infinitely often to a fixed compact region. They also observed that this property of *recurrence of the orbit implies unique ergodicity of the vertical geodesic flow on the surface,* and this is how they established the Keane conjecture.

Somewhat surprisingly, not all strata are connected. VeechVe90 showed that the connected components of the strata are in one-to-one correspondence with combinatorial objects called *extended Rauzy classes*. Then he exhibited two distinct classes corresponding to the same stratum, thus proving it has at least two connected components. Arnoux went further, exhibiting a stratum with at least three connected components. Recently, Kontsevich, Zorich [33] introduced two new invariants, hyperellipticity and spin parity, and showed that they suffice to catalog all connected components of strata of Abelian differentials. Lanneau [35, 34] carried out a corresponding classification of the connected components for strata of quadratic differentials.

Asymptotic flag phenomenon. Unique ergodicity implies that the vertical flow has a well-defined asymptotic cycle (Schwartzman [50]) in the first homology space of the surface . That is, there exists $c_1 \in H_1(M, \mathbb{R})$ such that

$$\frac{1}{l}[\gamma(p,l)] \to c_1 \quad \text{uniformly in } p \in M,$$

where $\gamma(p,l)$ is the vertical segment of length l starting from the point p in the upward direction, and $[\gamma(p,l)]$ represents the homology class of the closed curve obtained concatenating $\gamma(p,l)$ with some curve segment of bounded length that joins its endpoints. See Figure 3.

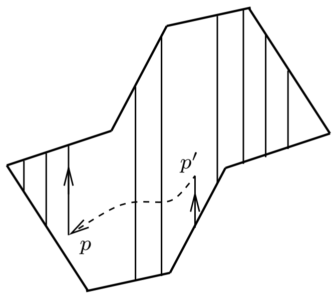


Figure 3:

This fact was observed by Zorich who, in fact, discovered that the asymptotic behavior in homology of long vertical geodesic segments admits a much more precise, and rather surprising description: numerical calculations suggested that there exist real numbers $1 > \nu_2 > \cdots > \nu_g > 0$ and linearly independent homology classes $c_1, \ldots, c_g$ such that dist $\big([\gamma(p,l)], \mathbb{R}c_1 \oplus \cdots \oplus \mathbb{R}c_g\big)$ is uniformly bounded and

$$\operatorname{dist}\big([\gamma(p,l)], \mathbb{R}c_1 \oplus \cdots \oplus \mathbb{R}c_j\big) \lesssim l^{\nu_{j+1}} \quad \text{for every } j = 1, \ldots, g-1, \tag{1}$$

meaning ν_{j+1} is the smallest exponent ν such that the left hand side is less than l^ν for every large l. Zorich [65] proved that this is indeed so, for almost all translation surfaces, conditioned to a conjecture on the Lyapunov spectrum of the Teichmüller flow that we shall discuss in a while.

Lyapunov exponents. Volume induces natural measures on the space of interval exchange transformations, which are absolutely continuous with respect to Lebesgue measure and are invariant under the Rauzy-Veech renormalization operator. These invariant measures are always infinite, but Zorich [63] explained how the renormalization operator can be "accelerated" so that the new operator admits an absolutely continuous invariant *probability*. This probability μ is even ergodic. The Zorich transformation may be seen as a higher-dimensional version of the classical continued fraction algorithm, with μ in the role of the classical Gauss measure.

Ergodicity ensures that the Teichmüller flow has well-defined Lyapunov exponents in each connected component of every stratum $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ of Abelian differentials. The Lyapunov spectrum has the form

$$2 \geq 1 + \nu_2 \geq \cdots \geq 1 + \nu_g \geq 1 = \cdots = 1 \geq 1 - \nu_g \geq \cdots \geq 1 - \nu_2 \geq 0 \geq$$
$$-1 + \nu_2 \geq \cdots \geq -1 + \nu_g \geq -1 = \cdots = -1 \geq -1 - \nu_g \geq \cdots \geq -1 - \nu_2 \geq -2$$

for some $\nu_2, \ldots, \nu_g$. Zorich and Kontsevich conjectured that all these inequalities are actually strict, and proved that the asymptotic flag phenomenon we described previously would follow from this conjecture.

Even before, Veech [56, 58] had shown that the Teichmüller flow is non-uniformly hyperbolic, which amounts to saying that $\nu_2 < 1$. Forni [14] proved the much deeper fact that $\nu_g > 0$ which, in particular, implies the genus 2 case. The full statement of the Zorich-Kontsevich conjecture $1 > \nu_2 > \cdots > \nu_g > 0$ was proved, even more recently, by Avila, Viana [5, 4].

Chapter 1

Interval Exchange Maps

In this chapter we initiate the study of interval exchange transformations. Such transformations arise naturally as Poincaré return maps of measured foliations and geodesic flows on translation surfaces. But they are also great examples of simple dynamical systems with very rich dynamics of parabolic type and, as such, they have been extensively studied for their own sake.

Firstly, we introduce the Rauzy-Veech induction operator, which assigns to each interval exchange transformation its first return map to a convenient subinterval. In terms of the geodesic flow, this corresponds to taking the Poincaré return map to a smaller cross-section. The Keane condition defines the largest set where the operator may be iterated for all times. Moreover, interval exchange transformations that satisfy the Keane condition are minimal.

In the early eighties, Masur and Veech proved that almost every interval exchange transformation is uniquely ergodic. The proof will be discussed in detail in Chapter 4. The starting point is the Rauzy-Veech renormalization operator, defined by composing the induction operator with a rescaling of the domain. The crucial step is to show that the renormalization operator admits a natural invariant measure which is ergodic. Interval exchange transformations that are typical relative to this measure are uniquely ergodic.

The Masur-Veech invariant measure is infinite. Zorich explained how the renormalization operator may be modified so that the new operator admits a natural invariant probability. This probability has an important role in Chapter 7. Moreover, the Zorich renormalization operator may be seen as a high-dimensional version of the usual continued fraction expansion.

1.1 Definitions

Let $I \subset \mathbb{R}$ be an interval[1] and $\{I_\alpha : \alpha \in \mathcal{A}\}$ be a partition of I into subintervals, indexed by some alphabet $\mathcal{A}$ with $d \geq 2$ symbols. An *interval exchange map* is

[1] All intervals will be bounded, closed on the left and open on the right. For notational simplicity, we take the left endpoint of I to coincide with 0.

a bijective map from I to I which is a translation on each subinterval I_α. Such a map f is determined by combinatorial and metric data as follows:

1. A pair $\pi = (\pi_0, \pi_1)$ of bijections $\pi_\varepsilon : \mathcal{A} \to \{1, \ldots, d\}$ describing the ordering of the subintervals I_α before and after the map is iterated. This will be represented as

$$\pi = \begin{pmatrix} \alpha_1^0 & \alpha_2^0 & \ldots & \alpha_d^0 \\ \alpha_1^1 & \alpha_2^1 & \ldots & \alpha_d^1 \end{pmatrix}$$

where $\alpha_j^\varepsilon = \pi_\varepsilon^{-1}(j)$ for $\varepsilon \in \{0, 1\}$ and $j \in \{1, 2, \ldots, d\}$.

2. A vector $\lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}$ with positive entries, where λ_α is the length of the subinterval I_α.

We call $p = \pi_1 \circ \pi_0^{-1} : \{1, \ldots, d\} \to \{1, \ldots, d\}$ the *monodromy invariant* of the pair $\pi = (\pi_0, \pi_1)$. Observe that our notation, that we borrow from Marmi, Moussa, Yoccoz [40], is somewhat redundant. Given any (π, λ) as above and any bijection $\phi : \mathcal{A}' \to \mathcal{A}$, we may define

$$\pi'_\varepsilon = \pi_\varepsilon \circ \phi, \quad \varepsilon \in \{0, 1\} \quad \text{and} \quad \lambda'_{\alpha'} = \lambda_{\phi(\alpha')}, \quad \alpha' \in \mathcal{A}'.$$

Then (π, λ) and (π', λ') have the same monodromy invariant and they define the same interval exchange transformation. This means one can always normalize the combinatorial data by choosing $\mathcal{A} = \{1, 2, \ldots, d\}$ and $\pi_0 = \mathrm{id}$, in which case π_1 coincides with the monodromy invariant p. However, this notation hides the symmetric roles of π_0 and π_1, and is not invariant under the induction and renormalization algorithms that we are going to present. On the contrary, the present notation $\pi = (\pi_0, \pi_1)$ allows for a very elegant formulation of these algorithms, as we are going to see.

Example 1.1. The interval exchange transformation described by Figure 1.1 corresponds to the pair $\pi = \begin{pmatrix} C & B & A & D \\ D & B & A & C \end{pmatrix}$. The monodromy invariant is equal to $p = (4, 2, 3, 1)$.

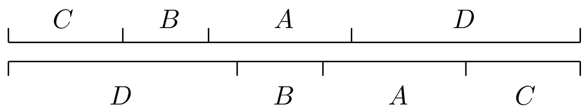


Figure 1.1:

Example 1.2. For $d = 2$ there is essentially only one combinatorics, namely

$$\pi = \begin{pmatrix} A & B \\ B & A \end{pmatrix}.$$

The interval exchange transformation associated to (π, λ) is given by

$$f(x) = \begin{cases} x + \lambda_B & \text{if } x \in I_A \\ x - \lambda_A & \text{if } x \in I_B. \end{cases}$$

Identifying I with the circle $\mathbb{R}/(\lambda_A + \lambda_B)\mathbb{Z}$, we get

$$f(x) = x + \lambda_B \qquad \mod (\lambda_A + \lambda_B)\mathbb{Z}. \tag{1.1}$$

That is, the transformation corresponds to the rotation of angle $\lambda_B/(\lambda_A + \lambda_B)$.

Example 1.3. The data (π, λ) is *not* uniquely determined by f. Indeed, let

$$\pi = \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}.$$

Given any λ, the interval exchange transformation f defined is

$$f(x) = \begin{cases} x + \lambda_B + \lambda_C & \text{for } x \in I_A \\ x - \lambda_A & \text{for } x \in I_B \cup I_C\,. \end{cases}$$

This shows that f is also the interval exchange transformation defined by either of the following data:

- (π, λ') for any other λ' such that $\lambda'_A = \lambda_A$ and $\lambda'_B + \lambda'_C = \lambda_B + \lambda_C$
- $(\tilde{\pi}, \tilde{\lambda})$ with $\tilde{\pi} = \begin{pmatrix} A & D \\ D & A \end{pmatrix}$ and $\lambda''_A = \lambda_A$ and $\lambda''_D = \lambda_B + \lambda_C$.

Translation vectors. Given $\pi = (\pi_0, \pi_1)$, define $\Omega_\pi : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ by

$$\Omega_\pi(\lambda) = w \quad \text{with} \quad w_\alpha = \sum_{\pi_1(\beta) < \pi_1(\alpha)} \lambda_\beta - \sum_{\pi_0(\beta) < \pi_0(\alpha)} \lambda_\beta. \tag{1.2}$$

Then the corresponding interval exchange transformation f is given by

$$f(x) = x + w_\alpha, \quad \text{for } x \in I_\alpha.$$

We call w the *translation vector* of f. Notice that the matrix [2] $(\Omega_{\alpha,\beta})_{\alpha,\beta \in \mathcal{A}}$ of Ω_π is given by

$$\Omega_{\alpha,\beta} = \begin{cases} +1 & \text{if } \pi_1(\alpha) > \pi_1(\beta) \text{ and } \pi_0(\alpha) < \pi_0(\beta) \\ -1 & \text{if } \pi_1(\alpha) < \pi_1(\beta) \text{ and } \pi_0(\alpha) > \pi_0(\beta) \\ 0 & \text{in all other cases.} \end{cases} \tag{1.3}$$

Example 1.4. In the case of Figure 1.1,

$$(w_A, w_B, w_C, w_D) = (\lambda_D - \lambda_C, \lambda_D - \lambda_C, \lambda_D + \lambda_B + \lambda_A, -\lambda_C - \lambda_B - \lambda_A).$$

[2] Except where otherwise stated, all matrices are with respect to the canonical basis of $\mathbb{R}^{\mathcal{A}}$.

The image of Ω_π is the 2-dimensional subspace

$$\{w \in \mathbb{R}^{\mathcal{A}} : w_A = w_B = w_C + w_D\}.$$

On the other hand, for $\pi = \begin{pmatrix} A & B & C & D \\ D & C & B & A \end{pmatrix}$ we have

$$(w_A, w_B, w_C, w_D) = (\lambda_D + \lambda_C + \lambda_B, \lambda_D + \lambda_C - \lambda_A, \lambda_D - \lambda_B - \lambda_A, -\lambda_C - \lambda_B - \lambda_A)$$

and Ω_π is a bijection from $\mathbb{R}^{\mathcal{A}}$ to itself.

Lemma 1.5. *We have $\lambda \cdot w = 0$.*

Proof. This is an immediate consequence of the fact that Ω_π is anti-symmetric. A detailed calculation follows. By definition

$$\lambda \cdot w = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha w_\alpha = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha \left(\sum_{\pi_1(\beta) < \pi_1(\alpha)} \lambda_\beta - \sum_{\pi_0(\beta) < \pi_0(\alpha)} \lambda_\beta \right)$$

and this is equal to

$$\sum_{\pi_1(\beta) < \pi_1(\alpha)} \lambda_\alpha \lambda_\beta - \sum_{\pi_0(\beta) < \pi_0(\alpha)} \lambda_\alpha \lambda_\beta = \frac{1}{2} \sum_{\alpha \neq \beta} \lambda_\alpha \lambda_\beta - \frac{1}{2} \sum_{\alpha \neq \beta} \lambda_\alpha \lambda_\beta = 0.$$

This proves the statement. □

The *canonical involution* is the operation in the space of (π, λ) corresponding to interchanging the roles of π_0 and π_1 while leaving λ unchanged. Clearly, under this operation the monodromy invariant p and the transformation f are replaced by their inverses. Moreover, Ω_π is replaced by $-\Omega_\pi$, and so the translation vector is also replaced by its symmetric.

1.2 Rauzy-Veech induction

Let (π, λ) represent an interval exchange transformation. For each $\varepsilon \in \{0, 1\}$, denote by $\alpha(\varepsilon)$ the last symbol in the expression of π_ε, that is

$$\alpha(\varepsilon) = \pi_\varepsilon^{-1}(d) = \alpha_d^\varepsilon$$

Let us assume the intervals $I_{\alpha(0)}$ and $I_{\alpha(1)}$ have different lengths. Then we say that (π, λ) has *type* 0 if $\lambda_{\alpha(0)} > \lambda_{\alpha(1)}$ and *type* 1 if $\lambda_{\alpha(0)} < \lambda_{\alpha(1)}$. In either case, the largest of the two intervals is called the *winner* and the shortest one is called the *loser* of (π, λ). Let J be the subinterval of I obtained by removing the loser, that is, the shortest of these two intervals:

$$J = \begin{cases} I \setminus f(I_{\alpha(1)}) & \text{if } (\pi, \lambda) \text{ has type 0} \\ I \setminus I_{\alpha(0)} & \text{if } (\pi, \lambda) \text{ has type 1.} \end{cases}$$

The *Rauzy-Veech induction* of f is the first return map $\hat{R}(f)$ to the subinterval J. This is again an interval exchange transformation, as we are going to explain.

If (π, λ) has type 0, take $J_\alpha = I_\alpha$ for $\alpha \neq \alpha(0)$ and $J_{\alpha(0)} = I_{\alpha(0)} \setminus f(I_{\alpha(1)})$. These intervals form a partition of J. Note that $f(J_\alpha) \subset J$ for every $\alpha \neq \alpha(1)$. This means that $\hat{R}(f) = f$ restricted these J_α. On the other hand,

$$f(J_{\alpha(1)}) = f(I_{\alpha(1)}) \subset I_{\alpha(0)}$$

and so,

$$f^2(J_{\alpha(1)}) \subset f(I_{\alpha(0)}) \subset J.$$

Consequently, $\hat{R}(f) = f^2$ restricted to $J_{\alpha(1)}$. See Figure 1.2.

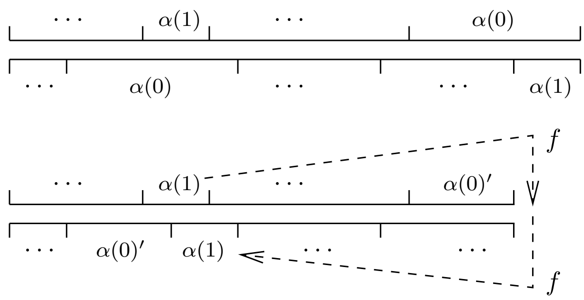


Figure 1.2:

If (π, λ) has type 1, define $J_{\alpha(0)} = f^{-1}(I_{\alpha(0)})$ and $J_{\alpha(1)} = I_{\alpha(1)} \setminus J_{\alpha(0)}$, and $J_\alpha = I_\alpha$ for all other values of α. See Figure 1.3. Then $f(J_\alpha) \subset J$ for every $\alpha \neq \alpha(0)$, and so $\hat{R}(f) = f$ restricted these J_α. On the other hand,

$$f^2(J_{\alpha(0)}) = f(I_{\alpha(0)}) \subset J,$$

and so $\hat{R}(f) = f^2$ restricted to $J_{\alpha(0)}$.

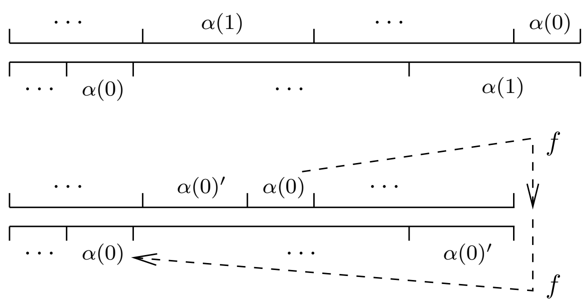


Figure 1.3:

The induction map $\hat{R}(f)$ is not defined when the two rightmost intervals $I_{\alpha(0)}$ and $I_{\alpha(1)}$ have the same length. We shall return to this point in Sections 1.3 and 1.5.

Remark 1.6. Suppose the n'th iterate $\hat{R}^n(f)$ is defined, for some $n \geq 1$, and let I^n be its domain. It follows from the definition of the induction algorithm that $\hat{R}^n(f)$ is the first return map of f to I^n. Similarly, $\hat{R}^n(f)^{-1} = \hat{R}^n(f^{-1})$ is the first return map of f^{-1} to I^n.

Let us express the map $f \mapsto \hat{R}(f)$ in terms of the coordinates (π, λ) in the space of interval exchange transformations. It follows from the previous description that if (π, λ) has type 0 then the transformation $\hat{R}(f)$ is described by (π', λ'), where

- $\pi' = \begin{pmatrix} \pi'_0 \\ \pi'_1 \end{pmatrix} = \begin{pmatrix} \alpha^0_1 & \cdots & \alpha^0_{k-1} & \alpha^0_k & \alpha^0_{k+1} & \cdots & \cdots & \alpha(0) \\ \alpha^1_1 & \cdots & \alpha^1_{k-1} & \alpha(0) & \alpha(1) & \alpha^1_{k+1} & \cdots & \alpha^1_{d-1} \end{pmatrix}.$

  or, in other words,
$$\alpha^{0'}_j = \alpha^0_j \quad \text{and} \quad \alpha^{1'}_j = \begin{cases} \alpha^1_j & \text{if } j \leq k \\ \alpha(1) & \text{if } j = k+1 \\ \alpha^1_{j-1} & \text{if } j > k+1, \end{cases} \tag{1.4}$$
  where $k \in \{1, \ldots, d-1\}$ is defined by $\alpha^1_k = \alpha(0)$.

- $\lambda' = (\lambda'_\alpha)_{\alpha \in \mathcal{A}}$ where
$$\lambda'_\alpha = \lambda_\alpha \quad \text{for } \alpha \neq \alpha(0), \quad \text{and} \quad \lambda'_{\alpha(0)} = \lambda_{\alpha(0)} - \lambda_{\alpha(1)}. \tag{1.5}$$

Analogously, if (π, λ) has type 1 then $\hat{R}(f)$ is described by (π', λ'), where

- $\pi' = \begin{pmatrix} \pi'_0 \\ \pi'_1 \end{pmatrix} = \begin{pmatrix} \alpha^0_1 & \cdots & \alpha^0_{k-1} & \alpha(1) & \alpha(0) & \alpha^0_{k+1} & \cdots & \alpha^0_{d-1} \\ \alpha^1_1 & \cdots & \alpha^1_{k-1} & \alpha^1_k & \alpha^1_{k+1} & \cdots & \cdots & \alpha(1) \end{pmatrix}.$

  or, in other words,
$$\alpha^{0'}_j = \begin{cases} \alpha^0_j & \text{if } j \leq k \\ \alpha(0) & \text{if } j = k+1 \\ \alpha^0_{j-1} & \text{if } j > k+1 \end{cases} \quad \text{and} \quad \alpha^{1'}_j = \alpha^1_j, \tag{1.6}$$
  where $k \in \{1, \ldots, d-1\}$ is defined by $\alpha^0_k = \alpha(1)$.

- $\lambda' = (\lambda'_\alpha)_{\alpha \in \mathcal{A}}$ where
$$\lambda'_\alpha = \lambda_\alpha \quad \text{for } \alpha \neq \alpha(1), \quad \text{and} \quad \lambda'_{\alpha(1)} = \lambda_{\alpha(1)} - \lambda_{\alpha(0)}. \tag{1.7}$$

Example 1.7. If $\pi = \begin{pmatrix} B & C & A & E & D \\ A & E & B & D & C \end{pmatrix}$ and $\lambda_D < \lambda_C$ (type 1 case) then
$$\pi' = \begin{pmatrix} B & C & D & A & E \\ A & E & B & D & C \end{pmatrix}$$
and $\lambda' = (\lambda_A, \lambda_B, \lambda_C - \lambda_D, \lambda_D, \lambda_E)$.

Operator Θ. Let us also compare the translation vectors w and w' of f and $\hat{R}(f)$, respectively. From Figure 1.2 we see that, if (π, λ) has type 0,

$$w'_\alpha = w_\alpha \text{ for } \alpha \neq \alpha(1), \quad \text{and} \quad w'_{\alpha(1)} = w_{\alpha(1)} + w_{\alpha(0)}.$$

Analogously, if (π, λ) has type 1,

$$w'_\alpha = w_\alpha \text{ for } \alpha \neq \alpha(0), \quad \text{and} \quad w'_{\alpha(0)} = w_{\alpha(0)} + w_{\alpha(1)}.$$

This may be expressed as

$$w' = \Theta(w) \tag{1.8}$$

where $\Theta = \Theta_{\pi,\lambda} : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ is the linear operator whose matrix $(\Theta_{\alpha,\beta})_{\alpha,\beta \in \mathcal{A}}$ is given by

$$\Theta_{\alpha,\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ 1 & \text{if } \alpha = \alpha(1) \text{ and } \beta = \alpha(0) \\ 0 & \text{in all other cases.} \end{cases} \tag{1.9}$$

if (π, λ) has type 0, and

$$\Theta_{\alpha,\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ 1 & \text{if } \alpha = \alpha(0) \text{ and } \beta = \alpha(1) \\ 0 & \text{in all other cases.} \end{cases} \tag{1.10}$$

if (π, λ) has type 1. Notice that Θ depends only on π and the type ε.

Observe that Θ is invertible and its inverse is given by

$$\Theta^{-1}_{\alpha,\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ -1 & \text{if } \alpha = \alpha(1) \text{ and } \beta = \alpha(0) \\ 0 & \text{in all other cases} \end{cases}$$

when (π, λ) has type 0, and

$$\Theta^{-1}_{\alpha,\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ -1 & \text{if } \alpha = \alpha(0) \text{ and } \beta = \alpha(1) \\ 0 & \text{in all other cases} \end{cases}$$

when (π, λ) has type 1. So, the relations (1.5) and (1.7) may be rewritten as

$$\lambda' = \Theta^{-1*}(\lambda) \qquad \text{or} \qquad \lambda = \Theta^*(\lambda') \tag{1.11}$$

where Θ^* denotes the adjoint operator of Θ, that is, the operator whose matrix is transposed of that of Θ.

Remark 1.8. The canonical involution does not affect the operator Θ: if $\tilde{\pi}$ is obtained by interchanging the lines of π, then $\Theta_{\tilde{\pi},\lambda} = \Theta_{\pi,\lambda}$. Notice that $(\tilde{\pi}, \lambda)$ and (π, λ) have opposite types.

1.3 Keane condition

Summarizing the previous section, the Rauzy-Veech induction is expressed by the transformation

$$\hat{R} : \hat{R}(\pi, \lambda) = (\pi', \lambda')$$

where π' is given by (1.4) and (1.6), and λ' is given by (1.5) and (1.7). Recall that $\hat{R}$ is not defined when the two rightmost intervals have the same length, that is, when $\lambda_{\alpha(0)} = \lambda_{\alpha(1)}$. We want to consider $\hat{R}$ as a dynamical system in the space of interval exchange transformations, but for this we must restrict the map to an invariant subset of (π, λ) such that the iterates $\hat{R}^n(\pi, \lambda)$ are defined for all $n \geq 1$.

Let us start with the following observation. We say that a pair $\pi = (\pi_0, \pi_1)$ is *reducible* if there exists $k \in \{1, \ldots, d-1\}$ such that

$$\pi_1 \circ \pi_0^{-1}(\{1, \ldots, k\}) = \{1, \ldots, k\}. \tag{1.12}$$

Then, for any choice of λ, the subinterval

$$J = \bigcup_{\pi_0(\alpha) \leq k} I_\alpha = \bigcup_{\pi_1(\alpha) \leq k} I_\alpha$$

is invariant under the transformation f, and so is its complement. This means that f splits into two interval exchange transformations, with simpler combinatorics. Moreover, $(\pi', \lambda') = \hat{R}(\pi, \lambda)$ is also reducible, with the same invariant subintervals. In what follows, we always restrict ourselves to irreducible data.

A natural possibility is to restrict the induction algorithm to the subset of *rationally independent* vectors $\lambda \in \mathbb{R}_+^{\mathcal{A}}$, that is, such that

$$\sum_{\alpha \in \mathcal{A}} n_\alpha \lambda_\alpha \neq 0 \quad \text{for all nonzero integer vectors } (n_\alpha)_{\alpha \in \mathcal{A}} \in \mathbb{Z}^{\mathcal{A}}. \tag{1.13}$$

It is clear that this condition is invariant under iteration of (1.5) and (1.7), and that it ensures that all iterates $\hat{R}^n(\pi, \lambda)$ are defined. Observe also that the set of rationally independent vectors has full Lebesgue measure in the cone $\mathbb{R}_+^{\mathcal{A}}$.

However, it was observed by Keane [26, 27] that rational independence is a bit too strong: depending on the combinatorial data, failure of (1.13) for certain integer vectors may not be an obstruction to further iteration of $\hat{R}$. Let ∂I_γ be the left endpoint of each subinterval I_γ. Recall that we take the left endpoint of I to coincide with the origin. Then

$$\partial I_\gamma = \sum_{\pi_0(\eta) < \pi_0(\gamma)} \lambda_\eta$$

represents the left endpoint of each subinterval I_γ. A pair (π, λ) satisfies the *Keane condition* if the orbits of these endpoints are as disjoint as they can possible be [3]:

$$f^m(\partial I_\alpha) \neq \partial I_\beta \quad \text{for all } m \geq 1 \text{ and } \alpha, \beta \in \mathcal{A} \text{ with } \pi_0(\beta) \neq 1. \tag{1.14}$$

[3] It is clear that if $\pi_0(\beta) = 1$ then $f(\partial I_\alpha) = \partial I_\beta$ for $\alpha = \pi_1^{-1}(1)$.

This ensures that π is irreducible and $(\pi', \lambda') = \hat{R}(\pi, \lambda)$ is well-defined. Moreover, property (1.14) is invariant under iteration of $\hat{R}$, because $\hat{R}(f)$-orbits are contained in f-orbits. Thus, the Keane condition is sufficient for all iterates $(\pi^n, \lambda^n) = \hat{R}^n(\pi, \lambda)$, $n \geq 0$ to be defined. We shall see in Corollary 1.22 that it is also *necessary*.

Remark 1.9. The Keane condition (1.14) is not affected if one restricts to the case $\pi_1(\alpha) > 1$. Indeed, suppose one has $f^m(\partial I_\alpha) = \partial I_\beta > 0$ with $\pi_1(\alpha) = 1$ and $m > 1$. Then $f(\partial I_\alpha) = 0 = \partial I_\gamma$ for some $\gamma \in \mathcal{A}$. Then, $f^{m-1}(\partial I_\gamma) = \partial I_\beta$. Moreover, $\pi_1(\gamma) > 1$ because π is irreducible and $\pi_0(\gamma) = 1$.

The next result shows that, assuming irreducibility, the Keane condition is indeed more general than rational independence. In particular, it also corresponds to full Lebesgue measure.

Proposition 1.10. *If λ is rationally independent and π is irreducible then (π, λ) satisfies the Keane condition.*

Proof. Assume there exist $m \geq 1$ and $\alpha, \beta \in \mathcal{A}$ such that $f^m(\partial I_\alpha) = \partial I_\beta$ and $\pi_0(\beta) > 1$. Define β_j, $0 \leq j \leq m$, by

$$f^j(\partial I_\alpha) \in I_{\beta_j}.$$

Notice that $\beta_0 = \alpha$ and $\beta_m = \beta$. Then

$$\partial I_\beta - \partial I_\alpha = \sum_{0 \leq j < m} w_{\beta_j}$$

where $w = (w_\gamma)_{\gamma \in \mathcal{A}}$ is the translation vector defined in (1.2). Equivalently,

$$\sum_{\pi_0(\gamma) < \pi_0(\beta_m)} \lambda_\gamma - \sum_{\pi_0(\gamma) < \pi_0(\beta_0)} \lambda_\gamma = \sum_{0 \leq j < m} \Big(\sum_{\pi_1(\gamma) < \pi_1(\beta_j)} \lambda_\gamma - \sum_{\pi_0(\gamma) < \pi_0(\beta_j)} \lambda_\gamma \Big).$$

This may be rewritten as $\sum_{\gamma \in \mathcal{A}} n_\gamma \lambda_\gamma = 0$, where

$$n_\gamma = \#\{0 \leq j < m : \pi_1(\beta_j) > \pi_1(\gamma)\} - \#\{0 < j \leq m : \pi_0(\beta_j) > \pi_0(\gamma)\}.$$

Since we assume rational independence, we must have $n_\gamma = 0$ for all $\gamma \in \mathcal{A}$. Now let D be the maximum of $\pi_0(\beta_j)$ over all $0 < j \leq m$ and $\pi_1(\beta_j)$ over all $0 \leq j < m$. Note that $D \geq \pi_0(\beta) > 1$. So, since we assume that π is irreducible, there exists $\gamma \in \mathcal{A}$ such that $\pi_0(\gamma) < D \leq \pi_1(\gamma)$. The last inequality implies that $\pi_1(\beta_j) \leq \pi_1(\gamma)$ for all $0 \leq j < m$. Since $n_\gamma = 0$, this implies that $\pi_0(\beta_j) \leq \pi_0(\gamma) < D$ for all $0 < j \leq m$. A symmetric argument shows that $\pi_1(\beta_j) < D$ for all $0 \leq j < m$. This contradicts the definition of D. This contradiction proves that (π, λ) satisfies the Keane condition, as stated. □

Example 1.11. Suppose $d = 2$. By (1.1), the interval exchange transformation is given by $f(x) = x + \lambda_B \mod (\lambda_A + \lambda_B)\mathbb{Z}$. So, the Keane condition means that, given any $m \geq 1$ and $n \in \mathbb{Z}$, both

$$m\lambda_B \neq \lambda_A + n(\lambda_A + \lambda_B) \quad \text{and} \quad \lambda_A + m\lambda_B \neq \lambda_A + n(\lambda_A + \lambda_B).$$

It is clear that this holds if and only if (λ_A, λ_B) is rationally independent.

Example 1.12. Starting from $d = 3$, the Keane condition may be strictly weaker than rational independence. Consider, for instance, $\pi = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}$. Then $f(x) = x + \lambda_C \mod (\lambda_A + \lambda_B + \lambda_C)\mathbb{Z}$ and the Keane condition means that

$$m\lambda_C \quad \text{and} \quad \lambda_A + m\lambda_C \quad \text{and} \quad \lambda_A + \lambda_B + m\lambda_C$$

are different from $\lambda_A + n(\lambda_A + \lambda_B + \lambda_C)$ and $\lambda_A + \lambda_B + n(\lambda_A + \lambda_B + \lambda_C)$, for all $m \geq 1$ and $n \in \mathbb{Z}$. This may be restated in a more compact form, as follows: given any $p \in \mathbb{Z}$ and $q \in \mathbb{Z}$,

$$p\lambda_C \neq q(\lambda_A + \lambda_B + \lambda_C) \quad \text{and} \quad p\lambda_C \neq \lambda_A + q(\lambda_A + \lambda_B + \lambda_C).$$

Clearly, this may hold even if (λ_A, λ_B) is rationally dependent.

1.4 Minimality

A transformation is called *minimal* if every orbit is dense in the whole domain of definition or, equivalently, the domain is the only nonempty closed invariant set.

Proposition 1.13. *If (π, λ) satisfies the Keane condition then f is minimal.*

For the proof, we begin by noting that the first return map of f to some interval $J \subset I_\alpha$ is again an interval exchange transformation:

Lemma 1.14. *Given any subinterval $J = [a, b)$ of some I_α, there exists a partition $\{J_j : 1 \leq j \leq k\}$ of J and integers $n_1, \ldots, n_k \geq 1$, where $k \leq d + 2$, such that*

1. *$f^i(J_j) \cap J = \emptyset$ for all $0 < i < n_j$ and $1 \leq j \leq k$;*
2. *each $f^{n_j} \mid J_j$ is a translation from J_j to some subinterval of J;*
3. *those subintervals $f^{n_j}(J_j)$, $1 \leq j \leq k$ are pairwise disjoint.*

Proof. Let A be the union of the boundary $\{a, b\}$ of J with the set of endpoints of all the intervals I_γ, $\gamma \in \mathcal{A}$, the endpoints of I excluded. Note that $\#A \leq d+1$. Let $B \subset J$ be the set of points $z \in J$ for which there exists some $m \geq 1$ such that $f^i(z) \notin J$ for all $0 < i < m$ and $f^m(z) \in A$. The map $B \ni z \mapsto f^m(z) \in A$ is injective, because f is injective and there are no iterates in J prior to time m. Consequently, $\#B \leq \#A$. Consider the partition of J determined by the points of B. This partition has at most $d + 2$ elements. By the Poincaré recurrence theorem, for each element $J_j = [a_j, b_j)$ there exists $n_j \geq 1$ such that $f^{n_j}(J_j)$ intersects J. Take n_j smallest. From the definition of B it follows that the restriction $f^{n_j} \mid J_j$ is a translation and its image is contained in J. Finally, the $f^{n_j}(J_j)$, $1 \leq j \leq k$ are pairwise disjoint because f is injective and the n_j are the first return times to J. $\square$

In fact, the statement is true for any interval $J \subset I$. See [53, § 3].

Corollary 1.15. *Under the assumptions of Lemma 1.14, the union $\hat{J}$ of all forward iterates of J is a finite union of intervals and a fully invariant set: $f(\hat{J}) = \hat{J}$.*

Proof. The first claim follows directly from the first part of Lemma 1.14:

$$\hat{J} = \bigcup_{n=0}^{\infty} f^n(J) = \bigcup_{j=1}^{k} \bigcup_{i=0}^{n_j-1} f^i(J_j).$$

Moreover, parts 2 and 3 of Lemma 1.14, together with the observation

$$\sum_{j=1}^{k} |f^{n_j}(J_j)| = \sum_{j=1}^{k} |J_j| = |J|$$

(we use $|\cdot|$ to represent length), give that J coincides with $\cup_{j=1}^{k} f^{n_j}(J_j)$. This implies that $\hat{J}$ is fully invariant. □

Lemma 1.16. *If (π, λ) satisfies the Keane condition then f has no periodic points.*

Proof. Suppose there exists $m \geq 1$ and $x \in I$ such that $f^m(x) = x$. Define β_j, $0 \leq j \leq m$ by the condition $f^j(x) \in I_{\beta_j}$. Let J be the set of all points $y \in I$ such that $f^j(y) \in I_{\beta_j}$ for all $0 \leq j < m$. Then J is an interval and f^m restricted to it is a translation. Since $f^m(x) = x$, we actually have $f^m \mid J = \mathrm{id}$. In particular, $f^m(\partial J) = \partial J$. The definition of J implies that there are $1 \leq k \leq m$ and $\beta \in \mathcal{A}$ such that $f^k(\partial J) = \partial I_\beta$. Then $f^m(\partial I_\beta) = \partial I_\beta$. If $\pi_0(\beta) > 1$, this contradicts the Keane condition. If $\pi_0(\beta) = 1$ then there exists $\alpha \in \mathcal{A}$ such that $f(\partial I_\alpha) = 0 = \partial I_\beta$. Note that $\alpha \neq \beta$, and so $\partial I_\alpha > 0$, because π is irreducible. Hence, $f^m(\partial I_\alpha) = \partial I_\alpha$ contradicts the Keane condition. These contradictions prove that there is no such periodic point x. □

Proof of Proposition 1.13. Suppose there exists $x \in I$ such that $\{f^n(x) : n \geq 0\}$ is not dense in I. Then we may choose a subinterval $J = [a, b)$ of some I_α that avoids the closure of the orbit. Let $\hat{J}$ be the union of all forward iterates of J. By Corollary 1.15, this is a finite union of intervals, fully invariant under f. We claim that $\hat{J}$ can not be of the form $[0, \hat{b})$. The proof is by contradiction. Let $\mathcal{B}$ be the subset of $\alpha \in \mathcal{A}$ such that I_α is contained in $\hat{J}$. Then $\pi_0(\mathcal{B}) = \{1, \ldots, k\}$ for some k. Since $\hat{J}$ is invariant, we also have $\pi_1(\mathcal{B}) = \{1, \ldots, k\}$. Hence,

$$\pi_0^{-1}(\{1, \ldots, k\}) = \mathcal{B} = \pi_1^{-1}(\{1, \ldots, k\}). \tag{1.15}$$

It is clear that $k < d$, because $\hat{J}$ avoids the closure of the orbit of x, and so it can not be the whole I. If $k = 0$ then $\hat{J}$ would be contained in I_α, where $\pi_0(\alpha) = 1$; by invariance, it would also be contained in $f(I_\alpha)$, implying that $\pi_1(\alpha) = 1$; this would contradict irreducibility (which is a consequence of the Keane condition). Thus, k must be positive. Then (1.15) contradicts irreducibility, and this contradiction proves our claim.

As a consequence, there exists some connected component $[\hat{a}, \hat{b})$ of $\hat{J}$ with $\hat{a} > 0$. If $f^n(\hat{a}) \neq \partial I_\beta$ for every $n \geq 0$ and $\beta \in \mathcal{A}$, then (by continuity of f and invariance of $\hat{J}$) every $f^n(\hat{a})$, $n \geq 0$ would be on the boundary of some connected component of $\hat{J}$. As there are finitely many components, f would have a periodic point, which is forbidden by Lemma 1.16. Similarly, if $f^n(\hat{a}) \neq f(\partial I_\alpha)$ for every $n \leq 0$ and $\alpha \in \mathcal{A}$, then every $f^n(\hat{a})$, $n \leq 0$ would be on the boundary of some connected component of $\hat{J}$. Just as before, this would imply the existence of some periodic point , which is forbidden by Lemma 1.16. This proves that there are $n_1 \leq 0 \leq n_2$ and $\alpha, \beta \in \mathcal{A}$ such that

$$f^{n_1}(\hat{a}) = f(\partial I_\alpha) \quad \text{and} \quad f^{n_2}(\hat{a}) = \partial I_\beta. \tag{1.16}$$

If $\partial I_\beta > 0$, this contradicts the Keane condition (take $m = n_2 - n_1 + 1$). If $\partial I_\beta = 0$ then $n_2 > 0$, because we have taken $\hat{a} > 0$. Moreover, $\partial I_\beta = f(\partial I_\gamma)$, where $\pi_1(\gamma) = 1$. This means that (1.16) remains valid if one replaces β by γ and n_2 by $n_2 - 1$. As $\gamma \neq \beta$, by irreducibility, we have $\partial I_\gamma > 0$ and this leads to a contradiction just as in the previous case. $\square$

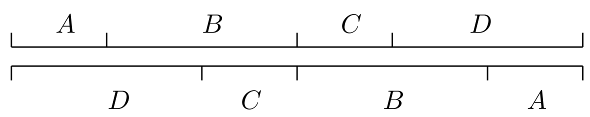


Figure 1.4:

Remark 1.17. The Keane condition is not necessary for minimality. Consider the interval exchange transformation f illustrated in Figure 1.4, where $\lambda_A = \lambda_C$, $\lambda_B = \lambda_D$, and $\lambda_A/\lambda_B = \lambda_C/\lambda_D$ is irrational. Then f does not satisfy the Keane condition, yet it is minimal.

Unique ergodicity. A transformation is called *uniquely ergodic* if it admits exactly one invariant probability (which is necessarily ergodic). See Mañé [39]. Then the transformation is minimal restricted to the support of this probability. Observe that interval exchange transformations always preserve the Lebesgue measure. Thus, in this context, unique ergodicity means that every invariant measure is a multiple of the Lebesgue measure.

Keane [26] conjectured that every minimal interval exchange transformation is uniquely ergodic, and checked that this is true for $d = 2, 3$. However, Keynes, Newton [31] gave an example with $d = 5$ and two ergodic invariant probabilities. In turn, they conjectured that rational independence should suffice for unique ergodicity. Again, a counterexample was given by Keane [27], with $d = 4$ and two ergodic invariant probabilities. He then went on to make the following

Conjecture 1.18. Almost every interval exchange transformation is uniquely ergodic.

This statement was proved by Masur [41] and Veech [54], independently, in the early eighties. The proof will be one of the main topics of Chapter 4. That unique ergodicity holds for a (Baire) residual subset had been proved by Keane, Rauzy [28].

1.5 Dynamics of the induction map

This section contains a number of useful facts on the dynamics of the induction algorithm in the space of interval exchange transformations. The presentation follows Section 4.3 of Yoccoz [61].

Let (π, λ) be such that the iterates $(\pi^n, \lambda^n) = \hat{R}^n(\pi, \lambda)$ are defined for all $n \geq 0$. For instance, this is the case if (π, λ) satisfies the Keane condition. For each $n \geq 0$, let $\varepsilon^n \in \{0, 1\}$ be the type and $\alpha^n, \beta^n \in \mathcal{A}$ be, respectively, the winner and the loser of (π^n, λ^n). In other words, α^n and β^n are the two rightmost symbols in the two lines of π^n, with $\lambda_{\alpha^n} > \lambda_{\beta^n}$. In yet another equivalent formulation, $\pi_{\varepsilon^n}(\alpha^n) = d = \pi_{1-\varepsilon^n}(\beta^n)$.

It is clear that the sequence $(\varepsilon^n)_n$ takes both values 0 and 1 infinitely many times. Indeed, suppose the type ε^n was eventually constant. Then α^n would also be eventually constant, and so would λ_α^n for all $\alpha \neq \alpha^n$. On the other hand,

$$\lambda_{\alpha^{n+1}}^{n+1} = \lambda_{\alpha^n}^{n+1} = \lambda_{\alpha^n}^n - \lambda_{\beta^n}^n$$

for all large n. Since the $\lambda_{\beta^n}^n$ are bounded from zero, the $\lambda_{\alpha^n}^n$ would be eventually negative, which is a contradiction.

Proposition 1.19. *Both sequences $(\alpha^n)_n$ and $(\beta^n)_n$ take every value $\alpha \in \mathcal{A}$ infinitely many times.*

Proof. Given any symbol $\alpha \in \mathcal{A}$, consider any maximal time interval $[p, q)$ such that $\alpha^n = \alpha$ for every $n \in [p, q)$. At the end of this interval the type must change:

$$\varepsilon^q = 1 - \varepsilon^{q-1} \quad \text{and} \quad \pi_{1-\varepsilon^q}^q(\alpha) = d.$$

In other words, $\alpha = \beta^q$. This shows that we only have to prove the statement for the sequence $(\alpha^n)_n$.

Let $\mathcal{B}$ be the subset of symbols $\beta \in \mathcal{A}$ that occur only finitely many times in the sequence $(\alpha^n)_n$. Up to replacing (π, λ) by some iterate, we may suppose that those symbols do not occur at all in $(\alpha_n)_n$. Then $\lambda_\beta^n = \lambda_\beta$ for all $\beta \in \mathcal{B}$ and $n \geq 0$. Since

$$\lambda_{\alpha^{n+1}}^{n+1} = \lambda_{\alpha^n}^n - \lambda_{\beta^n}^n,$$

this implies that every $\beta \in \mathcal{B}$ occurs only finitely many times in the sequence (β_n). Once more, up to replacing the initial point by an iterate, we may suppose they do not occur at all in (β_n). It follows that, for every $\beta \in \mathcal{B}$, the sequences

$$\pi_0^n(\beta) \quad \text{and} \quad \pi_1^n(\beta), \quad n \geq 0,$$

are non-decreasing. So, replacing (π, λ) by an iterate one more time, if necessary, we may suppose that these sequences are constant. We claim that

$$\pi_\varepsilon(\beta) < \pi_\varepsilon(\alpha) \quad \text{for every } \alpha \in \mathcal{A} \setminus \mathcal{B},\ \beta \in \mathcal{B}, \text{ and } \varepsilon = 0, 1. \tag{1.17}$$

Indeed, suppose there were α, β, and ε such that $\pi_\varepsilon(\alpha) < \pi_\varepsilon(\beta)$. Then, since the sequence $\pi_\varepsilon^n(\beta)$ in non-decreasing, so must be the sequence $\pi_\varepsilon^n(\alpha)$. In particular, $\pi_\varepsilon^n(\alpha) < d$ for all $n \geq 0$. Now, since $\alpha \notin \mathcal{B}$, this implies that $\pi_{1-\varepsilon}^n(\alpha) = d$ and $\varepsilon^n = 1 - \varepsilon$, for some value of n.

$$\begin{pmatrix} \cdots & \alpha & \cdots & \beta & \cdots & \gamma \\ \cdots & \cdots & \cdots & \cdots & \cdots & \alpha \end{pmatrix} \xrightarrow{\hat{R}} \begin{pmatrix} \cdots & \alpha & \gamma & \cdots & \beta & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \alpha \end{pmatrix}$$

Then $\pi_\varepsilon^{n+1}(\beta) = \pi_\varepsilon^n(\beta) + 1$, contradicting the previous conclusion that $\pi_\varepsilon^n(\beta)$ is constant. This contradiction proves our claim. Finally, (1.17) implies that

$$\pi_0(\mathcal{B}) = \{1, \ldots, k\} = \pi_1(\mathcal{B})$$

for some $k < d$. Since π is assumed to be irreducible, we must have $k = 0$, that is, $\mathcal{B}$ is the empty set. This proves the statement for the sequence $(\alpha^n)_n$ and, hence, completes the proof of the proposition. ◻

Corollary 1.20. *The length of the domain I^n of the transformation $\hat{R}^n(f)$ goes to zero when n goes to ∞.*

Proof. Since the sequences λ_α^n are non-increasing, for all $\alpha \in \mathcal{A}$, it suffices to show that they all converge to zero. Suppose there was $\beta \in \mathcal{A}$ and $c > 0$ such that $\lambda_\beta^n \geq c$ for every $n \geq 0$. For any value of n such that $\beta^n = \beta$, we have

$$\lambda_{\alpha^n}^{n+1} = \lambda_{\alpha^n}^n - \lambda_{\beta^n}^n \leq \lambda_{\alpha^n}^n - c.$$

By Proposition 1.19, this occurs infinitely many times. As the alphabet $\mathcal{A}$ is finite, it follows that there exists some $\alpha \in \mathcal{A}$ such that

$$\lambda_\alpha^{n+1} \leq \lambda_\alpha^n - c.$$

for infinitely many values of n. This contradicts the fact that $\lambda_\alpha^n > 0$. ◻

Corollary 1.21. *For each $m \geq 0$ there exists $n \geq 1$ such that*

$$\Theta_{\pi^m,\lambda^m}^{*n} > 0 \quad \textit{(all the entries of the matrix are positive).}$$

Proof. Given $\alpha, \beta \in \mathcal{A}$, $m \geq 0$, $n \geq 1$, we represent by $\Theta^*(\alpha, \beta, m, n)$ the entry on row α and column β of the matrix of $\Theta_{\pi^m,\lambda^m}^{*n}$. By definition (1.9)-(1.10),

$$\Theta^*(\alpha, \beta, m, 1) = 1 \text{ if either } \alpha = \beta \text{ or } (\alpha, \beta) = (\alpha^m, \beta^m), \tag{1.18}$$

and $\Theta^*(\alpha,\beta,m,1)=0$ in all other cases. Observe also that every $\Theta^*(\alpha,\beta,m,n)$ is non-decreasing on n:

$$\begin{aligned}\Theta^*(\alpha,\beta,m,n+1) &= \sum_{\gamma} \Theta^*(\alpha,\gamma,m,n)\Theta^*(\gamma,\beta,m+n,1)\\ &\geq \Theta^*(\alpha,\beta,m,n)\Theta^*(\beta,\beta,m+n,1) \geq \Theta^*(\alpha,\beta,m,n).\end{aligned} \tag{1.19}$$

Let α be fixed. We are going to construct an enumeration $\gamma_1, \gamma_2, \ldots, \gamma_d$ of $\mathcal{A}$ and integers $n_1, n_2, \ldots, n_d$ such that

$$\Theta^*(\alpha,\gamma_i,m,n) > 0 \quad \text{for every } n > n_i \text{ and } i = 1,2,\ldots,d. \tag{1.20}$$

It is clear that this implies the corollary, as β must be one of the γ_i.

For $i=1$ just take $\gamma_1=\alpha$ and $n_1=0$. The relations (1.18) and (1.19) immediately imply (1.20). Next, use Proposition 1.19 to find $m_2>m$ such that the winner α^{m_2} coincides with γ_1. Let $\gamma_2=\beta^{m_2}$ be the loser. Note that $\gamma_2\neq\gamma_1$, by irreducibility. Moreover, (1.18) gives that $\Theta^*(\gamma_1,\gamma_2,m_2,1)=1$, and this implies $\Theta^*(\gamma_1,\gamma_2,m,n)>0$ for every $n>m_2-m$. This gives (1.20) for $i=2$, with $n_2=m_2-m$. If $d=2$ then there is nothing left to prove, so assume $d>2$. Using Proposition 1.19 twice, one finds $p_2>m_2$ such that the winner α^{p_2} is neither γ_1 nor γ_2, and $m_3>p_2$ such that the winner $\alpha^{m_3}=\gamma_j$ for either $j=1$ or $j=2$. Consider the smallest such m_3, and let $\gamma_3=\beta^{m_3}$ be the loser. Notice that $\gamma_3=\alpha^{m_3-1}$ and so it is neither γ_1 nor γ_2. Moreover, (1.18) gives that $\Theta^*(\gamma_j,\gamma_3,m_3,1)=1$ and this implies

$$\Theta^*(\gamma_1,\gamma_3,m,n) \geq \Theta^*(\gamma_1,\gamma_j,m,m_3-m)\Theta^*(\gamma_j,\gamma_3,m_3,n-m_3+m) > 0$$

for $n>m_3-m$. Notice that $m_3-m>m_2-m=n_2$. This proves (1.20) for $i=3$ with $n_3=m_3-m$.

The general step of the enumeration is analogous. Assume we have constructed $\gamma_1,\ldots,\gamma_k\in\mathcal{A}$, all distinct, and integers $n_1,n_2,\ldots,n_k$ such that (1.20) holds for $1\leq i\leq k$. Assuming $k<d$, we may use Proposition 1.19 twice to find $p_k>m_k$ such that the winner α^{n_k} is not an element of $\{\gamma_1,\ldots,\gamma_k\}$ and $m_{k+1}>p_k$ such that the winner $\alpha^{m_{k+1}}=\gamma_j$ for some $j\in\{1,\ldots,k\}$. Choose the smallest such m_{k+1} and let $\gamma_{k+1}=\beta^{m_{k+1}}$ be the loser. Then $\gamma_{k+1}=\alpha^{m_{k+1}-1}$ and so it is not an element of $\{\gamma_1,\ldots,\gamma_k\}$. The relation (1.18) gives $\Theta^*(\gamma_j,\gamma_{k+1},m_{k+1},1)=1$, and then

$$\Theta^*(\gamma_1,\gamma_{k+1},m,n) \geq \Theta^*(\gamma_1,\gamma_j,m,m_{k+1}-m)\Theta^*(\gamma_j,\gamma_{k+1},m_{k+1},n-m_{k+1}+m)$$

is strictly positive for all $n>n_{k+1}=m_{k+1}-m$. This completes our recurrence construction and, thus, finishes the proof of the corollary. □

At this point we can prove that (π,λ) can be iterated indefinitely (if and) only if it satisfies the Keane condition:

Corollary 1.22. *If $(\pi^n,\lambda^n)=\hat{R}^n(\pi,\lambda)$ is defined for all $n\geq 0$ then (π,λ) satisfies the Keane condition.*

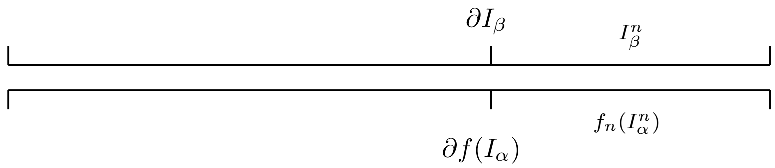


Figure 1.5:

Proof. Suppose that, for some $\alpha, \beta \in \mathcal{A}$, and $m \geq 1$,

$$f^{m-1}(\partial f(I_\alpha)) = \partial I_\beta. \tag{1.21}$$

Choose m minimum. In particular, by Remark 1.9, we have $\partial f(I_\alpha) > 0$. The definition of $f_n = \hat{R}^n(f)$ gives

$$\partial f(I_\alpha) = \partial f_n(I^n_\alpha), \quad \text{and} \quad \partial I_\beta = \partial I^n_\beta$$

for every n such that $\partial f(I_\alpha)$ and ∂I_β are in the domain I^n of f_n. Take n maximum such that both points are in I^n (Corollary 1.20). Since f_n is the first return map of f to I^n (Remark 1.6), the hypothesis (1.21) implies that

$$f^k_n(\partial f(I_\alpha)) = \partial I_\beta \quad \text{for some } k \leq m-1. \tag{1.22}$$

Moreover, either I_β or $f_n(I^n_\alpha)$ (or both) is a rightmost partition interval for f_n.

If $\partial f(I_\alpha) = \partial I_\beta$ then $f_n(I^n_\alpha) = I^n_\beta$, that is, the two rightmost intervals of f_n have the same length. See Figure 1.5. Hence, $f_{n+1} = \hat{R}^{n+1}(f)$ is not defined, which contradicts the hypothesis. This proves the statement in this case.

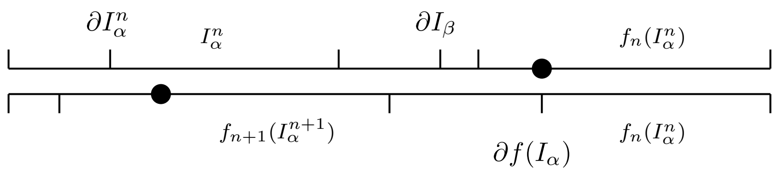


Figure 1.6:

Now suppose f_n has type 0, that is, $\partial I_\beta < \partial f(I_\alpha)$. By definition,

$$f_{n+1}(\partial I^{n+1}_\alpha) = f^2_n(\partial I^n_\alpha) = f_n(\partial f(I_\alpha)) \quad \text{and} \quad \partial I^{n+1}_\beta = \partial I^n_\beta = \partial I_\beta.$$

See Figure 1.6. Comparing with (1.22) we get

$$f^{k-1}_n(\partial f_{n+1}(I^{n+1}_\alpha)) = f^k_n(\partial I_\alpha) = \partial I_\beta = \partial I^{n+1}_\beta.$$

Since both points are in I^{n+1} and f_{n+1} is the return map of f_n to I^{n+1}, this may be rewritten as

$$f^{l-1}_{n+1}(\partial f_{n+1}(I^{n+1}_\alpha)) = \partial I^{n+1}_\beta \quad \text{for some } l \leq k < m. \tag{1.23}$$

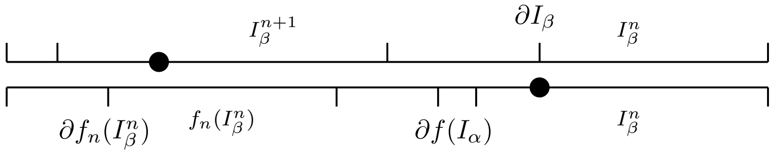


Figure 1.7:

Now suppose f_n has type 1, that is, $\partial I_\beta > \partial f(I_\alpha)$. By definition,

$$\partial f_{n+1}(I_\alpha^{n+1}) = \partial f_n(I_\alpha^n) = \partial f(I_\alpha) \quad \text{and} \quad \partial I_\beta^{n+1} = f_n^{-1}(\partial I_\beta^n) = f_n^{-1}(\partial I_\beta).$$

See Figure 1.7. Comparing with (1.22) we get

$$f_n^{k-1}(\partial f_{n+1}(I_\alpha^{n+1})) = f_n^{k-1}(\partial f(I_\alpha)) = f_n^{-1}(\partial I_\beta) = \partial I_\beta^{n+1}.$$

Since f_{n+1} is the return map of f_n to I^{n+1}, this may be rewritten as

$$f_{n+1}^{l-1}(\partial I_\alpha^{n+1}) = \partial I_\beta^{n+1} \quad \text{for some } l \leq k < m. \tag{1.24}$$

In both subcases, we have shown that (1.21) implies a similar relation, either (1.23) or (1.24), where f is replaced by some induced map f_{n+1}, and $m \geq 2$ is replaced by a *smaller* l. Iterating this procedure, we must eventually reach the case $m = 1$, which was treated previously. □

1.6 Rauzy classes

Given pairs π and π', we say that π' is a *successor* of π if there exist $\lambda, \lambda' \in \mathbb{R}_+^{\mathcal{A}}$ such that $\hat{R}(\pi, \lambda) = (\pi', \lambda')$. Any pair π has exactly two successors, corresponding to types 0 and 1. Similarly, each π' is the successor of exactly two pairs π, obtained by reversing the relations (1.4) and (1.6). Notice that π is irreducible if and only if π' is irreducible. Thus, this relation defines a partial order in the set of irreducible pairs, which we may represent as a directed graph G. We call *Rauzy classes* the connected components of this graph.

Lemma 1.23. *If π and π' are in the same Rauzy class then there exists an oriented path in G starting at π and ending at π'.*

Proof. Let $A(\pi)$ be the set of all pairs π' that can be attained through an oriented path starting at π. As we have just seen, each vertex of the graph G has exactly two outgoing and two incoming edges. By definition, every edge starting from a vertex of $A(\pi)$ must end at some vertex of $A(\pi)$. By a counting argument, it follows that every edge ending at a vertex of $A(\pi)$ starts at some vertex of $A(\pi)$. This means that $A(\pi)$ is a connected component of G, and so it coincides with the whole Rauzy class $C(\pi)$. □

A result of Kontsevich, Zorich [33] that we shall review in Chapter 6 yields a complete classification of the Rauzy classes. Here, let us calculate all Rauzy classes for the first few values of d. The results are summarized in the table at the end of this section.

For $d = 2$ there are two possibilities for the monodromy invariant, but only one is irreducible: $(2, 1)$. The Rauzy graph reduces to

$$\begin{tikzcd} \begin{pmatrix} A & B \\ B & A \end{pmatrix} \arrow[loop left, "0"] \arrow[loop right, "1"] \end{tikzcd}$$

For $d = 3$ there are six possibilities for the monodromy invariant, but only three are irreducible: $(2, 3, 1)$, $(3, 1, 2)$, $(3, 2, 1)$. They are all represented in the Rauzy class

$$\begin{tikzcd} \begin{pmatrix} A & C & B \\ C & B & A \end{pmatrix} \arrow[loop left, "0"] \arrow[r, shift left, "1"] & \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix} \arrow[l, shift left] \arrow[r, shift left, "0"] & \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix} \arrow[l, shift left] \arrow[loop right, "1"] \end{tikzcd}$$

So, there exists a unique Rauzy class for $d = 3$.

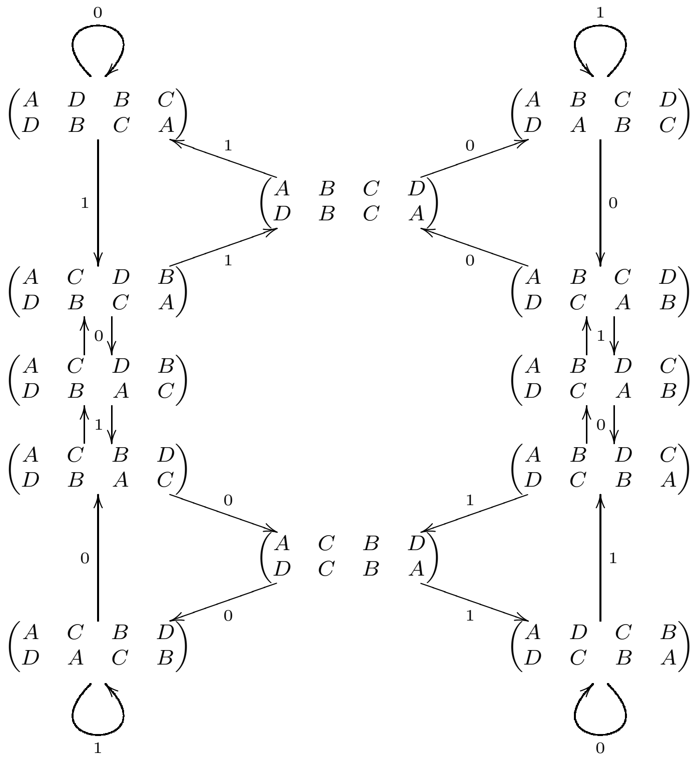
For $d = 4$ there are 24 possibilities for the monodromy invariant, 13 of which are irreducible:

$$\begin{array}{lllll} (4,3,2,1), & (4,1,3,2), & (3,1,4,2), & (4,2,1,3), & (2,4,3,1), \\ (3,2,4,1), & (2,4,1,3), & (4,2,3,1), & (4,1,2,3), & (4,3,1,2), \\ (3,4,1,2), & (2,3,4,1), & (3,4,2,1) & & \end{array}$$

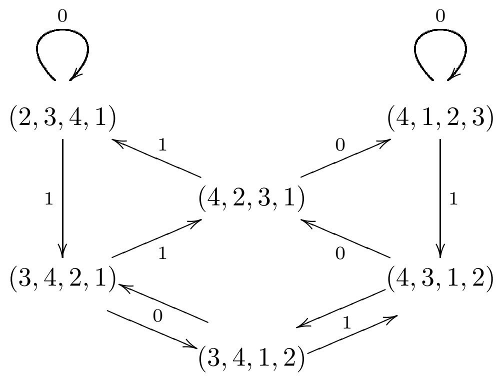
The following Rauzy class, with seven vertices, accounts for the first seven values of the monodromy invariant:

$$\begin{tikzcd} \begin{pmatrix} A & D & B & C \\ D & C & A & B \end{pmatrix} \arrow[loop above, "1"] \arrow[d, shift right, "0"'] & & & & \begin{pmatrix} A & B & D & C \\ D & A & C & B \end{pmatrix} \arrow[loop above, "0"] \arrow[d, shift right, "1"'] \\ \begin{pmatrix} A & D & B & C \\ D & C & B & A \end{pmatrix} \arrow[u, shift right] \arrow[dd, "1"] & & & & \begin{pmatrix} A & B & C & D \\ D & A & C & B \end{pmatrix} \arrow[u, shift right] \arrow[dd, "0"] \\ & & \begin{pmatrix} A & B & C & D \\ D & C & B & A \end{pmatrix} \arrow[ull, "1"] \arrow[urr, "0"'] & & \\ \begin{pmatrix} A & C & D & B \\ D & C & B & A \end{pmatrix} \arrow[loop below, "0"] \arrow[urr, "1"'] & & & & \begin{pmatrix} A & B & C & D \\ D & B & A & C \end{pmatrix} \arrow[loop below, "1"] \arrow[ull, "0"] \end{tikzcd}$$

The remaining six values of the monodromy invariant occur in the Rauzy class that we represent next. Observe that it has twice as many vertices: we shall return to this point in a while. Thus, there exactly two Rauzy classes for $d = 4$.



All these graphs are symmetric with respect to the vertical axis: this symmetry corresponds to the canonical involution, that is, to interchanging the roles of π_0 and π_1. The last graph has an additional central symmetry: pairs that are opposite relative to the center have the same monodromy invariant, and so they correspond to essentially the same interval exchange transformation. Identifying such pairs, one obtains the corresponding *reduced Rauzy class*:



The Rauzy classes for $d \leq 5$ are listed below:

d	representative	# vertices (full class)	# vertices (reduced)
2	(2,1)	1	1
3	(3,2,1)	3	3
4	(4,3,2,1)	7	7
4	(4,2,3,1)	12	6
5	(5,4,3,2,1)	15	15
5	(5,3,2,4,1)		11
5	(5,4,2,3,1)		35
5	(5,2,3,4,1)		10

Standard pairs. A pair $\pi = (\pi_0, \pi_1)$ is called *standard* if the last symbol in each line coincides with the first symbol in the other line. In other words, the monodromy invariant satisfies

$$\pi_1 \circ \pi_0^{-1}(1) = d \quad \text{and} \quad \pi_1 \circ \pi_0^{-1}(d) = 1.$$

Inspection of the examples of Rauzy classes in Section 1.6 shows that they all contain some standard pair. This turns out to be a general fact:

Proposition 1.24. *Every Rauzy class contains some standard pair.*

Notice that the Rauzy-Veech operator leaves the first symbols $\alpha_1^\varepsilon = \pi_\varepsilon^{-1}(1)$, $\varepsilon \in \{0,1\}$ in both top and bottom lines unchanged throughout the entire Rauzy class $C(\pi)$. The proof of Proposition 1.24 is based on the auxiliary lemma that we state below. The lemma can be easily deduced from Proposition 1.19, but we also give a short direct proof.

Lemma 1.25. *Given any $\varepsilon \in \{0,1\}$ and any $\beta \in \mathcal{A}$ such that $\pi_\varepsilon(\beta) \neq 1$, there exists some pair π' in the Rauzy class $C(\pi)$ such that $\pi'_\varepsilon(\beta) = d$, that is, β is the last symbol in the line ε of π'.*

Proof. For each $\varepsilon \in \{0,1\}$ let $\mathcal{A}_\varepsilon$ be the subset of all $\beta \in \mathcal{A}$ such that $\pi'_\varepsilon(\beta) < d$ for every π' in the Rauzy class. In view of the previous remarks, $\alpha_1^\varepsilon \in \mathcal{A}_\varepsilon$. Let $\kappa(\varepsilon)$ be the rightmost position ever attained by these symbols, that is, the maximum value of $\pi'_\varepsilon(\beta)$ over all π' in $C(\pi)$ and $\beta \in \mathcal{A}_\varepsilon$. By definition, $\kappa(\varepsilon) < d$. Our goal is to prove that $\kappa(\varepsilon) = 1$, and so $\mathcal{A}_\varepsilon = \{\alpha_1^\varepsilon\}$, for both $\varepsilon \in \{0,1\}$.

Fix any $\beta_\varepsilon \in \mathcal{A}_\varepsilon$ for which the maximum is attained. Then $\pi'_\varepsilon(\beta_\varepsilon) = \kappa(\varepsilon)$ for every π' in $C(\pi)$. That is because symbols γ with $\pi_\varepsilon(\gamma) < d$ can only move to the right under the Rauzy-Veech iteration and, were that to happen, it would contradict the assumption that $\kappa(\varepsilon)$ is maximum. Recall also Lemma 1.23. The same argument shows that all the symbols to the left of β_ε are also constant on the Rauzy class:

$$(\pi'_\varepsilon)^{-1}(i) = \pi_\varepsilon^{-1}(i) \quad \text{for all } 1 \le i \le \kappa(\varepsilon). \tag{1.25}$$

In particular, no symbol to the left of β_ε on the line ε can ever reach the last position in the line $1-\varepsilon$:

$$\pi_\varepsilon(\alpha) < \kappa(\varepsilon) \quad \Rightarrow \quad \pi'_{1-\varepsilon}(\alpha) < d \quad \Rightarrow \quad \pi'_{1-\varepsilon}(\alpha) \le \kappa(1-\varepsilon), \tag{1.26}$$

for any pair π' in $C(\pi)$. Let us write

$$\pi' = \begin{pmatrix} \alpha_1^0 & \cdots & \alpha_{\kappa(0)}^0 & \cdots & \cdots & \alpha_d^0 \\ \alpha_1^1 & \cdots & \cdots & \alpha_{\kappa(1)}^1 & \cdots & \alpha_d^1 \end{pmatrix}, \quad \alpha_i^\varepsilon = (\pi'_\varepsilon)^{-1}(i).$$

In view of (1.25), the relation (1.26) implies

$$\{\alpha_1^\varepsilon, \cdots, \alpha_{\kappa(\varepsilon)-1}^\varepsilon\} \subset \{\alpha_1^{1-\varepsilon}, \cdots, \alpha_{\kappa(1-\varepsilon)}^{1-\varepsilon}\} \quad \text{for } \varepsilon \in \{0,1\}. \tag{1.27}$$

In particular, $\kappa(\varepsilon) - 1 \le \kappa(1-\varepsilon) \le \kappa(\varepsilon) + 1$. There are four possibilities:

1. $\kappa(0) = \kappa(1) + 1$: then the case $\varepsilon = 0$ of (1.27) implies $\{\alpha_1^0, \cdots, \alpha_{\kappa(1)}^0\} = \{\alpha_1^1, \cdots, \alpha_{\kappa(1)}^1\}$, and this contradicts the assumption of irreducibility.
2. $\kappa(0) = \kappa(1) - 1$: this is analogous to the first case, using the case $\varepsilon = 1$ in (1.27) instead.
3. $\kappa(0) = \kappa(1)$ and $\{\alpha_1^0, \cdots, \alpha_{\kappa(0)-1}^0\} = \{\alpha_1^1, \cdots, \alpha_{\kappa(1)-1}^1\}$: this also contradicts irreducibility, unless $\kappa(0) = \kappa(1) = 1$.
4. $\kappa(0) = \kappa(1)$ and there exists $1 \le i < \kappa(0)$ such that $\alpha_i^0 = \alpha_{\kappa(1)}^1$: together with the case $\varepsilon = 1$ of (1.27), this gives

   $$\{\alpha_1^1, \cdots, \alpha_{\kappa(1)-1}^1, \alpha_{\kappa(1)}^1\} = \{\alpha_1^0, \cdots, \alpha_{\kappa(0)}^0\}$$

   and this implies that the two sets coincide (hence, there exists $1 \le j < \kappa(1)$ such that $\alpha_j^1 = \alpha_{\kappa(0)}^0$). Once more, this contradicts irreducibility.

This completes the proof of the lemma. □

Now we can give the proof of Proposition 1.24:

Proof. As observed before, the first symbols α_1^ε in both lines remain unchanged under Rauzy-Veech iteration. By irreducibility, they are necessarily distinct. So, using Lemma 1.25, we may find a pair π' in $C(\pi)$ such that $\pi'_0(\alpha_1^1) = d$, that is, the last symbol in the top line coincides with the first one in the bottom line. Now, iterating π' under type 0 Rauzy-Veech map, we keep the top line unchanged, while rotating all the symbols in the bottom line to the right of α_1^1. So, we eventually reach a pair π'' which satisfies $\pi''_1(\alpha_1^0) = d$, in addition to $\pi''_0(\alpha_1^1) = d$. Then π'' is standard. □

1.7 Rauzy-Veech renormalization

We are especially interested in a variation of the induction algorithm where one scales the domains of all interval exchange transformations to length 1.

Let π and π' be irreducible pairs such that π' is the type ε successor of π, for $\varepsilon \in \{0,1\}$. For each $\lambda \in \mathbb{R}_+^{\mathcal{A}}$ satisfying

$$\lambda_{\alpha(\varepsilon)} > \lambda_{\alpha(1-\varepsilon)} \tag{1.28}$$

we have

$$\hat{R}(\pi,\lambda) = (\pi',\lambda') \quad \text{with} \quad \lambda'_\alpha = \begin{cases} \lambda_\alpha & \text{if } \alpha \neq \alpha(\varepsilon) \\ \lambda_{\alpha(\varepsilon)} - \lambda_{\alpha(1-\varepsilon)} & \text{if } \alpha = \alpha(\varepsilon). \end{cases}$$

The map $\lambda \mapsto \lambda'$ thus defined is a bijection from the set of length vectors satisfying (1.28) to the whole $\mathbb{R}_+^{\mathcal{A}}$: the inverse is given by

$$\lambda_\alpha = \begin{cases} \lambda'_\alpha & \text{if } \alpha \neq \alpha(\varepsilon) \\ \lambda'_{\alpha(\varepsilon)} + \lambda'_{\alpha(1-\varepsilon)} & \text{if } \alpha = \alpha(\varepsilon). \end{cases}$$

Take the interval I to have unit length, that is, $\sum_{\alpha\in\mathcal{A}} \lambda_\alpha = 1$. The induction $\hat{R}(f)$ is defined on a shorter interval, with length $1-\lambda_{\alpha(1-\varepsilon)}$, but after appropriate rescaling we may see it as a map $R(f)$ on a unit interval. This means we are now considering

$$R : (\pi,\lambda) \mapsto (\pi',\lambda''), \quad \text{where} \quad \lambda'' = \frac{\lambda'}{1-\lambda_{\alpha(1-\varepsilon)}}, \tag{1.29}$$

that we refer to as the *Rauzy-Veech renormalization map.* Let $\Lambda_{\mathcal{A}}$ be the set of all length vectors $\lambda \in \mathbb{R}_+^{\mathcal{A}}$ with $\sum_{\alpha\in\mathcal{A}} \lambda_\alpha = 1$, and let

$$\Lambda_{\pi,\varepsilon} = \{\lambda \in \Lambda_{\mathcal{A}} : \lambda_{\alpha(\varepsilon)} > \lambda_{\alpha(1-\varepsilon)}\} \quad \text{for } \varepsilon \in \{0,1\}.$$

The previous observations mean that $(\pi,\lambda) \mapsto (\pi',\lambda'')$ maps $\{\pi\}\times\Lambda_{\pi,\varepsilon}$ bijectively onto $\{\pi'\}\times\Lambda_{\mathcal{A}}$. Figure 1.8 illustrates the case $d=3$:

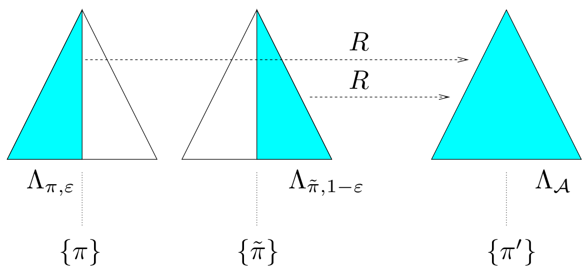


Figure 1.8:

For each Rauzy class C we have a map $R : (\pi,\lambda) \mapsto (\pi',\lambda'')$ from $C \times \Lambda_{\mathcal{A}}$ to itself [4], with the following Markov property: R sends each $\{\pi\}\times\Lambda_{\pi,\varepsilon}$ bijectively onto $\{\pi'\}\times\Lambda_{\mathcal{A}}$, where π' is the type ε successor of π. Note that

$$\lambda'' = \frac{\Theta^{-1*}(\lambda)}{1-\lambda_{\alpha(1-\varepsilon)}} \tag{1.30}$$

and the operator Θ depends only on π and the type ε, that is, it is constant on each $\{\pi\}\times\Lambda_{\pi,\varepsilon}$.

[4]More precisely, this map is defined on the full Lebesgue measure subset of length vectors λ that satisfy the Keane condition.

Example 1.26. For $d = 2$ there is only one pair, $\pi = \begin{pmatrix} A & B \\ B & A \end{pmatrix}$. We have

$$\Lambda_{\mathcal{A}} = \{(\lambda_A, \lambda_B) : \lambda_A > 0, \lambda_B > 0, \text{ and } \lambda_A + \lambda_B = 1\} \sim (0, 1),$$

where $\sim$ refers to the bijective correspondence $(\lambda_A, \lambda_B) \mapsto x = \lambda_A$. Under this correspondence, $\Lambda_{\pi,0} \sim (0, 1/2)$ and $\Lambda_{\pi,1} \sim (1/2, 1)$, and the Rauzy-Veech renormalization $(\pi, \lambda) \mapsto (\pi, \lambda'')$ is given by (see Figure 1.9)

$$r(x) = \begin{cases} x/(1-x) & \text{for } x \in (0, 1/2) \\ 2 - 1/x & \text{for } x \in (1/2, 1). \end{cases}$$

Observe that r has a tangency of order 1 with the identity at $x = 0$ and $x = 1$.

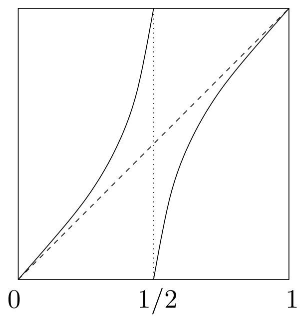


Figure 1.9:

Let $d\pi$ denote the counting measure in the set of pairs π, and Leb be the Lebesgue measure (of dimension $d - 1$) in the simplex $\Lambda_{\mathcal{A}}$. It was proven by Masur [41] and Veech [54], independently, that for each Rauzy class C, the Rauzy-Veech renormalization map $R : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$ admits an invariant measure ν which is absolutely continuous with respect to $d\pi \times \text{Leb}$. Moreover, this measure ν is unique, up to product by a scalar, and ergodic. A proof of this theorem will be given in Chapter 4.

1.8 Zorich transformations

In general, the measures ν in Theorem 4.1 have infinite mass. For instance, it is well-known that for maps with neutral fixed points such as the one in Example 1.26, absolutely continuous invariant measures are necessarily infinite. Zorich [63] introduced an accelerated version of the Rauzy-Veech algorithm for which there exists a (unique) invariant *probability* absolutely continuous with respect to Lebesgue measure on each simplex $\Lambda_{\mathcal{A}}$. This is defined as follows.

Let C be a Rauzy class, $\pi = (\pi_0, \pi_1)$ be a vertex of C, and $\lambda \in \mathbb{R}_+^{\mathcal{A}}$ satisfy the Keane condition. Let $\varepsilon \in \{0, 1\}$ be the type of (π, λ) and, for each $j \geq 1$, let

ε^j be the type of the iterate $(\pi^j, \lambda^j) = \hat{R}^j(\pi, \lambda)$. Then define $n = n(\pi, \lambda) \geq 1$ to be smallest such that $\varepsilon^j \neq \varepsilon$. The *Zorich induction map* is defined by

$$\hat{Z}(\pi, \lambda) = (\pi^n, \lambda^n) = \hat{R}^n(\pi, \lambda).$$

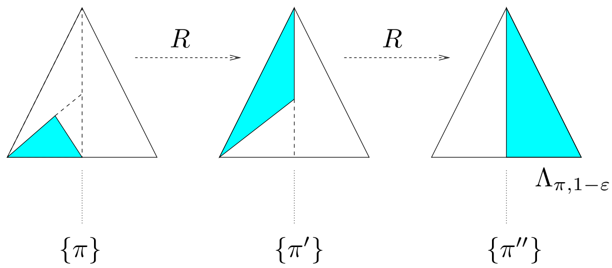


Figure 1.10:

We also consider the *Zorich renormalization map*

$$Z : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}, \quad Z(\pi, \lambda) = R^n(\pi, \lambda).$$

This map admits a Markov partition, into countably many domains. Indeed, for each π in the Rauzy class and $\varepsilon \in \{0, 1\}$, let

$$\Lambda^*_{\pi,\varepsilon,n} = \{\lambda \in \Lambda_{\pi,\varepsilon} : \varepsilon^1 = \cdots = \varepsilon^{n-1} = \varepsilon \neq \varepsilon^n\}.$$

Then Z maps every $\{\pi\} \times \Lambda^*_{\pi,\varepsilon,n}$ bijectively onto $\{\pi^n\} \times \Lambda_{\pi^n,1-\varepsilon}$. See Figure 1.10. Moreover, by (1.30),

$$\lambda^n = c_n \Theta^{-n*}(\lambda) \tag{1.31}$$

where $c_n > 0$ and Θ^{-n*} depend only on π, ε, n, that is, they are constant on each $\{\pi\} \times \Lambda^*_{\pi,\varepsilon,n}$.

Example 1.27. For $d = 2$ (recall Example 1.26), the Zorich transformation Z is described by the map $z(x) = r^n(x)$ where $n = n(x) \geq 1$ is the smallest integer such that

$$r^n(x) \in (1/2, 1), \quad \text{if} \quad x \in (0, 1/2) \quad \text{or} \quad r^n(x) \in (0, 1/2), \quad \text{if} \quad x \in (1/2, 1).$$

See Figure 1.11. This map is Markov and uniformly expanding (the latter is specific to $d = 2$). It is well-known that such maps admit absolutely continuous invariant probabilities.

In Chapter 4 we shall prove, following Zorich [63], that for each Rauzy class C, the Zorich renormalization map $Z : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$ admits an invariant probability measure μ which is absolutely continuous with respect to $d\pi \times \text{Leb}$. Moreover, his probability μ is unique and ergodic.

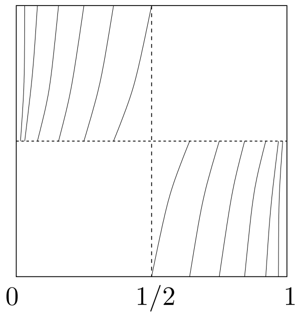


Figure 1.11:

Continued fractions. The classical continued fraction algorithm associates to each irrational number $x_0 \in (0,1)$ the sequences of integers

$$n_k = \left[\frac{1}{x_{k-1}}\right] \quad \text{and} \quad x_k = \frac{1}{x_{k-1}} - n_k\,,$$

where $[\cdot]$ denotes the integer part. Observe that

$$x_0 = \frac{1}{n_1 + x_1} = \frac{1}{n_1 + \frac{1}{n_2+x_2}} = \frac{1}{n_1 + \frac{1}{n_2+\frac{1}{n_3+x_3}}} = \cdots$$

The algorithm may also be written as

$$x_k = G^k(x_0) \quad \text{and} \quad n_k = \left[\frac{1}{x_{k-1}}\right]$$

where G is the Gauss map (see Figure 1.12)

$$G : (0,1) \to [0,1], \quad G(x) = \frac{1}{x} - \left[\frac{1}{x}\right]$$

The Gauss map is very much equivalent to the Zorich transformation for $d = 2$ (and so the cases $d > 2$ of the Zorich transformation may be seen as higher dimensional generalizations of the classical continued fraction expansion). To see this, consider the bijection

$$\phi : (\lambda_A, \lambda_B) \mapsto y = \frac{\lambda_A}{\lambda_B}$$

from $\Lambda_{\mathcal{A}}$ to $(0,\infty)$. Moreover, let P be the bijection of $\Lambda_{\mathcal{A}}$ defined by $P : (\lambda_A, \lambda_B) \mapsto (\lambda_B, \lambda_A)$. Consider (λ_A, λ_B) in $\Lambda_{\pi,0}$, that is, such that $\lambda_A < \lambda_B$. Then $y = \phi(\lambda_A, \lambda_B) \in (0,1)$. By definition,

$$\hat{Z} \circ P(\lambda_A, \lambda_B) = \hat{Z}(\lambda_B, \lambda_A) = (\lambda_B - n\lambda_A, \lambda_A)$$

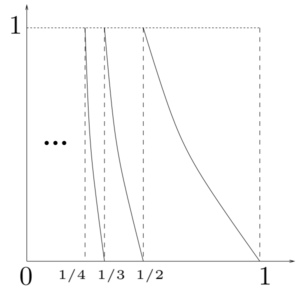


Figure 1.12:

where n is the integer part of λ_B/λ_A. In terms of the variable y, this corresponds to

$$y \mapsto \frac{1}{y} - n = G(y).$$

In other words, we have just shown that ϕ conjugates $Z \circ P$, restricted to $\Lambda_{\pi,0}$, to the Gauss map G. Consequently, ϕ conjugates $(Z \circ P)^n$, restricted to $\Lambda_{\pi,0}$, to G^n, for every $n \geq 1$. Observe that $P^2 = \mathrm{id}$ and Z commutes [5] with P. Hence, we have shown that $Z^{2k} \mid \Lambda_{\pi,0}$ is conjugate to G^{2k}, and $Z^{2k-1} \circ P \mid \Lambda_{\pi,0}$ is conjugate to G^{2k-1}, for every $k \geq 1$.

1.9 Symplectic form

It is clear from (1.3) that the operator $\Omega_\pi : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ is anti-symmetric:

$$\Omega_\pi^* = -\Omega_\pi \tag{1.32}$$

where Ω_π^* is the adjoint operator, relative to the Euclidean metric $\cdot$ on $\mathbb{R}^{\mathcal{A}}$. Thus,

$$\tilde{\omega}_\pi : \mathbb{R}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}} \to \mathbb{R}, \quad \tilde{\omega}_\pi(u, v) = -u \cdot \Omega_\pi(v)$$

defines an alternate bilinear form on $\mathbb{R}^{\mathcal{A}}$. In general, this form is degenerate: $\tilde{\omega}_\pi(u, v) = 0$ for every $u \in \mathbb{R}^{\mathcal{A}}$ if (and only if) $v \in \ker \Omega_\pi$. On the other hand, it can always be turned into a *symplectic form*, that is, a non-degenerate alternate bilinear form, in the following two ways. Firstly, we may consider

$$\omega_\pi : H_\pi \times H_\pi \to \mathbb{R}, \quad \omega_\pi(\Omega_\pi(u), \Omega_\pi(v)) = -u \cdot \Omega_\pi(v) \tag{1.33}$$

on the image subspace $H_\pi = \Omega_\pi(\mathbb{R}^{\mathcal{A}})$. Secondly, we may consider

$$\omega'_\pi : \mathbb{R}^{\mathcal{A}}/\ker \Omega_\pi \to \mathbb{R}^{\mathcal{A}}/\ker \Omega_\pi, \quad \omega'_\pi([u], [v]) = -u \cdot \Omega_\pi(v) \tag{1.34}$$

on the quotient subspace $\mathbb{R}^{\mathcal{A}}/\ker \Omega_\pi$.

[5] In other words, P conjugates the restriction of Z to $\Lambda_{\pi,0}$ to the restriction of Z to $\Lambda_{\pi,1}$.

Lemma 1.28. *The relations* (1.33) *and* (1.34) *define symplectic forms* ω_π *and* ω'_π *on the corresponding spaces.*

Proof. The relation (1.32) implies that the orthogonal complement $H_\pi^\perp$ coincides with $\ker \Omega_\pi$. Suppose $\Omega_\pi(u) = \Omega_\pi(u')$ or, equivalently, $u - u' \in \ker \Omega_\pi$. Then

$$u \cdot \Omega_\pi(v) = u' \cdot \Omega_\pi(v) \quad \text{for every } v \in \mathbb{R}^{\mathcal{A}}.$$

This shows that ω_π and ω'_π are well-defined. It is clear that they are bilinear. The fact that they are alternate is an immediate consequence of (1.32):

$$-v \cdot \Omega_\pi(u) = -u \cdot \Omega^*_\pi(v) = u \cdot \Omega_\pi(v).$$

Finally, it is also easy to see that they are non-degenerate:

$$-u \cdot \Omega_\pi(v) = 0 \quad \text{for all } v \quad \Leftrightarrow \quad u \in H_\pi^\perp = \ker \Omega_\pi.$$

For $u \in H_\pi$, this can only happen if u vanishes. In general, it means that $[u]$ vanishes in the quotient space $\mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi$. □

The following simple relation will be used several times:

Lemma 1.29. *If* $(\pi', \lambda') = \hat{R}(\pi, \lambda)$ *then* $\Theta\, \Omega_\pi\, \Theta^* = \Omega_{\pi'}$*, where* $\Theta = \Theta_{\pi,\lambda}$*.*

Proof. Let $\lambda' \in \mathbb{R}^{\mathcal{A}}$ be given by $\lambda = \Theta^*(\lambda')$, and then define $w = \Omega_\pi(\lambda)$ and $w' = \Omega_{\pi'}(\lambda')$. Compare (1.11) and (1.2). We have seen in (1.8) that $w' = \Theta(w)$. This means that $\Omega_{\pi'}(\lambda') = \Theta\, \Omega_\pi \Theta^*(\lambda')$ for all $\lambda' \in \mathbb{R}^{\mathcal{A}}$, as claimed. □

Corollary 1.30. *The operator* Θ *induces a symplectic isomorphism from* H_π *to* $H_{\pi'}$*, relative to the symplectic forms in the two spaces. Analogously, the operator* Θ^{-1*} *induces a symplectic isomorphism from* $\mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi$ *to* $\mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi$*, relative to the symplectic forms in the two spaces.*

Proof. Lemma 1.29, together with the fact that Θ and Θ^* are invertible, implies that $u \in H_\pi$ if and only if $\Theta(u)$ is in $H_{\pi'}$. This means that $\Theta : H_\pi \to H_{\pi'}$ is a well defined isomorphism. Moreover, it is symplectic:

$$\begin{aligned}\omega_{\pi'}(\Theta\Omega_\pi(u), \Theta\Omega_\pi(v)) &= \omega_{\pi'}(\Omega_{\pi'}\Theta^{-1*}(u), \Theta\Omega_\pi(v)) \\ &= -\Theta^{-1*}(u) \cdot \Theta\Omega_\pi(v) = -u \cdot \Omega_\pi(v) = \omega_\pi(\Omega_\pi(u), \Omega_\pi(v)),\end{aligned}$$

for any vectors $u, v \in \mathbb{R}^{\mathcal{A}}$. For the same reasons, $v \in \ker \Omega_\pi$ if and only if $\Theta^{-1*}(v) \in \ker \Omega_{\pi'}$ and that ensures $\Theta^{-1*} : \mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi \to \mathbb{R}^{\mathcal{A}} / \ker \Omega_{\pi'}$ is a well defined isomorphism. Moreover, it is symplectic:

$$\begin{aligned}\omega_{\pi'}([\Theta^{-1*}(u)], [\Theta^{-1*}(v)]) &= -\Theta^{-1*}(u) \cdot \Omega_{\pi'}\Theta^{-1*}(v) \\ &= -\Theta^{-1*}(u) \cdot \Theta\Omega_\pi(v) = -u \cdot \Omega_\pi(v) = \omega_\pi([u], [v]).\end{aligned}$$

This completes the proof. □

Remark 1.31. Lemma 1.28 implies that the spaces H_π and $\mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi$ have even dimension: we write $\dim H_\pi = \dim \mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi = 2g$. Since Θ is always an isomorphism from H_π to H'_π, it follows that the dimension is constant on the whole Rauzy class. We shall later interpret g as the genus of an orientable surface canonically associated to the Rauzy class. We shall also see that ω_π and ω'_π may be interpreted as intersection forms in the homology and cohomology spaces of that surface. Incidentally, this is the reason for the minus sign in the definitions.

Notes

The main original sources are Rauzy [47], Masur [41], and Veech [53, 54, 55, 56, 57, 58], as well as Keane [26, 27] for Sections 1.3 and 1.4, and Zorich [63, 65] for Section 1.8. Our presentation owes much to Marmi, Moussa, Yoccoz [40] and Yoccoz [61].

Chapter 2

Translation Surfaces

The structure of translation surface may be approached from several points of view: analytically, it corresponds to a holomorphic complex 1-form on a Riemann surface; geometrically, it is given by a flat Riemannian metric with conic singularities, together with a parallel unit vector field; topologically, it may be viewed as a pair of transverse measured foliations on the surface. We are going to exploit these different points of view, here and in Chapter 3.

Translation surfaces are useful to us because they provide a natural setting for defining the suspensions of interval exchange transformations, and introducing invertible versions of the induction and renormalization operators defined in the previous chapter. We describe the suspension construction and explain how the resulting translation surface may be computed from the combinatorial and metric data of the exchange transformation.

Suspensions can also be defined in terms of zippered rectangles, a notion introduced by Veech [54]. In fact both languages are useful for our purposes, and so we spend sometime explaining how one relates to the other. In particular, we define invertible induction and renormalization operators in the two languages, and describe their mutual relations.

Another important dynamical system in the space of translation surfaces, or of zippered rectangles, is the Teichmüller flow, that we introduce in Section 2.10. It is related to the Rauzy-Veech renormalization in that the latter may be seen as the Poincaré return map of the Teichmüller flow to the cross-section corresponding to length vectors with total length 1.

2.1 Definitions

An *Abelian differential* α is a holomorphic complex 1-form on a Riemann surface. We assume the Riemann surface is compact, and α is not identically zero. Then it has a finite number of zeroes. In local coordinates, $\alpha_z = \varphi(z)\,dz$ for some $\varphi(z) \in \mathbb{C}$ that depends holomorphically on the point z. Near any non-singular point p, one can always find so-called *adapted coordinates* ζ relative to which the Abelian differential takes the form $\alpha_\zeta = d\zeta$: it suffices to take

$$\zeta = \int_p^z \varphi(\xi)\, d\xi \tag{2.1}$$

If p is a zero of α, with multiplicity $m \geq 1$ say, then one considers instead

$$\zeta = \Big(\int_p^z \varphi(\xi)\, d\xi \Big)^{1/m+1} : \tag{2.2}$$

in these coordinates

$$\alpha_\zeta = (m+1)\zeta^m\, d\zeta = d\left(\zeta^{m+1}\right). \tag{2.3}$$

Notice that all changes of adapted coordinates near a regular point are given by translations: if ζ and ζ' are adapted coordinates then $d\zeta' = d\zeta$, and so $\zeta' = \zeta + \text{const}$. We say that the adapted coordinates form a *translation atlas*, and call the resulting structure a *translation surface*. Coordinate changes near zeroes are more subtle. If ζ' is a *regular adapted coordinate*, as in (2.1), and ζ is a *singular adapted coordinate*, as in(2.2), then $d\zeta' = \zeta^m d\zeta$ or, in other words, $(m+1)\zeta' = \zeta^{m+1} + \text{const}$. Figure 2.1 illustrates this relation between the two types of coordinates.

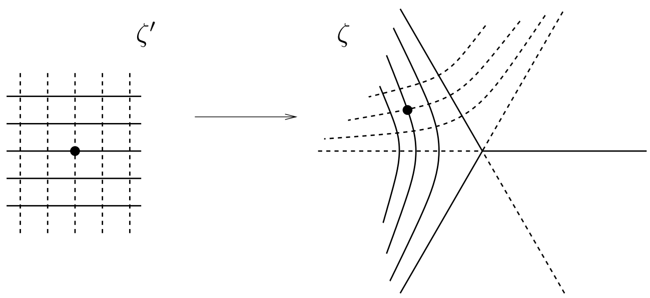


Figure 2.1:

The translation atlas defines a flat (zero curvature) Riemannian metric on the surface minus the zeroes, transported from the complex plane through the adapted charts. The form of (2.3) shows that the zeroes of the Abelian differential correspond to singularities of this flat metric, of a special kind: the geometry near the zero corresponds to the pull-back of the usual metric on the plane by a branched covering $\zeta \mapsto \zeta^{m+1}$. See Figure 2.2. This means that in appropriate polar coordinates $(\rho, \theta) \in \mathbb{R}_+ \times S^1$ centered at the singularity, the Riemannian metric is given by

$$ds^2 = d\rho^2 + (c\rho d\theta)^2, \tag{2.4}$$

where the "horizon angle" at the singularity $c = 2\pi(m+1)$. Singularities of type (2.4) are called *conical.* In addition, the translation atlas defines a parallel

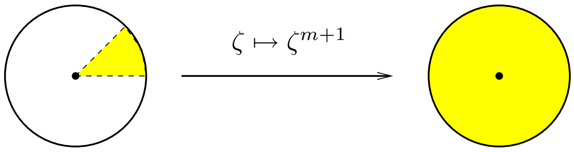


Figure 2.2:

unit vector field on the complement of the singularities, namely, the pull-back of the vertical vector field under the local charts.

Conversely, a flat metric with finitely many singularities, of conical type, together with a parallel unit vector field X, completely determine a translation structure. Indeed, the neighborhood of any regular point p is isometric to an open subset of $\mathbb{C}$. Choose the isometry so that it sends the vector X_p to the vertical vector $(0,1)$. Then the isometry is uniquely determined, and sends X to the constant vector field $(0,1)$. In particular, these isometries coincide in the intersection of their domains, and so they define a Riemann surface atlas on the complement of the singularities. Moreover, they transport the canonical Abelian differential dz from $\mathbb{C}$ to the surface.

Construction of translation surfaces. Let us describe a simple construction of translation surfaces. Later, in Section 3.5, we shall see that this construction is general: every translation surface can be obtained in this way.

Consider a polygon in $\mathbb{R}^2$ having an even number $2d \geq 4$ of sides

$$s_1, \ldots, s_d, s_1', \ldots, s_d'$$

such that s_i and s_i' are parallel (non-adjacent) and have the same length, for every $i = 1, \ldots, d$. See Figure 2.3 for an example with $d = 4$. Identifying s_i with

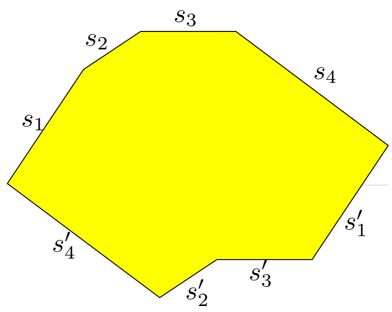


Figure 2.3:

s_i' by translation, for each $i = 1, \ldots, d$, we obtain a translation surface M: the singularities correspond to the points obtained by identification of the vertices of the polygon; the Abelian differential and the flat metric are inherited from $\mathbb{R}^2 = \mathbb{C}$, and the vertical vector field $X = (0, 1)$ is parallel.

Let $a_1, \ldots, a_\kappa$, $\kappa = \kappa(\pi)$ be the singularities. The *angle* of a singularity a_i is the topological index around zero

$$\operatorname{angle}(a_i) = 2\pi \operatorname{ind}(\beta, 0) = \frac{1}{i} \int_0^1 \frac{\dot\beta(t)}{\beta(t)}\, dt$$

of the curve $\beta(t) = \alpha_{\gamma(t)}(\dot\gamma(t))$, where $\gamma : [0, 1] \to M$ is any small simple closed curve around a_i. It is clear that

$$\operatorname{angle}(a_i) = 2\pi(m_i + 1) \tag{2.5}$$

where m_i denotes the order of the zero of α at a_i. We call the singularity *removable* if the angle is exactly 2π, that is, if a_i is actually not a zero of α.

Let the translation surface be constructed from a planar polygon with $2d$ sides, as described above. Then the sum of all angles at the singularities coincides with the sum of the internal angles of the $2d$-gon, that is

$$\sum_{i=1}^{\kappa} \operatorname{angle}(a_i) = 2\pi(d - 1). \tag{2.6}$$

Using (2.5) we deduce that

$$\sum_{i=1}^{\kappa} m_i = d - \kappa - 1. \tag{2.7}$$

The angles are also related to the genus $g(M)$ and the Euler characteristic $\mathcal{X}(M) = 2 - 2g(M)$ of the surface M. To this end, consider a decomposition into $4d$ triangles as described in Figure 2.4: a central point is linked to the vertices of the polygon and to the midpoint of every side.

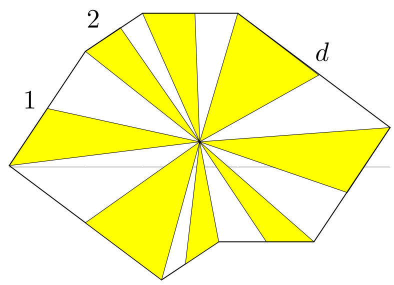


Figure 2.4:

Recall that the sides of the polygon are identified pairwise. So, this decomposition has $6d$ edges, $2d$ of them corresponding to segments inside the sides of the polygon. Moreover, there are $d + \kappa + 1$ vertices: the central one, plus d vertices coming from the midpoints of the polygon sides, and κ more sitting at the singularities. Therefore,

$$2 - 2g(M) = \mathcal{X}(M) = \kappa + 1 - d. \tag{2.8}$$

From (2.6) and (2.8), we obtain a kind of Gauss-Bonnet theorem for these flat surfaces:

$$\sum_{i=1}^{\kappa} \big[2\pi - \operatorname{angle}(a_i)\big] = -2\pi \sum_{i=1}^{\kappa} m_i = 2\pi(\kappa + 1 - d) = 2\pi\mathcal{X}(M). \tag{2.9}$$

We shall see in Corollary 6.26 that this is the only restriction imposed on the orders of the singularities by the topology of the surface: given any $g \geq 1$ and integers $m_i \geq 0$, $i = 1, \ldots, \kappa$ with $\sum_{i=1}^{\kappa} m_i = 2g - 2$, there exists some translation surface with κ singularities of orders $m_1, \ldots, m_\kappa$.

2.2 Suspending interval exchange maps

Let π be an irreducible pair and $\lambda \in \mathbb{R}_+^{\mathcal{A}}$ be a length vector. We denote by T_π^+ the subset of vectors $\tau = (\tau_\alpha)_{\alpha \in \mathcal{A}} \in \mathbb{R}^{\mathcal{A}}$ such that

$$\sum_{\pi_0(\alpha) \leq k} \tau_\alpha > 0 \quad \text{and} \quad \sum_{\pi_1(\alpha) \leq k} \tau_\alpha < 0 \tag{2.10}$$

for all $1 \leq k \leq d - 1$. Clearly, T_π^+ is a convex cone. We say that τ has *type* 0 if the total sum $\sum_{\alpha \in \mathcal{A}} \tau_\alpha$ is positive and type 1 if the total sum is negative. Define $\zeta_\alpha = (\lambda_\alpha, \tau_\alpha) \in \mathbb{R}^2$ for each $\alpha \in \mathcal{A}$. Then consider the closed curve $\Gamma = \Gamma(\pi, \lambda, \tau)$ on $\mathbb{R}^2$ formed by concatenation of

$$\zeta_{\alpha_1^0}, \zeta_{\alpha_2^0}, \ldots, \zeta_{\alpha_d^0}, -\zeta_{\alpha_d^1}, -\zeta_{\alpha_{d-1}^1}, \ldots, -\zeta_{\alpha_1^1}$$

with starting point at the origin. Condition (2.10) means that the endpoints of all $\zeta_{\alpha_1^0} + \cdots + \zeta_{\alpha_k^0}$ are on the upper half plane, and the endpoints of all $\zeta_{\alpha_1^1} + \cdots + \zeta_{\alpha_k^1}$ are in the lower half plane, for every $1 \leq k \leq d - 1$. See Figure 2.5.

Assume, for the time being, that this closed curve Γ is simple. Then it defines a planar polygon with $2d$ sides organized in pairs of parallel segments with the same length, as considered in the previous section. The *suspension surface* $M = M(\pi, \lambda, \tau)$ is the translation surface obtained by identification of the sides in each of the pairs. Let $I \subset M$ be the horizontal segment of length $\sum_{\alpha \in \mathcal{A}} \lambda_\alpha$ with the origin as left endpoint, that is,

$$I = \big[0, \sum_{\alpha \in \mathcal{A}} \lambda_\alpha\big) \times \{0\}. \tag{2.11}$$

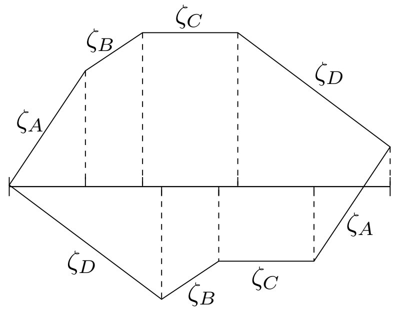


Figure 2.5:

The interval exchange transformation f defined by (π, λ) corresponds to the first return map to I of the vertical flow on M. To see this, for each $\alpha \in \mathcal{A}$, let

$$I_\alpha = \big[\sum_{\pi_0(\beta)<\pi_0(\alpha)} \lambda_\beta, \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \lambda_\beta \big) \times \{0\}$$

Consider the vertical segment starting from $(x, 0) \in I_\alpha$ and moving upwards. It hits the side represented by ζ_α at some point (x, z). This is identified with the point (x', z') in the side represented by $-\zeta_\alpha$, given by

$$\begin{aligned} x' &= x - \sum_{\pi_0(\beta)<\pi_0(\alpha)} \lambda_\beta + \sum_{\pi_1(\beta)<\pi_1(\alpha)} \lambda_\beta = x + w_\alpha \\ z' &= z - \sum_{\pi_0(\beta)<\pi_0(\alpha)} \tau_\beta + \sum_{\pi_1(\beta)<\pi_1(\alpha)} \tau_\beta = z - h_\alpha \end{aligned} \tag{2.12}$$

($h_\alpha > 0$ is *defined* by the last equality). Continuing upwards from (x', z') we hit I back at the point $(x', 0)$. This shows that the return map does coincide with $f(x) = x + w_\alpha$ on each I_α.

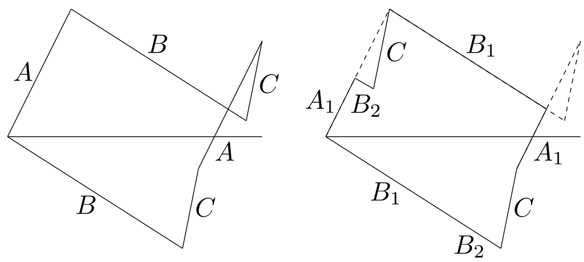


Figure 2.6:

In some fairly exceptional situations, such as in Figure 2.6, the closed curve Γ may have self-intersections. It is easy to extend the definition of the suspension surface to this case: just consider the simple polygon obtained by removing the self-intersections in the way described in the figure, and then take the translation surface M obtained by identification of parallel sides of this polygon. The horizontal segment I may still be viewed as a cross-section to the vertical flow on M, and the corresponding first return map coincides with the interval exchange transformation f (even if the data (π, λ) may not be the same).

We are going to focus our presentation on the case when Γ is simple and, in general, let the reader to adapt the arguments to the case when there are self-intersections. In some sense, the non-simple case can be avoided altogether:

Remark 2.1. The curve $\Gamma(\pi, \lambda, \tau)$ can have self-intersections only if either

$$\sum_{\alpha\in\mathcal{A}} \tau_\alpha > 0 \quad \text{and} \quad \lambda_{\alpha(0)} < \lambda_{\alpha(1)}, \quad \text{i.e. } \tau \text{ has type 0 and } (\pi, \lambda) \text{ has type 1,}$$

as is the case in Figure 2.6, or

$$\sum_{\alpha\in\mathcal{A}} \tau_\alpha < 0 \quad \text{and} \quad \lambda_{\alpha(0)} > \lambda_{\alpha(1)}, \quad \text{i.e. } \tau \text{ has type 1 and } (\pi, \lambda) \text{ has type 0.}$$

In other words, if (π, λ) and τ and have the same type then the curve $\Gamma(\pi, \lambda, \tau)$ is necessarily simple. Using this observation, we shall see in Remark 2.15 that by Rauzy-Veech induction one eventually finds data $(\pi^n, \lambda^n, \tau^n)$ that represents the same translation surface and for which the curve $\Gamma(\pi^n, \lambda^n, \tau^n)$ has no self-intersections.

2.3 Some translation surfaces

We shall see later that the type (genus and singularities) of the translation surface $M = M(\pi, \lambda, \tau)$ depends only on the Rauzy class of π. Here we consider a representative of each Rauzy class with $d \leq 5$, and we exhibit the corresponding translation surface for generic vectors λ and τ. The conclusions are summarized in the table near the end of this section.

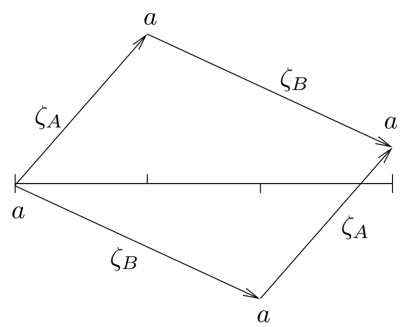


Figure 2.7:

For $d = 2$ and $\pi = \begin{pmatrix} A & B \\ B & A \end{pmatrix}$, corresponding to monodromy invariant $p = (2, 1)$, the four vertices are identified to a single point a, and angle $(a) = 2\pi$. Using (2.9) we conclude that M is the torus, (and the singularity is removable). See Figure 2.7.

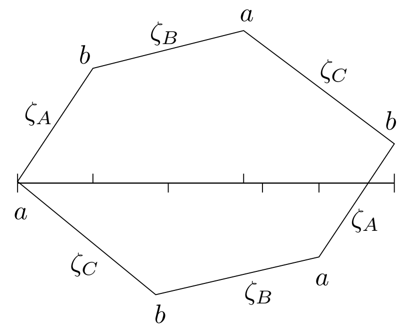


Figure 2.8:

For $d = 3$ and $\pi = \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix}$, corresponding to $p = (3, 2, 1)$, the six vertices are identified to two different points, with angle $(a) =$ angle $(b) = 2\pi$. Thus, M is the torus, and both singularities are removable. See Figure 2.8.

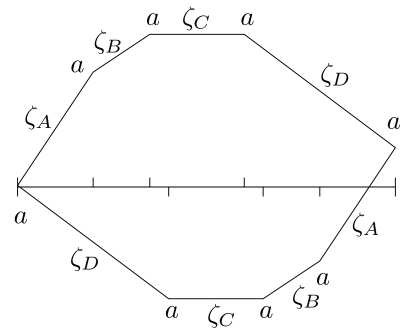


Figure 2.9:

For $d = 4$ and $\pi = \begin{pmatrix} A & B & C & D \\ D & C & B & A \end{pmatrix}$, corresponding to $p = (4, 3, 2, 1)$, the eight vertices are identified to a single point, with angle $(a) = 6\pi$. Thus, M has genus 2 (bitorus). See Figure 2.9.

For $d = 4$ and $\pi = \begin{pmatrix} A & B & C & D \\ D & B & C & A \end{pmatrix}$, hence $p = (4, 2, 3, 1)$, the vertices are identified to three different points, with angle $(a) =$ angle $(b) =$ angle $(c) = 2\pi$. M is the torus, and all singularities are removable. See Figure 2.10.

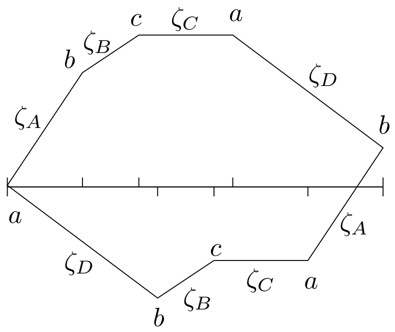


Figure 2.10:

For $d = 5$ and $\pi = \begin{pmatrix} A & B & C & D & E \\ E & D & C & B & A \end{pmatrix}$, hence $p = (5, 4, 3, 2, 1)$, the ten vertices are identified to two different points, a and b, with angle $(a) =$ angle $(b) = 4\pi$. Thus, M is the bitorus ($g = 2$).

For $d = 5$ and $\pi = \begin{pmatrix} A & B & C & D & E \\ E & C & B & D & A \end{pmatrix}$, hence $p = (5, 3, 2, 4, 1)$, the vertices are identified to two different points, a and b, with angle $(a) = 2\pi$ and angle $(b) = 6\pi$. M is, again, the bitorus.

For $d = 5$ and $\pi = \begin{pmatrix} A & B & C & D & E \\ E & D & B & C & A \end{pmatrix}$, hence $p = (5, 4, 2, 3, 1)$, the vertices are identified to two different points, a and b, with angle $(a) = 6\pi$ and angle $(b) = 2\pi$. M is, once more, the bitorus.

For $d = 5$ and $\pi = \begin{pmatrix} A & B & C & D & E \\ E & B & C & D & A \end{pmatrix}$, hence $p = (5, 2, 3, 4, 1)$, the vertices are identified to four different points, with angle $(a) =$ angle $(b) =$ angle $(c) =$ angle $(d) = 2\pi$. M is the torus and all singularities are removable.

Summarizing, we have:

d	representative	# vertices	angles	orders	genus	$\mathcal{X}$
2	(2,1)	1	2π	0	1	0
3	(3,2,1)	3	2π, 2π	0, 0	1	0
4	(4,3,2,1)	7	6π	2	2	-2
4	(4,2,3,1)	8	$2\pi, 2\pi, 2\pi$	0, 0, 0	1	0
5	(5,4,3,2,1)	15	$4\pi, 4\pi$	1, 1	2	-2
5	(5,3,2,4,1)	11	$6\pi, 2\pi$	2, 0	2	-2
5	(5,4,2,3,1)	35	$6\pi, 2\pi$	2, 0	2	-2
5	(5,2,3,4,1)	10	$2\pi, 2\pi, 2\pi, 2\pi$	0, 0, 0, 0	1	0

Remark 2.2. Starting from $d = 5$, different Rauzy classes may give rise to translation surfaces with the same number and orders of singularities. The relation between Rauzy classes and strata of translation surfaces, corresponding to fixed types of singularities, will be explained in Sections 2.5 and 6.6.

2.4 Computing the suspension surface

Let us explain how the number κ and the orders m_i of the singularities may be computed from π, in general. Consider the set of all pairs (α, S) with $\alpha \in \mathcal{A}$ and $S \in \{L, R\}$. We think of (α, L) and (α, R) as representing, respectively, the origin (left endpoint) and the end (right endpoint) of the sides of the polygon labeled by α. Then, under the identifications that define the suspension surface, one must identify

$$(\alpha, R) \sim (\beta, L) \quad \text{if} \quad \pi_0(\alpha) + 1 = \pi_0(\beta) \tag{2.13}$$

$$(\alpha, R) \sim (\beta, L) \quad \text{if} \quad \pi_1(\alpha) + 1 = \pi_1(\beta) \tag{2.14}$$

and also

$$(\alpha, L) \sim (\beta, L) \quad \text{if} \quad \pi_0(\alpha) = 1 = \pi_1(\beta) \tag{2.15}$$

$$(\alpha, R) \sim (\beta, R) \quad \text{if} \quad \pi_0(\alpha) = d = \pi_1(\beta). \tag{2.16}$$

Extend $\sim$ to an equivalence relation in the set of pairs (α, S). Then the number κ of singularities is, precisely, the number of equivalence classes for this relation.

Figure 2.11 describes a specific case with $d = 7$:

$$\pi = \begin{pmatrix} A & B & C & D & E & F & G \\ G & F & E & D & C & B & A \end{pmatrix}.$$

There are two equivalence classes:

$$(A, L) \sim (B, R) \sim (C, L) \sim (D, R) \sim (E, L) \sim (F, R) \sim (G, L) \sim (A, L)$$

and

$$(A, R) \sim (B, L) \sim (C, R) \sim (D, L) \sim (E, R) \sim (F, L) \sim (G, R) \sim (A, R)$$

It is also easy to guess what the angles of these singularities are. For instance,

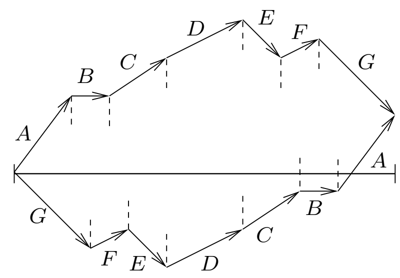


Figure 2.11:

consider the singularity a associated to the first equivalence class (the other one is analogous). The angle corresponds to the sum of the internal angles of the polygon at the 9 vertices that are identified to a. This sum is readily computed by noting that the arcs describing these internal angles cut the vertical direction exactly 6 times: one for each vertex, except for the exceptional $(A, L) = (G, L)$. See Figure 2.11. Thus, angle $(a) = 6\pi$ and the singularity has order 2.

The general rule can be formulated as follows. Let us call *irregular pairs* to

$$(\pi_0^{-1}(1), L), \quad (\pi_1^{-1}(1), L), \quad (\pi_0^{-1}(d), R), \quad (\pi_1^{-1}(d), R).$$

All other pairs are called *regular*. Then there is an even number $2k$ of regular pairs in each equivalence class (one half above the horizontal axis and the other half below), and the angle of the corresponding singularity is equal to $2k\pi$.

This calculation remains valid when the closed curve $\Gamma(\pi, \lambda, \tau)$ has self-intersections. Let us explain this in the case when τ has type 0, the other one being symmetric. Then (π, λ) has type 1, according to Remark 2.1. Begin by writing

$$\pi = \begin{pmatrix} \cdots & \cdots & \cdots & A & \cdots & B & C_1 & \cdots & C_s \\ \cdots & B & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & A \end{pmatrix},$$

where $A = \alpha(1)$ and B is the leftmost symbol on the top row such that the side ζ_B contains some self-intersection. Recall that the suspension surface is defined from the simple polygon obtained by removing self-intersections in the way described in Figure 2.6. Combinatorially, this polygon corresponds to the permutation pair

$$\tilde{\pi} = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & A_1 & B_2 & C_1 & \cdots & C_s & \cdots & B_1 \\ \cdots & B_1 & B_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & A_1 \end{pmatrix},$$

and so the number and orders of the singularities are determined by the equivalence classes of $\tilde{\pi}$, according to the calculation described above. Our claim is that the same is true for the original permutation pair π. This can be seen as follows. Going from π to $\tilde{\pi}$ one replaces A, B by the symbols A_1, B_1, B_2. Consider the map ϕ defined by

$$\phi(A, L) = (A_1, L),\ \phi(A, R) = (A_1, R),\ \phi(B, L) = (B_1, L),\ \phi(B, R) = (B_2, R),$$

and $\phi(\alpha, S) = (\alpha, S)$ for any other (α, S). This projects down to a map ψ from the set of equivalence classes of π to the set of equivalence classes of $\tilde{\pi}$ (for the corresponding equivalence relations $\sim$). Moreover, ψ is injective and leaves invariant the number of regular pairs in each class. The map ψ is *not* surjective: the image avoids, exactly, the equivalence class

$$(B_1, R) \sim (A_1, R) \sim (B_2, L)$$

of $\tilde{\pi}$. However, this equivalence class contains exactly two regular pairs, and so it corresponds to a removable singularity. For consistency, we do remove this singularity from the structure of the suspension surface M. Thus, the number and order of the singularities of M can be obtained from the equivalence classes of π, as we claimed.

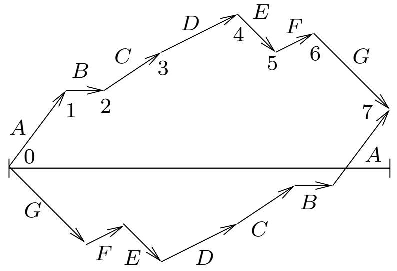


Figure 2.12:

Permutation σ. For computations, it is useful to introduce the following alternative terminology. Let us label the pairs (α, S) by integer numbers in the range $\{0, 1, \ldots, d\}$ as follows:

$$(\alpha, L) \leftrightarrow \pi_0(\alpha) - 1 \quad \text{and} \quad (\alpha, R) \leftrightarrow \pi_0(\alpha).$$

See Figure 2.12. Notice that this labeling incorporates (2.13). The remaining identifications can be expressed in terms of the monodromy invariant p:

$$j \sim k \quad \text{if } p(j) + 1 = p(k+1), \quad j \notin \{0, p^{-1}(d)\},$$

corresponding to (2.14), and $0 \sim p^{-1}(1) - 1$, corresponding to (2.15), and $p^{-1}(d) \sim d$, corresponding to (2.16). Moreover, these relations may be condensed into

$$j \sim \sigma(j) \quad \text{for every } 0 \leq j \leq d, \tag{2.17}$$

where $\sigma : \{0, 1, \ldots, d\} \to \{0, 1, \ldots, d\}$ is the transformation defined by

$$\sigma(j) = \begin{cases} p^{-1}(1) - 1 & \text{if } j = 0 \\ d & \text{if } j = p^{-1}(d) \\ p^{-1}\big(p(j) + 1\big) - 1 & \text{otherwise.} \end{cases} \tag{2.18}$$

It is clear from the construction that σ is a bijection of $\{0, 1, \ldots, d\}$, but that can also be checked directly, as follows. Extend p to a bijection P of the set $\{0, 1, \ldots, d, d+1\}$, simply, by defining $P(0) = 0$ and $P(d+1) = d+1$. Then (2.18) becomes

$$\sigma(j) = P^{-1}(P(j) + 1) - 1 \quad \text{for all } 0 \leq j \leq d. \tag{2.19}$$

This implies that σ is injective, because P is, and it is also clear that σ takes values in $\{0, 1, \ldots, d\}$. Thus, it is a bijection, as claimed.

In view of (2.17), the orbits of σ are in 1–to–1 correspondence to the equivalence classes of $\sim$. Therefore, the number κ of singularities coincides with the number of distinct orbits of σ. The rule for calculating the angles also translates easily to this terminology. Let us call $1, 2, \ldots, d-1$ *regular*, and 0 and d *irregular* vertices. Then the angle of each singularity a_i is given by

$$\text{angle}\,(a_i) = 2k_i\pi \tag{2.20}$$

where k_i is the number of regular vertices in the corresponding orbit of σ.

Remark 2.3. We have shown that κ and the a_i are determined by σ and, hence, by the monodromy invariant p. In particular, they are independent of λ and τ. This can be understood geometrically by noting that these integer invariants are locally constant on the parameters λ and τ and the domains $\mathbb{R}_+^{\mathcal{A}}$ and T_π^+ are connected, since they are convex cones.

Remark 2.4. Under the canonical involution $(\pi_0, \pi_1) \mapsto (\pi_1, \pi_0)$, the monodromy invariant is replaced by its inverse. Thus, the permutation σ is replaced by

$$\tilde{\sigma}(j) = P(P^{-1}(j) + 1) - 1 \quad \text{for all } 0 \le j \le d. \tag{2.21}$$

This is not quite the same as $\sigma^{-1}(j) = P^{-1}(P(j+1) - 1)$, but the two transformations are conjugate:

$$\tilde{\sigma} \circ P(j) = P(j+1) - 1 = P \circ \sigma^{-1}(j).$$

Thus, P maps each orbit of σ to an orbit of $\tilde{\sigma}$. In particular, both permutations have the same number of orbits and corresponding orbits contain the same number of vertices. This correspondence under the conjugacy P preserves the locations of the irregular vertices: the σ-orbit of 0 corresponds to the $\tilde{\sigma}$ orbit of 0, because $P(0) = 0$, and the σ-orbit of d corresponds to the $\tilde{\sigma}$ orbit of d, because the former is also the σ-orbit of $P^{-1}(d)$. It follows that corresponding orbits for σ and $\tilde{\sigma}$ so have the same number of *regular* vertices. This shows that the number and orders of the singularities are preserved by the canonical involution.

Proposition 2.5. *The number and the orders of the singularities are constant on each Rauzy class and, consequently, so is the genus.*

Proof. It suffices to prove that the number and the orders of the singularities corresponding to (π, λ) and $(\pi', \lambda') = \hat{R}(\pi, \lambda)$ always coincide. To this end, let p and p' be the monodromy invariants of (π, λ) and (π', λ'), respectively, and σ and σ' be the corresponding permutations of $\{0, 1, \ldots, d\}$ given by (2.18)–(2.19). Suppose first that (π, λ) has type 0. Then

$$p'(j) = \begin{cases} p(j) & \text{if } p(j) \le p(d) \\ p(j) + 1 & \text{if } p(d) < p(j) < d \\ p(d) + 1 & \text{if } p(j) = d \end{cases}$$

or, equivalently,

$$(p')^{-1}(j) = \begin{cases} p^{-1}(j) & \text{if } j \le p(d) \\ p^{-1}(d) & \text{if } j = p(d) + 1 \\ p^{-1}(j-1) & \text{if } p(d) + 1 < j \le d \end{cases}$$

(we suppose $p(d) \neq d - 1$, for otherwise $p' = p$ and so $\sigma' = \sigma$). This gives

$$\sigma'(j) = \begin{cases} p^{-1}(d) - 1 & \text{if } j = d \\ d & \text{if } p(j) = d - 1 \\ \sigma(d) & \text{if } p(j) = d \\ \sigma(j) & \text{in all other cases.} \end{cases}$$

This means that after Rauzy-Veech induction we have

$$p^{-1}(d-1) \xrightarrow{\sigma'} d \xrightarrow{\sigma'} p^{-1}(d) - 1 \quad \text{and} \quad p^{-1}(d) \xrightarrow{\sigma'} \sigma(d) \tag{2.22}$$

whereas, beforehand,

$$p^{-1}(d-1) \xrightarrow{\sigma} p^{-1}(d) - 1 \quad \text{and} \quad p^{-1}(d) \xrightarrow{\sigma} d \xrightarrow{\sigma} \sigma(d). \tag{2.23}$$

In other words, replacing σ by σ' means that d is displaced from the orbit of $p^{-1}(d)$ to the orbit of $p^{-1}(d-1)$ and $p^{-1}(d) - 1$, but the orbit structure is otherwise unchanged. Consequently, the two permutations have the same number of orbits, and corresponding orbits have the same number of *regular* vertices. It follows that the number and orders of the singularities remain the same. Now suppose (π, λ) has type 1. Let $\tilde{\pi}$ and $\tilde{\pi}'$ be obtained from π and π' by canonical involution. Then $(\tilde{\pi}, \lambda)$ has type zero, and $(\tilde{\pi}', \lambda') = \hat{R}(\tilde{\pi}, \lambda)$. So, by the previous paragraph, the number and orders of the singularities are the same for $(\tilde{\pi}, \lambda)$ and for $(\tilde{\pi}', \lambda')$. By Remark 2.4, the same is true about (π, λ) and $(\tilde{\pi}, \lambda)$, and about (π', λ') and $(\tilde{\pi}', \lambda')$. Thus, the number and orders of the singularities for (π, λ) and (π', λ) are also the same, as claimed. □

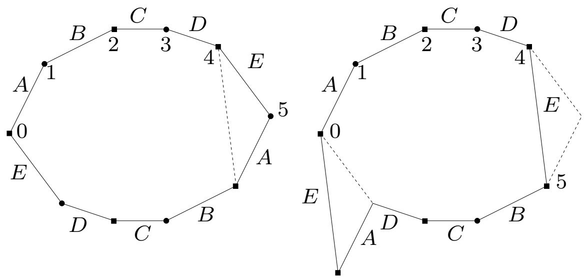


Figure 2.13:

Example 2.6. Figure 2.13 illustrates the vertex displacement in type 0 case of the proof of the proposition. One has $0 \to 4 \to 2 \to 0$ and $1 \to 5 \to 3 \to 5$ before inducing, and $0 \to 4 \to 2 \to 5 \to 0$ and $1 \to 3 \to 1$ afterwards.

In Section 2.7, we shall extend the Rauzy-Veech induction $\hat{R}(\pi, \lambda) = (\pi', \lambda')$ to an operator $\hat{\mathcal{R}}(\pi, \lambda, \tau) = (\pi', \lambda', \tau')$ in the space of translation surfaces, in such a way that the data (π, λ, τ) and (π', λ', τ') always define the same translation surface. As the number and the orders of the singularities depend only on the combinatorial data, by Remark 2.3, that will provide an alternative proof of Proposition 2.5.

2.5 Zippered rectangles

We are going to describe a useful alternative construction of the suspension of an interval exchange transformation, due to Veech [54]. Given an irreducible pair π and a vector $\tau \in \mathbb{R}^{\mathcal{A}}$, define $h \in \mathbb{R}^{\mathcal{A}}$ by

$$h_\alpha = -\sum_{\pi_1(\beta)<\pi_1(\alpha)} \tau_\beta + \sum_{\pi_0(\beta)<\pi_0(\alpha)} \tau_\beta = -\Omega_\pi(\tau)_\alpha. \tag{2.24}$$

Observe that if $\tau \in T_\pi^+$, that is, if it satisfies (2.10) then

$$\sum_{\pi_0(\beta)<\pi_0(\alpha)} \tau_\beta > 0 > \sum_{\pi_1(\beta)<\pi_1(\alpha)} \tau_\beta,$$

and so $h_\alpha > 0$, for all $\alpha \in \mathcal{A}$. We shall consider the convex cones inside the subspace $H_\pi = W_\pi = \Omega_\pi(\mathbb{R}^{\mathcal{A}})$ defined by

$$W_\pi^+ = \Omega_\pi(\mathbb{R}_+^{\mathcal{A}}) \quad \text{and} \quad H_\pi^+ = -\Omega_\pi(T_\pi^+). \tag{2.25}$$

Suppose $\tau \in T_\pi^+$. For each $\alpha \in \mathcal{A}$, consider the rectangles of width λ_α and height h_α defined by (see Figure 2.14)

$$R_\alpha^0 = \left(\sum_{\pi_0(\beta)<\pi_0(\alpha)} \lambda_\beta, \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \lambda_\beta \right) \times [0, h_\alpha]$$
$$R_\alpha^1 = \left(\sum_{\pi_1(\beta)<\pi_1(\alpha)} \lambda_\beta, \sum_{\pi_1(\beta)\le\pi_1(\alpha)} \lambda_\beta \right) \times [-h_\alpha, 0]$$

and consider also the vertical segments

$$S_\alpha^0 = \left\{ \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \lambda_\beta \right\} \times \left[0, \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \tau_\beta \right]$$
$$S_\alpha^1 = \left\{ \sum_{\pi_1(\beta)\le\pi_1(\alpha)} \lambda_\beta \right\} \times \left[\sum_{\pi_1(\beta)\le\pi_1(\alpha)} \tau_\beta, 0 \right].$$

That is, S_α^ε joins the horizontal axis to the endpoint of the vector

$$\sum_{\pi_\varepsilon(\beta)\le\pi_\varepsilon(\alpha)} \zeta_\beta = \sum_{\pi_\varepsilon(\beta)\le\pi_\varepsilon(\alpha)} (\lambda_\beta, \tau_\beta).$$

Notice that

$$S_{\alpha(0)}^0 = S_{\alpha(1)}^1 = \left\{ \sum_{\beta\in\mathcal{A}} \lambda_\beta \right\} \times \left[0, \sum_{\beta\in\mathcal{A}} \tau_\beta \right]. \tag{2.26}$$

Figure 2.14 describes two situations where this last segment is above and below the horizontal axis, respectively, depending on the type of τ.

The suspension surface $M = M(\pi, \lambda, \tau, h)$ is the quotient of the union

$$\bigcup_{\alpha\in\mathcal{A}} \bigcup_{\varepsilon=0,1} R_\alpha^\varepsilon \cup S_\alpha^\varepsilon$$

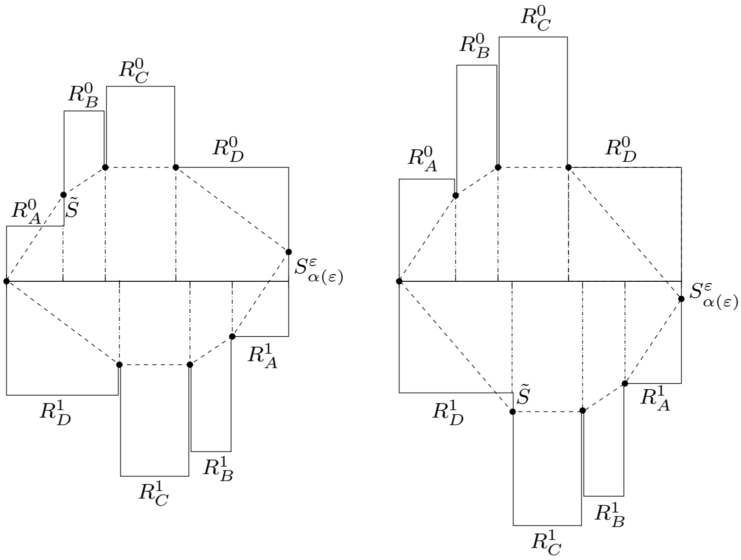


Figure 2.14:

of these objects by certain identifications, that we are going to describe. First, we identify each R^0_α to R^1_α through the translation

$$(x, z) \mapsto (x + w_\alpha, z - h_\alpha),$$

that maps one to the other. Note that this is just the same map we used before to identify the two sides of the polygon corresponding to the vector $\zeta_\alpha = (\lambda_\alpha, \tau_\alpha)$: recall (2.12).

We may think of the segments S^ε_α as "zipping" adjacent rectangles together *up to a certain height*. Observe that, in most cases, S^ε_α is shorter than the heights of both adjacent rectangles (compare Figure 2.14):

Lemma 2.7. *For any $\varepsilon \in \{0, 1\}$ and $\alpha \in \mathcal{A}$,*

1. $(-1)^\varepsilon \displaystyle\sum_{\pi_\varepsilon(\beta) \le \pi_\varepsilon(\alpha)} \tau_\beta < h_\alpha$ *except, possibly, if $\pi_{1-\varepsilon}(\alpha) = d$.*

2. $(-1)^\varepsilon \displaystyle\sum_{\pi_\varepsilon(\beta) \le \pi_\varepsilon(\alpha)} \tau_\beta \le h_\gamma$*, where $\gamma \in \mathcal{A}$ is defined by $\pi_\varepsilon(\gamma) = \pi_\varepsilon(\alpha) + 1$ and we suppose $\pi_\varepsilon(\alpha) < d$. The inequality is strict unless $\pi_{1-\varepsilon}(\gamma) = 1$.*

Proof. For $\varepsilon = 0$ the relations (2.10) and (2.24) give

$$h_\alpha - \sum_{\pi_0(\beta) \le \pi_0(\alpha)} \tau_\beta = - \sum_{\pi_1(\beta) \le \pi_1(\alpha)} \tau_\beta > 0 \tag{2.27}$$

except, possibly, if $\pi_1(\alpha) = d$, that is, $\alpha = \alpha(1)$. This takes care of the rectangle to the left of S^0_α. The rectangle to the right (when it exists) is handled similarly: suppose $\pi_0(\alpha) < d$ and let $\gamma \in \mathcal{A}$ be such that $\pi_0(\gamma) = \pi_0(\alpha) + 1$. Then

$$h_\gamma - \sum_{\pi_0(\beta) \le \pi_0(\alpha)} \tau_\beta = - \sum_{\pi_1(\beta) < \pi_1(\gamma)} \tau_\beta \ge 0,$$

and the inequality is strict unless $\pi_1(\gamma) = 1$. The case $\varepsilon = 1$ is analogous. $\square$

On the other hand, the calculation in (2.27) also shows that for $\alpha = \alpha(1)$ the length of S^0_α may exceed the height of R^0_α: this happens if the sum of all τ_β is positive. In that case, let

$$\tilde{S} = \left\{ \sum_{\pi_0(\beta) \le \pi_0(\alpha)} \lambda_\beta \right\} \times \left[h_\alpha, \sum_{\pi_0(\beta) \le \pi_0(\alpha)} \tau_\beta \right], \tag{2.28}$$

that is, $\tilde{S}$ is the subsegment of length $\sum_{\beta \in \mathcal{A}} \tau_\beta$ at the top of S^0_α. Dually, if the sum of all τ_β is negative then, for $\alpha = \alpha(0)$, the length of S^1_α exceeds the height of R^1_α. In this case, define

$$\tilde{S} = \left\{ \sum_{\pi_0(\beta) \le \pi_0(\alpha)} \lambda_\beta \right\} \times \left[\sum_{\pi_1(\beta) \le \pi_1(\alpha)} \tau_\beta, -h_\alpha \right], \tag{2.29}$$

instead. That is, $\tilde{S}$ is the subsegment of length $-\sum_{\beta \in \mathcal{A}} \tau_\beta$ at the bottom of S^1_α. In either case, we identify $\tilde{S}$ with the vertical segment $S^0_{\alpha(0)} = S^1_{\alpha(1)}$, by translation. This completes the definition of the suspension surface.

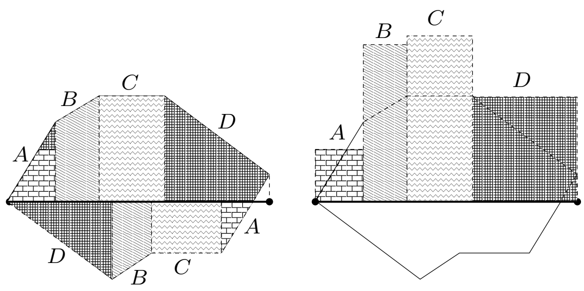


Figure 2.15:

We say that two translation surfaces are *isomorphic* if there exists an isometry between the two that preserves the vertical direction. The construction we have just described is equivalent to the one in Section 2.2, in the sense that they give rise to suspension surfaces that are isomorphic. This is clear from the previous observations, at least when the closed curve $\Gamma(\pi, \lambda, \tau)$ is simple. See

Figure 2.15. We leave it to the reader to check that the same is true when there are self-intersections.

There is a natural notion of *area* for zippered rectangles defined in terms of (π, λ, τ, h), namely

$$\operatorname{area}(\pi, \lambda, \tau, h) = \lambda \cdot h = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha \, h_\alpha. \tag{2.30}$$

Sometimes we write $\operatorname{area}(\pi, \lambda, \tau)$ to mean $\operatorname{area}(\pi, \lambda, \tau, h)$ with $h = -\Omega_\pi(\tau)$.

2.6 Genus and dimension

We have seen in Remark 1.31 that the vector space $H_\pi = \Omega_\pi(\mathbb{R}^{\mathcal{A}})$ has even dimension. We can now interpret this dimension in terms of the genus of the suspension surface:

Proposition 2.8. *The dimension of H_π coincides with $2g(M)$, where $g(M)$ is the genus of the suspension surface M.*

Proof. Rename the intervals I_α so that the permutation pair π becomes normalized to $\mathcal{A} = \{1, \ldots, d\}$, $\pi_0 = \mathrm{id}$ and, thus, $\pi_1 = p =$ monodromy invariant. Write the translation vector as $w = \Omega_\pi(\lambda)$, that is

$$w_j = \sum_{p(i) < p(j)} \lambda_i - \sum_{i<j} \lambda_i \quad \text{for each } 1 \le j \le d.$$

It is convenient to extend the definition to $j = 0$ and $j = d+1$, simply, by replacing p by its extension P in (2.19). Since $P(0) = 0$ and $P(d+1) = d+1$, by definition, this just means we take $w_0 = w_{d+1} = 0$. Define $a_j = \sum_{i \le j} \lambda_i$ for $1 \le j \le d$, and $a_0 = 0$.

Lemma 2.9. *We have $w_{\sigma(j)+1} - w_j = a_j - a_{\sigma(j)}$ for every $0 \le j \le d$.*

Proof. As we have see in (2.19), $\sigma(j) = P^{-1}(P(j)+1) - 1$, and so

$$w_{\sigma(j)+1} = \sum_{P(i) < P(\sigma(j)+1)} \lambda_i - \sum_{i < \sigma(j)+1} \lambda_i = \sum_{P(i) \le P(j)} \lambda_i - \sum_{i \le \sigma(j)} \lambda_i.$$

It follows that

$$w_{\sigma(j)+1} - w_j = \lambda_j - \sum_{i \le \sigma(j)} \lambda_i + \sum_{i<j} \lambda_i = \sum_{i \le j} \lambda_i - \sum_{i \le \sigma(j)} \lambda_i = a_j - a_{\sigma(j)}$$

as claimed. □

Recall that the number of orbits of σ is equal to the number κ of singularities.

Lemma 2.10. *A vector λ is in $\ker \Omega_\pi$ if and only if the $(d+1)$-dimensional vector $(0, a_1, \ldots, a_d)$ is constant on the orbits of σ. Hence, $\dim \ker \Omega_\pi = \kappa - 1$.*

Proof. The only if part is a direct consequence of Lemma 2.9: if $w = 0$ then $a_{\sigma(j)} - a_j = 0$ for every $0 \le j \le d$. To prove the converse, let λ be such that $(0, a_1, \ldots, a_d)$ is constant on orbits of σ. Then, by Lemma 2.9,

$$w_{P^{-1}(P(j)+1)} = w_{\sigma(j)+1} = w_j \quad \text{for all } 0 \le j \le d.$$

Writing $P(j) = i$, this relation becomes

$$w_{P^{-1}(i+1)} = w_{P^{-1}(i)} \quad \text{for all } 0 \le i \le d.$$

It follows that $w_{P^{-1}(i)}$ is constant on $\{0, 1, \ldots, d+1\}$. Then, since it vanishes for $i = 0$, it must vanish for every $1 \le i \le d$. Consequently, $w = (w_1, \ldots, w_d)$ vanishes, and this means that $\lambda \in \ker \Omega_\pi$. This proves the first part of the lemma.

To prove the second one, consider the linear isomorphism

$$\psi : \mathbb{R}^d \to \mathbb{R}^d, \quad (\lambda_1, \ldots, \lambda_d) \mapsto (a_1, \ldots, a_d), \quad a_j = \sum_{i=1}^{j} \lambda_i. \tag{2.31}$$

Let K_π be the subspace of all $(a_1, \ldots, a_d) \in \mathbb{R}^d$ such that $(0, a_1, \ldots, a_d)$ is constant on the orbits of σ. The dimension of K_π is $\kappa - 1$, because the value of a_j on the orbit of 0 is predetermined by $a_0 = 0$. The previous paragraph shows that

$$\ker \Omega_\pi = \psi^{-1}(K_\pi).$$

Consequently, the dimension of the kernel is $\kappa - 1$, as claimed. □

Using Lemma 2.10 and the relation (2.8), we find

$$\dim \Omega_\pi(\mathbb{R}^{\mathcal{A}}) = d - \dim \ker \Omega_\pi = d - \kappa + 1 = 2g(M).$$

This proves Proposition 2.8. □

It is possible to give an explicit description of $\ker \Omega_\pi$ and H_π, as follows. For each orbit $\mathcal{O}$ of σ not containing zero, and for each $1 \le j \le d$, define

$$\lambda(\mathcal{O})_j = \mathcal{X}_{\mathcal{O}}(j) - \mathcal{X}_{\mathcal{O}}(j-1) = \begin{cases} 1 & \text{if } j \in \mathcal{O} \text{ but } j-1 \notin \mathcal{O} \\ -1 & \text{if } j \notin \mathcal{O} \text{ but } j-1 \in \mathcal{O} \\ 0 & \text{in all other cases.} \end{cases} \tag{2.32}$$

Lemma 2.11. *Define $a(\mathcal{O}) = \psi(\lambda(\mathcal{O}))$, that is, $a(\mathcal{O})_j = \sum_{i \le j} \lambda(\mathcal{O})_i$. Then*

$$a(\mathcal{O})_j = \mathcal{X}_{\mathcal{O}}(j) = \begin{cases} 1 & \textit{if } j \in \mathcal{O} \\ 0 & \textit{if } j \notin \mathcal{O}. \end{cases}$$

Proof. For $j = 1$ this follows from a simple calculation: $a(\mathcal{O})_1 = 1$ if $1 \in \mathcal{O}$ (and $0 \notin \mathcal{O}$) and $a(\mathcal{O})_1 = 0$ if $1 \notin \mathcal{O}$ (and $0 \notin \mathcal{O}$). The proof proceeds by induction: if $a(\mathcal{O})_{j-1} = \mathcal{X}_{\mathcal{O}}(j-1)$ then

$$a(\mathcal{O})_j = a(\mathcal{O})_{j-1} + \lambda(\mathcal{O})_j = \mathcal{X}_{\mathcal{O}}(j-1) + \mathcal{X}_{\mathcal{O}}(j) - \mathcal{X}_{\mathcal{O}}(j-1) = \mathcal{X}_{\mathcal{O}}(j).$$

The argument is complete. □

Clearly, the $a(\mathcal{O})$ form a basis of the subspace K_π of vectors $(a_1, \ldots, a_d)$ such that $(0, a_1, \ldots, a_d)$ is constant on orbits of σ. It follows that

$$\{\lambda(\mathcal{O}) : \mathcal{O} \text{ is an orbit of } \sigma \text{ not containing } 0\}$$

is a basis of $\ker \Omega_\pi$. Moreover, since Ω_π is anti-symmetric, the range H_π is just the orthogonal complement of the kernel. In other words, $w \in H_\pi$ if and only if $w \cdot \lambda(\mathcal{O}) = 0$ for every orbit $\mathcal{O}$ of σ not containing zero.

Hyperelliptic Rauzy classes. Let us carry out the previous calculation in a specific case. Let $d \geq 2$ be fixed. We call *hyperelliptic* the Rauzy class that contains the pair,

$$\pi = \begin{pmatrix} A_1 & A_2 & \cdots & \cdots & A_d \\ A_d & \cdots & \cdots & A_2 & A_1 \end{pmatrix}$$

that is, which corresponds to the monodromy invariant $p = (d, d-1, \cdots, 2, 1)$ defined by $p(i) = d + 1 - i$ for all i.

Lemma 2.12.

1. *If d is even then the number of singularities $\kappa(\pi) = 1$, the singularity has order $d - 2$ and the surface M has genus $g(M) = d/2$. Moreover, the operator $\Omega_\pi : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ is an isomorphism.*

2. *If d is odd then there are $\kappa(\pi) = 2$ singularities, and they both have order $(d-3)/2$. The surface M has genus $g(M) = (d-1)/2$. Moreover, the kernel of Ω_π has dimension 1.*

Proof. Observe that $p^{-1}(i) = p(i) = d + 1 - i$ for all $1 \leq i \leq d$. From (2.18) we find that the permutation σ is given by

$$\sigma(j) = \begin{cases} d-1 & \text{for } j = 0 \\ d & \text{for } j = 1 \\ j-2 & \text{in all other cases.} \end{cases}$$

That is, σ is the right rotation by two units

$$\sigma = (d-1, d, 0, 1, \ldots, d-2).$$

If d is even, then this rotation has a unique orbit in $\{0, 1, \ldots, d\}$. It follows that $\kappa = 1$ and, by (2.6) the singularity has angle $(2d-2)\pi$, that is, order $d - 2$. Moreover, (2.8) gives $g(M) = d/2$. If d is odd then the rotation has exactly two orbits:

$$0 \to d-1 \to d-3 \to \cdots \to 2 \to 0 \quad \text{and} \quad d \to d-2 \to d-4 \to \cdots \to 1 \to d.$$

Each one involves $(d-1)/2$ regular elements (that is, different from 0 and d). Using (2.20) we get that they both have angle $(a_i) = (d-1)\pi$, and so their order is $(d-3)/2$. Moreover, (2.8) gives $g(M) = (d-1)/2$.

The statement about Ω_π is now an immediate consequence of Proposition 2.8, but it may also be proved directly. To this end, let us normalize the permutation pair π (rename the intervals) so that $\mathcal{A} = \{1, \ldots, d\}$, $\pi_0 = \mathrm{id}$ and, thus, $\pi_1 =$ monodromy invariant p. Then $\Omega_\pi(\lambda) = w$ is given by

$$w_j = \sum_{\pi_1(i)<\pi_1(j)} \lambda_i - \sum_{\pi_0(i)<\pi_0(j)} \lambda_i = \sum_{i>j} \lambda_i - \sum_{i<j} \lambda_i.$$

This gives $w_j - w_{j+1} = \lambda_j + \lambda_{j+1}$ for $j = 1, \ldots, d-1$, and also $w_d + w_1 = \lambda_d - \lambda_1$. Suppose λ is in the kernel, that is, $w = 0$. Then the λ_j must be alternately symmetric, and the first and the last one must coincide: $\lambda_1 = \lambda_d$. If d is even this can only happen for $\lambda = 0$: thus, Ω_π is an isomorphism. If d is odd, it means that $\lambda = (x, -x, x, -x, \cdots, -x)$ for some real number x. It is easy to check that vectors of this form are, indeed, in the kernel. This proves that the kernel of Ω_π has dimension 1 in this case. □

The relation (2.8) shows that d and κ always have opposite parities. So, the situation described in Lemma 2.12 corresponds to the smallest possible number of singularities.

2.7 Invertible Rauzy-Veech induction

We are going to define a counterpart $\hat{\mathcal{R}} : (\pi, \lambda, \tau) \mapsto (\pi', \lambda', \tau')$ of the Rauzy-Veech induction $\hat{R} : (\pi, \lambda) \mapsto (\pi', \lambda')$ at the level of suspension data (π, λ, τ). Recall that $\hat{R}$ corresponds to replacing the original interval exchange map by its first return to a conveniently chosen subinterval of the domain. Similarly, this map $\hat{\mathcal{R}}$ we are introducing corresponds to replacing the horizontal cross-section in (2.11) by a shorter one. The Poincaré return map of the vertical flow to this new cross-section is precisely the interval exchange map described by (π', λ'), and we want to rewrite the ambient surface as a suspension over this map: the coordinate τ' is chosen with this purpose in mind. Thus, the data (π, λ, τ) and (π', λ', τ') are really different presentations of the same translation surface. We shall check that the transformation $\hat{\mathcal{R}}$ is invertible almost everywhere and has a Markov property. Later, we shall see that it is a realization of the inverse limit (natural extension) of $\hat{R}$.

Let $\hat{\mathcal{H}} = \hat{\mathcal{H}}(C) = \{(\pi, \lambda, \tau) : \pi \in C,\ \lambda \in \mathbb{R}^{\mathcal{A}}_+,\ \tau \in T^+_\pi\}$. The transformation $\hat{\mathcal{R}}$ is defined on $\hat{\mathcal{H}}$ by $\hat{\mathcal{R}}(\pi, \lambda, \tau) = (\pi', \lambda', \tau')$, where $(\pi', \lambda') = \hat{R}(\pi, \lambda)$ and

$$\tau'_\alpha = \begin{cases} \tau_\alpha & \alpha \neq \alpha(\varepsilon) \\ \tau_{\alpha(\varepsilon)} - \tau_{\alpha(1-\varepsilon)} & \alpha = \alpha(\varepsilon) \end{cases} \qquad \varepsilon = \text{type of } (\pi, \lambda).$$

In other words (compare (1.11) for the definition of λ'),

$$\tau' = \Theta^{-1*}(\tau). \tag{2.33}$$

Figure 2.16 (case $\varepsilon = 0$) and Figure 2.17 (case $\varepsilon = 1$) provide a geometric interpretation of this Rauzy-Veech induction, in terms of the polygon defining the suspension surface: one cuts from the polygon the triangle determined

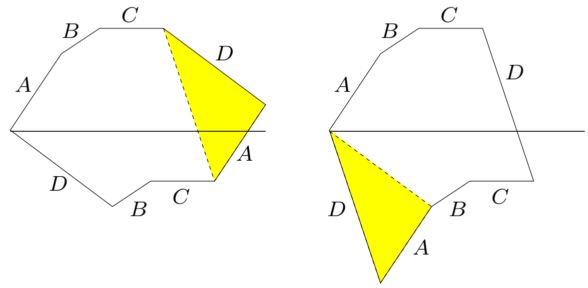


Figure 2.16:

by the sides $\zeta_{\alpha(0)}$ and $-\zeta_{\alpha(1)}$ and pastes it back, adjacently to the other side labeled by $\alpha(\varepsilon)$, where $\varepsilon=$ type of (π,λ). Observe that the surface itself remains unchanged or, rather, the translation surfaces determined by (π,λ,τ) and (π',λ',τ') are isomorphic: there exists an isometry between the two that preserves the vertical direction. We leave it to the reader to check how this geometric interpretation extends to the case when the closed curve $\gamma(\pi,\lambda,\tau)$ has self-intersections. An equivalent formulation of the Rauzy-Veech induction in terms of zippered rectangles will be given in Section 2.8.

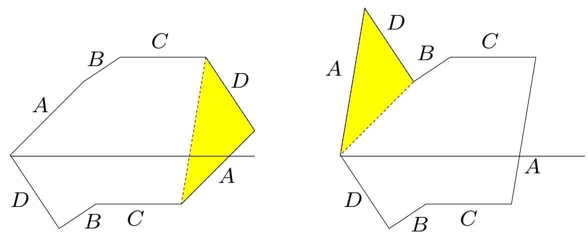


Figure 2.17:

Recall that we defined the type of τ to be 0 if the sum of τ_α over all $\alpha\in\mathcal{A}$ is positive and 1 if the sum is negative. Figures 2.16 and 2.17 immediately suggest that

$$(\pi,\lambda)\text{ has type }\varepsilon \quad\Leftrightarrow\quad \tau'\text{ has type }1-\varepsilon. \tag{2.34}$$

This observation is also contained in the next, more precise, lemma. See also Figure 2.18, that describes the action of $\hat{\mathcal{R}}$ on both variables λ and τ.

Lemma 2.13. *The linear transformation Θ^{-1*} sends T^+_π injectively inside $T^+_{\pi'}$ and, denoting $\varepsilon=$ type of (π,λ), the image coincides with the set of $\tau'\in T^+_{\pi'}$ whose type is $1-\varepsilon$.*

Proof. Suppose $\varepsilon=0$, as the other case is analogous. We begin by checking that the image of Θ^{-1*} is contained in $T^+_{\pi'}$, that is, τ' satisfies (2.10) if τ does.

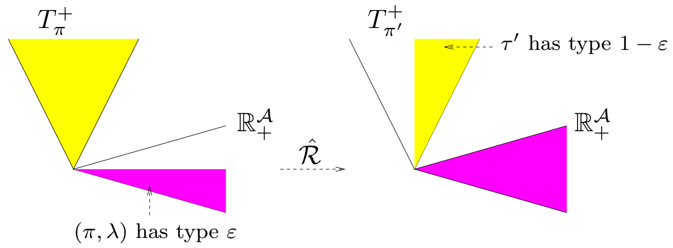


Figure 2.18:

Firstly, $\pi_0' = \pi_0$ and $\tau_\alpha' = \tau_\alpha$ for every $\alpha \neq \alpha(0)$ imply

$$\sum_{\pi_0'(\alpha)\le k} \tau_\alpha' = \sum_{\pi_0(\alpha)\le k} \tau_\alpha > 0 \tag{2.35}$$

for every $k < d$. Now let $l = \pi_1(\alpha(0))$ be the position of $\alpha(0)$ in the bottom line of π. Recall that π_1' and π_1 coincide to the left of l. So, just as before,

$$\sum_{\pi_1'(\alpha)\le k} \tau_\alpha' = \sum_{\pi_1(\alpha)\le k} \tau_\alpha < 0 \tag{2.36}$$

for every $k < l$. The case $k = l$ is more interesting: using $\tau_{\alpha(0)}' = \tau_{\alpha(0)} - \tau_{\alpha(1)}$

$$\sum_{\pi_1'(\alpha)\le l} \tau_\alpha' = \sum_{\pi_1(\alpha)\le l} \tau_\alpha - \tau_{\alpha(1)}.$$

To prove that this is less than zero, rewrite the right hand side as (recall the definition (2.24) of h)

$$-h_{\alpha(0)} + \sum_{\pi_0(\alpha)\le l} \tau_\alpha - \tau_{\alpha(1)} = -h_{\alpha(0)} + \sum_{\alpha\in\mathcal{A}} \tau_\alpha - \tau_{\alpha(1)} = -h_{\alpha(0)} + \sum_{\pi_1(\alpha)<\pi_1(\alpha(1))} \tau_\alpha.$$

Both terms in the last expression are negative, because the entries of h are positive and τ satisfies (2.10). This deals with the case $k = l$. Next, for $k = l+1$, we use the fact that $\pi_1'(\alpha(1)) = l+1$ to obtain

$$\sum_{\pi_1'(\alpha)\le l+1} \tau_\alpha' = \sum_{\pi_1(\alpha)\le l} \tau_\alpha < 0. \tag{2.37}$$

More generally, for $l < k \le d$ we have

$$\sum_{\pi_1'(\alpha)\le k} \tau_\alpha' = \sum_{\pi_1(\alpha)\le k-1} \tau_\alpha < 0. \tag{2.38}$$

This proves that the image of T_π^+ is indeed contained in $T_{\pi'}^+$. Moreover, the case $k = d$ gives that every τ' in the image has type 1,

$$\sum_{\alpha\in\mathcal{A}} \tau_\alpha' < 0, \tag{2.39}$$

as claimed. To complete the proof we only have to check that if $\tau' \in T^+_{\pi'}$ satisfies (2.39) then $\tau = \Theta^*(\tau')$ is in T^+_π. This is easily seen from the relations (2.35)-(2.38). The hypothesis (2.39) is needed only when $k = d-1$. $\square$

Recall that the Rauzy-Veech induction $\hat{R} : (\pi, \lambda) \mapsto (\pi', \lambda')$ for interval exchange transformations is 2-to-1 on its domain, the two pre-images corresponding to the two possible values of the type ε. For each $\varepsilon \in \{0, 1\}$, let us denote

$$\mathbb{R}^{\mathcal{A}}_{\pi,\varepsilon} = \{\lambda \in \mathbb{R}^{\mathcal{A}}_+ : (\pi, \lambda) \text{ has type } \varepsilon\} \quad \text{and} \quad T^\varepsilon_\pi = \{\tau \in T^+_\pi : \tau \text{ has type } \varepsilon\}$$

From the previous lemma we obtain

Corollary 2.14. *The transformation $\hat{\mathcal{R}} : \hat{\mathcal{H}} \to \hat{\mathcal{H}}$ is an (almost everywhere) invertible Markov map, and it preserves the natural area:*

1. $\hat{\mathcal{R}}\big(\{\pi\} \times \mathbb{R}^{\mathcal{A}}_{\pi,\varepsilon} \times T^+_\pi\big) = \{\pi'\} \times \mathbb{R}^{\mathcal{A}}_+ \times T^{1-\varepsilon}_{\pi'}$ *for every π and ε;*
2. *every (π', λ', τ') such that $\sum_{\alpha \in \mathcal{A}} \tau'_\alpha \neq 0$ has exactly one preimage for $\hat{\mathcal{R}}$;*
3. *if $\hat{\mathcal{R}}(\pi, \lambda, \tau) = (\pi', \lambda', \tau')$ then* $\operatorname{area}(\pi, \lambda, \tau) = \operatorname{area}(\pi', \lambda', \tau')$.

Proof. The first claim is contained in Lemma 2.13. The second one follows from the injectivity in that lemma, together with the observation that the sets $\{\pi'\} \times \mathbb{R}^{\mathcal{A}}_+ \times T^{1-\varepsilon}_{\pi'}$ are pairwise disjoint. Finally, Lemma 1.29 and the relations (1.11) and (2.40) give

$$-\lambda' \cdot \Omega_{\pi'}(\tau') = \lambda' \cdot h' = \Theta^{-1*}(\lambda) \cdot \Theta(h) = \lambda \cdot h = -\lambda \cdot \Omega_\pi(\tau).$$

and this proves the third claim. $\square$

Remark 2.15. Let ε be the type of (π, λ). If τ also has type ε then the curve $\Gamma(\pi, \lambda, \tau)$ is simple, according to Remark 2.15. Otherwise, let $n \geq 1$ be minimum such that the type of (π^n, λ^n) is $1 - \varepsilon$. By (2.34), the type of τ^n is also $1 - \varepsilon$. It follows that the curve $\Gamma(\pi^n, \lambda^n, \tau^n)$ has no self-intersections. Recall that (π, λ, τ) and $(\pi^n, \lambda^n, \tau^n)$ represent the same translation surface, up to an isometry that preserves the vertical direction.

2.8 Induction for zippered rectangles

The definition of the induction operator $\hat{\mathcal{R}}$ is, perhaps, more intuitive in the language of zippered rectangles. Indeed, as explained previously, the idea behind the definition is to rewrite the translation surface as a suspension of the Poincaré return map of the vertical flow to a shorter cross-section. In terms of zippered rectangles this is achieved by an especially simple geometric procedure, described in Figure 2.19: one removes a rightmost subrectangle from the rectangle corresponding to the symbol $\alpha(\varepsilon)$ and pastes it back on top of the

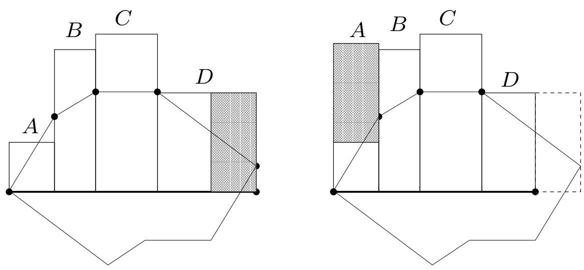


Figure 2.19:

rectangle corresponding to the symbol $\alpha(1-\varepsilon)$. The precise definition goes as follows.

Let $\tilde{\mathcal{H}} = \tilde{\mathcal{H}}(C)$ be the set of (π, λ, τ, h) such that $\pi \in C$, $\lambda \in \mathbb{R}^{\mathcal{A}}_{+}$, $\tau \in T^{+}_{\pi}$, and $h = -\Omega_\pi(\tau) \in H^{+}_{\pi}$. Then define $\hat{\mathcal{R}}(\pi, \lambda, \tau, h) = (\pi', \lambda', \tau', h')$, where

$$h'_\alpha = \begin{cases} h_\alpha & \alpha \neq \alpha(1-\varepsilon) \\ h_{\alpha(1-\varepsilon)} + h_{\alpha(\varepsilon)} & \alpha = \alpha(1-\varepsilon). \end{cases}$$

Compare Figure 2.19. Equivalently (recall (1.8)),

$$h' = \Theta(h). \tag{2.40}$$

Let us relate this to the definition $\hat{\mathcal{R}}(\pi, \lambda, \tau) = (\pi', \lambda', \tau')$ with $\tau' = \Theta^{-1*}(\tau)$ that was given in the previous section.

By Lemma 1.29, we have $\Theta\,\Omega_\pi\,\Theta^* = \Omega_{\pi'}$. Since Θ is an isomorphism, this gives that $\tau \in \ker \Omega_\pi$ if and only if $\tau' \in \ker \Omega_{\pi'}$. In other words,

$$\{\ker \Omega_\pi : \pi \in C\}$$

defines an invariant subbundle for $\tau \mapsto \tau' = \Theta^{-1*}(\tau)$. As we have seen before, H_π is the orthogonal complement of the kernel, because Ω_π is anti-symmetric. Hence

$$\{H_\pi : \pi \in C\}$$

is an invariant subbundle for the adjoint cocycle Θ. The map defined by (2.40) is just the restriction of the adjoint to this invariant subbundle. The relation $\Theta\,\Omega_\pi\,\Theta^* = \Omega_{\pi'}$ also says that there is a conjugacy

$$\begin{tikzcd} \mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi \arrow[r, "\Theta^{-1*}"] \arrow[d, "\Omega_\pi"'] & \mathbb{R}^{\mathcal{A}}/\ker\Omega_{\pi'} \arrow[d, "\Omega_{\pi'}"] \\ H_\pi \arrow[r, "\Theta"] & H_{\pi'} \end{tikzcd} \tag{2.41}$$

In fact, not much information is lost by passing to the quotient $\mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi$, because the dynamics of Θ^{-1*} on the invariant subbundle $\{\ker\Omega_\pi : \pi \in C\}$ is rather trivial: it just maps the elements of the bases $\{\lambda(\mathcal{O})\}$ defined in (2.32) to one another, as we are going to see in the next lemma.

Lemma 2.16. *Given π and $\varepsilon \in \{0,1\}$, let π' be the type ε successor of π. Let σ and σ' be the permutations of $\{0,1,\dots,d\}$ corresponding to π and π', respectively. Then for each σ-orbit $\mathcal{O}$ not containing 0 there exists a σ'-orbit not containing zero such that $\Theta^{-1*}\lambda(\mathcal{O}) = \lambda(\mathcal{O}')$.*

Proof. Consider first $\varepsilon = 0$. According to (2.22)–(2.23), to each orbit $\mathcal{O}$ of σ we may associate an orbit $\mathcal{O}'$ of σ' that contains exactly the same regular vertices, that is, the same elements different from 0 and d. Notice that $\mathcal{O}$ contains 0 if and only $\mathcal{O}'$ contains 0; we are only interested in the case where this does not happen. Denote $\lambda' = \Theta^{-1*}(\lambda(\mathcal{O}))$ and also $a(\mathcal{O}) = \psi(\lambda(\mathcal{O}))$ and $a' = \psi(\lambda')$, where ψ is the linear operator defined in (2.31). Lemma 2.11 gives that $a(\mathcal{O}) = \mathcal{X}_{\mathcal{O}}$ and then, to conclude that $\lambda' = \lambda(\mathcal{O}')$, we only have to show that $a' = \mathcal{X}_{\mathcal{O}'}$. To this end, recall that

$$\lambda'_j = \lambda(\mathcal{O})_j = \mathcal{X}_{\mathcal{O}}(j) - \mathcal{X}_{\mathcal{O}}(j-1) \quad \text{if } j \neq d, \quad \text{and} \quad \lambda'_d = \lambda(\mathcal{O})_d - \lambda(\mathcal{O})_{p^{-1}(d)}.$$

The first relation gives

$$a'_j = a(\mathcal{O})_j = \mathcal{X}_{\mathcal{O}}(j) = \mathcal{X}_{\mathcal{O}'}(j) \quad \text{for all } 1 \leq j < d.$$

Then the second one gives $a'_d = a(\mathcal{O})_d - \lambda(\mathcal{O})_{p^{-1}(d)}$. There are several subcases to consider, according to the relations (2.22)–(2.23):

- If $d \in \mathcal{O}' \setminus \mathcal{O}$ then $a'_d = 0 + 1 = \mathcal{X}_{\mathcal{O}'}(d)$ because $p^{-1}(d) - 1 \in \mathcal{O}$ and $p^{-1}(d) \notin \mathcal{O}$, and so $\lambda(\mathcal{O})_{p^{-1}(d)} = -1$.
- If $d \in \mathcal{O} \setminus \mathcal{O}'$ then $a'_d = 1 - 1 = \mathcal{X}_{\mathcal{O}'}(d)$ because $p^{-1}(d) - 1 \notin \mathcal{O}$ and $p^{-1}(d) \in \mathcal{O}$, and so $\lambda(\mathcal{O})_{p^{-1}(d)} = 1$.
- If $d \notin \mathcal{O} \cup \mathcal{O}'$ then $a'_d = 0 + 0 = \mathcal{X}_{\mathcal{O}'}(d)$ because neither $p^{-1}(d) - 1$ nor $p^{-1}(d)$ are in $\mathcal{O}$.
- If $d \in \mathcal{O} \cap \mathcal{O}'$ then $\mathcal{O} = \mathcal{O}'$ and both $p^{-1}(d) - 1$ and $p^{-1}(d)$ belong to it. Consequently, $a'_d = 1 - 0 = \mathcal{X}_{\mathcal{O}'}(d)$ also in this case.

This settles the type 0 case.

The case $\varepsilon = 1$ can now be deduced as follows. Let $\tilde{\pi}$ and $\tilde{\pi}'$ be obtained from π and π' by canonical involution, that is, by exchanging the top and bottom lines. Let $\tilde{\sigma}$ and $\tilde{\sigma}'$ be the permutations of $\{0,1,\dots,d\}$ corresponding to $\tilde{\pi}$ and $\tilde{\pi}'$, respectively. According to Remark 2.4,

$$\tilde{\sigma} = P \circ \sigma^{-1} \circ P^{-1} \quad \text{and} \quad \tilde{\sigma}' = (P') \circ (\sigma')^{-1} \circ (P')^{-1}$$

where P and P' are the monodromy invariants of π and π', respectively. Then, given any orbit $\mathcal{O}$ of σ we have that $\tilde{\mathcal{O}} = P(\mathcal{O})$ is an orbit of $\tilde{\sigma}$, and the latter contains 0 if and only if the former does. Notice that $\tilde{\pi}'$ is the type 0 successor of $\tilde{\pi}$. Hence, we may associate to $\tilde{\mathcal{O}}$ an orbit $\tilde{\mathcal{O}}'$ of $\tilde{\sigma}'$ containing exactly the same regular vertices. Moreover, $\mathcal{O}' = (P')^{-1}(\tilde{\mathcal{O}}')$ is an orbit of σ'. By case $\varepsilon = 0$ of the present lemma, we have

$$\tilde{\Theta}^{-1*}(a(\tilde{\mathcal{O}})) = a(\tilde{\mathcal{O}}') \tag{2.42}$$

where $\tilde{\Theta} = \Theta_{\tilde{\pi},\lambda}$ for any λ such that $(\tilde{\pi},\lambda)$ has type 0. We are going to deduce that

$$\Theta^{-1*}(a(\mathcal{O})) = a(\mathcal{O}'). \tag{2.43}$$

To this end, begin by observing that the definitions and Lemma 2.11 imply

$$a(\mathcal{O})_j = a(\tilde{\mathcal{O}})_{P(j)} \quad \text{and} \quad a(\mathcal{O}')_j = a(\tilde{\mathcal{O}}')_{P'(j)} \quad \text{for all } j = 1, \dots, d. \tag{2.44}$$

Recall also, from Remark 1.8, that the operators Θ^{-1*} and $\tilde{\Theta}^{-1*}$ coincide. However, their matrices that are relevant in this context are not the same, because they are with respect to *different bases*. The reason is that, with the notations we are using, canonical involution involves reordering of the alphabet:

$$\begin{array}{lll} \tilde{\pi} = \begin{pmatrix} 1 & 2 & \cdots & d \\ \alpha_1 & \alpha_2 & \cdots & \alpha_d \end{pmatrix} & \mapsto & \qquad\qquad \alpha_j = P(j) \\ \pi = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_d \\ 1 & 2 & \cdots & d \end{pmatrix} & \sim \begin{pmatrix} 1 & 2 & \cdots & d \\ \beta_1 & \beta_2 & \cdots & \beta_d \end{pmatrix} & \qquad \beta_j = P^{-1}(j), \end{array}$$

and similarly for π'. Consequently, the relevant matrix $(\tilde{\Theta}_{i,j})$ of the operator $\tilde{\Theta}^{-1*}$ is obtained from the matrix $(\Theta_{i,j})$ of the operator Θ^{-1*} by reordering of columns and lines, as determined by P and P', respectively:

$$\Theta^{-1*}_{i,j} = \tilde{\Theta}^{-1*}_{P'(i),P(j)} \quad \text{for } i,j = 1, \dots, d. \tag{2.45}$$

Therefore, combining (2.44) and (2.45) with (2.42), we find that

$$\Theta^{-1*}(a(\mathcal{O}))_i = \sum_j \Theta^{-1*}_{i,j} a(\mathcal{O})_j = \sum_j \tilde{\Theta}^{-1*}_{P'(i),P(j)} a(\tilde{\mathcal{O}})_{P(j)} = a(\tilde{\mathcal{O}}')_{P'(i)} = a(\mathcal{O}')_i$$

for all $i = 1, \dots, d$. This proves (2.43). As a direct consequence we get that $\Theta^{-1*}(\lambda(\mathcal{O})) = \lambda(\mathcal{O}')$. The proof of the lemma is now complete. $\square$

2.9 Homological interpretation

We are going to interpret some of the previous objects and conclusions in terms of the first homology and cohomology groups of the surface M. The necessary background in homology theory may be found, for instance, in Fulton [15]. The conclusions are summarized in the following dictionary:

$$\begin{aligned} \text{space } \mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi &\longleftrightarrow \text{homology } H_1(M,\mathbb{R}) \\ \text{space } H_\pi &\longleftrightarrow \text{cohomology } H^1(M,\mathbb{R}) \\ \text{inner product in } \mathbb{R}^{\mathcal{A}} &\longleftrightarrow \text{duality } H^1(M,\mathbb{R}) \approx H_1(M,\mathbb{R})^* \\ \text{operator } \Omega_\pi &\longleftrightarrow \text{Poincaré duality} \\ \text{symplectic form } \omega'_\pi \text{ on } \mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi &\longleftrightarrow \text{intersection form on } H_1(M,\mathbb{R}) \\ \text{symplectic form } \omega_\pi \text{ on } H_\pi &\longleftrightarrow \text{intersection form on } H^1(M,\mathbb{R}) \end{aligned}$$

Let M be represented in the form of zippered rectangles, corresponding to data (π, λ, τ, h), as introduced in Section 2.5. We identify the horizontal segment $\sigma = I \times \{0\}$ with the interval I. For each symbol $\alpha \in \mathcal{A}$, let $[v_\alpha]$ be the homology class of any closed curve v_α formed by a vertical segment crossing the rectangle R_α from bottom to top together with a horizontal segment that joins its endpoints. See Figure 2.20. We are going to see that these homology classes generate the first homology group $H_1(M, \mathbb{R})$.

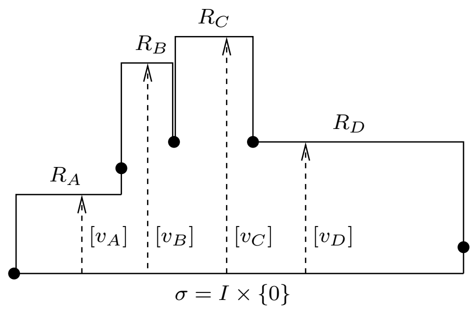


Figure 2.20:

For definiteness, let us consider the case when τ has type 0, that is, the sum of its coefficients is positive (the type 1 case is analogous). For each $\alpha \in \mathcal{A}$, consider the following line segments on the boundary of R_α (some reduce to points, see Figure 2.21):

$$
\begin{aligned}
H_b(\alpha) &= \Big[\sum_{\pi_0(\beta)<\pi_0(\alpha)} \lambda_\beta, \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \lambda_\beta \Big] \times \{0\}, \\
H_t(\alpha) &= \Big[\sum_{\pi_0(\beta)<\pi_0(\alpha)} \lambda_\beta, \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \lambda_\beta \Big] \times \{h_\alpha\}, \\
V_{b,l}(\alpha) &= \Big\{ \sum_{\pi_0(\beta)<\pi_0(\alpha)} \lambda_\beta \Big\} \times \Big[0, \sum_{\pi_0(\beta)<\pi_0(\alpha)} \tau_\beta \Big] \\
V_{t,l}(\alpha) &= \Big\{ \sum_{\pi_0(\beta)<\pi_0(\alpha)} \lambda_\beta \Big\} \times \Big[\sum_{\pi_0(\beta)<\pi_0(\alpha)} \tau_\beta, h_\alpha \Big] \\
V_{b,r}(\alpha) &= \Big\{ \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \lambda_\beta \Big\} \times \Big[0, \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \tau_\beta \Big] \\
V_{t,r}(\alpha) &= \Big\{ \sum_{\pi_0(\beta)\le\pi_0(\alpha)} \lambda_\beta \Big\} \times \Big[\sum_{\pi_0(\beta)\le\pi_0(\alpha)} \tau_\beta, h_\alpha \Big]
\end{aligned}
$$

The case $\alpha = \alpha(1)$ is a bit special. As observed following Lemma 2.7, the sum of τ_β over all $\pi_0(\beta) \le \pi_0(\alpha(1))$ is larger than h_α if τ has type 0. Hence, $V_{t,r}(\alpha(1))$ is oriented downwards, is contained in $V_{b,r}(\alpha(1))$, and is not really on the boundary of R_α. Note that $V_{t,r}(\alpha(1))$ coincides with the segment $\tilde{S}$ defined in (2.28)–(2.29). We call $V_{b,l}(\alpha) \cup H_b(\alpha) \cup V_{b,r}(\alpha)$ the *lower boundary* and $V_{t,l}(\alpha) \cup H_t(\alpha) \cup V_{t,r}(\alpha)$ the *upper boundary* of R_α, for any $\alpha \in \mathcal{A}$.

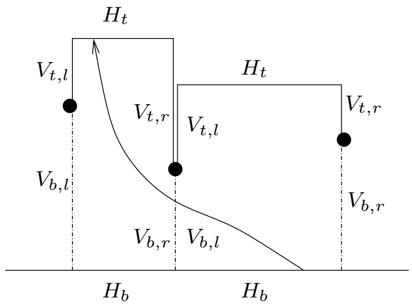


Figure 2.21:

Lemma 2.17. *The map* $\Phi : \mathbb{R}^{\mathcal{A}} \to H_1(M, \mathbb{R})$, $\Phi(\tau) = \sum_{\alpha \in \mathcal{A}} \tau_\alpha [v_\alpha]$ *is onto.*

Proof. It suffices to prove that the homology class $[\gamma]$ represented by any closed curve γ in M is a linear combination of the $[v_\alpha]$, with integer coefficients. It is no restriction to suppose that γ is transverse to the boundaries of the rectangles R_α. In addition, we may suppose that γ does not intersect the special segment $V_{t,r}(\alpha(1)) = \tilde{S}$: such intersections may be removed by isotopy as described in Figure 2.22. Recall that $\tilde{S}$ is identified to the segment $S^0_{\alpha(0)}$ defined in (2.26). Then γ may be partitioned into a finite number of segments each of which is contained in the closure of some rectangle R_α and does not intersect $\tilde{S}$. We call such a segment *essential* if one endpoint is in the upper boundary and the other one is in the lower boundary, and *inessential* if both endpoints are either in the upper boundary or in the lower boundary of R_α. Both possibilities are illustrated in Figure 2.21. Then γ may be deformed by isotopy in such a

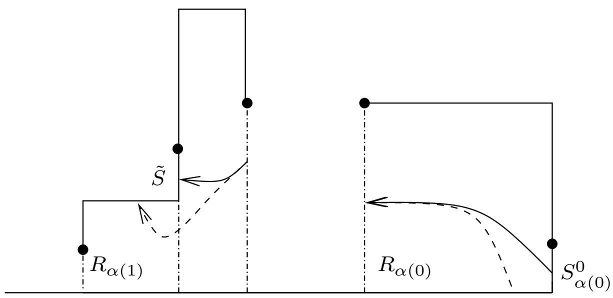


Figure 2.22:

way that every inessential segment is replaced by one contained in I, and every essential segment is replaced by one of the form

$$\eta' \pm \Big(V_{b,l}(\alpha) + V_{t,l}(\alpha)\Big) + \eta'',$$

where η' and η'' are contained in I (the symbol $+$ denotes concatenation). By definition, v_α is homologous to $V_{b,l}(\alpha)+V_{t,l}(\alpha)+\eta(\alpha)$ for some segment $\eta(\alpha) \subset I$. It follows that γ is homologous to some

$$(\pm v_{\alpha_1}) + \eta_1 + \cdots + (\pm v_{\alpha_N}) + \eta_N \quad \text{with all } \eta_i \subset I.$$

As I is contractible, we get that $[\gamma] = \sum_{i=1}^{N} \pm [v_{\alpha_i}]$ and this proves the claim. $\square$

Lemma 2.18. *The map Φ induces an isomorphism from $\mathbb{R}^{\mathcal{A}}/\ker \Omega_\pi$ to the homology group:*

$$\begin{tikzcd} \mathbb{R}^{\mathcal{A}} \arrow[r, "\Phi"] \arrow[d, "quotient"'] & H_1(M,\mathbb{R}) \\ \mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi \arrow[ur, "\cong"'] & \end{tikzcd} \qquad (2.46)$$

Proof. We claim that the kernel of Φ contains $\ker \Omega_\pi$. This ensures that the map Ψ is well defined in the quotient space $\mathbb{R}^{\mathcal{A}}/\ker \Omega_\pi$. Since the dimension of the latter is equal to $\dim H_\pi = 2g = \dim H_1(M,\mathbb{R})$, and we already know Φ is surjective, it will follow that the map induced by Φ in the quotient space is an isomorphism onto the homology group, and the kernels of Φ and Ω_π must coincide. Therefore, we are left to prove the claim.

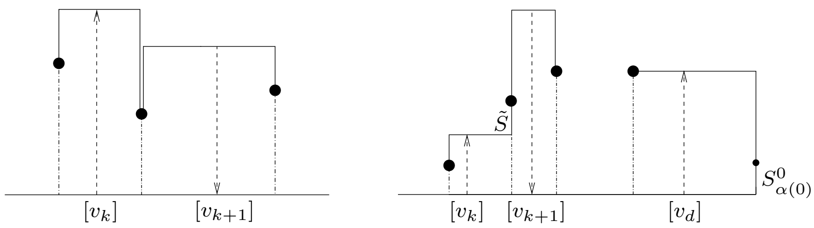


Figure 2.23:

In what follows we assume τ has type zero, that is, $\sum_{\delta\in\mathcal{A}} \tau_\delta > 0$. The type 1 case is analogous. Let us normalize π so that $\mathcal{A} = \{1,\ldots,d\}$ and $\pi_0 = \mathrm{id}$. Let $\{\lambda(\mathcal{O})\}$ be the basis of $\ker \Omega_\pi$ defined in (2.32), where $\mathcal{O}$ runs over the orbits of σ not containing 0. From the definition we get that

$$\Phi(\lambda(\mathcal{O})) = \sum_{k=1}^{d} \lambda(\mathcal{O})_k [v_k] = \sum_{k\in\mathcal{O}} [v_k] - \sum_{k\in\mathcal{O}, k<d} [v_{k+1}]. \qquad (2.47)$$

Let us suppose, first, that $\mathcal{O}$ does not contain d. As illustrated on the left hand side of Figure 2.23,

$$\begin{aligned} [v_k] - [v_{k+1}] &= [V_{b,l}(k) + V_{t,l}(k) - V_{b,l}(k+1) - V_{t,l}(k+1) + \eta'(k)] \\ &= [V_{b,r}(k) + V_{t,r}(k) - V_{b,l}(k+1) - V_{t,l}(k+1) + \eta''(k)] \end{aligned} \qquad (2.48)$$

with $\eta'(k)$ and $\eta''(k)$ contained in I. Clearly, $V_{b,l}(k+1) = V_{b,r}(k)$. Moreover, $V_{t,l}(k+1) = V_{t,r}(\sigma^{-1}(k))$: the definition (2.18) gives

$$p(k+1) = p(\sigma^{-1}(k)) + 1$$

and this means that $\sigma^{-1}(k)$ is the symbol immediately to the left of $k+1$ on the bottom row of π. Replacing these equalities in (2.48), we obtain

$$[v_k] - [v_{k+1}] = [V_{t,r}(k) - V_{t,r}(\sigma^{-1}(k)) + \eta''(k)] \tag{2.49}$$

Noting that σ^{-1} permutes the elements of $\mathcal{O}$, we conclude that

$$\Phi(\mathcal{O}) = \sum_{k\in\mathcal{O}} [v_k] - [v_{k+1}] = \sum_{k\in\mathcal{O}} [V_{t,r}(k) - V_{t,r}(\sigma^{-1}(k)) + \eta''(k)] = [\eta] \tag{2.50}$$

for some segment $\eta \subset I$. Since I is contractible, this proves that $\Phi(\mathcal{O})$ vanishes in the homology, as claimed.

Now let us explain how the argument can be adapted to deal with the case when $d \in \mathcal{O}$. There are two differences. Firstly, in this case (2.47) gives

$$\Phi(\lambda(\mathcal{O})) = [v_d] + \sum_{k\in\mathcal{O},k<d} [v_k] - [v_{k+1}].$$

Secondly, for $k = p^{-1}(d) = \sigma^{-1}(d)$ the relation (2.49) must be replaced by (see the right hand side of Figure 2.23)

$$[v_k] - [v_{k+1}] = [-\tilde{S} - V_{t,l}(k+1) + \eta''(k)] = [-\tilde{S} - V_{t,r}(\sigma^{-1}(k)) + \eta''(k)].$$

Consequently, using also that $\tilde{S} = S^0_{\alpha(0)} = V_{b,r}(d)$ and $k = \sigma^{-1}(d)$,

$$\begin{aligned}
[v_d] + [v_k] - [v_{k+1}] &= [V_{b,r}(d) + V_{t,r}(d) - \tilde{S} - V_{t,l}(\sigma^{-1}(k)) + \eta] \\
&= [V_{t,r}(d) - V_{t,r}(\sigma^{-1}(k)) + \eta] \\
&= [V_{t,r}(d) - V_{t,r}(\sigma^{-1}(d)) + V_{t,r}(k) - V_{t,r}(\sigma^{-1}(k)) + \eta].
\end{aligned}$$

This, together with (2.49) for the values of $k \notin \{d, p^{-1}(d)\}$, means that (2.50) remains valid, and so the claim follows just as in the previous case. □

Now let $H^1(M,\mathbb{R})$ be the first (de Rham) cohomology of the surface M: its elements are the equivalence classes of closed 1-forms, for the equivalence relation $\phi_1 \sim \phi_2 \;\Leftrightarrow\; \phi_1 - \phi_2$ is exact. The homology and cohomology spaces are dual to each other through

$$H_1(M,\mathbb{R}) \times H^1(M,\mathbb{R}) \to \mathbb{R}, \quad ([c],[\phi]) \mapsto [c]\cdot[\phi] = \int_c \phi \tag{2.51}$$

(the integral is independent of the choices of representatives of $[c]$ and $[\phi]$).

Lemma 2.19. *The map* $\Psi : H^1(M, \mathbb{R}) \to \mathbb{R}^{\mathcal{A}}$, $\Psi([\phi]) = \sum_{\alpha \in \mathcal{A}} e_\alpha \int_{v_\alpha} \phi$ *sends the cohomology space isomorphically onto* H_π*:*

$$\begin{tikzcd} H^1(M,\mathbb{R}) \arrow[r, "\Psi"] \arrow[dr, "\cong"'] & \mathbb{R}^{\mathcal{A}} \\ & H_\pi \arrow[u, "inclusion"'] \end{tikzcd} \tag{2.52}$$

Proof. Since the $[v_\alpha]$, $\alpha \in \mathcal{A}$ generate the homology, the map Ψ is injective. Thus, keeping in mind that $\dim H_\pi = 2g = \dim H_1(M, \mathbb{R})$, it suffices to prove that the image of Ψ is contained in H_π. Recall also that H_π is the orthogonal complement of $\ker \Omega_\pi$. So, all we need to do is to prove that every $\Psi([\phi])$, $[\phi] \in H^1(M, \mathbb{R})$ is orthogonal to every $\tau \in \ker \Omega_\pi$. To this end, observe that

$$\Phi(\tau) = \sum_{\alpha \in \mathcal{A}} \tau_\alpha [v_\alpha]$$

is equal to zero in $H_1(M, \mathbb{R})$, according to Lemma 2.18. It follows that

$$\Psi([\phi]) \cdot \tau = \sum_{\alpha \in \mathcal{A}} \tau_\alpha \int_{v_\alpha} \phi = \int_{\sum_\alpha \tau_\alpha v_\alpha} \phi$$

is equal to zero, as claimed, and so the proof is complete. □

Identifying the two spaces $\mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi$ and $H_1(M, \mathbb{R})$ through (2.46), we may think of Θ^{-1*} as acting on the homology space $H_1(M, \mathbb{R})$. Analogously, identifying $H^1(M, \mathbb{R})$ and H_π through (2.52), we may think of Θ as acting on the cohomology space $H^1(M, \mathbb{R})$. Then the diagram (2.41) becomes

$$\begin{tikzcd} H_1(M,\mathbb{R}) \arrow[r, "\Theta^{-1*}"] \arrow[d, "\mathcal{P}"'] & H_1(M,\mathbb{R}) \arrow[d, "\mathcal{P}"] \\ H^1(M,\mathbb{R}) \arrow[r, "\Theta"] & H^1(M,\mathbb{R}). \end{tikzcd} \tag{2.53}$$

We are going to see that the isomorphism $\mathcal{P}$ represented by the vertical arrows in (2.53) corresponds to the Poincaré duality between homology and cohomology. Before recalling this notion (see Chapters 18 and 24 of Fulton [15]) let us prove

Lemma 2.20. *We have* $[c] \cdot \Theta([\phi]) = \Theta^*([c]) \cdot [\phi]$ *for every* $[c] \in H_1(M, \mathbb{R})$ *and every* $[\phi] \in H^1(M, \mathbb{R})$*, where* $\cdot$ *represents the duality* (2.51).

Proof. It suffices to check that, under the identifications (2.46) and (2.52), the duality (2.51) corresponds to

$$\mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi \times H_\pi \mapsto \mathbb{R}, \quad ([\tau], w) \mapsto \tau \cdot w,$$

where $\cdot$ is the canonical inner product in $\mathbb{R}^{\mathcal{A}}$. By linearity, it suffices to consider $[c] \in H_1(M,\mathbb{R})$ of the form $[c] = [v_\alpha]$, for some $\alpha \in \mathcal{A}$. This corresponds to $[\tau] = [e_\alpha] \in \mathbb{R}^{\mathcal{A}} / \ker \Omega_\pi$ under (2.46). Moreover, $[\phi] \in H^1(M,\mathbb{R})$ corresponds to

$$w = \sum_{\beta\in\mathcal{A}} e_\beta \int_{v_\beta} \phi \in H_\pi$$

under (2.52). Then,

$$[c]\cdot[\phi] = [v_\alpha]\cdot[\phi] = \int_{v_\alpha} \phi = e_\alpha \cdot \sum_{\beta\in\mathcal{A}} e_\beta \int_{v_\beta} \phi = \tau \cdot w$$

as claimed. □

The vector space $H^1(M,\mathbb{R})$ comes with a bilinear *intersection form*

$$H^1(M,\mathbb{R}) \times H^1(M,\mathbb{R}) \to \mathbb{R}, \quad ([\phi_1],[\phi_2]) \mapsto [\phi_1]\wedge[\phi_2] = \int_M \phi_1 \wedge \phi_2. \quad (2.54)$$

Observe that the sign of the integral depends on the choice of an orientation n M. The Poincaré duality theorem states that (2.54) defines an isomorphism between the cohomology space and its dual: for every linear map $g : H^1(M,\mathbb{R}) \to \mathbb{R}$ there exists a unique $[\phi_g] \in H^1(M,\mathbb{R})$ such that $g([\psi]) = [\psi]\wedge[\phi_g]$ for every $[\psi] \in H^1(M,\mathbb{R})$. On the other hand, (2.51) defines an isomorphism between the homology space and the dual to the cohomology space: to any $[c] \in H_1(M,\mathbb{R})$ we may associate the linear map $g_c : H^1(M,\mathbb{R}) \to \mathbb{R}$, $g_c([\psi]) = \int_c \psi$. Composing these two maps, we obtain an isomorphism associating to each $[c] \in H_1(M,\mathbb{R})$ the unique class $[\phi_c] \in H^1(M,\mathbb{R})$ such that

$$\int_c \psi = g_c(\psi) = [\psi]\wedge[\phi_c] = \int_M \psi \wedge \phi_c \quad \text{for all } [\psi] \in H^1(M,\mathbb{R}). \quad (2.55)$$

We say $[c] \in H_1(M,\mathbb{R})$ and $[\phi_c] \in H^1(M,\mathbb{R})$ are *Poincaré dual* to each other.

Through this isomorphism, we may transport the intersection form to the homology: the *intersection form* in $H_1(M,\mathbb{R})$ is defined by

$$H_1(M,\mathbb{R}) \times H_1(M,\mathbb{R}) \to \mathbb{R}, \quad ([c_1],[c_2]) \mapsto [c_1]\wedge[c_2] = [\phi_{c_1}]\wedge[\phi_{c_2}]. \quad (2.56)$$

It has the following geometric interpretation. Suppose $[c_1]$ and $[c_2]$ are represented by closed curves in M: such classes generate $H_1(M,\mathbb{R})$. Choose representatives c_1 and c_2 that intersect transversely. Let $\iota_j \in \{+1,-1\}$ be the *intersection sign* at each intersection point p_j: the sign is positive if the tangent vectors to c_1 and c_2 form a positive basis, relative to the orientation of M, and it is negative otherwise. Then

$$[c_1]\wedge[c_2] = \sum_j \iota_j\,. \quad (2.57)$$

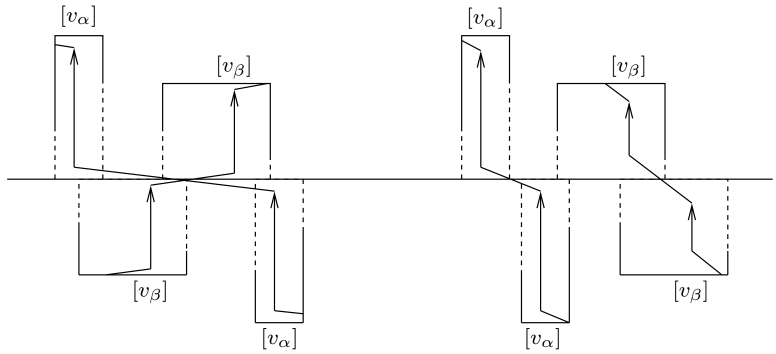


Figure 2.24:

Example 2.21. Figure 2.24 shows how, for any $\alpha, \beta \in \mathcal{A}$, we can find representatives of $[v_\alpha]$ and $[v_\beta]$ that intersect transversely. Using (2.57), we immediately see that

$$[v_\alpha] \wedge [v_\beta] = \left\{ \begin{array}{ll} -1 & \text{if } \pi_0(\alpha) < \pi_0(\beta) \text{ and } \pi_1(\alpha) > \pi_1(\beta) \\ +1 & \text{if } \pi_0(\alpha) > \pi_0(\beta) \text{ and } \pi_1(\alpha) < \pi_1(\beta) \\ 0 & \text{in all other cases.} \end{array} \right\} = -\Omega_{\alpha,\beta} \quad (2.58)$$

for all $\alpha, \beta \in \mathcal{A}$, and relative to the orientation of the translation surface.

Lemma 2.22. *The map $\mathcal{P}$ in* (2.53) *coincides with the isomorphism $[c] \mapsto [\phi_c]$ defined by Poincaré duality* (2.55)*.*

Proof. Since the operator $\mathcal{P}$ is defined by the commuting diagram

$$\begin{tikzcd} \mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi \arrow[r, "\Phi"] \arrow[d, "\Omega_\pi"'] & H_1(M,\mathbb{R}) \arrow[d, "\mathcal{P}"] \\ H_\pi & H^1(M,\mathbb{R}) \arrow[l, "\Psi"] \end{tikzcd} \quad (2.59)$$

we only have to check that the Poincaré dual of $\Phi([\tau])$ is sent to $\Omega([\tau])$ by the map Ψ, for every $[\tau] \in \mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi$. By linearity, it suffices to consider $[\tau] = [e_\alpha]$, for each $\alpha \in \mathcal{A}$. Then $\Phi([\tau]) = [v_\alpha]$ and, by (2.55) and (2.56), its Poincaré dual $[\phi_\alpha]$ satisfies

$$\int_{v_\beta} \phi_\alpha = [\phi_\alpha] \wedge [\phi_\beta] = [v_\alpha] \wedge [v_\beta] = \Omega_{\beta,\alpha} \quad \text{for every } \beta \in \mathcal{A}.$$

In the last equality we used (2.58). It follows that

$$\Psi([\phi_\alpha]) = \sum_{\beta\in\mathcal{A}} e_\beta \int_{v_\beta} \phi_\alpha = \sum_{\beta\in\mathcal{A}} e_\beta \Omega_{\beta,\alpha} = \Omega_\pi([e_\alpha])$$

for every $\alpha \in \mathcal{A}$, as we wanted to prove. $\square$

By abuse of language, we also represent by Φ and Ψ the isomorphisms in (2.46) and (2.52). The next lemma means that up to these isomorphisms, the intersection forms in homology and cohomology correspond precisely to the symplectic forms ω_π and ω'_π defined in (1.33) and (1.34):

Lemma 2.23. *We have*

1. $\Phi([\tau_1]) \wedge \Phi([\tau_2]) = \omega'_\pi([\tau_1],[\tau_2])$ *for all* $[\tau_1]$, $[\tau_2]$ *in* $\mathbb{R}^{\mathcal{A}}/\ker \Omega_\pi$

2. $[\phi_1] \wedge [\phi_2] = \omega_\pi(\Psi([\phi_1]), \Psi([\phi_2]))$ *for all* $[\phi_1]$, $[\phi_2]$ *in* $H^1(M,\mathbb{R})$.

Proof. By linearity, it suffices to prove the equality in part 1 for vectors $\tau_1 = e_\alpha$ and $\tau_2 = e_\beta$ in the canonical basis of $\mathbb{R}^{\mathcal{A}}$. Notice that $\Phi([\tau_1]) = [v_\alpha]$ and $\Phi([\tau_2]) = [v_\beta]$. Figure 2.24 shows how we can find representatives of the two homology classes that intersect transversely. From (2.58) we get

$$[v_\alpha] \wedge [v_\beta] = -\Omega_{\alpha,\beta} = -e_\alpha \cdot \Omega_\pi(e_\beta) = \omega'_\pi([e_\alpha],[e_\beta])$$

for every $\alpha, \beta \in \mathcal{A}$. This proves part 1 of the lemma.

Analogously, for part 2, it suffices to consider $[\phi_1] = [\phi_\alpha]$ and $[\phi_1] = [\phi_\beta]$, the Poincaré duals of $[v_\alpha]$ and $[v_\beta]$. On the one hand, using (2.58),

$$[\phi_\alpha] \wedge [\phi_\beta] = [v_\alpha] \wedge [v_\beta] = -\Omega_{\alpha,\beta} = -e_\alpha \cdot \Omega_\pi(e_\beta). \tag{2.60}$$

On the other hand, using (2.58) once more,

$$\Psi([\phi_\alpha]) = \sum_{\gamma\in\mathcal{A}} e_\gamma \int_{v_\gamma} \phi_\alpha = \sum_{\gamma\in\mathcal{A}} e_\gamma\, [\phi_\alpha] \wedge [\phi_\gamma] = \sum_{\gamma\in\mathcal{A}} -e_\gamma \Omega_{\gamma,\alpha} = -\Omega_\pi(e_\alpha),$$

and analogously for $\Psi([\phi_\beta])$. So,

$$\omega_\pi(\Psi([\phi_\alpha]), \Psi([\phi_\beta])) = \omega_\pi(\Omega_\pi(e_\alpha), \Omega_\pi(e_\beta)) = -e_\alpha \cdot \Omega_\pi(e_\beta). \tag{2.61}$$

Part 2 of the lemma follows from combining (2.60) and (2.61). $\square$

2.10 Teichmüller flow

Let C be any Rauzy class. We defined $\hat{\mathcal{H}} = \hat{\mathcal{H}}(C)$ to be the set of all (π, λ, τ) such that $\pi \in C$, $\lambda \in \mathbb{R}^{\mathcal{A}}_+$, and $\tau \in T^+_\pi$. The *Teichmüller flow* on $\hat{\mathcal{H}}$ is the natural action $\mathcal{T} = (\mathcal{T}^t)_{t\in\mathbb{R}}$ of the diagonal subgroup

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \quad t \in \mathbb{R}$$

defined by

$$\mathcal{T}^t : \hat{\mathcal{H}} \to \hat{\mathcal{H}}, \quad (\pi, \lambda, \tau) \mapsto (\pi, e^t\lambda, e^{-t}\tau) \tag{2.62}$$

This is well defined because both $\mathbb{R}_+^{\mathcal{A}}$ and T_π^+ are invariant under product by positive scalars. It is clear that the Teichmüller flow commutes with the Rauzy-Veech induction map $\hat{\mathcal{R}}$ and preserves the natural area (2.30). For each $\lambda \in \mathbb{R}_+^{\mathcal{A}}$, define the total length $|\lambda| = \sum_{\alpha\in\mathcal{A}} \lambda_\alpha$. Given any $c > 0$, the affine subset

$$\mathcal{H}_c = \{(\pi,\lambda,\tau) \in \hat{\mathcal{H}} : |\lambda| = c\} \qquad \text{(we denote } \mathcal{H} = \mathcal{H}_1\text{)}$$

is a global cross-section for the Teichmüller flow $\mathcal{T}$: each trajectory intersects $\mathcal{H}_c$ exactly once. In particular, the map

$$\Psi : \mathcal{H} \times \mathbb{R} \to \hat{\mathcal{H}}, \quad \Psi(\pi,\lambda,\tau,s) = \mathcal{T}^s(\pi,\lambda,\tau) = (\pi, e^s\lambda, e^s\tau) \tag{2.63}$$

is a diffeomorphism onto $\hat{\mathcal{H}}$. In these new coordinates, the Teichmüller flow is described, simply, by

$$\mathcal{T}^t : \mathcal{H} \times \mathbb{R} \to \mathcal{H} \times \mathbb{R}, \quad (\pi,\lambda,\tau,s) \mapsto (\pi,\lambda,\tau,s+t). \tag{2.64}$$

For each $(\pi,\lambda,\tau) \in \hat{\mathcal{H}}$, define the *Rauzy renormalization time* [1]

$$t_R = t_R(\pi,\lambda) = -\log\Big(1 - \frac{\lambda_{\alpha(1-\varepsilon)}}{|\lambda|}\Big), \quad \varepsilon = \text{ type of } (\pi,\lambda). \tag{2.65}$$

Notice that if $(\pi',\lambda') = \hat{R}(\pi,\lambda)$ then $|\lambda'| = e^{-t_R}|\lambda|$. This means that the transformation

$$\mathcal{R} = \hat{\mathcal{R}} \circ \mathcal{T}^{t_R} : (\pi,\lambda,\tau) \mapsto \hat{\mathcal{R}}\big(\pi, e^{t_R}\lambda, e^{-t_R}\tau\big) \tag{2.66}$$

maps each cross-section $\mathcal{H}_c$ back to itself. We call the restriction $\mathcal{R} : \mathcal{H} \to \mathcal{H}$ to $\mathcal{H} = \mathcal{H}_1$ the *invertible Rauzy-Veech renormalization map*. Observe that for any $(\pi,\lambda,\tau) \in \mathcal{H}$ we have $\mathcal{R}(\pi,\lambda,\tau) = (\pi',\lambda'',\tau'')$ where

$$(\pi',\lambda',\tau') = \hat{\mathcal{R}}(\pi,\lambda,\tau), \quad \lambda'' = \lambda'/(1-\lambda_{\alpha(1-\varepsilon)}), \quad \tau'' = \tau'(1-\lambda_{\alpha(1-\varepsilon)}).$$

In particular, $\mathcal{R}$ is a lift of the map $R(\pi,\lambda) = (\pi',\lambda'')$ introduced in Section 1.7. From Corollary 2.14 one obtains

Corollary 2.24. *The transformation $\mathcal{R} : \mathcal{H} \to \mathcal{H}$ is an (almost everywhere) invertible Markov map, and it preserves the natural area.*

We call *pre-stratum* $\hat{\mathcal{S}} = \hat{\mathcal{S}}(C)$ associated to C the quotient of the fundamental domain $\{(\pi,\lambda,\tau) \in \hat{\mathcal{H}} : 0 \le \log|\lambda| \le t_R(\pi,\lambda)\}$ by the equivalence relation

$$\mathcal{T}^{t_R(\pi,\lambda)}(\pi,\lambda,\tau) \sim \mathcal{R}(\pi,\lambda,\tau) \quad \text{for all } (\pi,\lambda,\tau) \in \mathcal{H}. \tag{2.67}$$

See Figure 2.25. Equivalently, the pre-stratum may be seem as the quotient of the whole $\hat{\mathcal{H}}$ by the equivalence relation generated by $\mathcal{T}^{t_R}(\pi,\lambda,\tau) \sim \mathcal{R}(\pi,\lambda,\tau)$. We denote by $\mathcal{S}$ the (injective) image of $\mathcal{H}$ under the quotient map. Observe that the dimension of the pre-stratum is given by

$$\dim \hat{\mathcal{S}}(C) = 2d = 4g + 2\kappa - 2. \tag{2.68}$$

[1] The renormalization time depends only on π and $\lambda/|\lambda|$.

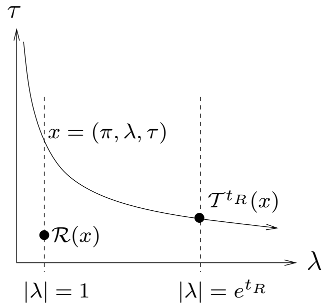


Figure 2.25:

Since $\mathcal{R}$ commutes with $\mathcal{T}^t$, the latter induces a flow $\mathcal{T} = (\mathcal{T}^t)_{t\in\mathbb{R}}$ on the pre-stratum, that we also call *Teichmüller flow.* The invertible Rauzy-Veech renormalization is naturally identified with the Poincaré return map of this flow to the cross-section $\mathcal{S} \subset \hat{\mathcal{S}}$. Notice that the Teichmüller flow preserves the natural volume measure on $\hat{\mathcal{S}}$, inherited from $\hat{\mathcal{H}}$. We shall see in Chapter 4 that this volume is finite, if one restricts to $\{\text{area}\,(\pi, \lambda, \tau) \le 1\}$.

Invertible Zorich maps. We also use accelerated versions of $\hat{\mathcal{R}}$ and $\mathcal{R}$, that we call *invertible Zorich induction* and *invertible Zorich renormalization*, respectively, defined by

$$\hat{\mathcal{Z}}(\pi, \lambda, \tau) = \hat{\mathcal{R}}^n(\pi, \lambda, \tau) \quad \text{and} \quad \mathcal{Z}(\pi, \lambda, \tau) = \mathcal{R}^n(\pi, \lambda, \tau), \tag{2.69}$$

where $n = n(\pi, \lambda) \ge 1$ is the first time the type of $(\pi^n, \lambda^n) = \hat{R}^n(\pi, \lambda)$ differs from the type of (π, λ). See Section 1.8. The domain of $\hat{\mathcal{Z}}$ is a subset $\hat{Z}_*$ of $\hat{\mathcal{H}}$ that we describe in the sequel. Begin by recalling (2.34):

- if (π, λ) has type 0, that is, $\lambda_{\alpha(0)} > \lambda_{\alpha(1)}$ then τ' has type 1, that is, $\sum_{\alpha\in\mathcal{A}} \tau'_\alpha < 0$;
- if (π, λ) has type 1, that is, $\lambda_{\alpha(0)} < \lambda_{\alpha(1)}$ then τ' has type 0, that is, $\sum_{\alpha\in\mathcal{A}} \tau'_\alpha > 0$.

Define $\hat{Z}_* = \hat{Z}_0 \cup \hat{Z}_1$ where, for each $\varepsilon \in \{0, 1\}$,

$$\hat{Z}_\varepsilon = \{(\pi, \lambda, \tau) \in \hat{\mathcal{H}} : (\pi, \lambda) \text{ has type } \varepsilon \text{ and } \tau \text{ has type } \varepsilon\}.$$

Then $n = n(\pi, \lambda)$ is just the first positive iterate for which $\hat{\mathcal{R}}^n(\pi, \lambda, \tau)$ hits $\hat{Z}_*$. Thus, we consider $\hat{\mathcal{Z}}$ defined on the domain $\hat{Z}_*$. The previous observations mean that $\hat{\mathcal{Z}} : \hat{Z}_* \to \hat{Z}_*$ is the first return map of $\hat{\mathcal{R}}$ to the domain $\hat{Z}_*$. It follows that $\hat{\mathcal{Z}}$ is invertible: the inverse is the first return map to $\hat{Z}_*$ of the map $\hat{\mathcal{R}}^{-1}$.

Analogously, we consider $\mathcal{Z} : Z_* \to Z_*$ where Z_* is the set of $(\pi, \lambda, \tau) \in \hat{Z}_*$ such that $|\lambda| = 1$. Then $\mathcal{Z}$ is the first return map of $\mathcal{R}$ to Z_*. Let $\mathcal{S}_* \subset \mathcal{S}$ be the (injective) image of Z_* under the quotient map $\hat{\mathcal{H}} \mapsto \hat{\mathcal{S}}$. Then the invertible

Zorich renormalization $\mathcal{Z}$ is naturally identified with the Poincaré return map of the Teichmüller flow to the cross-section $\mathcal{S}_*$. Notice that it may be written as

$$\mathcal{Z}(\pi,\lambda,\tau)=(\pi^1,e^{t_Z}\lambda^1,e^{-t_Z}\tau^1),\quad \text{where } (\pi^1,\tau^1,\lambda^1)=\hat{\mathcal{Z}}(\pi,\lambda,\tau)$$

* and the *Zorich renormalization time* $t_Z=t_Z(\pi,\lambda)$ is defined by $e^{t_Z}|\lambda^1|=1$.

Notes

The main references are Masur [41] and Veech [54] as well as previous papers of Veech quoted therein. Section 2.10 is mostly from Zorich [63, 65] and Kontsevich-Zorich [33]. As in the previous chapter, our presentation is greatly inspired by Marmi, Moussa, Yoccoz [40].

Chapter 3

Measured Foliations

Let β be a smooth real closed 1-form with finitely many zeros $z_1, \ldots, z_\kappa$, $\kappa \geq 0$ on some smooth surface M. The *measured foliation* $\mathcal{F}_\beta$ defined by β is the foliation on the complement of the set of zeros whose leaves are everywhere tangent to the kernel of β. The crucial feature of such foliations is the existence of a transverse measure, obtained by integrating the 1-form on cross-sections to the foliation, that is invariant under *holonomy maps*, that is, under all projections from one cross-section to another along the leaves of the foliation.

Measured foliations are an important ingredient in Thurston's classification of surface diffeomorphisms (Thurston [52] and Fathi, Laudenbach, Poenaru [13]). They also arise naturally in our context, as horizontal/vertical foliations of Abelian differentials and quadratic differentials (Strebel [51]), and allow for more geometric views of certain constructions we have been presenting. In this chapter we give a brief introduction to the subject and explore some of its connections.

The structure theorem of Maier [38] provides a very satisfactory picture of the global behavior of a measured foliation: there exists a finite decomposition of the surface into periodic regions, where all leaves are closed (homeomorphic to the circle), and minimal regions, where all leaves are dense; these regions are separated by saddle-connections. A proof is given in Sections 3.3 and 3.4. In Section 3.5, we deduce that every translation surface may be represented geometrically as a planar polygon modulo identification of parallel sides.

Another important result, obtained in slightly different forms by Calabi [9], Katok [24], Hubbard, Masur [22], Kontsevich, Zorich [33, 65], characterizes which measured foliations can be realized as horizontal/vertical foliations of some translation surface. A proof is given in Section 3.8. It turns out that typical measured foliations are indeed realizable (Corollary 3.47).

The theory also predicts (Corollary 3.26) that typical measured foliations are minimal, meaning that all leaves are dense in the whole ambient surface. We shall later see, in Section 5.3, that the horizontal/vertical foliations of almost all Abelian differentials are even uniquely ergodic: the leaves are uniformly distributed on the surface.

3.1 Definitions

Recall that a closed 1-form is, locally, the derivative of some function φ which is uniquely defined up to an additive constant. Thus, the leaves of the corresponding measured foliation $\mathcal{F}_\beta$ coincide with the level sets of φ. Also, there is a favorite orientation on each curve γ transverse to the foliation (*cross-section*): γ is positively, or *increasingly*, oriented if the primitives φ of the 1-form β are increasing relative to the orientation. It is clear that holonomy maps of the foliation preserve this orientation.

Near any regular point, that is, any $z \in M$ such that β_z is different from zero, one can choose local coordinates that are *adapted* to the foliation, in the sense that the range of the local chart is a rectangle in the plane, the leaves of $\mathcal{F}_\beta$ correspond to horizontal segments, and the positive transverse orientation corresponds to the upward direction on the plane. See Figure 3.1.

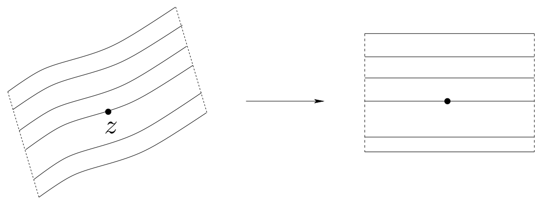


Figure 3.1:

Example 3.1. Suppose the surface M is orientable, and let ω be an area form, that is, a non-degenerate 2-form on M. Given a vector field Z and a 1-form β on M, we say that Z and β are *dual* relative to ω if they are related by

$$\beta_z(v) = \omega_z(Z(z), v) \quad \text{for every vector } v \in T_z M.$$

Since ω is non-degenerate, this relation also determines Z uniquely from β. Notice that β and Z have the same zeros and $\ker \beta$ coincides with the direction of Z at every regular point. Thus, the regular trajectories of the vector field Z are precisely the leaves of the measured foliation defined by the 1-form β. Furthermore, *β is closed if and only if Z preserves area.* This can be easily seen as follows. By Darboux's theorem one may find local coordinates (x, y) on the surface relative to which $\omega = dx \wedge dy$. Then, writing $Z = X\partial_x + Y\partial_y$, we have $\beta = X dy - Y dx$. It follows that

$$d\beta = (\partial_x X + \partial_y Y)\, dx \wedge dy = (\operatorname{div} Z)\, dx \wedge dy.$$

Then $d\beta = 0$ if and only if the divergence vanishes, that is, if and only if the flow of Z preserves area.

Example 3.2. Let α be a non-zero Abelian differential on some Riemann surface. Then the imaginary part $\beta^h = \Im(\alpha)$ and the real part $\beta^v = \Re(\alpha)$ are real analytic closed 1-forms and it is clear that their singularities are of the type described previously. To see that these 1-forms are indeed closed, write $z = x+iy$ and $\alpha(z) = u(x,y) + iv(x,y)$ in some local chart of the Riemann surface. Then

$$\alpha_z = \alpha(z)\,dz = \big(u(x,y)\,dx - v(x,y)\,dy\big) + i\big(u(x,y)\,dy + v(x,y)\,dx\big)$$

where (Cauchy-Riemann conditions)

$$\partial_x u = \partial_y v \quad \text{and} \quad \partial_y u = -\partial_x v.$$

The first inequality means that the imaginary part is closed, whereas the second one means that the real part is closed. These two closed 1-forms define a pair of measured foliations that are transverse to each other at every regular point. We call them the *horizontal foliation* and the *vertical foliation* of α, respectively, and denote them $\mathcal{F}^h_\alpha$ and $\mathcal{F}^v_\alpha$. Observe that the leaves of either foliation are geodesics for the flat metric defined by the Abelian differential. In Section 3.8 we shall see that typical measured foliations are of this kind.

Transversely invariant measures. The terminology "measured foliation" is motivated by the observation that $\mathcal{F}_\beta$ comes with a *transversely invariant measure*. By such we mean a length element on any curve γ transverse to the foliation which is invariant under the holonomy maps of $\mathcal{F}_\beta$. This length element is defined by

$$\ell(\gamma) = \int_\gamma |\beta| = \int_0^1 |\beta_{\gamma(t)}(\dot\gamma(t))|\,dt, \tag{3.1}$$

where $\gamma : [0,1] \to M \setminus \{z_1, \ldots, z_\kappa\}$ is any parametrized curve (the integral does not depend on the choice of the parametrization). Notice that if γ is transverse to $\mathcal{F}_\beta$ then $\beta_{\gamma(t)}(\dot\gamma(t))$ has constant sign, and so $\ell(\gamma) = |\int_\gamma \beta|$. To see that ℓ is invariant under holonomy, consider any cross-sections γ_1 and γ_2 such that there exists a holonomy map of $\mathcal{F}_\beta$ sending γ_1 homeomorphically onto γ_2. See Figure 3.2. Then γ_1 and γ_2, together with the leaf segments connecting their

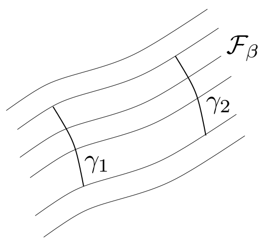


Figure 3.2:

endpoints, determine a simply connected domain D in M. Since β is closed and vanishes identically on the leaves,

$$\int_{\gamma_1} \beta + \int_{\gamma_2} \beta = \int_{\partial D} \beta = 0$$

and so $\ell(\gamma_1) = \ell(\gamma_2)$, as we claimed.

Singularities. The zeros z_i of β correspond to singularities of the foliation. We require all singularities to be of saddle type, possibly degenerate. More precisely, we assume that there are smooth local coordinates mapping each z_i to $0 \in \mathbb{C}$ and relative to which

$$\beta_z = \Im(z^{m_i} dz) \quad \text{for some } m_i \geq 1, \tag{3.2}$$

where m_i is the *order*, or *multiplicity*, of the singularity. In such *adapted singular coordinates*, we have $\beta = d\varphi$ with $\varphi = \Im(z^{m_i+1}/(m_i+1))$. The level set of φ through the origin is characterized by

$$\Im(z^{m_i+1}) = 0 \quad \Leftrightarrow \quad z \text{ is an } (m_i+1)\text{-root of a real number}$$

and so it consists of $2(m_i+1)$ half lines that separate the plane into the same number of sectors. Figure 3.3 illustrates the cases $m_i = 1$ and $m_i = 2$. Let us point out that all these half lines are contained in distinct leaves of the foliation $\mathcal{F}_\beta$, that we call the *separatrices* of z_i.

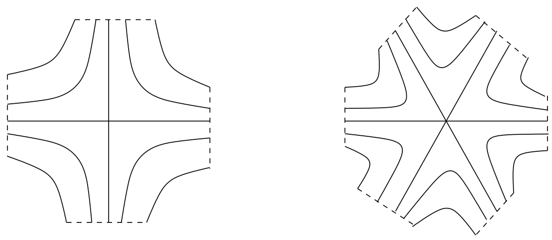


Figure 3.3:

Given singularities z_i and z_j, it may be that some separatrix of one coincides with some separatrix of the other. We call this a *saddle-connection.* Notice that we do not exclude the possibility $z_i = z_j$, in which case the saddle-connection is called a *homoclinic loop.*

Non-transversely orientable foliations. The horizontal and vertical foliations of an Abelian differential α (Example 3.2) are characterized by

$$\alpha(z)dz \in \mathbb{R} \quad \text{and} \quad \alpha(z)dz \in i\mathbb{R},$$

respectively. In terms of the quadratic differential $q = \alpha^2$, this may be written as

$$q(z)dz^2 > 0 \quad \text{and} \quad q(z)dz^2 < 0, \tag{3.3}$$

respectively. Now, given a non-orientable quadratic differential q, that is, one that is *not* the square of an Abelian differential, one may still use (3.3) to define a pair of foliations on the surface M. These foliations are tangent to the kernels of $\beta^h = \Im(\sqrt{q})$ and $\beta^v = \Re(\sqrt{q})$ where β^h and β^v depart from being real closed 1-forms only in that they are defined up to sign (corresponding to the choice of the square root). Moreover, these foliations also admit transversely invariant measures defined as in (3.1): it is clear that the choice of a sign does not affect the definition.

More generally, the notions of measured foliation and transversely invariant measure make sense when the smooth 1-form β is defined up to sign only: $\ker \beta_z$ does not depend on the choice of the sign, nor does the definition of the transversely invariant length element. The local model near a singularity is

$$\beta_z = \Im(z^{n_i/2} dz) \quad \text{with } n_i \geq 1, \tag{3.4}$$

where sign ambiguity corresponds to the choice of the square root. Figure 3.4 illustrates the cases $n_i = 1$ and $n_i = 3$. Notice that, in general, the singularity has $n_i + 2$ separatrices. A measured foliation defined in this more general way

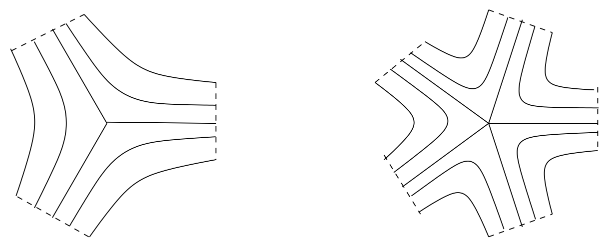


Figure 3.4:

is of the type introduced before (in other words, there is a globally consistent choice of the sign of β) if and only if it is transversely orientable, that is, if and only if there exists a choice of an orientation on each cross-section relative to which all holonomy maps are orientation preserving. For this it is necessary, but not sufficient, that all n_i be even: observe Figure 3.4.

Remark 3.3. If q is a *meromorphic* quadratic differential, then $\beta^h = \Im(\sqrt{q})$ and $\beta^v = \Re(\sqrt{q})$ define measured foliations on the surface M with singularities at the poles as well as at the zeros of q. The local forms of these foliations near general poles are described in Strebel [51, § III.7]. We are especially interested on *integrable* quadratic differentials, that is, whose norm

$$\|q\| = \int |q(z)|\, dz\, d\bar{z}$$

is finite. In that case, all poles of q must be simple, that is, of order 1. The local normal form (3.4) remains valid, with $n_i = -1$. The corresponding behavior of the measured foliation is described in Figure 3.5.

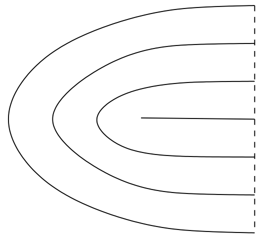


Figure 3.5:

Canonical branched double cover. On the other hand, this more general situation can be easily reduced to the transversely orientable one, using the following construction of a double cover $\tilde{M} \to M$ branched over the singularities z_i of odd order n_i, where the two preimages of each regular point $p \in M$ correspond to the two possible signs of β at that point.

Proposition 3.4. *Let β be a smooth closed 1-form defined up to sign on M whose zeros are as in (3.4). Then there is a branched double cover $\pi : \tilde{M} \to M$ and a smooth closed 1-form $\tilde{\beta}$ on the compact surface $\tilde{M}$ such that*

$$(\pi_* \tilde{\beta})^2 = \beta^2$$

and whose branching points are precisely the preimages of the zeros of β with odd order. The surface $\tilde{M}$ is disconnected if and only if β is orientable.

Proof. Consider an atlas $\{(U_i, \phi_i)\}$ of the complement of the zeros of β in M whose domains U_i are connected and simply connected. For each i, let $\pm\beta_i$ be the two determinations [1] of β restricted to U_i and consider two copies $U_i^\pm$ of the chart domain. For each i, j and any choices $\circ, \bullet \in \{+, -\}$ of the signs, identify the part of U_i° corresponding to $U_i \cap U_j$ with the part of $U_j^\bullet$ corresponding to $U_i \cap U_j$ whenever $\circ\beta_i = \bullet\beta_j$. In this way we get a punctured Riemann surface and a 1-form $\tilde{\beta}$ defined on it by

$$\tilde{\beta} \mid U_i^\pm = \pm\beta_i \quad \text{for each } i \text{ and either choice of the sign.} \tag{3.5}$$

It is clear that $\tilde{\beta}$ is closed, because the property of being closed is local. From the local form (3.4) of β near the zeros, one easily gets that the punctures can be filled in, to get a compact surface $\tilde{M}$. Moreover, $\tilde{\beta}$ extends to a closed 1-form on the whole $\tilde{M}$, vanishing at the punctures. Now, by construction, there exists a double cover $\pi : \tilde{M} \to M$ possibly branched at the punctures, which are the preimages of the zeros. The relation (3.5) means that $(\pi_* \tilde{\beta})^2 = \beta^2$ at every point. It is a simple exercise to see that if a zero has even order (Figure 3.3) then there exist well-defined determinations of the sign on a neighborhood. Thus, such a zero has two distinct preimages, which are not branching points. Similarly, if the zero has odd order (Figure 3.4) then the sign is not well-defined on

[1] The assignment of the signs $\pm$ is not canonical.

any neighborhood. Thus, such a zero has a unique preimage, which is a branching point. By definition, β is orientable if and only if there exists a continuous global determination of the sign. Clearly, this is the same as saying that the domains $U_i^{\pm}$ split into two classes such that any two domains in different classes are disjoint. In other words, this happens if and only if $\tilde{M}$ is not connected. $\square$

This proposition means that we can always lift β to a genuine 1-form $\tilde{\beta}$ on $\tilde{M}$: the sign ambiguity is resolved in the lift. This 1-form $\tilde{\beta}$ defines a transversely orientable foliation $\mathcal{F}_{\tilde{\beta}}$ on $\tilde{M}$ whose leaves project down to the leaves of $\mathcal{F}_{\beta}$. Properties of either of these two foliations can be easily translated to properties of the other: see Remark 3.10 and Example 3.48 for a couple of applications. This construction shows that not much is lost by focussing on the transversely orientable case, and we shall do so, except where specified otherwise.

3.2 Basic properties

In this section we collect some useful basic properties of measured foliations.

Limit sets. Let F be the leaf through a regular point p. Then F admits locally injective (periodic if the leaf is closed, and globally injective otherwise) parametrizations $\sigma : \mathbb{R} \to F$ with $\sigma(0) = p$. We call *rays* of p the images of $[0, +\infty)$ and $[0, -\infty)$ under any such parametrization. Then we define the *limit set* of each ray R to be the set $\omega(R)$ of accumulation points of $\sigma(t)$ as $t \to \infty$. Clearly, these definitions do not depend on the choice of the parametrization. If the leaf is closed then both rays and their limits set coincides with F itself.

Lemma 3.5. *For any ray R, the limit set $\omega(R)$ is non-empty, saturated, and connected.*

Proof. The fact that $\omega(R)$ is non-empty is an immediate consequence of the assumption that M is compact. It is clear that $\omega(F)$ is closed, and so its intersection with any leaf is a closed subset of the leaf. So, to prove that the limit set is saturated, we only have to show that the intersection is also open in the leaf. This is easily seen using adapted coordinates as in Figures 3.1 and 3.3: it is clear in such coordinates that if a ray accumulates a point p then it accumulates the whole segment of the leaf of p inside the chart domain. Finally, given any pair of disjoint open sets intersecting $\omega(R)$, the ray R must intersect both infinitely often. So, there must also be a sequence $t_n \to \infty$, corresponding to transitions from one open set to the other, for which $\sigma(t_n)$ does not belong to either of the open sets. Taking the limit as $n \to \infty$ one obtains a point in $\omega(R)$ outside the union of the two open sets. This shows that the limit set can not be split into two open subsets: in other words, $\omega(R)$ is connected. $\square$

We call a ray R *singular* if the limit set $\omega(R)$ is reduced to one singularity; otherwise we call the ray *regular*. Observe that if $\omega(R)$ contains a singularity but is not restricted to it, then it also contains at least two of the associated separatrices. See Figure 3.6.

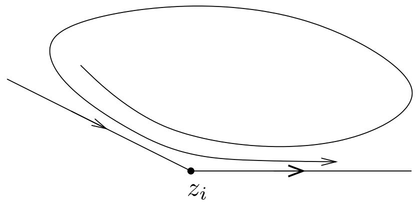


Figure 3.6:

Adapted atlas and holonomy maps. We have seen that every point $p \in M$ is contained in the domain V_p of some adapted local chart, either a regular chart as in Figure 3.1 or a singular chart as in Figure 3.3. We have defined regular domains to be bounded by two leaf segments and two cross-sections to the foliation. Similarly, let us settle that we take the singular domains around a singularity z_i to be bounded by $2(m_i + 1)$ leaf segments and an equal number of cross-sections. By compactness, there exist finite families of adapted local coordinates, either regular or singular, such that their domains cover the whole surface M. We call any such family an *adapted atlas* to the foliation.

Using these notions we are going to prove that holonomy maps can be extended for as long as their domains avoids singular rays. Before we give the precise statement (see also Figure 3.7), we introduce some useful terminology. Let γ be some small cross-section to the foliation. Let R_i, $i = 1, 2$ be rays with parametrizations $\sigma_i : [0, +\infty) \to R_i$ such that $\sigma_i(0) \in \gamma$ for $i = 1, 2$. We say the two rays leave the cross-section *in the same direction* if the tangent vectors $\dot{\sigma}_i(0)$, $i = 1, 2$ belong to the same transverse orientation of γ.

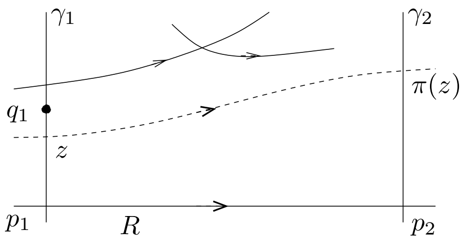


Figure 3.7:

Lemma 3.6. *Let γ_1 and γ_2 be cross-sections to the foliation and $p_1 \in \gamma_1$ be such that one of its rays R intersects γ_2 at some point p_2. Consider any segment $[p_1, q_1] \subset \gamma_1$ satisfying*

(a) $\ell([p_1, q_1])$ is less than the lengths of the connected components [2] *of $\gamma_2 \setminus \{p_2\}$*

[2] It will be clear from the proof that we only need $[p_1, q_1]$ to be shorter than one of the connected components, determined by the orientation.

(b) *there is no ray leaving from* $[p_1, q_1]$ *in the same direction as* R *which is singular and does not intersect* γ_2.

Then there is a well-defined holonomy map $\pi : [p_1, q_1] \to \gamma_2$ *with* $\pi(p_1) = p_2$.

Proof. We claim that there exists $\delta > 0$, depending only on the foliation, such that if $[p, q] \subset [p_1, q_1]$ is such that $\ell([p, q]) < \delta$ and the holonomy map π is defined at p then π extends to the whole $[p, q]$. The lemma is an immediate consequence of this claim, since we assume the holonomy map to be defined for $p = p_1$ and then the claim implies that the largest segment it extends to has to contain $[p_1, q_1]$.

To prove this claim, let us fix some adapted atlas, with local chart domains V_j, $j = 1, \dots, k$. By compactness, we may find (slightly smaller) domains U_j, $j = 1, \dots, k$ that still cover M and such that the closure of every U_j is contained in the corresponding V_j. Then, since we are dealing with finite families, we may find $\delta > 0$ such that whenever a point q belongs to the intersection of domains U_i and U_j there exists some cross-section $\sigma \subset V_i \cap V_j$ at the point q such that both connected components of $\sigma \setminus \{q\}$ have length δ. Next, we decompose the

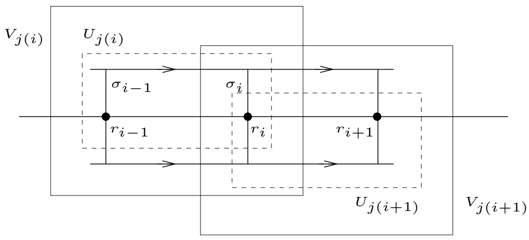


Figure 3.8:

leaf segment $[p, \pi(p)]$ into subsegments $[r_{i-1}, r_i]$, $i = 1, \dots, n$ such that $r_0 = p$, $r_n = \pi(p)$, and each $[r_{i-1}, r_i]$ is contained in some domain $U_{j(i)}$. For each $1 \le i \le n-1$, let $\sigma_i \subset V_{j(i)} \cap V_{j(i+1)}$ be a cross-section with length δ around r_i. Moreover, let $\sigma_0 \subset \gamma_1$ and $\sigma_n \subset \gamma_2$ be cross-sections with radius δ around r_0 and r_n, respectively. We are going to obtain the extended holonomy $\pi : [p, r] \to \gamma_2$ as a composition of holonomy maps from each σ_{i-1} to the next σ_i. Suppose first that V_i is a regular adapted domain. Then there exists a well-defined holonomy map $\sigma_{i-1} \to \sigma_i$ along the leaf segments inside $V_{j(i)}$. See Figure 3.8. That is also true when $V_{j(i)}$ is a singular domain, *unless the cross-section* σ_{i-1} *cuts some separatrix*. See Figure 3.9. However, the hypothesis (b) of the lemma ensures that we never encounter this situation before reaching the final cross-section γ_2, at least not on the forward image of the segment $[p, q]$. This means that we may proceed all the way to the end and, thus, obtain the extension of the holonomy map to $[p, q]$ as stated. □

Euler-Poincaré formula. The next lemma is a rephrasing of (2.9) in a more topological guise:

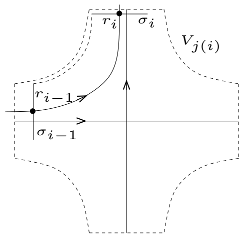


Figure 3.9:

Lemma 3.7. *Let $m_i \geq 1$ be the order and $s_i = 2(m_i + 1)$ be the number of separatrices of each singularity z_i of $\mathcal{F}_\beta$, for $i = 1, \ldots, \kappa$. Then*

$$\sum_{i=1}^{\kappa} s_i = 2\kappa - 2\mathcal{X}(M). \tag{3.6}$$

Proof. Suppose first that M is orientable. Then, as in Example 3.1, we may consider a vector field Z tangent to the foliation. By definition, the index $\operatorname{ind}(Z, z_i)$ of Z at each singularity z_i is the degree of the map from S^1 to S^1 defined by

$$t \mapsto \frac{Z(\xi(t))}{\|Z(\xi(t))\|},$$

where $\xi : S^1 \to M$ is any small simple closed curve around z_i. It is easy to see that $\operatorname{ind}(Z, z_i) = 1 - s_i/2$. So, by the Poincaré-Hopf theorem (see [43]),

$$2\mathcal{X}(M) = 2\sum_{i=1}^{\kappa} \operatorname{ind}(Z, z_i) = \sum_{i=1}^{\kappa} (2 - s_i),$$

as claimed. If M is not orientable, we may consider its (unbranched) double cover $\hat{M} \to M$, where $\hat{M}$ is an orientable surface. Notice that $\mathcal{X}(\hat{M}) = 2\mathcal{X}(M)$. Then we may lift $\mathcal{F}_\beta$ to the measured foliation $\mathcal{F}_{\hat{\beta}}$ defined on $\hat{M}$ by the lift of the 1-form β. Each singularity of $\mathcal{F}_\beta$ is reproduced twice as a singularity of $\mathcal{F}_{\hat{\beta}}$, and so $\mathcal{F}_{\hat{\beta}}$ has twice as many singularities and separatrices as $\mathcal{F}_\beta$. So, from the fact that the Euler-Poincaré formula is true for $\mathcal{F}_{\hat{\beta}}$ we get that it is also true for $\mathcal{F}_\beta$. This completes the proof. $\square$

Remark 3.8. The Euler-Poincaré formula (3.6) is valid restricted to any open domain D bounded by closed leaves of the foliation:

$$\sum_{i:z_i \in D} s_i = 2\#\{i : z_i \in D\} - 2\mathcal{X}(D). \tag{3.7}$$

That is because the Poincaré-Hopf theorem also holds for manifolds with boundary, assuming the vector field is everywhere tangential and non-vanishing on the boundary. More generally, and for similar reasons, if the boundary of D consists of leaves together with singularities then

$$\sum_{i:z_i\in D} s_i + \sum_{j:z_j\in\partial D} s_j(D) = 2\#\{i : z_i \in D\} - 2\mathcal{X}(D) \tag{3.8}$$

where $s_j(D)$ is the number of separatrices of $z_j \in \partial D$ contained in D.

Corollary 3.9. *A homoclinic loop γ is never homotopically trivial.*

Proof. Suppose there does exist a homoclinic loop that bounds a disk $D \subset M$. Let z_j be the associated singularity. Then, (3.8) becomes

$$\sum_{i:z_i\in D} s_i + s_j(D) = 2\#\{i : z_i \in D\} - 2.$$

This relation is impossible, because $s_i \geq 4$ for all i and $s_j(D) \geq 0$. This contradiction proves the corollary. □

Remark 3.10. The Euler-Poincaré formula (3.6) remains valid, as stated, for non-transversely orientable foliations. This can be seen by considering the branched double cover $\tilde{M} \to M$ introduced in Proposition 3.4. Indeed, by construction, every singularity $z_i \in M$ with odd n_i is a degree 2 branching point: the covering map is locally equivalent to $w \mapsto z = w^2$ near the singularity. Replacing this in the normal form (3.4) we see that z_i gives rise to a singularity $\tilde{z}_i \in \tilde{M}$ of the type (3.2) with $m_i = n_i + 1$. Notice that the number of separatrices of $\tilde{z}_i$ is $2(m_i + 1) = 2(n_i + 2)$. All the other points, including the singularities with even n_i, are simply reproduced twice in $\tilde{M}$. Therefore, $\tilde{S} = 2S$ where S and $\tilde{S}$ are the total number of separatrices of $\mathcal{F}_\beta$ and $\mathcal{F}_{\tilde{\beta}}$, respectively. In addition, $\tilde{\kappa} = 2\kappa - \iota$ where κ and $\tilde{\kappa}$ are the numbers of singularities of $\mathcal{F}_\beta$ and $\mathcal{F}_{\tilde{\beta}}$, respectively, and ι is the number of singularities of $\mathcal{F}_\beta$ with odd n_i. Since the covering map is 2-to-1 except at those ι singularities, we also have that $\mathcal{X}(\tilde{M}) = 2\mathcal{X}(M) - \iota$. Thus, using that the Euler-Poincaré formula holds for the transversely orientable foliation $\mathcal{F}_{\tilde{\beta}}$ we conclude that

$$2\kappa - S = \tilde{\kappa} + \iota - \frac{1}{2}\tilde{S} = \mathcal{X}(\tilde{M}) + \iota = 2\mathcal{X}(M).$$

In other words, the formula also holds for $\mathcal{F}_\beta$ as we had claimed.

Measured foliations in zero Euler characteristic. We call *rigid* foliation on the torus T^2 the integral foliation of any constant vector field. We are going to see that every measured foliation on the torus can be mapped to a rigid one by some diffeomorphism. The same argument gives that there are no (transversely orientable) measured foliations on the Klein bottle K^2.

Lemma 3.11. *Let $\mathcal{F}$ be a measured foliation on some surface M with Euler characteristic equal to zero. Then M is diffeomorphic to T^2 and there exists a diffeomorphism that maps $\mathcal{F}$ to a rigid foliation. In particular, either all leaves are closed or else all leaves are dense on the whole surface.*

Proof. The first step is to construct a simple closed cross-section to the foliation. To this end, fix some finite atlas of adapted charts as in Figure 3.2. Consider any adapted chart domain V_0 in the atlas and let γ be an increasing cross-section going across V_0 from bottom to top. Next, consider a chart domain V_1 containing the top endpoint of γ and extend the cross-section across V_1 in the increasing direction until it reaches the top boundary. Repeat this procedure, extending the cross-section across chart domains V_2, V_3, and so on. Since the atlas is finite, the extended cross-section must eventually return to some chart domain it has crossed previously. The first time this happens, modify the construction in the way described in Figure 3.10, to obtain a simple closed curve γ transverse to the foliation. Let $\sigma(t)$, $t \in S^1$ be an increasing parametrization of γ by arc-length.

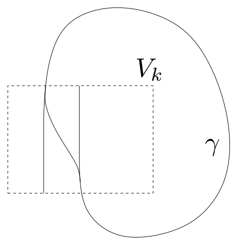


Figure 3.10:

Suppose first that $M = T^2$. As in Example 3.1, we may find an area preserving vector field X tangent to the leaves of $\mathcal{F}$. By the Poincaré recurrence theorem [39], it follows that the leaf through *some* point of γ returns to the cross-section. Next, the Euler-Poincaré formula implies that the foliation has no singularities and, hence, no singular rays. So, we may use Lemma 3.6 to conclude that the orbits of *all* points of γ return to it: there exists a well-defined holonomy (first return) map $\pi : \gamma \to \gamma$. Multiplying X by an appropriate scalar smooth function we obtain a vector field Y (possibly not conservative) such that the first return time relative to the flow $Y^t, t \in \mathbb{R}$ is constant equal to 1. In other words, $\pi(\sigma(t)) = Y^1(\sigma(t))$ for all $t \in S^1$. Since π preserves both orientation and the transverse measure, it must be of the form $\pi(\sigma(t)) = \sigma(t + \theta)$ for some constant $\theta \in S^1$. Now define $\psi : \mathbb{R} \times S^1 \to M$ by $\psi(s,t) = Y^s(\sigma(t - s\theta))$. Then ψ is smooth and it is periodic on s:

$$\psi(s+1,t) = Z^s(\pi(\sigma(t - s\theta - \theta))) = Z^s(\sigma(t - s\theta)) = \psi(s,t).$$

The minimum period is 1 because π is the first return map. So, ψ induces a smooth embedding $\phi : S^1 \times S^1 \to T^2$. As the image is both open and compact, it must coincide with the whole surface. So, ϕ is a diffeomorphism from $S^1 \times S^1$ to $M = T^2$ whose inverse map sends $\mathcal{F}_\beta$ to the rigid foliation of slope θ.

Finally, suppose there exists a measured foliation on $M = K^2$. Let $\hat{\mathcal{F}}$ be the lift of $\mathcal{F}$ to the double cover $\hat{M} \to M$. The cross-section γ lifts to cross-sections γ^+ and γ^- of $\hat{\mathcal{F}}$, corresponding to the two orientations of γ. Then we may construct, exactly as before, a holonomy map $\gamma^+ \to \gamma^+$ and a diffeomorphism $S^1 \times S^1 \to \hat{M}$. The fact that $\hat{\phi}$ is surjective implies that the leaves of $\hat{\mathcal{F}}$ through γ^+ intersect the dual cross-section γ^-. Then there exists a holonomy map $\gamma^+ \to \gamma^-$. Projecting down to M we obtain an orientation reversing holonomy map for $\mathcal{F}$. This contradicts the assumption that $\mathcal{F}$ is transversely orientable. Thus, there are no such foliations on K^2. □

Remark 3.12. This analysis extends to non-transversely orientable foliations. The construction of a simple closed cross-section is analogous. In the case when the holonomy map π reverses orientation we get that ϕ is a diffeomorphism from the Klein bottle K^2 to M. Moreover, all points of γ are periodic for the map π (minimum period 1 or 2). This shows that all measured foliations on the torus are transversely orientable, and non-transversely orientable foliations on K^2 have only closed leaves.

3.3 Stability and periodic components

We are going to prove a global structure theorem for measured foliations, due to Maier [38] (see also Strebel [51]). Having dealt with the case $\mathcal{X} = 0$ in the previous section, here we may assume that the Euler characteristic of the surface M is different from zero. Given a measured foliation $\mathcal{F}_\beta$, a subset of M is called *saturated* if it contains every leaf of $\mathcal{F}_\beta$ it intersects.

Theorem 3.13. *Let M be a surface with non-zero Euler characteristic. Given any measured foliation $\mathcal{F}_\beta$ on M, there exist pairwise disjoint saturated open domains $D_1, \ldots, D_N$ whose closures cover the surface M, such that for every $j = 1, \ldots, N$*

1. *either D_j consists of closed leaves and is homeomorphic to the cylinder*
2. *or D_j consists of non-closed leaves all of which are dense in D_j*

and, in either case, the boundary of D_j consists of saddle-connections and their limit singularities.

In the first case we call D_j a *periodic component* and in the second one we call it a *minimal component* of the foliation. The proof of Theorem 3.13 occupies Sections 3.3 and 3.4. The periodic case is dealt with by the following proposition:

Proposition 3.14. *The union $\mathcal{C}$ of all closed leaves of $\mathcal{F}_\beta$ is an open set. Moreover, $\mathcal{C}$ has finitely many connected components, which are homeomorphic to the open cylinder, and their boundaries consist of saddle-connections and their limit singularities.*

We begin by proving that $\mathcal{C}$ is an open subset of M (*stability lemma*):

Lemma 3.15. *Let p be a point in some closed leaf F of $\mathcal{F}_\beta$. Then the leaf through any point in a neighborhood of p is also closed.*

Proof. Let γ be any cross-section to $\mathcal{F}_\beta$ at the point p. Since the leaf F is closed, there exists a holonomy map $\pi : \gamma' \to \gamma$ of $\mathcal{F}_\beta$ defined on some neighborhood γ' of p inside γ and such that $\pi(p) = p$. For any point $z \in \gamma$, let $[p, z] \subset \gamma$ denote the segment of the cross-section having p and z as endpoints. Then we have $\ell([p, z]) = \ell(\pi([p, z])) = \ell([p, \pi(z)])$. Since π preserves the orientation of γ, this implies $\pi(z) = z$ for all $z \in \gamma'$. In other words, all leaves through γ' are closed. This proves the lemma. $\square$

Next, we analyze the boundary of $\mathcal{C}$. As a first step we characterize the endpoints of the connected components of its intersection with an arbitrary cross-section:

Lemma 3.16. *Let γ be a cross-section to $\mathcal{F}_\beta$ and $p, r \in \gamma$ be such that every leaf through the segment $(p, r] \subset \gamma$ is closed but the leaf through p is not closed. Then p belongs to a saddle-connection.*

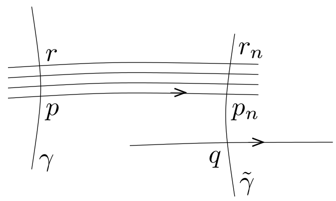


Figure 3.11:

Proof. Suppose some of the rays R of p is regular, that is, there exists some regular point q in its limit set $\omega(R)$. Let us consider some cross-section $\tilde{\gamma}$ at q. Then there exists an injective sequence points $p_n \in \tilde{\gamma} \cap R$ converging to q as $n \to \infty$. Replacing r by a point of γ closer to p, if necessary, we may suppose that $\ell([p, r])$ is smaller than the lengths of the connected components of $\tilde{\gamma} \backslash \{q\}$. Then it is also smaller than the lengths of the connected components of $\tilde{\gamma} \backslash [q, p_n]$ for all large n. This ensures that condition (a) of Lemma 3.6 is satisfied for $p_1 = p$ and $q_1 = r$. Condition (b) is also satisfied, since we assume that all leaves through $(p, r]$ are closed and the ray R is regular. Thus, we may apply Lemma 3.6 to conclude that, for every large n, there is a well defined holonomy map $\pi_n : [p, r] \to \tilde{\gamma}$ with $\pi_n(p) = p_n$. See Figure 3.11. Denote $r_n = \pi_n(r)$. Then every leaf through $(p_n, r_n]$ is closed and $\ell([p_n, r_n]) = \ell([p, r]) > 0$ for all large n. The latter property implies that the segments $[p_n, r_n]$ can not be pairwise disjoint. Then, there exist $m \neq n$ such that $p_m \in [p_n, r_n]$. Consequently, the leaf through p_m is closed, which is just the same as saying that the leaf through p is closed. This contradicts the assumptions, and so both rays of p must be singular. In other words, p belongs to some saddle-connection. $\square$

Lemma 3.16 implies that the boundary of $\mathcal{C}$ consists of saddle-connections and their limit singularities:

Corollary 3.17. *If p is a regular point on the boundary of $\mathcal{C}$ then p lies in some saddle-connection and, locally, all points to at least one side of the saddle-connection belong to $\mathcal{C}$.*

Proof. Let γ be a cross-section to the foliation at p. Suppose first that $\mathcal{C}$ intersects γ at an interval (p, q) having p as one of its endpoints. Then we may use Lemma 3.16 to conclude that p belongs to a saddle-connection. Moreover, all points to one side of the saddle-connection inside a neighborhood of p belong to $\mathcal{C}$. Therefore, the conclusion holds in this case. Otherwise, there exists a sequence of connected components (p_n, q_n) of the intersection $\mathcal{C} \cap \gamma$ converging to p. Then, using Lemma 3.16 once more, all the points p_n and q_n belong to saddle-connections. However, there are finitely many saddle-connections and the intersection of each one of them with the cross-section γ contains finitely many points. This means that this second situation is actually not possible. $\square$

It follows that the union $\mathcal{C}$ of all closed leaves has finitely many connected components, since the number of saddle-connections is finite and each one of them is on the boundary of not more than 2 connected components. All that is to left to do to prove Proposition 3.14 is to show that every connected component of $\mathcal{C}$ is homeomorphic to the open cylinder. We do this with the aid of

Lemma 3.18. *Given any point p in a closed leaf, there exists a cross-section $\bar{\gamma}$ to the foliation at p such that all the leaves through the interior of $\bar{\gamma}$ are closed but the leaves through the endpoints of $\bar{\gamma}$ are not closed.*

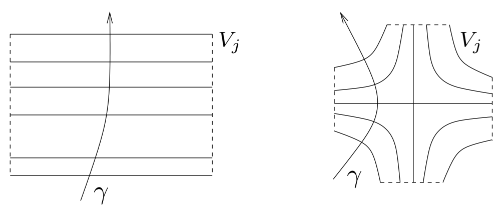


Figure 3.12:

Proof. Let us fix some adapted atlas and some adapted chart domain V_0 in the atlas containing p. Let γ be an increasing cross-section to the foliation at p going across V_0 from bottom to top. Figure 3.12 illustrates what we mean by this, both when the chart domain V_0 is regular and when it is singular. Of course, the extension is far from being unique. Next, consider a chart domain V_1 containing the top endpoint of γ and extend the cross-section across V_1 until it reaches the top boundary, in the same way as before. Repeat this procedure, thus finding successive extensions of the cross-section in the upward direction contained in chart domains V_j, $j \geq 0$. Then carry out the same construction in the opposite direction extending the cross-section downward from p across chart domains V_j, $j \leq 0$.

We claim that after finitely many steps, in either direction, the extended cross-section must intersect some non-closed leaf [3]. Assume this fact for a while. Then we just take $\bar{\gamma}$ to be the segment of γ containing p and having as endpoints the first intersection points with non-closed leaves, in the upward direction and the downward direction. It is clear that $\bar{\gamma}$ is as in the statement of the lemma. To conclude the proof we only have to justify our claim. Suppose the cross-section may be extended indefinitely, in the upward direction say, without ever meeting a non-closed leaf. Then, since the atlas is finite, it must return to some previous chart domain: there are $j < k$ such that $V_j = V_k$. At this point one can modify the construction as described in Figure 3.10 to obtain a closed curve transverse to the foliation whose points all lie in closed leaves. Let $\sigma(t)$, $t \in S^1$ be a parametrization of this closed cross-section and, for each t, let $F(t,s)$, $s \in S^1$ be a parametrization of the leaf through $\sigma(t)$, depending continuously on t. Then $\phi(s,t) = (F(t,s), \sigma(t))$ defines a covering map from the torus $T^2 = S^1 \times S^1$ to some subset of M. The image of ϕ must be the whole M, because it is both compact and open. This implies that the Euler characteristic of M is zero, which contradicts the assumptions. This proves our claim, and completes the proof of the lemma. □

This last kind of reasoning also shows that γ crosses the closed leaves through it at exactly one point. Indeed, suppose γ crosses the same closed leaf twice, at points r_0 and r_1 say. By continuity, there exists a holonomy map π between neighborhoods of r_0 and r_1 inside γ with $\pi(r_0) = r_1$. Since all the leaves through $[r_0, r_1]$ are closed, we can use Lemmas 3.15 and 3.16 to extend π to the whole segment $[r_0, r_1] \subset \gamma$. The image is a segment $[r_1, r_2] \subset \gamma$ such that all the leaves through it are closed. Then, for the same reason as before, we may extend the holonomy map to the segment $[r_1, r_2]$. Repeating this procedure, we find segments $[r_{i-1}, r_i] \subset \gamma$, $i \geq 1$ such that all leaves through them are closed. Since all these segments have the same length, this implies that the cross-section can be extended indefinitely inside $\mathcal{C}$, which contradicts the conclusion of the previous lemma. This shows that γ crosses the closed leaves through it at exactly one point, as we claimed.

Let $\bar{\sigma}(t)$, $t \in [0,1]$ be a parametrization of a cross-section as in Lemma 3.18. Moreover, for each $t \in (0,1)$ let $F(s,t)$, $s \in S^1$ be a parametrization of the leaf through $\bar{\sigma}(t)$ depending continuously on t. Then $\phi(s,t) = (F(s,t), \bar{\sigma}(t))$ defines a continuous map from the cylinder $S^1 \times (0,1)$ to a subset of $\mathcal{C}$. It is clear that ϕ is a local homeomorphism and, by the observation in the previous paragraph, it is also injective. Therefore, ϕ is a homeomorphism onto its image. We are left to show that the image coincides with the connected component of $\mathcal{C}$ that contains it. Since the image of ϕ is an open set, it suffices to prove that it is also closed in $\mathcal{C}$. Let z_n be any sequence of points in the image, converging to some point $z \in \mathcal{C}$. Since the leaves through the z_n intersect $\bar{\gamma}$, and their Hausdorff limit coincides with the leaf through z, the latter must intersect $\bar{\gamma}$, its endpoints included. Actually, the intersection can not be at the endpoints, because the

[3] This happens not later than the first time one encounters a singular domain, because in this case the extension crosses a separatrix. See Figure 3.12.

leaf through z is closed. Hence, the leaf through z intersects the interior of $\bar{\gamma}$, and so z belongs to the image of ϕ. This proves our claim that the image of ϕ coincides with the connected component of $\mathcal{C}$ that contains it.

The proof of Proposition 3.14 is now complete. We close this section by observing that closed leaves in different periodic components are never homotopic. The converse is, clearly, also true.

Corollary 3.19. *The closed leaves of $\mathcal{F}_\beta$ are never homotopically trivial. Moreover, any two closed leaves that are homotopic to each other must belong to the same connected component of $\mathcal{C}$.*

Proof. Suppose a closed leaf bounds a disk $D \subset M$. Applying the Euler-Poincaré formula (Remark 3.8) to this domain, we find that

$$\sum_{i:z_i \in D} (2 - s_i) = 2\mathcal{X}(D) = 2$$

where the sum is over all singularities inside D. This is impossible because $s_i = 2(m_i + 1) \geq 4$ for all i. Therefore, a closed leaf can not be homotopic to a point.

Next, consider any two closed leaves which bound a cylinder $C \subset M$. Applying the Euler-Poincaré formula to C, we find that

$$\sum_{i:z_i \in C} (2 - s_i) = 2\mathcal{X}(C) = 0.$$

This can only happen if the sum is void, that is, if there are no singularities and, hence, no saddle-connections in C. This shows that the boundary of $\mathcal{C}$ does not intersect C, and so the whole cylinder is contained in a single connected component of $\mathcal{C}$. In particular, the two boundary closed leaves belong to the same periodic component, as we claimed. □

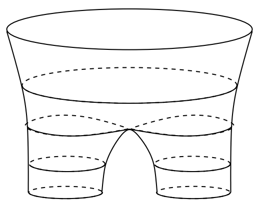


Figure 3.13:

Remark 3.20. If M is an orientable surface of genus g then any family of disjoint, non-homotopic, simple closed curves has at most $3g-3$ elements (see [51, Theorem 2.6]). Consequently, for any measured foliation in M the set $\mathcal{C}$ has at most $3g - 3$ connected components. This upper bound is always attained. This can be seen by considering a decomposition of M into pairs of pants. On each pair of pants one can define a foliation whose leaves are the level sets of a Morse function as described in Figure 3.13: all leaves are closed except for the existence of a double homoclinic associated to an order singularity. By gluing the foliations on all pairs of pants one obtains a measured foliation on M whose leaves are all closed, except for the presence of $\kappa = 2g - 2$ (by Euler-Poincaré) singularities, each one carrying a double loop. The corresponding set $\mathcal{C}$ has $3g - 3$ components, since the homoclinic loops separate all $3g - 3$ curves in M originating from the boundary components of the pairs of pants.

3.4 Recurrence and minimal components

The construction of minimal components, in the context of Theorem 3.13, is provided by the following

Proposition 3.21. *Let R be a regular non-closed ray. Then the interior of $\omega(R)$ is a saturated non-empty set and its boundary consists of saddle-connections and their limit singularities. Moreover, the limit set of every regular ray contained in the interior coincides with the whole $\omega(R)$.*

By Proposition 3.14, a regular non-closed ray can not accumulate a closed leaf; this fact will be used implicitly on a number of occasions. The starting point in the proof of Proposition 3.14 is the *recurrence lemma*:

Lemma 3.22. *Let R be a regular non-closed ray and γ be a cross-section at some point $p \in R$. Then R intersects every non-trivial segment $[p, q] \subset \gamma$.*

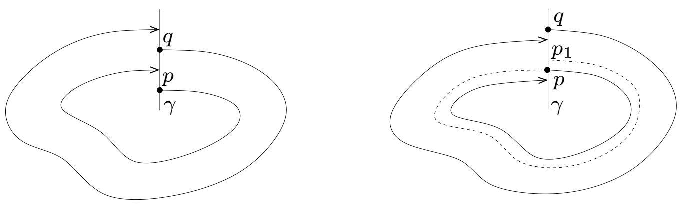


Figure 3.14:

Proof. Reduce $[p, q]$ if necessary to ensure that there is no singular ray leaving from this segment in the same direction as R and never returning to γ; compare condition (b) in Lemma 3.6. We claim that the ray leaving from *some point* in (p, q) must intersect (p, q). Let us consider the case when the surface M is orientable; in the non-orientable case it suffices to apply the same arguments to the orientable double cover $\hat{M}$. Then we may fix an area form on M and, as we saw in Example 3.1, we may find an area preserving vector field Z tangent to the leaves. Up to reversing the orientation of M, we may take Z compatible with the transverse orientation of γ defined by R. Consider the union E of all rays leaving from (p, q) in the same direction as R. This is a positive area set, since it has non-empty interior, and it is forward invariant under the flow Z^t of the vector field Z. If the claim were false then, since (p, q) is transverse to the vector field, the difference $E \setminus Z^t(E)$ would have positive area for $t > 0$. That would contradict the fact that Z is area preserving. This contradiction proves that the ray leaving from some $z \in (p, q)$ intersects (p, q) at some point w, as we claimed.

Now, in view of our initial assumption on $[p, q]$, we may use Lemma 3.6 to conclude that there exists a well-defined holonomy map $\pi : [p, q] \to \gamma$ with $\pi(z) = w$: use the lemma twice, first with $p_1 = z$ and $q = p$ and then with $p_1 = z$ and $q_1 = q$. Since $\ell(\pi[p, q]) = \ell([p, q])$ there are two possibilities for the relative position of $[p, q]$ and its image, that are illustrated in Figure 3.14. In the case on the left hand side of the figure, $\pi(p) \in [p, q]$ and so the conclusion of the lemma holds. In the case on the right hand side, there exists some point $p_1 \in (p, q)$ such that $\pi(p_1) = p$. Choose a point $q_1 \in (p, q)$ close to p_1 and such that $\pi(q_1)$ is also in (p, q), as described in Figure 3.15. Denote $q_2 = \pi(q_1)$. Taking q_1 close

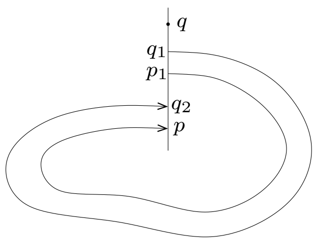


Figure 3.15:

enough to p_1, we may assume that $[p, q_2]$ and $[p_1, q_1]$ are disjoint. Considering the union E' of all rays leaving from (p, q_2), and arguing with a conservative vector field Z in the same way as before, we conclude that the ray leaving from some point in (p, q_2) must intersect (p_1, q_1). Hence, still as before, there exists a well-defined holonomy map $\pi : [p, q_2] \to \gamma$ whose image intersects $[p_1, q_1]$. By further reducing $\ell([p_1, q_1])$ if necessary, the latter implies that $\pi([p, q_2])$ is contained in $[p, q]$. In particular, the ray R leaving from p intersects $[p, q]$, as claimed. $\square$

This lemma means that the limit set of a regular non-closed ray R contains the ray itself. Then, as $\omega(R)$ is saturated, the whole leaf that contains R is contained in $\omega(R)$. Next, we analyze the boundary of the limit set.

Lemma 3.23. *Let R be a regular non-closed ray and q be a regular point on the boundary of $\omega(R)$. Then q belongs to some saddle-connection and, locally, all points to one side of the saddle-connection are in the interior of $\omega(R)$.*

Proof. Let γ be a cross-section to the foliation at q. If $\omega(R)$ intersects γ on a whole neighborhood of q inside the cross-section then the conclusion of the lemma follows. Otherwise, $\gamma \setminus \omega(R)$ contains (maximal) segments $(q_n, r_n) \subset \gamma$ arbitrarily close to q. We claim that the endpoints p_n and q_n belong to saddle-connections. Indeed, let R_n be any ray leaving from p_n or q_n. Suppose R_n was a regular ray. Then, by Lemma 3.22, it would intersect (p_n, q_n). Since R_n is contained in the limit set of R, it would follow that R intersects (p_n, q_n). This is a contradiction, because this interval was chosen in the complement of the limit set of R. This contradiction shows that the points p_n and q_n do belong to saddle-connections. However, there are finitely many saddle-connections and the intersection of each one of them with the cross-section γ contains finitely many points. So, this second possibility can not actually occur. The proof is complete. □

It follows that the interior of the limit set $\omega(R)$ of a regular non-closed ray is non-empty: every point of R belongs to $\omega(R)$, by Lemma 3.22, and these points are interior to the limit set, by Lemma 3.23. It is also clear that these interiors are saturated sets (use the same argument as in Lemma 3.5). The interiors of the limits sets of the different regular non-closed rays are the minimal components in Theorem 3.13. In order to show they do satisfy all the claims in the theorem we also need

Corollary 3.24. *We have $\omega(R') = \omega(R)$ for any regular ray R' inside $\omega(R)$.*

Proof. Since $\omega(R)$ is closed and saturated, it is clear that it contains $\omega(R')$. To prove the other inclusion observe that, by Lemmas 3.22 and 3.23, every point $q \in R'$ is in the interior of $\omega(R')$. The ray R accumulates on q, because $q \in \omega(R') \subset \omega(R)$, and so it must intersect $\omega(R')$. Since the latter is closed and saturated, it follows that $\omega(R) \subset \omega(R')$. This completes the argument. □

Consequently, the limit sets $\omega(R_1)$ and $\omega(R_2)$ of two regular non-closed rays either coincide or have disjoint interiors. This implies that there are finitely many such limit sets, since the number of saddle-connections is finite and each one of them is on the boundary of not more than 2 limit sets.

At this point we proved Proposition 3.21 and we are ready to finish the

Proof of Theorem 3.13. We summarize the conclusions from Propositions 3.14 and 3.21. Let $D_1, \ldots, D_N$ be the connected components of the union of the closed leaves together with the interiors of the limit sets of regular non-closed rays. The former consist of closed leaves, whereas the latter consist of non-closed leaves such that any of their regular rays is dense in the component. All of them are saturated open sets whose boundaries consist of saddle-connections and their limit singularities, and they are finitely many. By construction, they are pairwise disjoint and, since their union contains every leaf which is not a saddle-connection, their closures cover the whole surface. □

From the previous arguments we also get an upper bound for the total number of components of $\mathcal{F}_\beta$ (an upper bound on the number of periodic components was obtained in Remark 3.20):

Corollary 3.25. *The total number N of periodic and minimal components does not exceed $-2\mathcal{X}(M) = 4g(M) - 4$.*

Proof. Each saddle-connection is on the boundary of not more than 2 components. We may suppose $N > 1$, for otherwise there is nothing to prove (recall we consider $\mathcal{X}(M) < 0$). Then we claim that the boundary of every component contains at least 2 saddle-connections. Indeed, suppose there is some component whose boundary consists of a unique saddle-connection γ, together with singularities of the foliation. If γ joins two different singularities then, clearly, it does does not disconnect M. The same is true when γ is a homoclinic loop, by Corollary 3.9. Thus, in either case, the component must be unique, contradicting our assumption that $N > 1$. This proves our claim. It follows that the number N of components is bounded by the number of saddle-connections, which is itself bounded by half the total number of separatrices. Using the Euler-Poincaré formula (3.6) we conclude that

$$N \leq \frac{1}{2}\sum_{i=1}^{\kappa} s_i = \kappa - \mathcal{X}(M) \leq -2\mathcal{X}(M). \tag{3.9}$$

Concerning the last inequality, observe that the Euler-Poincaré formula also implies $\kappa \leq -\mathcal{X}(M)$, because $s_i \geq 4$ for all i. ☐

We call a foliation *minimal* if all regular rays are dense in the whole surface. We also conclude from the previous observations that minimality is typical among measured foliations:

Corollary 3.26. *If a measured foliation has no saddle-connections then it is minimal.*

Proof. From Theorem 3.13 we get that, under the assumptions of the corollary, there exists a unique component and so it coincides with the whole M. If the component were periodic then the foliation would have no singularities and so, by the Euler-Poincaré formula (3.6) and Lemma 3.11, the Euler characteristic of M would be zero. Since we assume otherwise, the component can only be minimal. ☐

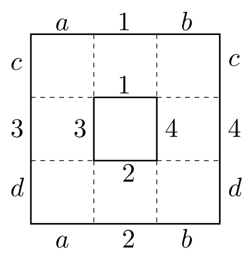


Figure 3.16:

Example 3.27 (Zorich). Consider the translation surface obtained from the annular region in Figure 3.16 by identification of equally labeled sides. The 4 corners, where the sides labeled by letters meet, are identified to a removable singularity. The other 12 vertices, which are endpoints of sides labeled by digits, are identified to a unique singularity, whose angle is 14π. Using (2.9), it follows that the surface has genus 4. Vertical segments connecting vertices of the annular region correspond to homoclinic loops of the vertical foliation. All the other leaves of the vertical foliation are closed. Suppose one rotates the translation structure by an angle θ, that is, one replaces the Abelian differential $\alpha = dz$ by $e^{i\theta}\alpha$. For all but a countable set of values of θ, the saddle-connections are broken, and so the new vertical foliation is minimal. Indeed, this is the case whenever θ is not a rational multiple of π. Analogous remarks apply to the horizontal foliation as well.

We close this section with the observation that if a measured foliation is minimal then there is a well-defined first return map to the cross-section, and this map is conjugate to an interval exchange transformation. More generally,

Lemma 3.28. *Let D be a minimal component of a measured foliation $\mathcal{F}_\beta$ and $\gamma \subset D$ be a cross-section* [4] *to the foliation. Then there exists a holonomy map $\pi : \gamma \to \gamma$ such that $\phi^{-1} \circ \pi \circ \phi$ is an interval exchange transformation, where $\phi : [0, \ell(\gamma)) \to \gamma$ is the arc-length parametrization.*

Proof. Fix some transverse orientation of γ. Let a be the initial endpoint and b be the final endpoint of γ. Let $b_1, \ldots, b_k$, $k \geq 0$ be the points of γ whose rays leaving the cross-section in the chosen (forward) direction are singular and never return to γ. We are going to define two more special points $c, d \in \gamma$, as follows. Consider any sequence of points $y_n \in \gamma$ converging to a such that their rays leaving γ in the opposite (backward) direction to the chosen one are regular. By minimality, these rays intersect γ. Let x_n be the first intersection points, that is so that there is no other intersection in the leaf segment (x_n, y_n), and let c be their limit as $n \to \infty$. Notice that either c is one of the b_i or else the forward ray of c intersects γ, and a is the first intersection point. The definition of d is similar, starting with a sequence $y_n \in \gamma$ that converges to b. Let B the set formed by a, b, $b_1, \ldots, b_k$, c, d; then B has from $k+2$ to $k+4$ elements. Consider any of the segments $I_\alpha \subset \gamma$ determined by the points of B. By minimality and Lemma 3.6, there exists a holonomy map f from the interior of each I_α to some open interval in γ such that $f(x)$ is the first point of intersection of the forward ray of x with γ. Extend f to the initial endpoint ∂I_α by continuity. Notice that the images $f(I_\alpha)$ are pairwise disjoint, because every $f : I_\alpha \to \gamma$ is a first return map. Since all these maps preserve orientation and transverse length, it follows that $\phi^{-1} \circ f \circ \phi$ is an interval exchange transformation, as claimed. □

Remark 3.29. Assume c and d do coincide with points in $\{b_1, \ldots, b_k\}$, as in the configuration of Figure 3.17. Then we can give an interpretation of the Keane

[4] For consistency with the terminology in Chapter 1, here we suppose that γ contains its initial endpoint but not the final one.

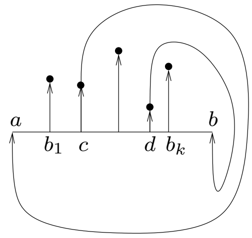


Figure 3.17:

condition for the interval exchange map in terms of the measured foliation. Suppose f does not satisfy the Keane condition: there exist α and β such that $f^m(\partial I_\alpha) = \partial I_\beta$ and $\partial I_\beta \neq a$. It is no restriction to suppose $\partial I_\alpha \neq a$: otherwise, just replace a by c and m by $m+1$. Then both ∂I_α and ∂I_β are points in $\{b_1, \ldots, b_k\}$. By construction, every b_i is connected to $f(b_i)$ by a path formed by leaf segments and at least one singularity. Then $f^m(\partial I_\alpha) = \partial I_\beta$ implies that the points ∂I_α and ∂I_β are contained in a path formed by leaf segments one of which, at least, is a saddle-connection. Compare Corollary 3.26.

3.5 Representation of translation surfaces

We have seen in Chapter 2 that every triple (π, λ, τ) defines a translation surface, that is, a pair (M, α) where M is a Riemann surface and α is an Abelian differential α on M. Actually, the triple determines an additional structure on the surface, namely, the choice of an *adapted cross-section*, that is, a horizontal segment whose left endpoint is a singularity (possibly removable), whose right endpoint belongs to a vertical separatrix, and which intersects every regular vertical leaf. Recall (2.11). We call *distinguished separatrix* the one that contains this horizontal cross-section.

In the present section we describe an inverse construction where, given a translation surface whose vertical foliation $\mathcal{F}_\alpha^v$ has no saddle-connections, *and* given a choice of an adapted cross-section to $\mathcal{F}_\alpha^v$, one obtain a representation of (M, α) through some triple (π, λ, τ). In Corollary 3.34 we deduce that every translation surface may be represented in the form of a simple planar polygon with parallel sides identified by translations, as introduced in Chapter 2. This representation is, of course, not unique: different choices of an adapted cross-section lead to different values of (π, λ, τ). We are going to see in Section 3.7 that all the permutation pairs π one can obtain in this way, for a given translation surface, belong to the same *extended* Rauzy class.

We assume that some finite non-empty set of "singular points" has been chosen on the translation surface, containing all the singularities of the Abelian differential; some elements may removable singularities, that is, marked points where α is actually regular.

Proposition 3.30. *Let (M, α) be translation surface such that either the vertical foliation $\mathcal{F}^v_\alpha$ or the horizontal foliation $\mathcal{F}^h_\alpha$ have no saddle-connections. Then (M, α) may be represented in the form of zippered rectangles, involving $d = 2g + \kappa - 1$ rectangles, where g is the genus of M and $\kappa \geq 1$ is the number of singular points.*

For clearness, we divide the construction into 4 main steps that are detailed in the sequel. Notice that the leaves of the horizontal (respectively, vertical) foliation of α come with a natural orientation, relative to which the real part (respectively, the imaginary part) of α are increasing on each leaf. We call a horizontal/vertical separatrix *incoming* if its natural orientation points toward the associated singularity, otherwise we call it *outgoing*. Recall (Figure 3.3) that a singularity of order m_i has $2(m_i + 1)$ horizontal (respectively, vertical) separatrices, and they are alternately incoming and outgoing.

Step 1: We choose an adapted cross-section σ to the vertical foliation. Let z_1 be any singularity and σ_1 be a segment in any of its outgoing separatrices having z_1 as an endpoint. The assumption that at least one of the two foliations, $\mathcal{F}^v_\alpha$ or $\mathcal{F}^h_\alpha$, has no saddle-connections ensures that σ_1 may be chosen intersecting every regular ray of the vertical foliation. Indeed, if the vertical foliation has no saddle-connections then, by Corollary 3.26, all regular vertical rays are dense, and so they do intersect σ_1 (in this case σ_1 can be arbitrarily short). If the horizontal foliation has no saddle-connections then, for the same reason, all regular horizontal rays are dense. In particular, taking σ_1 sufficiently long, it must intersect every regular vertical ray.

Let z_2 be any singularity (possible $z_1 = z_2$) and ν be any vertical separatrix of z_1, either incoming or outgoing. The first case is illustrated in Figure 3.18; the second one is analogous, with z_2 lying below σ_1. The construction of σ_1 implies, in particular, that ν intersects σ_1. Let w be the first intersection point, that is, the unique point in $\nu \cap \sigma_1$ such that the segment $[z_2, w] \subset \nu$ contains no other point of σ_1. Then denote by σ the horizontal segment $[z_1, w] \subset \sigma_1$.

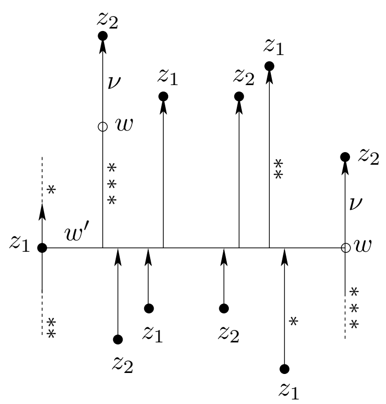


Figure 3.18:

Step 2: We identify the points where the vertical separatrices first meet σ. More precisely, we introduce the finite subsets σ^+ and σ^- of the cross-section, having exactly the following elements (Figure 3.18 corresponds to an example where $\kappa = 2$ with $m_1 = m_2 = 1$ and the separatrix ν is incoming):

(i) The first intersection point with σ of every incoming vertical separatrix of every singularity is an element of σ^+. Analogously, the first intersection point with σ of every outgoing vertical separatrix of every singularity is an element of σ^-. In particular, $w \in \sigma^+$ if ν is incoming and $w \in \sigma^-$ is ν is outgoing.

(ii) The point w is always an element of both σ^+ and σ^-. Moreover, extending ν past r until it intersects σ again, the second intersection point w' is an element of σ^+ if ν is incoming and is an element of σ^- id ν is outgoing. Finally, the singularity z_1 is an element of both σ^+ and σ^-.

Using the Euler-Poincaré formula (3.6),

$$\#\sigma^+ = 2 + \#\{\text{incoming separatrices}\} = 2 + \sum_{i=1}^{\kappa}(m_i + 1) = 2g + \kappa \tag{3.10}$$

and, analogously, $\#\sigma^- = 2g + \kappa$. The term 2 in the middle terms of (3.10) corresponds to the points included in part (ii) of the definition.

Let $|\lambda|$ be the length of σ and $\phi : [0, |\lambda|] \to M$ be the parametrization of σ by arc-length with $\phi(0) = z_1$. Let $d = 2g + \kappa - 1$. Write

$$\sigma^\pm = \{z_1 = s_0^\pm < s_1^\pm < \cdots < s_{d-1}^\pm < s_d^\pm = w\}$$

where $<$ refers to the natural orientation on σ. Then there are numbers

$$0 = a_0^\pm < a_1^\pm < \cdots < a_{d-1}^\pm < a_d^\pm = |\lambda|$$

such that $\phi(a_j^\pm) = s_j^\pm$. Denote $\lambda_j^\pm = |I_j^\pm| = a_j^\pm - a_{j-1}^\pm$ for every $j = 1, \ldots, d$. Notice that $|\lambda| = \sum_{j=1}^{d} \lambda_j^\pm$.

Step 3: We define an interval exchange transformation associated to the translation surface and the cross-section. Since the vertical foliation is minimal, every regular vertical ray leaving σ in the upward (or downward) direction must return to σ. Then, using Lemma 3.6, for each $j = 1, 2, \ldots, d$ the first-return map $f_j : (s_{j-1}^+, s_j^+) \to \sigma$ for the vertical foliation is well-defined and continuous. Note that these maps preserve the arc-length on σ, because the latter corresponds to the transverse measure of $\mathcal{F}_\alpha^v$. In particular, f_j extends continuously to s_{j-1}^+. Define

$$f : \sigma \to \sigma, \quad f \mid [s_{j-1}^+, s_j^+) = f_j, \quad \text{for each } j = 1, \ldots, d.$$

Recalling that the points s_i^+ belong to separatrices, and using the local form of the foliation at the singularities (see (3.2) and Figure 3.3) one immediately gets that the endpoints of every image $f([s_{j-1}^+, s_j^+))$ are elements $s_k^- < s_l^-$ of

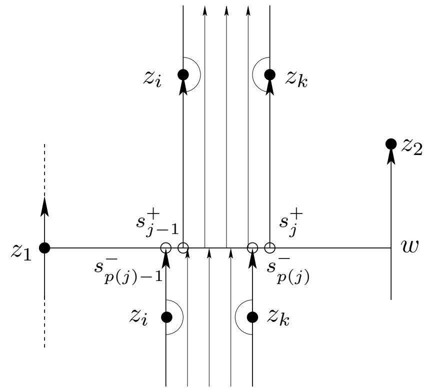


Figure 3.19:

σ^-. See Figure 3.19. Furthermore, they must be consecutive points, that is, we must have $k = l - 1$, because no element of σ^- can be the image under a holonomy map of any point in $\sigma \setminus \sigma^+$. Writing $l = p(j)$, we conclude that

$$f \text{ maps each } [s^+_{j-1}, s^+_j) \text{ isometrically to } [s^-_{p(j)-1}, s^-_{p(j)}). \tag{3.11}$$

Notice that p is a bijection of the set $\{1, \dots, d\}$ to itself.

Transporting this construction from σ to $[0, |\lambda|)$ via the parametrization ϕ, we obtain an interval exchange transformation on $[0, |\lambda|)$, that we also denote as f, mapping each interval $[a^+_{j-1}, a^+_j)$ isometrically to $[a^-_{p(j)-1}, a^-_{p(j)})$. This transformation is defined by the length vector $\lambda = (\lambda^+_j)_j$, and the permutation pair $\pi = (\pi_0, \pi_1)$ with $\pi_0 = \mathrm{id}$ and $\pi_1 = p$ on the alphabet is $\mathcal{A} = \{1, \dots, d\}$. Compare Lemma 3.28.

Step 4: We identify the suspension data h and exhibit the zippered rectangles representation. For each $z \in \sigma$, let $\psi_z : J_z \to M$ be the arc-length parametrization of the (oriented) vertical leaf with $\psi_z(0) = z$. If the leaf is regular, ψ_z is defined on the whole $\mathbb{R}$, otherwise the domain J_z is some subinterval. For $z \in \sigma \setminus \sigma^+$, let $h(z) > 0$ be the return time of z to the cross-section σ under the vertical flow:

$$f(z) = \psi_z(h(z)).$$

To each point $s^\pm_j \in \sigma^\pm$ we associate a number $b^\pm_j \geq 0$ as follows (compare the definition of $\sigma^\pm$ above):

(i) If $s^\pm_j$ is the first intersection point with σ some separatrix of a singularity, then $b^\pm_j$ is the length of the vertical segment connecting $s^\pm_j$ to the singularity.

(ii) If ν is incoming then $w = s^-_d$ and $w' = s^+_k$ for some $k < d$. By definition, $b^-_d = b^+_k$ = length of the vertical segment connecting the two points. If

ν is outgoing then $w = s_d^+$ and $w' = s_k^-$ for some $k < d$. By definition, $b_d^+ = b_k^-$ = length of the vertical segment connecting the two points. Finally, corresponding to the singularity z_1, define $b_0^\pm = 0$ in either case.

Lemma 3.31. *For every $z \in (s_{j-1}^+, s_j^+)$ and every $j = 1, \ldots, d$,*

$$h(z) = b_{j-1}^+ + b_{p(j)-1}^- = b_j^+ + b_{p(j)}^-.$$

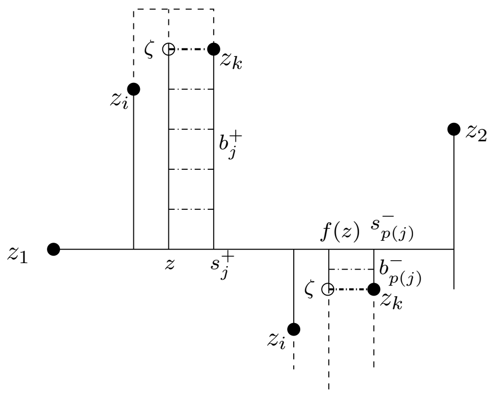


Figure 3.20:

Proof. Let $[z, s_j^+] \subset \sigma$ be the horizontal segment determined by z and s_j^+ and let $[s_j^+, z_k]$ be the vertical segment determined by s_j^+ and the singularity z_k associated to it. Let φ_z be the horizontal holonomy from $[s_j^+, z_k)$ to the vertical leaf through z, with $\varphi_z(s_j^+) = z$. See Figure 3.20. Observe that φ_z is indeed defined on the whole $[s_j^+, z_k)$: just notice that the horizontal leaf segment connecting each $\xi \in [s_j^+, z_k)$ to the vertical leaf through z coincides with

$$\{\psi_\eta(t_\xi) : \eta \in [z, s_j^+]\}, \qquad t_\xi = \text{length of } [s_j^+, \xi]$$

(because horizontal holonomy preserves arc-length). Then φ_z extends to z_k by continuity: the image $\zeta = \varphi_z(z_k)$ is the first intersection point of a horizontal separatrix of z_k with the vertical leaf through z. Since horizontal holonomy preserves arc-length, we conclude that length of $[z, \zeta]$ = length of $[s_j^+, z_k] = b_j^+$. Analogously, length of $[\zeta, f(z)]$ = length of $[z_k, s_{p(j)}^-] = b_{p(j)}^-$. See Figure 3.20. These two relations yield

$$h(z) = \text{length of } [z, \zeta] + \text{ length of } [\zeta, f(z)] = b_j^+ + b_{p(j)}^-.$$

The proof of the other equality in the statement is analogous. $\square$

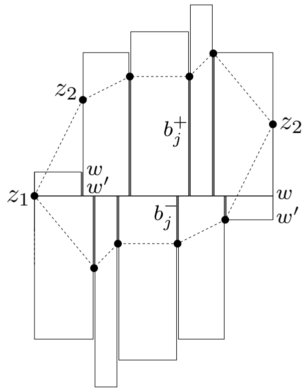


Figure 3.21:

We are ready to exhibit the zippered rectangles model for the translation surface (M, α). Let $h_j = h(z)$ for any $z \in (s_{j-1}^+, s_j^+)$ and $j = 1, \ldots, d$. Let Z be the union of the rectangles $[a_{j-1}^+, a_j^+) \times [0, h_j)$ over $1 \le j \le d$ and then define $\Psi : Z \to M$ by (see Figure 3.21 and compare also Section 2.5)

$$\Psi(x, y) = \begin{cases} \psi_{s^-_{p(j)}}(y - h_j) & \text{if } x = a_j^+ \text{ and } y > b_j^+ \\ \psi_{\phi(x)}(y) & \text{in all other cases.} \end{cases} \tag{3.12}$$

Lemma 3.32. *Denote $Z_0 = \{(a_j^+, b_j^+),\ j = 0, 1, \ldots, d\}$.*

1. *Ψ is a differentiable local isometry on $Z \setminus Z_0$.*

2. *Ψ is injective on $Z \setminus Z_0$ and it is surjective.*

Proof. By construction, Ψ maps every vertical segment $\{x\} \times [0, h_j]$, $x \notin \{a_j^+\}$ to a vertical segment of α, preserving arc-length. The same is true for the vertical segments $\{a_j^+\} \times [0, b_j^+)$ and $\{a_j^+\} \times (b_j^+, h_j]$. Moreover, Ψ maps every horizontal segment $[a_{j-1}^+, a_j^+] \times \{y\}$, $y \in [0, h_j]$ to a horizontal segment of α, also preserving arc-length. For $y = 0$ this is an immediate consequence of the definitions: $\Psi(x, 0) = \phi(x)$ for every x, and $\Psi([a_{j-1}^+, a_j^+] \times \{0\}) = [s_{j-1}^+, s_j^+] \subset \sigma$. It follows for any other y, since the holonomies of the horizontal foliation and the foliation preserve arc-length. In addition, for any $y \neq b_j^+$, the image of a horizontal segment $(a_j^+ - \delta, a_j^+ + \delta) \times \{y\}$ is also a horizontal segment of α, since it is a connected union of two horizontal segments. Once more, arc-length is preserved. In summary, at every point in the complement of $\{(a_j^+, b_j^+)\}$ the map Ψ sends vertical segments to vertical segments of α and horizontal segments to horizontal segments of α, preserving arc-length. This implies Ψ is C^1 and a local isometry, as stated in the first part of the lemma.

Next, we prove surjectivity. Consider any point $\zeta \in M$. By construction, every singularity of α can be written as $\Psi(a_j^+, b_j^+)$ for some $1 \le j < d$. Hence, it is no restriction to suppose ζ is a regular point. Suppose first that the backward vertical ray of ζ meets σ, that is, there exists some point in the intersection of σ with the vertical leaf through ζ, preceding ζ relative to the upward orientation on the leaf. Take $z \in \sigma$ to be the last intersection point, that is, the unique point in $F \cap \sigma$ such that the vertical segment $[z, \zeta]$ contains no other point of σ. If this last intersection point is the endpoint w then take $z = w'$ instead. Then $z = \phi(x)$ for some $x \in (a_0^+, a_d^+)$. Moreover, either

$$x \in (a_{j-1}^+, a_j^+) \quad \text{and} \quad y < h_j \quad \text{or} \quad x = a_j^+ \quad \text{and} \quad y < b_j^+,$$

for some $j = 1, \dots, d$, where $y =$ length of $[z, \zeta]$. In either case, $\zeta = \Psi(x, y)$. Next, suppose the backward ray of ζ does not intersect σ. Then, since the vertical foliation is minimal, ζ must belong to an outgoing separatrix of some singularity. Let z' be the first intersection point of the separatrix with σ, with the convention that if this happens to be the endpoint w then we take $z' = w'$ instead. Then $z' = s_k^-$ for some $0 < k < d$, and $\zeta = \psi_{z'}(-t_\zeta)$ where $t_\zeta =$ length of $[\zeta, z']$. Let $j \in \{1, \dots, d\}$ be defined by $k = p(j)$ and let $y = h_j - t_\zeta$. Notice that $y \in (0, h_j)$ because, according to Lemma 3.31, we have $0 < t_\zeta < b_k^- < h_j$. Then, by the definition of Ψ,

$$\Psi(a_j^+, y) = \psi_{s_k^-}(h_j - y) = \psi_{s_k^-}(-t_\zeta) = z.$$

Now we prove injectivity. By construction, each rectangle $(a_{j-1}^+, a_j^+) \times [0, h_j)$ is mapped injectively to a domain in M, consisting of points whose backward vertical rays intersect σ for the last time at the segment (s_{j-1}^+, s_j^+) and whose forward vertical rays also intersect σ. Each segment $\{a_j^+\} \times [0, b_j^+)$ is mapped injectively to a vertical segment consisting of points whose backward vertical ray intersects σ for the last time at s_j^+ and whose forward vertical ray does not intersect σ. Finally, each segment $\{a_j^+\} \times [b_j^+, h_j)$ is mapped injectively to a vertical segment consisting of points whose backward vertical ray does not intersect σ and whose forward vertical ray intersects σ for the last time at $s_{p(j)}^-$. This description shows that the images of these various sets are all pairwise disjoint. The injectivity statement follows. $\square$

This lemma completes the proof of Proposition 3.30. From the argument we also get a representation of the translation surface in terms of a planar polygon, as introduced in Section 2.2, with data (π, λ, τ) given by

- $\pi = (\pi_0, \pi_1)$ with $\pi_0 = \mathrm{id}$ and $\pi_1 = p$ on the alphabet $\mathcal{A} = \{1, \dots, d\}$
- $\lambda = (\lambda_j)_j$ where $\lambda_j =$ length of $[s_{j-1}^+, s_j^+] =$ length of $[s_{p(j)-1}^-, s_{p(j)}^-]$
- $\tau = (\tau_j)_j$ where $\tau_j = b_j^+ - b_{j-1}^+ = -b_{p(j)}^- + b_{p(j)-1}^-$.

Notice that τ belongs to the cone T_π: for any $1 \le k < d$, we have

$$\sum_{j=1^k} \tau_j = b_k^+ > 0 \quad \text{and} \quad \sum_{j=1^k} \tau_p(j) = -b_{p(k)}^- < 0.$$

Moreover, we are nearly ready to prove that every translation surface admits such a polygon representation. All we need is the following simple observation:

Lemma 3.33. *Given any Abelian differential α on a Riemann surface M, the horizontal/vertical foliation of $e^{i\theta}\alpha$ has no saddle-connections for all but countably many values of θ.*

Proof. Let $S \subset M$ be the set of singularities of α. Consider the flat metric defined on $M \setminus S$ by the Abelian differential α. The geodesics $\gamma(t)$ are characterized by the property that $\alpha_{\gamma(t)}(\dot\gamma(t))$ is constant. The horizontal (respectively, vertical) leaves of $e^{i\theta}\alpha$ coincide with the geodesics such that

$$e^{i\theta}\alpha_{\gamma(t)}(\dot\gamma(t)) \in \mathbb{R} \quad \text{(respectively, } \in i\mathbb{R}). \tag{3.13}$$

Let $\mathcal{C}$ denote the set of connections, that is, C^1 curves $\gamma : [0,1] \to M$ such that $\gamma(0)$, $\gamma(1) \in S$ and $\gamma \mid (0,1)$ is a geodesic for the flat metric on $M \setminus S$. The horizontal (respectively, vertical) foliation of $e^{i\theta}\alpha$ has saddle-connections if and only if (3.13) holds for some $\gamma \in \mathcal{C}$. Hence, it suffices to prove that $\mathcal{C}$ is countable and, for this, it is enough to show that the subset of $\gamma \in \mathcal{C}$ with length $|\gamma| \le L$ is finite, for any $L > 0$. Indeed, suppose there exists an injective sequence γ_j of connections with lengths bounded by L. Up to replacing it by a subsequence, we may suppose that $p = \gamma_j(0)$ is independent of j, the tangent $\dot\gamma_j(0)$ converges to some vector v, and $|\gamma_j|$ also converges. Then $\gamma_j(1)$ converges to some $q \in M$. Since $\gamma_j(1) \in S$ for all j and S is a discrete set, it follows that $\gamma_j(1) = q$ for all large j. This implies that the γ_j coincide for all large j, contradicting the assumption that the sequence is injective. $\square$

Corollary 3.34. *Every translation surface (M, α) may be represented as a simple planar polygon with $2d$ sides, $d = 2g + \kappa - 1$, modulo identification by translation of pairs of parallel sides.*

Proof. Using Lemma 3.33, we may find $\theta \in S^1$ such that $e^{i\theta}\alpha$ has no saddle-connections. By the previous observations (recall also Remarks 2.1 and 2.15), $e^{i\theta}\alpha$ is isomorphic to the canonical Abelian differential dz on some simple planar polygon P with $2d$ sides. This means that α itself is isomorphic to $e^{-i\theta}dz$ on P or, equivalently, to the canonical form dz on the rotated polygon $e^{i\theta}P$. $\square$

Remark 3.35. The sides of the rotated polygon $e^{i\theta}P$ are congruent to the vectors $(\tilde\lambda_\beta, \tilde\tau_\beta)$ defined in complex notation by

$$\tilde\lambda_\beta + i\tilde\tau_\beta = e^{i\theta}(\lambda_\beta + i\tau_\beta) \quad \text{for } \beta \in \mathcal{A},$$

where (π, λ, τ) are the data describing the polygon P. In general, $\tilde\lambda$ needs not belong to $\Lambda_{\mathcal{A}}$. For instance, we may have $\tilde\lambda_\beta = 0$ for some $\beta \in \mathcal{A}$, corresponding to a rotated polygon with vertical sides, corresponding to saddle-connections of the vertical foliation.

Example 3.36 (Zorich). The translation surface in Figure 3.16 is non-generic: there are 7 homoclinic saddle-connections for the vertical foliation, involving all 14 separatrices of the singularity. All other leaves are closed, contained in 4 distinct periodic components. Moreover, the same is true for the horizontal foliation. In particular, there is no horizontal cross-section (contained in a single leaf) which intersects all regular vertical rays, and so the previous construction can not be carried out in this case.

3.6 Representation changes

The construction $(M, \alpha) \mapsto (\pi, \lambda, \tau, h)$ presented in the previous section depends on the choice of a horizontal adapted cross-section $\sigma = [z_1, w]$ such that

(a) σ is contained in some horizontal separatrix σ_0 of a singularity z_1

(b) w is a point of first intersection of σ with some vertical separatrix.

Here and in the next section we analyze this dependence. By *zippered rectangles representation* we always mean the construction in Proposition 3.30 (but a slight generalization will be introduced in Section 3.7). In particular, it is implicit that the representation involves $d = 2g + \kappa - 1$ rectangles, where κ is the number of "singular points" (all the zeros of the Abelian differential plus, possibly, some marked points). In what follows we assume that the translation surface is *generic*, meaning that the are no saddle-connections neither for the vertical foliation nor for the horizontal foliation.

Here we take the distinguished separatrix σ_0 to be fixed, and we study the effect of changing the endpoint w. In our present notation, the Rauzy-Veech induction operator $\hat{\mathcal{R}}$ can be described as follows. Let (π, λ, τ, h) be a representation of the translation surface associated to some adapted cross-section σ. Let σ and σ^- be the sets of intersection points of σ with vertical separatrices defined in Section 3.5 (step 2 in the proof of Proposition 3.30). Let w^1 be the point of $\sigma^+ \cup \sigma^- \setminus \{w\}$ closest to w. Then $(\pi^1, \lambda^1, \tau^1, h^1)$ given by $(\pi^1, \lambda^1, \tau^1) = \hat{\mathcal{R}}(\pi, \lambda, \tau)$ is the representation of the translation surface associated to the adapted cross-section $\sigma^1 = [z_1, w^1]$. Observing also that $\hat{\mathcal{R}}$ is invertible, we get that $(\pi^n, \lambda^n, \tau^n, h^n)$ given by $(\pi^n, \lambda^n, \tau^n) = \hat{\mathcal{R}}^n(\pi, \lambda, \tau)$ is a representation of the translation surface, corresponding to the same distinguished separatrix, for every $n \in \mathbb{Z}$. The converse is also true:

Proposition 3.37. *Let (π, λ, τ, h) and $(\tilde{\pi}, \tilde{\lambda}, \tilde{\tau}, \tilde{h})$ be representations of the same translation surface (M, α) in the form of zippered rectangles, with the same distinguished separatrix. Then there exists $n \in \mathbb{Z}$ such that $(\tilde{\pi}, \tilde{\lambda}, \tilde{\tau}) = \hat{\mathcal{R}}^n(\pi, \lambda, \tau)$.*

Proof. Up to reversing the roles of the two representations, it is no restriction to suppose $(\tilde{\pi}, \tilde{\lambda}, \tilde{\tau}, \tilde{h})$ originates from an adapted cross-section $\tilde{\sigma} = [z_1, \tilde{w}]$ which is contained in $\sigma = [z_1, w]$. Let $\sigma^n = [z_1, w^n]$, $n \geq 0$ be the decreasing sequence of horizontal segments obtained by successive application of the Rauzy-Veech induction operator, starting from $w^0 = w$. We want to prove that $\tilde{w} = w^n$

for some n. Let σ^{n+} and σ^{n-} be the sets of special points in the adapted cross-section σ^n as constructed in step 2 of Proposition 3.30. Let $\tilde{w}_1$ be the first point of intersection with σ of the vertical separatrix $\tilde{\nu}_0$ that contains $\tilde{w}$. Then $\tilde{w}_1 \in \sigma^{n+} \cup \sigma^{n-}$ for all $n \geq 0$ such that $\tilde{w}_1 \in \sigma^n$. Notice also that the length of σ^n goes to zero as $n \to \infty$, by Corollary 1.20. Thus, there must exist $n(1) \geq 1$ such that $w^{n(1)} = \tilde{w}_1$. Let $\tilde{w}_2$ be the second point of intersection of $\tilde{\nu}_0$ with the cross-section $\sigma^{n(1)}$. Notice that $\tilde{w}_2 \in \sigma^{n+} \cup \sigma^{n-}$ for $n = n(1)$, corresponding to step (ii) in the definition of these sets, and then the same is true for every $n \geq n(1)$ such that $\tilde{w}_2 \in \sigma^n$. So, there must be $n(2) > n(1)$ such that $w^{n(2)} = \tilde{w}_2$. Repeating this procedure, we eventually reach $n(k) \geq 1$ such that $\tilde{w} = \tilde{w}_k = w^{n(k)}$. The proof of the proposition is complete. $\square$

In particular, for all representations (π, λ, τ, h) of a translation surface (M, α) associated to a given distinguished separatrix, the permutation pair belongs to the same Rauzy class. We are going to see that a similar statement is true if one also allows for the distinguished separatrix to vary, but we must consider *extended* Rauzy classes instead.

3.7 Extended Rauzy classes

At this point it is convenient to generalize somewhat the zippered rectangles representation construction. So far, we have always taken the distinguished separatrix to be outgoing, that is, to point to the right of the corresponding singularity z_1. We drop this restriction, and consider incoming horizontal separatrices as well. The advantage of doing this will be apparent in the proof of Proposition 3.39.

The theory extends immediately to the incoming case, by symmetry. In particular, in this setting the Rauzy-Veech induction operator is defined in terms of the lengths of the two *leftmost* intervals. More precisely, it is given by

$$\hat{\mathcal{R}}^{\mathrm{ad}} = \mathrm{ad} \circ \hat{\mathcal{R}} \circ \mathrm{ad}$$

where ad is the involution $\mathrm{ad}(\pi, \lambda, \tau) = (\mathrm{ad}(\pi), \lambda, \tau)$ that acts on π by reversing the order of the terms in both rows. Figure 3.22 describes an example of such dual zippered rectangles (notice the singularity z_1 is at the *right* endpoint), with

$$\pi = \begin{pmatrix} A & B & C & D & E \\ D & C & A & E & B \end{pmatrix}$$

and $\lambda_A > \lambda_D$. Hence A is the winner and D is the loser. We have

$$\mathrm{ad}(\pi) = \begin{pmatrix} E & D & C & B & A \\ B & E & A & C & D \end{pmatrix}$$

and the induction operator maps $(\pi, \lambda, \tau) \mapsto (\pi', \lambda', \tau')$ with

$$\pi' = \begin{pmatrix} B & C & A & D & E \\ D & C & A & E & B \end{pmatrix}, \quad \lambda'_A = \lambda_A - \lambda_D \quad \text{and} \quad \tau'_A = \tau_A - \tau_D.$$

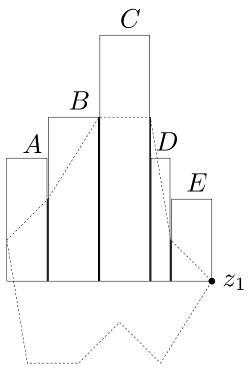


Figure 3.22:

Remark 3.38. The monodromy invariants p and p' of the permutation pairs π and $\pi' = \mathrm{ad}(\pi)$ are conjugated by the involution $p_0(j) = d + 1 - j$. Let σ and σ' be the associated permutations of $\{0, 1, \dots, d\}$, defined as in (2.19). Direct computation gives (recall the expression of σ^{-1} in Remark 2.4)

$$p_0(\sigma'(p_0(j))) = P^{-1}(P(j) - 1) + 1 = t(\sigma^{-1}(t^{-1}(j)))$$

for all j, where $t(j) = j + 1$. Denoting $q_0(j) = d - j$, we conclude that

$$\sigma' = q_0 \circ \sigma^{-1} \circ q_0 \quad \text{for all } j = 0, 1, \dots, d. \tag{3.14}$$

Proposition 3.39. *Let (π, λ, τ, h) and $(\tilde{\pi}, \tilde{\lambda}, \tilde{\tau}, \tilde{h})$ be zippered rectangles representations of the same translation surface. Then there exist $n \geq 0$ and*

$$\hat{\mathcal{R}}_1, \dots, \hat{\mathcal{R}}_n \in \{\hat{\mathcal{R}}, \hat{\mathcal{R}}^{-1}, \hat{\mathcal{R}}^{\mathrm{ad}}, (\hat{\mathcal{R}}^{\mathrm{ad}})^{-1}\}$$

such that $(\tilde{\pi}, \tilde{\lambda}, \tilde{\tau}) = \hat{\mathcal{R}}_n \circ \cdots \circ \hat{\mathcal{R}}_1(\pi, \lambda, \tau)$.

Proof. Let us begin with the following observation. Let $[z_1, w_1]$ and $[w_2, z_2]$ be adapted cross-sections, one outgoing and the other incoming, such that the vertical holonomy $h : (z_1, w_1) \mapsto (w_2, z_2)$, where $h(z)$ is the point where the vertical leaf through z first intersects $[w_2, z_2]$, is well-defined and a homeomorphism. We say that the two cross-sections *match*. See Figure 3.23 for an example.

Lemma 3.40. *If two adapted cross-sections match then the corresponding zippered rectangles representations of the translation surface coincide.*

Proof. The map h may be written as $h(z) = V^\theta(z)$, where $(V^t)_t$ denotes the vertical flow: the hitting time $\theta \in \mathbb{R}$ is constant because horizontal holonomies preserve arc-length. Moreover, h is an isometry, because vertical holonomies preserve arc-length, and it extends to $h : [z_1, w_1] \mapsto [w_2, z_2]$. Clearly, h conjugates the first return maps f_1 and f_2 of the vertical flow to the adapted cross-sections $[z_1, w_1]$ and $[w_2, z_2]$, respectively. It follows that f_1 and f_2 correspond

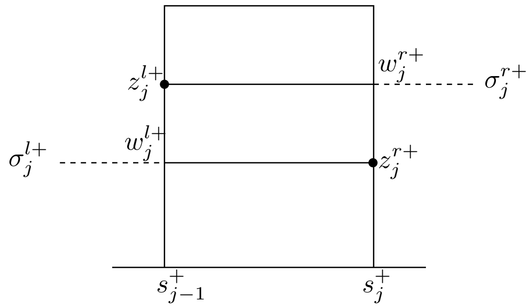


Figure 3.23:

to the same data π and λ (as long as the alphabets are chosen consistently). The finite subsets $\sigma_1^+ \subset [z_1, w_1]$ and $\sigma_2^+ \subset [w_2, z_2]$ constructed in Step 1 of the proof of Proposition 3.30 are in one-to-one correspondence, and the heights of the associated vertical separatrices are related by $b_{j,1}^+ = b_{j,2}^+ + \theta$, and so $\tau_j = b_{j,1} - b_{j-1,1} = b_{j,2} - b_{j-1,2}$ is the same in both cases. This means that the two zippered rectangle representations also have the same data τ. It follows that the data h is also the same, since it is determined by π and τ. Thus, the zippered rectangle representations associated to the two cross-sections coincide. □

We say that two horizontal separatrices are *coupled* if they contain matching cross-sections. This is a symmetric relation. Extend it to an equivalence relation by forcing transitivity. We still say two horizontal cross-sections are coupled if they belong to the same equivalence class for this extended relation. In view of Proposition 3.37 and Lemma 3.40, in order to prove Proposition 3.39 we only have to show that there exists a unique equivalence class, that is, all separatrices are coupled.

Let us fix some (arbitrary) adapted cross-section σ and consider the corresponding zippered rectangle representation of the translation surface. For each $j = 1, \ldots, d$, let z_j^{l+} and z_j^{r+} be the singularities whose separatrices hit the cross-section at the distinguished points s_{j-1}^+ and s_j^+, respectively. Then let

$$[z_j^{l+}, w_j^{r+}] \quad \text{and} \quad [w_j^{l+}, z_j^{r+}]$$

be the horizontal segments going across the rectangle associated to the interval $[s_{j-1}^+, s_j^+]$: both segments project to this interval under the vertical flow. See Figure 3.23. These two segments are adapted cross-sections, and they match. Consequently, the separatrices σ_j^{l+} and σ_j^{r+} that contain them are coupled.

In a dual way, corresponding to the lower half of Figure 3.21, we consider singularities z_i^{l-} and z_i^{r-} associated to the points s_{i-1}^- and s_i^-, as well as adapted cross-sections $[z_i^{l-}, w_i^{r-}]$ and $[w_i^{l-}, z_i^{r-}]$ contained in separatrices σ_i^{l-} and σ_i^{r-}, and these cross-sections match. Notice that

$$z_j^{*+} = z_{p(j)}^{*-}, \quad \sigma_j^{*+} = \sigma_{p(j)}^{*-}, \quad w_j^{*+} = w_{p(j)}^{*-}, \quad \text{for } * \in \{l, r\} \text{ and } j = 1, \ldots, d.$$

In particular, σ_i^{l-} and σ_i^{r-} are always coupled. Recall that the return map $f : \sigma \to \sigma$ under the vertical flow sends each I_j^+ onto $I_{p(j)}^-$.

Lemma 3.41. *The separatrices σ_j^{l+}, σ_j^{r+} are coupled to σ_k^{l+}, σ_k^{r+}, for every k such that $I_{p(j)}^-$ intersects I_k^+.*

Proof. First, let us suppose neither interval contains the other. The subcase

$$I_{p(j)}^- \cup I_k^+ = [s_{k-1}^+, s_{p(j)}^-]$$

is described in Figure 3.24. Extend the vertical separatrix at $z_{p(j)}^{l-}$ upwards until it meets $[w_k^{l+}, z_k^{r+}]$ at some point w^+, and extend the vertical separatrix at z_k^{r+} downwards until it meets $[z_{p(j)}^{l-}, w_{p(j)}^{r-}]$ at some point w^-. The adapted cross-sections $[w^+, z_k^{r+}]$ and $[z_{p(j)}^{l-}, w^-]$ thus obtained match, and so the separatrices $\sigma_{p(j)}^{l-} = \sigma_j^{l+}$ and σ_k^{r+} are coupled. The subcase

$$I_{p(j)}^- \cup I_k^+ = [s_{p(j)-1}^-, s_k^+]$$

is analogous: we get that the separatrices $\sigma_{p(j)}^{r-} = \sigma_j^{r+}$ and σ_k^{l+} are coupled. This proves the lemma when neither of the two intervals contains the other.

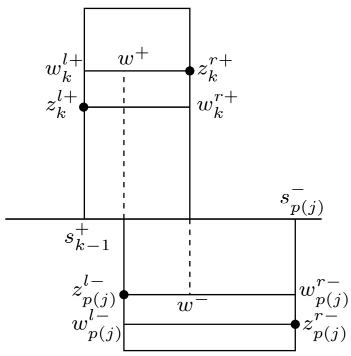


Figure 3.24:

Next, suppose $I_{p(j)}^-$ contains I_k^+. Consider first the subcase when $b_{k-1}^+ < b_k^+$, that is, z_k^{l+} is lower than z_k^{r+}. See Figure 3.25. If there are no points

$$s_i^+ \in \sigma^+ \cap (s_k^+, s_{p(j)}^-) \tag{3.15}$$

such that the corresponding singularity is lower than z_k^{l+} (the grey region in Figure 3.25 contains no singularities) then the vertical trajectories of all points in $[s_{k-1}^+, s_{p(j)}^-]$ extend for, at least, the length b_{k-1}^+ of the vertical segment $[s_{k-1}^+, z_k^{l+}]$. In particular, we may extend the vertical separatrix at $z_{p(j)}^{r-}$ upwards, and the horizontal separatrix σ_k^{l+} rightwards, until they meet at some point w_0. We may also extend the vertical separatrix at z_l^{r+} downwards until it meets $[w_{p(j)}^{l-}, z_{p(j)}^{r-}]$ at some point w^-. The adapted cross-sections $[z_k^{l+}, w_0]$ and $[w^-, z_{p(j)}^{r-}]$ thus obtained match, and so the conclusion follows in this case.

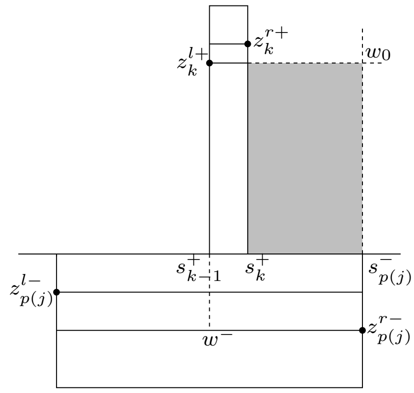


Figure 3.25:

Now suppose that there does exist s_i as in (3.15) such that the corresponding singularity is lower than z_k^{l+}. Take the smallest i with that property. Then the vertical trajectories of all points in $[s_{k-1}^+, s_i^+]$ extend for, at least, the length b_i^+ of the vertical segment $[s_i^+, z_i^{r+}]$ (the grey region in Figure 3.26 contains no singularities). Extend the separatrix σ_k^{l+} rightwards until it cross the rectangle over $[s_{i-1}^+, s_i^+]$: let $w_k \in \sigma_k^{l+}$ be such that the segment $[z_k^{l+}, w_k]$ projects to $[s_{k-1}^+, s_i^+]$ under the vertical flow. Extend σ_i^{r+} leftwards until it crosses the rectangle over $[s_{k-1}^+, s_k^+]$: let $w_i \in \sigma_i^{r+}$ be such that the segment $[w_i, z_i^{r+}]$ projects to $[s_{k-1}^+, s_i^+]$ under the vertical flow. These adapted cross-sections $[z_k^{l+}, w_k]$ and $[w_i, z_i^{r+}]$ match, and so the separatrices σ_k^{l+} and σ_i^{r+} are coupled. This means that, for the purpose of proving the lemma, one may replace σ_k^{l+}, σ_k^{r+} by σ_i^{l+}, σ_i^{r+}. Observe that, by construction,

$$\min\{b_{i-1}^+, b_i^+\} = b_i^+ < b_{k-1}^+ = \min\{b_{k-1}^+, b_k^+\}. \tag{3.16}$$

When $b_{k-1}^+ > b_k^+$, the same arguments show that σ_k^{r+} is coupled to $\sigma_{p(j)}^{l-}$ (similarly to Figure 3.25), or there exists a singularity z_i^{l+} to the *left* of z_k^{l+} such that the separatrices σ_k^{r+} and σ_i^{l+} are coupled (similarly to Figure 3.26). In the latter case

$$\min\{b_{i-1}^+, b_i^+\} = b_{i-1}^+ < b_k^+ = \min\{b_{k-1}^+, b_k^+\} \tag{3.17}$$

and one may replace σ_k^{l+}, σ_k^{r+} by σ_i^{l+}, σ_i^{r+}. The relations (3.16) and (3.17) ensure that this replacement procedure never leads to a loop and, thus, must eventually comes to a stop. So, we have proven the lemma whenever $I_{p(j)}^-$ contains I_k^+.

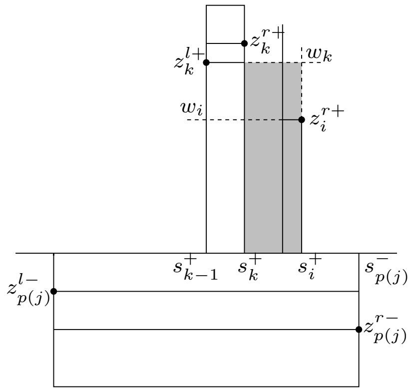


Figure 3.26:

The case when $I_{p(j)}^-$ is contained in I_k^+ is analogous, just carrying the previous construction upwards instead of downwards. The proof of Lemma 3.41 is complete. □

Under our assumptions the first return map $f : \sigma \to \sigma$ is minimal. In particular, the orbit of any point visits all intervals I_j^+. Thus, we may use Lemma 3.41 successively to prove that *all* horizontal separatrices are coupled. This proves Proposition 3.39. □

Definition 3.42. The *extended Rauzy class* of an irreducible permutation pair π is the orbit of π under the two (top=type 0 and bottom=type 1) Rauzy operations together with the involution ad.

It follows from Proposition 3.39 that for all representations (π, λ, τ) of a translation surface, corresponding to all choices of an adapted cross-section, the permutation pair π belongs to the same extended Rauzy class. Thus, we proved

Corollary 3.43. *There is a canonical map assigning to each translation surface without saddle-connections* (M, α) *an extended Rauzy class* $\mathfrak{R}(M, \alpha)$.

Clearly, extended Rauzy classes are unions of entire Rauzy classes. For small alphabets the two notions actually coincide: the first examples of a Rauzy class strictly contained in the corresponding extended Rauzy class occur for $d = 6$. We shall return to these topics in Chapter 6.

3.8 Realizable measured foliations

We call a measured foliation $\mathcal{F}$ on an orientable surface M *realizable* if it can be represented as the horizontal/vertical foliation of some Abelian differential, relative to some conformal structure on M. In other words, $\mathcal{F}$ is realizable if there exists some Abelian differential α, for some complex structure on the surface, such that $\mathcal{F} = \mathcal{F}_{\Im(\alpha)}$.

It follows immediately from Lemma 3.11 that every measurable foliation on the torus T^2 is realizable. On the other hand, we shall see in Example 3.48 that higher genus surfaces do admit non-realizable foliations. Nevertheless, we are going to see that typical measured foliations on any orientable surface are realizable. More precisely, one has the following characterization of realizability, proved independently by Calabi [9] and Hubbard, Masur [22].

Theorem 3.44. *A measured foliation $\mathcal{F}$ is realizable if and only if for any pair of regular points x and y in M there exists an increasing cross-section to $\mathcal{F}$ going from x to y.*

Proof. For proving the 'only if' half of the statement, we use the following fact:

Lemma 3.45. *Given any regular point $x \in M$, the set $A(x)$ of regular points y such that there exists an increasing cross-section from x to y is open and its boundary consists of closed leaves, saddle-connections and their limit singularities. Moreover, locally, $A(x)$ contains all points to the positive side of every leaf on the boundary.*

Proof. Considering an adapted local coordinate at any point $y \in A(x)$, as described on the left hand side of Figure 3.27, one immediately sees that y is interior to V. Thus, $A(x)$ is open. Now let p be any regular point on the

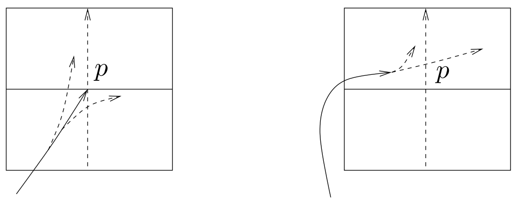


Figure 3.27:

boundary of $A(x)$. Then $A(x)$ contains every point to the positive side of the leaf of p inside an adapted neighborhood; see the right hand of Figure 3.27. In particular, the boundary contains a neighborhood of p inside the corresponding leaf. This implies that the boundary of $A(x)$ is a saturated set. Now, suppose the leaf of p was neither closed nor a saddle-connection. Then we could apply the recurrence lemma (Lemma 3.22) to conclude that some ray of p intersects the positive side of any cross-section to the foliation at p. Then, in view of the previous observations, the leaf would contain interior points of $A(x)$. This is a contradiction, because the whole leaf is contain on the boundary. So, the leaf of p must be either closed or a saddle-connection. □

Now assume $\mathcal{F}$ is realizable, that is, there exists some Abelian differential α whose horizontal foliation $\mathcal{F}_\alpha^h$ coincides with $\mathcal{F}$. Let $\mathcal{F}_\alpha^v$ be the vertical foliation of α. Fix any regular point x and let $A(x)$ be as in Lemma 3.45. We want to prove that $A(x)$ contains all regular points. For this, it suffices to prove that the boundary contains singularities only. Fix any area form ω on M, for instance $\omega = \alpha\bar{\alpha}$. As in Example 3.1, we may find an area preserving vector field W tangent to the leaves of $\mathcal{F}_\alpha^h$ in the positive transverse direction to $\mathcal{F}_\alpha^v$. Assuming the boundary of $A(x)$ contains regular points. Then W points to the interior of $A(x)$ at every regular boundary point. Hence, $A(x)$ is strictly invariant under the corresponding flow W^t: $W^t(A(x)) \subset A(x)$ and $A(x) \setminus W^t(A(x))$ has non-empty interior for all $t > 0$. This implies that W^t decreases the area of $A(x)$, which contradicts the fact that W preserves area. This contradiction proves that the boundary of $A(x)$ contains no regular points, and so $A(x)$ must coincide with the set of all regular points. In other words, every regular point is reachable from x through an increasing cross-section, as claimed in the 'only if' part of the statement of Theorem 3.44.

Now we prove the 'if' half of the theorem. Consider the sectors determined (locally) by the separatrices of all the singularities. We call a sector *outgoing* if cross-sections from the singularity to a point in the sector are increasing, and we call it *incoming* otherwise. Pair each outgoing sector to an incoming one. By hypothesis, for each such pair we can find an increasing cross-section from any point in the outgoing sector to any point in the incoming sector. Then this may be extended to an increasing cross-section leaving one singularity through the outgoing sector and arriving at the other singularity through the incoming sector; the fact that we are considering oriented (increasing) curves is important at this point. Let G_0 be the graph formed by all these cross-sections together with the singularities. It is clear that we may choose the edges of G_0 transverse to each other. Moreover, breaking and then reconnecting cross-sections as illustrated in Figure 3.28, we may remove all intersection points: we obtain a new graph G whose edges are pairwise disjoint except for their endpoints.

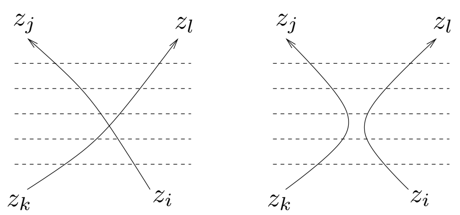


Figure 3.28:

Then, cutting the surface M along the graph G we obtain a finite family of (connected) surfaces M_j bounded by piecewise smooth simple curves. Note that the boundary curves have corners, corresponding to the singularities of the foliation. The restriction of $\mathcal{F}$ to each one of these surfaces M_j has no singularities and is transverse to the boundary, including at the corners. See Figure 3.29.

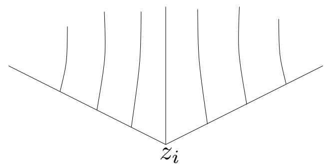


Figure 3.29:

Let M_j be some of these surfaces and γ be any of its boundary components. Consider the rays leaving from γ inside M_j. Observe that every such ray returns to the boundary of M_j: this is clear for any singular ray, because there are no saddles in the interior of M; for a regular ray it is a consequence of the fact that the ray accumulates its entire leaf (recall Lemma 3.22). Moreover, if some ray leaving from γ hits a boundary component γ' then so do all rays leaving from nearby points in γ. Then, there exists a holonomy map $\pi : \gamma \to \gamma'$ defined on the whole boundary component. Let $\sigma(t)$, $t \in S^1$ be a parametrization of γ and let $F(s,t)$, $s \in [0,1]$ be a parametrization of the ray segment from $\sigma(t)$ to $\pi(\sigma(t))$, depending continuously on t. Then

$$\phi_j : S^1 \times [0,1] \to M_j, \quad \phi_j(s,t) = (F(s,t), \sigma(t))$$

is an embedding of the cylinder $S^1 \times [0,1]$. Since the image is both compact and open in M_j, we get that ϕ_j is a homeomorphism onto M_j. Moreover, ϕ may be chosen smooth and with bounded norm $\sup\{\|D\phi\|, \|D\phi^{-1}\|\}$ away from the rays leaving from or arriving at singularities on the boundary of M_j. Then we may use ϕ to push the canonical complex structure and the canonical Abelian differential $ds+i\,dt$ from the cylinder forward to M_j. Observe that the restriction of $\mathcal{F}$ to each M_j is the measured foliation defined by the push-forward of dt, which is the horizontal foliation of the push-forward of $ds + i\,dt$. The proof of Theorem 3.44 is complete. □

Remark 3.46. An alternative characterization is proposed in [65, Theorem 1]: a measured foliation on an orientable surface is realizable if and only no cycle obtained as the union of closed paths formed by increasing saddle-connections is homologous to zero.

The following statement first appeared in Katok [24]. At this point we can easily deduce it from the previous arguments.

Corollary 3.47. *If a measured foliation $\mathcal{F}_\beta$ has no saddle-connections then it is realizable.*

Proof. All measured foliations on the torus are realizable, as we observed near the beginning of this section. So, we may suppose $M \neq T^2$. Then, by Corollary 3.26, the foliation $\mathcal{F}_\beta$ is minimal. In particular, it has no closed orbits.

Since we assume it has no saddle-connections either, Lemma 3.45 implies that the boundary of any attainability set $A(x)$ contains only singularities. This implies that condition (c) in Theorem 3.44 is satisfied, and so the foliation is realizable. □

Let us also present an example of non-realizable measured foliation:

Example 3.48. (Hubbard, Masur) Consider two cylinders foliated by circles, with singularities on the boundary as described in Figure 3.30; the boundary

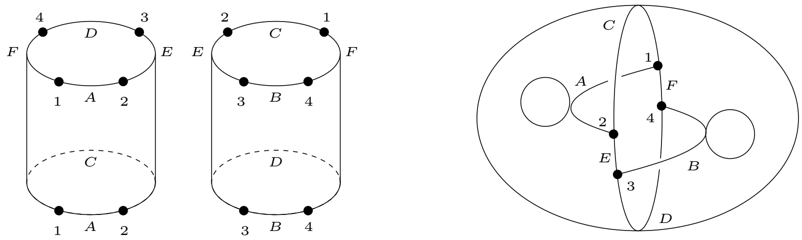


Figure 3.30:

segments determined by the singularities are saddle-connections for the foliation on each of the cylinders. Let us glue the two cylinders together by identifying boundary segments labeled with the same letter. In this way we obtain a genus 2 surface M, endowed with a measured foliation by closed curves, with 4 singularities and 6 saddle-connections. This measured foliation $\mathcal{F}$ is non-transversely orientable [5]: the singularities have an odd number $s_i = 3$ of separatrices. Hubbard, Masur [22] observe that it can not be realized by a quadratic differential: If there was such a quadratic differential then it would define a flat metric on each of the two cylinders, relative to which all closed leaves would have the same length. Then, by continuity, the boundary components would also have the same length. In terms of the lengths λ_α of the saddle-connections, this would mean that

$$\lambda_A + \lambda_C = \lambda_A + \lambda_D + \lambda_E + \lambda_F \quad \text{and} \quad \lambda_B + \lambda_D = \lambda_B + \lambda_C + \lambda_E + \lambda_F.$$

However, this system of equations has no positive solutions.

One can also use the criterium in Theorem 3.44 to show that the double cover of $\mathcal{F}$ branched over the singularities is not realizable (by Abelian differentials). The branched double cover may be represented by considering two copies of each of the two cylinders and identifying boundary segments with the same label. See Figure 3.31. For the sake of clearness, in the figure we represent cylinders as rectangles: left and right sides of each rectangle are also identified. The lift $\hat{\mathcal{F}}$ of $\mathcal{F}$ is the foliation by horizontal circles, and it is clear that it is transversely orientable: consider the upward orientation on each of the cylinders. It is easy to check that any increasing cross-section starting in any of the two last

[5] Recall the comments on non-transversely orientable measured foliations in Section 3.1.

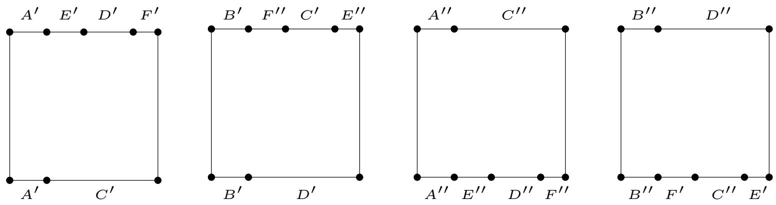


Figure 3.31:

cylinders is contained in the union of those two cylinders. Thus, the condition in Theorem 3.44 is not satisfied, and so $\hat{\mathcal{F}}$ is not realizable.

Remark 3.49. The original statements of Calabi [9] and Katok [24] are about characterizing harmonic 1-forms on M. $\mathcal{F}_\beta$ is realizable if and only if the 1-form β is harmonic relative to some Riemannian metric on the surface.

Notes

Theorem 3.13 first appeared in Maier [38], in the context of vector fields on surfaces, and has been rediscovered in other situations. Our presentation is close to Strebel [51], which contains a version for trajectories of quadratic differentials. Much more information can be found in Fathi, Laudenbach, Poenaru [13]. Sections 3.5 and 3.6 are based on Veech [59]. Versions of Theorem 3.44 were proved, independently, by Calabi [9] and by Hubbard-Masur [22]. Our presentation is closer to the latter. Corollary 3.47 is due to Katok [24].

Chapter 4

Invariant Measures

In the first two chapters we introduced certain dynamical systems acting on the spaces of interval exchange maps and translation surfaces: induction and renormalization operators, Teichmüller flows. Here we are going to construct natural volume measures that are invariant under these dynamical systems.

The starting point is the observation in Section 2.10 that each pre-stratum $\mathcal{S}(C)$ comes with a natural volume measure $\hat{m}$, inherited from $\mathcal{H}(C)$, and this measure is invariant under the Teichmüller flow. A crucial fact, established in Theorem 4.13 below, is that this volume is finite if one normalizes the area. Finiteness allows for Poicaré recurrence arguments that, eventually, prove that the Teichmüller flow is ergodic for $\hat{m}$, restricted to each hypersurface of constant area (Corollary 4.28).

The invertible Rauzy-Veech renormalization transformation $\mathcal{R}$ may be seen as a Poincaré return map of the Teichmüller flow to a cross-section $\{|\lambda| = 1\}$. Restricting the natural volume $\hat{m}$ to the cross-section one obtains an absolutely continuous invariant measure m for $\mathcal{R}$. On the way to proving ergodicity of the Teichmuüller flow, one shows that $\mathcal{R}$ is ergodic for this measure, restricted to each hypersurface of constant area (Corollary 4.28).

Projecting m down to the space of pairs (π, λ) one obtains the following result of Masur [41] and Veech [54], which is a crucial step in their proof of the Keane conjecture (Conjecture 1.18), as we are going to see in Chapter 5.

Theorem 4.1. *For each Rauzy class C, the Rauzy-Veech renormalization transformation $R : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$ admits an invariant measure ν which is absolutely continuous with respect to $d\pi \times$ Leb. This measure ν is unique, up to product by a scalar, and ergodic. Moreover, its density with respect to Lebesgue measure is given by a homogeneous rational function of degree $-d$ and bounded away from zero.*

In general, the Masur-Veech measure ν has infinite mass, which means that many tools from ergodic theory can not be applied immediately to it. The accelerated renormalization algorithm proposed by Zorich [63] and introduced in Section 1.8 circumvents this difficulty:

Theorem 4.2. *For each Rauzy class C, the Zorich renormalization transformation $Z : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$ admits an invariant probability measure μ which is absolutely continuous with respect to $d\pi \times \mathrm{Leb}$. This probability μ is unique and ergodic. Moreover, its density with respect to Lebesgue measure is given by a homogeneous rational function of degree $-d$ and bounded away from zero.*

The proofs of these results occupy most of the present chapter.

4.1 Volume measure

For translation surfaces. Let C be a Rauzy class. The space $\hat{\mathcal{H}} = \hat{\mathcal{H}}(C)$ has a natural volume measure $\hat{m} = d\pi\, d\lambda\, d\tau$, where $d\pi$ is the counting measure on C, and $d\lambda$ and $d\tau$ are the restrictions to $\mathbb{R}_+^{\mathcal{A}}$ and T_π^+, respectively, of the Lebesgue measure on $\mathbb{R}^{\mathcal{A}}$. Clearly, $\hat{m}$ is invariant under the Teichmüller flow

$$\mathcal{T}^t : (\pi, \lambda, \tau) \mapsto (\pi, e^t \lambda, e^{-t} \tau).$$

Let us consider the coordinate change $\mathcal{H} \times \mathbb{R} \to \hat{\mathcal{H}}$, $(\pi, \lambda, \tau, s) \mapsto (\pi, e^s \lambda, e^s \tau)$ introduced in (2.64). Observe that

$$d\lambda = e^{s(d-1)} d_1\lambda\, e^s ds = e^{sd} d_1\lambda\, ds,$$

where $d_1\lambda$ denotes the Lebesgue measure induced on $\Lambda_{\mathcal{A}} = \{\lambda \in \mathbb{R}_+^{\mathcal{A}} : |\lambda| = 1\}$ by the Riemannian metric of $\mathbb{R}^{\mathcal{A}}$. See Figure 4.1. Thus, $\hat{m} = e^{sd}\, d\pi\, d_1\lambda\, d\tau\, ds$. We denote $m = d\pi\, d_1\lambda\, d\tau$, and view it as a measure on $\mathcal{H} = \mathcal{H}(C)$.

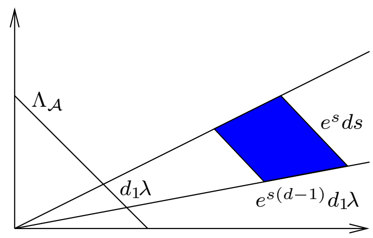


Figure 4.1:

Lemma 4.3. *The measure $\hat{m}$ is invariant under the Rauzy-Veech maps $\hat{\mathcal{R}}$ and $\mathcal{R}$. Moreover, m is invariant under the restriction $\mathcal{R} : \mathcal{H} \to \mathcal{H}$.*

Proof. Recall that $\hat{\mathcal{R}}(\pi, \lambda, \tau) = (\pi', \lambda', \tau')$ where $\lambda = \Theta^*_{\pi,\lambda}(\lambda')$ and $\tau = \Theta^*_{\pi,\lambda}(\tau')$. Since $\hat{\mathcal{R}}$ is injective and

$$\det \Theta^*_{\pi,\lambda} = \det \Theta_{\pi,\lambda} = 1,$$

it follows that $\hat{\mathcal{R}}$ preserves $\hat{m} = d\pi\, d\lambda\, d\tau$, as claimed. Now, in view of the definition (2.66), to prove that $\hat{m}$ is preserved by $\mathcal{R}$ we only have to show that it is preserved by

$$(\pi, \lambda, \tau) \mapsto \mathcal{T}^{t_R(\pi,\lambda)}(\pi, \lambda, \tau).$$

Using coordinates (π, λ, τ, s), this corresponds to showing that the measure $e^{sd}\, d_1\lambda\, d\tau\, ds$ is invariant under the map

$$\Phi : (\lambda, \tau, s) \mapsto (\lambda, e^{-t_R(\pi,\lambda)}\tau, s + t_R(\pi, \lambda)).$$

The Jacobian matrix of Φ has the form

$$D\Phi = \begin{pmatrix} I_{d-1} & 0 & 0 \\ * & e^{-t_R} I_d & 0 \\ * & 0 & 1 \end{pmatrix}$$

(I_j denotes the j-dimensional identity matrix) and so its determinant is $e^{-t_R d}$. Hence,

$$e^{(s+t_R)d}\, d_1\lambda\, d\tau\, ds\, |\det D\Phi| = e^{sd}\, d_1\lambda\, d\tau\, ds,$$

which means that Φ does preserve $e^{sd}\, d_1\lambda\, d\tau\, ds$.

Finally, $\mathcal{R}$ preserves every $\mathcal{H}_c = \{(\pi, \lambda, \tau, s) \in \hat{\mathcal{H}} : e^s = c\}$ and the measure $\hat{m} = e^{sd}\, d\pi\, d_1\lambda\, d\tau\, ds$ disintegrates to conditional measures $c^d d\pi\, d_1\lambda\, d\tau$ on each $\mathcal{H}_c$. So, the previous conclusion that $\mathcal{R}$ preserves $\hat{m}$ means that it preserves these conditional measures for almost every c. From the definition (2.66) we get that $\lambda \mapsto c\lambda$ conjugates the restrictions of $\mathcal{R}$ to $\mathcal{H}$ and to $\mathcal{H}_c$, respectively. Consequently, $\mathcal{R} \mid \mathcal{H}_c$ preserves $c^d d\pi\, d_1\lambda\, d\tau$ if and only if the renormalization map $\mathcal{R} : \mathcal{H} \to \mathcal{H}$ preserves $m = d\pi\, d_1\lambda\, d\tau$. It follows that $\mathcal{R} : \mathcal{H} \to \mathcal{H}$ does preserve m, as claimed. □

Given any $c > 0$, we denote by $\hat{m}_c$ the restriction of $\hat{m}$ to the region $\{(\pi, \lambda, \tau) \in \hat{\mathcal{H}} : \text{area}\,(\lambda, \tau) \le c\}$. Since this region is invariant under $\hat{\mathcal{R}}$, $\mathcal{R}$, and $\mathcal{T}$, so are all these measures $\hat{m}_c$. Similarly, we denote by m_c the restriction of m to the region $\{(\pi, \lambda, \tau) \in \mathcal{H} : \text{area}\,(\lambda, \tau) \le c\}$. Then every m_c is invariant under the restriction of the Rauzy-Veech renormalization $\mathcal{R}$.

Recall that the pre-stratum $\hat{\mathcal{S}} = \hat{\mathcal{S}}(C)$ is the quotient of the space $\hat{\mathcal{H}}$ by the equivalence relation generated by

$$\mathcal{T}^{t_R(\pi,\lambda)}(\pi, \lambda, \tau) = (\pi, e^{t_R(\pi,\lambda)}\lambda, e^{-t_R(\pi,\lambda)}\tau) \sim \mathcal{R}(\pi, \lambda, \tau).$$

Since the Teichmüller flow commutes with $\mathcal{R}$, it projects down to a flow on $\hat{\mathcal{S}}$, that we also denote by $\mathcal{T}$. The (injective) image $\mathcal{S} \subset \hat{\mathcal{S}}$ of $\mathcal{H}$ under the quotient map is a global cross-section to this flow. Moreover, the restriction of $\hat{m}$ to the fundamental domain

$$\{(\pi, \lambda, \tau) \in \hat{\mathcal{H}} : 0 \le \log|\lambda| \le t_R(\pi, \lambda)\}$$

defines a volume measure on $\hat{\mathcal{S}}$, that we also denote by $\hat{m}$. It is easy to check that $\hat{m}$ is invariant under the Teichmüller flow $\mathcal{T}^t$ on the pre-stratum $\hat{\mathcal{S}}$. Finally, since area is invariant under the equivalence relation above, it is well defined in the pre-stratum. Sometimes, we denote by hm_c the restriction of $\hat{m}$ to the subset of elements of the pre-stratum with area $(\pi, \lambda, \tau) \le c$. All these measures are invariant under the Teichmüller flow on $\hat{\mathcal{S}}$. *

For interval exchange maps. Let $P : \hat{\mathcal{H}} \to C \times \mathbb{R}_+^{\mathcal{A}}$ be the canonical projection $P(\pi, \lambda, \tau) = (\pi, \lambda)$. Then let $\hat{\nu} = P_*(\hat{m}_1)$ be the measure obtained by projecting $\hat{m}_1$ down to $C \times \mathbb{R}_+^{\mathcal{A}}$:

$$\hat{\nu}(E) = \hat{m}_1\big(P^{-1}(E)\big) = \hat{m}\big(\{(\pi, \lambda, \tau) : (\pi, \lambda) \in E \text{ and } \operatorname{area}(\lambda, \tau) \le 1\}\big).$$

Let $\hat{R}$ and R be the Rauzy-Veech transformations at the level of interval exchange maps, introduced in Sections 1.2 and 1.7. Likewise, let T^t be the projected Teichmüller flow $T^t(\pi, \lambda) = (\pi, e^t\lambda)$. Since

$$P \circ \mathcal{T}^t = T^t \circ P \quad \text{and} \quad P \circ \hat{\mathcal{R}} = \hat{R} \circ P \quad \text{and} \quad P \circ \mathcal{R} = R \circ P,$$

the measure $\hat{\nu}$ is invariant under $\hat{R}$, R, and T. Moreover, let $\nu = P_*(m_1)$ be the measure obtained by projecting m_1 down to $C \times \Lambda_{\mathcal{A}}$:

$$\nu(E) = m_1\big(P^{-1}(E)\big) = m\big(\{(\pi, \lambda, \tau) : (\pi, \lambda) \in E \text{ and } \operatorname{area}(\lambda, \tau) \le 1\}\big).$$

Then ν is invariant under Rauzy-Veech renormalization $R : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$.

Let $\hat{S}$ be the quotient of $C \times \mathbb{R}_+^{\mathcal{A}}$ by the equivalence relation generated on $C \times \mathbb{R}_+^{\mathcal{A}}$ by $T^{t_R(\pi,\lambda)}(\pi, \lambda) \sim R(\pi, \lambda)$. We represent by S the (injective) image of $C \times \Lambda_{\mathcal{A}}$ under this quotient map. The flow T^t induces a semi-flow $T^t : \hat{S} \to \hat{S}$, $t > 0$ which admits S as a global cross-section and whose first return map to this cross-section is the Rauzy-Veech renormalization $R : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$.

The projection $P : \hat{\mathcal{H}} \to C \times \mathbb{R}_+^{\mathcal{A}}$ induces a projection $P : \hat{\mathcal{S}} \to \hat{S}$ such that $P \circ \mathcal{T}^t = T^t \circ P$. The absolutely continuous measure $\hat{\nu}$ restricted to the fundamental domain

$$\{(\pi, \lambda) \in C \times \mathbb{R}_+^{\mathcal{A}} : 0 \le \log|\lambda| \le t_R(\pi, \lambda)\}$$

induces an absolutely continuous measure on $\hat{S}$, that we also denote as $\hat{\nu}$. It may also be obtained as $\hat{\nu} = P_*(\hat{m})$ where $\hat{m}$ denotes the volume measure on $\hat{\mathcal{S}}$ introduced previously. It follows from $P \circ \mathcal{T}^t = T^t \circ P$ that $\hat{\nu}$ is invariant under the semi-flow T^t.

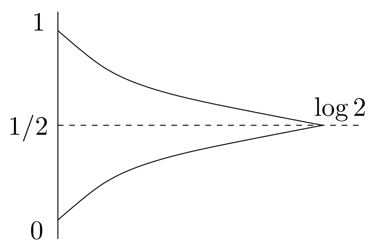


Figure 4.2:

Example 4.4. For $d = 2$, the domain $\mathbb{R}_+^{\mathcal{A}}$ may be identified with $\mathbb{R} \times (0, 1)$, through

$$(\lambda_A, \lambda_B) \mapsto (\log|\lambda|, \lambda_A).$$

Note that the simplex $\Lambda_{\mathcal{A}}$ is identified with the open interval $(0,1)$, through $(\lambda_A, \lambda_B) \mapsto x = \lambda_A$. Then $d_1\lambda$ corresponds to the measure dx, and the Rauzy renormalization time is

$$t_R(x) = \begin{cases} -\log(1-x) & \text{if } x < 1/2 \\ -\log x & \text{if } x > 1/2. \end{cases} \tag{4.1}$$

$\hat{S}$ is the quotient of the domain $\{(s,x) : 0 \le s \le t_R(x)\}$ by an identification of the boundary segment on the left with each of the two boundary curves on the right. See Figure 4.2. The semi-flow T^t is horizontal, pointing to the right, and its return map to $\{0\} \times (0,1)$ is the renormalization map R as presented in Example 1.26. The pre-stratum $\hat{\mathcal{S}} = \hat{S} \times T_\pi^+$, where T_π^+ is the set of pairs (τ_A, τ_B) such that $\tau_A > 0 > \tau_B$. The measure m_1 is the restriction of $m = dx\, d\tau_A\, d\tau_B$ to the domain

$$\mathcal{H}_1 = \{(x, \tau_A, \tau_B) : \operatorname{area}(x,\tau) = x\tau_A - (1-x)\tau_B \le 1\}.$$

Observe that the total mass is infinite:

$$m_1(\mathcal{H}_1) = \int_0^1 \frac{dx}{2x(1-x)} = \infty. \tag{4.2}$$

Invariant densities. Since P is a submersion, the measure $\hat{\nu}$ is absolutely continuous with respect to $d\lambda$ (or, more precisely, $d\pi \times d\lambda$), with density

$$\frac{d\hat{\nu}}{d\lambda}(\pi,\lambda) = \operatorname{vol}(\{\tau \in T_\pi^+ : \operatorname{area}(\lambda,\tau) \le 1\}) \quad \text{for} \quad (\pi,\lambda) \in C \times \mathbb{R}_+^{\mathcal{A}},$$

where $\operatorname{vol}(\cdot)$ represents d-dimensional volume in T_π^+. Analogously, ν is absolutely continuous with respect to $d_1\lambda$ (or, more precisely, $d\pi \times d_1\lambda$), with density

$$\frac{d\nu}{d_1\lambda}(\pi,\lambda) = \operatorname{vol}(\{\tau \in T_\pi^+ : \operatorname{area}(\lambda,\tau) \le 1\}) \quad \text{for } (\pi,\lambda) \in C \times \Lambda_{\mathcal{A}}.$$

The right hand side in these expressions may be calculated as follows (an explicit example will be worked out in Section 4.2).

The polyhedral cone T_π^+ may be written, up to a codimension 1 subset, as a finite union of *simplicial cones* $T^1, \ldots, T^k$, that is, subsets of $\mathbb{R}^{\mathcal{A}}$ of the form

$$T^i = \{\sum_{\beta \in \mathcal{A}} c_\beta \tau^{i,\beta} : c_\beta > 0 \text{ for each } \beta \in \mathcal{A}\},$$

for some basis $(\tau^{i,\beta})_{\beta \in \mathcal{A}}$ of $\mathbb{R}^{\mathcal{A}}$. We always assume that this basis has been chosen with volume 1, that is, it is the image of some orthonormal basis by a linear operator with determinant 1. The volume of each domain

$$\{\tau \in T^i : \operatorname{area}(\lambda,\tau) \le 1\} = \{\tau \in T^i : -\lambda \cdot \Omega_\pi(\tau) \le 1\}$$

may be calculated using the following elementary fact:

Lemma 4.5. *Let $T \subset \mathbb{R}^{\mathcal{A}}$ be a simplicial cone, $(\tau^\beta)_{\beta\in\mathcal{A}}$ be a volume 1 basis of generators of T, and $L : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ be a linear operator. Then, for any λ satisfying $\lambda \cdot L(\tau^\beta) > 0$ for all $\beta \in \mathcal{A}$, we have*

$$\operatorname{vol}(\{\tau \in T : \lambda \cdot L(\tau) \le 1\}) = \frac{1}{d!} \prod_{\beta\in\mathcal{A}} \frac{1}{\lambda \cdot L(\tau^\beta)}.$$

Proof. Let $M : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ be a linear operator mapping the canonical basis $(e^\beta)_{\beta\in\mathcal{A}}$ of $\mathbb{R}^{\mathcal{A}}$ to the basis $(\tau^\beta)_{\beta\in\mathcal{A}}$. Then let $\tilde{T} = M^{-1}(T)$ and $\tilde{L} = LM$. Then

$$\operatorname{vol}(\{\tau \in T : \lambda \cdot L(\tau) \le 1\}) = \operatorname{vol}(\{v \in \tilde{T} : \lambda \cdot \tilde{L}(v) \le 1\}). \tag{4.3}$$

Since T is a simplicial cone, $\tilde{T}$ is the cone of vectors $v = \sum_{\beta\in\mathcal{A}} c_\beta e^\beta$ with entries $c_\beta > 0$ relative to the orthonormal basis. Then the set on the right hand side of (4.3) is the simplex with vertices at the origin and at each one of the points

$$\frac{e^\beta}{\lambda \cdot \tilde{L}(e^\beta)} = \frac{e^\beta}{\lambda \cdot L(\tau^\beta)}, \quad \beta \in \mathcal{A}.$$

Therefore,

$$\operatorname{vol}(\{v \in \tilde{T} : \lambda \cdot \tilde{L}(v) \le 1\}) = \prod_{\beta\in\mathcal{A}} \frac{1}{\lambda \cdot L(\tau^\beta)} \operatorname{vol}(\Sigma_{\mathcal{A}})$$

where $\Sigma_{\mathcal{A}}$ is the canonical d-dimensional simplex, with vertices at the origin and at each of the points e^β, $\beta \in \mathcal{A}$. The latter has volume $1/d!$, and so the proof is complete. □

Applying this lemma to each $T = T^i$ with $L = -\Omega_\pi$, we obtain

Proposition 4.6. *The density of $\hat{\nu}$ relative to Lebesgue measure is*

$$\frac{d\hat{\nu}}{d\lambda}(\pi, \lambda) = \operatorname{vol}(\{\tau \in T_\pi^+ : \operatorname{area}(\lambda, \tau) \le 1\}) = \frac{1}{d!} \sum_{i=1}^{k} \prod_{\beta\in\mathcal{A}} \frac{1}{\lambda \cdot h^{i,\beta}}$$

where $h^{i,\beta} = -\Omega_\pi(\tau^{i,\beta})$. Moreover, the same formula holds for $d\nu/d_1\lambda$. In particular, all these densities are homogeneous rational functions with degree $-d$ and bounded away from zero.

Example 4.7. Let $d = 2$ and $\pi = \begin{pmatrix} A & B \\ B & A \end{pmatrix}$. The conditions (2.10) defining T_π^+ reduce to $\tau_A > 0 > \tau_B$. The operator Ω_π is given by

$$\Omega_\pi(\tau_A, \tau_B) = (\tau_B, -\tau_A)$$

and $\operatorname{area}(\lambda, \tau) = \lambda_B \tau_A - \lambda_A \tau_B$. The operator Θ is given by

$$\Theta = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{if type } = 0 \quad \text{and} \quad \Theta = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{if type } = 1$$

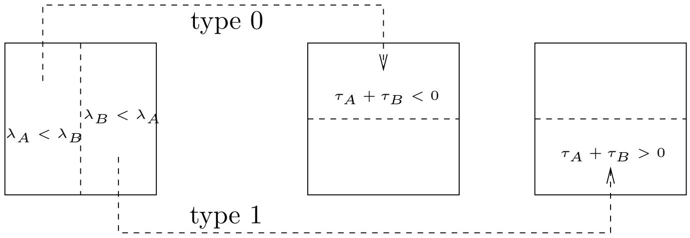


Figure 4.3:

Figure 4.3 illustrates the action of the Rauzy transformation $\hat{\mathcal{R}}$ on the space of translation surfaces.

The measure $\hat{m} = d\lambda d\tau$ on $\hat{\mathcal{H}} = \{(\lambda, \tau) : \lambda_A > 0,\ \lambda_B > 0,\ \tau_A > 0 > \tau_B\}$ projects down to a measure $\hat{\nu}$ on $\mathbb{R}_+^{\mathcal{A}}$ which is absolutely continuous with respect to Lebesgue measure $d\lambda$, with density

$$\begin{aligned}\frac{d\hat{\nu}}{d\lambda}(\lambda) &= \operatorname{vol}(\{\tau \in T_\pi^+ : \operatorname{area}(\lambda, \tau) \le 1\}) \\ &= \operatorname{vol}(\{\tau \in \mathbb{R}^{\mathcal{A}} : \tau_A > 0 > \tau_B \text{ and } \lambda_B \tau_A - \lambda_A \tau_B \le 1\}),\end{aligned}$$

that is,

$$\frac{d\hat{\nu}}{d\lambda}(\lambda) = \frac{1}{2\lambda_A \lambda_B}. \tag{4.4}$$

The same expression holds for $d\nu/d_1\lambda$, restricted to $\Lambda_{\mathcal{A}}$. Notice that the measure ν is infinite. Indeed, identifying $\Lambda_{\mathcal{A}}$ with $(0,1)$ and $d_1\lambda$ with dx, through $(\lambda_A, \lambda_B) \mapsto x = \lambda_A$,

$$\nu(\Lambda_{\mathcal{A}}) = \int_{\Lambda_{\mathcal{A}}} \frac{1}{2\lambda_A \lambda_B}\, d_1\lambda = \int_0^1 \frac{1}{2x(1-x)}\, dx = \infty;$$

compare (4.2). However, $\hat{\nu}$ is *finite* on $\hat{S}$. Indeed (recall Example 4.4)

$$\hat{\nu}(\hat{S}) = \int_0^1 \int_0^{t_R(x)} \frac{1}{2\, e^s x\, e^s (1-x)}\, e^{2s}\, dx\, ds = \int_0^1 t_R(x) \frac{1}{2x(1-x)}\, dx.$$

Using the expression (4.1), this becomes

$$\hat{\nu}(\hat{S}) = 2\int_0^{1/2} -\log(1-x)\frac{1}{2x(1-x)}\, dx \le 2\int_0^{1/2} -\log(1-x)\frac{1}{x}\, dx < \infty.$$

The corresponding property holds in general for all $d \ge 2$, as we shall see.

4.2 Hyperelliptic pairs

We are going to compute an explicit expression for the density in the case when

$$\pi_1 \circ \pi_0^{-1}(j) = d - j + 1 \quad \text{for } j = 1, \ldots, d. \tag{4.5}$$

Denote

$$b_\alpha^\varepsilon = \sum_{\pi_\varepsilon(\beta)\le\pi_\varepsilon(\alpha)} \tau_\beta \quad \text{for each } \alpha\in\mathcal{A} \text{ and } \varepsilon\in\{0,1\}. \tag{4.6}$$

Note that $\sum_{\alpha\in\mathcal{A}}\tau_\alpha = b_{\alpha(\varepsilon)}^\varepsilon$ for $\varepsilon = 0,1$. The cone T_π^+ is defined by

$$b_\alpha^0 > 0 \quad \text{for } \alpha\neq\alpha(0) \quad \text{and} \quad b_\alpha^1 < 0 \quad \text{for } \alpha\neq\alpha(1), \tag{4.7}$$

which is just a reformulation of (2.10). Let T_π^0 and T_π^1 be the subsets of T_π^+ defined by

$$\tau\in T_\pi^0 \Leftrightarrow \sum_{\alpha\in\mathcal{A}}\tau_\alpha > 0 \quad \text{and} \quad \tau\in T_\pi^1 \Leftrightarrow \sum_{\alpha\in\mathcal{A}}\tau_\alpha < 0. \tag{4.8}$$

Clearly, $T_\pi^+ = T_\pi^0\cup T_\pi^1$, up to a codimension 1 subset.

Given $\alpha\in\mathcal{A}$ and $\varepsilon\in\{0,1\}$, denote by α_ε^- the symbol to the left and by α_ε^+ the symbol to the right of α in line ε. That is,

$$\begin{aligned} \alpha_\varepsilon^- &= \pi_\varepsilon^{-1}(\pi_\varepsilon(\alpha)-1) \quad \text{if } \pi_\varepsilon(\alpha) > 1 \\ \alpha_\varepsilon^+ &= \pi_\varepsilon^{-1}(\pi_\varepsilon(\alpha)+1) \quad \text{if } \pi_\varepsilon(\alpha) < d. \end{aligned} \tag{4.9}$$

Lemma 4.8. *T_π^ε is a simplicial cone for every $\varepsilon\in\{0,1\}$.*

Proof. We treat the case $\varepsilon = 0$, the other one being entirely analogous. For notational simplicity, let $b_\alpha = b_\alpha^0$ for every $\alpha\in\mathcal{A}$. Note that, because of (4.5),

$$b_\alpha^0 + b_\alpha^1 = \sum_{\beta\in\mathcal{A}}\tau_\beta + \tau_\alpha.$$

Equivalently,

$$b_{\alpha_0^-}^0 + b_\alpha^1 = \sum_{\beta\in\mathcal{A}}\tau_\beta = b_\alpha^0 + b_{\alpha_1^-}^1$$

(the first equality is for $\alpha\neq\alpha(1)$, the second one for $\alpha\neq\alpha(0)$). In particular,

$$b_\alpha^1 = \sum_{\beta\in\mathcal{A}}\tau_\beta - b_{\alpha_0^-}^0 = b_{\alpha(0)}^0 - b_{\alpha_0^-}^0 = b_{\alpha(0)} - b_{\alpha_0^-}.$$

Notice that when α varies in $\mathcal{A}\setminus\{\alpha(1)\}$ the symbol α_0^- varies in $\mathcal{A}\setminus\{\alpha(0)\}$:

$$\begin{pmatrix} \alpha(1) & \cdots & \alpha_0^- & \alpha & \cdots & \alpha(0) \\ \alpha(0) & \cdots & \cdots & \cdots & \cdots & \alpha(1) \end{pmatrix}.$$

Then (4.7) becomes

$$b_\alpha > 0 \quad \text{for } \alpha\neq\alpha(0) \quad \text{and} \quad b_{\alpha(0)} - b_\beta < 0 \quad \text{for } \beta\neq\alpha(0),$$

and (4.8) gives that the cone T_π^0 is described by

$$b_\alpha > 0 \quad \text{for all } \alpha\in\mathcal{A} \quad \text{and} \quad 0 < b_{\alpha(0)} < \min_{\beta\neq\alpha(0)} b_\beta. \tag{4.10}$$

Now it is easy to exhibit a basis of generators: take $b^\alpha = (b^\alpha_\beta)_{\beta\in\mathcal{A}}$ with

$$\begin{aligned} b^\alpha_\beta &= \begin{cases} 1 & \text{if } \beta = \alpha \\ 0 & \text{otherwise} \end{cases} \quad \text{if } \alpha \neq \alpha(0) \\ b^\alpha_\beta &= 1 \quad \text{for every } \beta \in \mathcal{A} \quad \text{if } \alpha = \alpha(0). \end{aligned} \tag{4.11}$$

A vector $b = (b_\beta)_{\beta\in\mathcal{A}}$ satisfies (4.10) if and only if it can be written in the form $b = \sum_{\alpha\in\mathcal{A}} c_\alpha b^\alpha$ with $c_\alpha > 0$ for all $\alpha \in \mathcal{A}$. It follows that T^0_π is a simplicial cone admitting the basis $\tau^\alpha = (\tau^\alpha_\beta)_{\beta\in\mathcal{A}}$ given by $\tau^\alpha_\beta = b^\alpha_\beta - b^\alpha_{\beta_0^-}$, that is,

$$\begin{aligned} \tau^\alpha_\beta &= \begin{cases} 1 & \text{if } \beta = \alpha \\ -1 & \text{if } \beta = \alpha_0^+ \\ 0 & \text{in all other cases} \end{cases} \quad \text{if } \alpha \neq \alpha(0) \\ \tau^\alpha_\beta &= \begin{cases} 1 & \text{if } \pi_0(\beta) = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{if } \alpha = \alpha(0). \end{aligned} \tag{4.12}$$

This completes the proof. □

Let $h^\alpha = -\Omega_\pi(\tau^\alpha)$, where $(\tau^\alpha)_{\alpha\in\mathcal{A}}$ is the basis of T^0_π we found in (4.12), that is,

$$\begin{aligned} h^\alpha_\beta &= \begin{cases} 1 & \text{if } \beta = \alpha \text{ or } \beta = \alpha_0^+ \\ 0 & \text{otherwise} \end{cases} \quad \text{if } \alpha \neq \alpha(0) \\ h^\alpha_\beta &= \begin{cases} 0 & \text{if } \pi_0(\beta) = 1 \text{ or } \beta = \alpha(1) \\ 1 & \text{otherwise} \end{cases} \quad \text{if } \alpha = \alpha(0). \end{aligned}$$

It is clear that the basis $(b^\alpha)_{\alpha\in\mathcal{A}}$ defined by (4.11) has volume 1. Since the map

$$b \mapsto \tau, \quad \tau_\beta = b_\beta - b_{\beta_0^-}$$

has determinant 1, it follows that $(\tau^\alpha)_{\alpha\in\mathcal{A}}$ also has volume 1. Hence, by Lemma 4.5, the contribution of the cone T^0_π to the density is

$$\frac{1}{d!}\prod_{\alpha\in\mathcal{A}} \frac{1}{\lambda\cdot h^\alpha} = \frac{1}{d!}\prod_{\alpha\neq\alpha(0)}\left(\frac{1}{\lambda_\alpha + \lambda_{\alpha_0^+}}\right)\cdot\frac{1}{\sum_{\beta\neq\alpha(1)}\lambda_\beta}.$$

There is a completely symmetric calculation for T^1_π. In this way, we get the following formula for the density in this case:

Proposition 4.9. *If π satisfies* (4.5) *then the invariant density is*

$$\frac{d\hat{\nu}}{d\lambda}(\pi,\lambda) = \sum_{\varepsilon=0,1}\frac{1}{d!}\prod_{\alpha\neq\alpha(\varepsilon)}\left(\frac{1}{\lambda_\alpha + \lambda_{\alpha_\varepsilon^+}}\right)\cdot\frac{1}{\sum_{\beta\neq\alpha(1-\varepsilon)}\lambda_\beta}$$

and $d\nu/d_1\lambda$ is given by the same expression, restricted to $C\times\Lambda_{\mathcal{A}}$.

Example 4.10. Let $d = 5$ and $\mathcal{A} = \{A, B, C, D, E\}$. Then

$$\pi = \begin{pmatrix} A & B & C & D & E \\ E & D & C & B & A \end{pmatrix}.$$

The cone T^0_π is described by

$$\begin{aligned} &b^0_A > 0,\ b^0_B > 0,\ b^0_C > 0,\ b^0_D > 0, b^0_E > 0, \\ &b^1_E < 0,\ b^1_D < 0,\ b^1_C < 0,\ b^1_B < 0, \end{aligned}$$

that is,

$$\begin{aligned} &b_A > 0,\ b_B > 0,\ b_C > 0,\ b_D > 0,\ b_E > 0, \\ &b_E - b_D < 0,\ b_E - b_C < 0,\ b_E - b_B < 0,\ b_E - b_A < 0, \end{aligned}$$

or, equivalently,

$$b_A > 0,\ b_B > 0,\ b_C > 0,\ b_D > 0,\ 0 < b_E < \min\{b_A, b_B, b_C, b_D\}.$$

As a basis take

$$\begin{aligned} &b^A = (1, 0, 0, 0, 0),\ b^B = (0, 1, 0, 0, 0),\ b^C = (0, 0, 1, 0, 0), \\ &b^D = (0, 0, 0, 1, 0),\ b^E = (1, 1, 1, 1, 1) \end{aligned}$$

or, equivalently,

$$\begin{aligned} &\tau^A = (1, -1, 0, 0, 0),\ \tau^B = (0, 1, -1, 0, 0),\ \tau^C = (0, 0, 1, -1, 0), \\ &\tau^D = (0, 0, 0, 1, -1),\ \tau^E = (1, 0, 0, 0, 0). \end{aligned}$$

We may write any $\tau = (\tau_A, \tau_B, \tau_C, \tau_D, \tau_E) \in T^0_\pi$ as

$$\tau = (b_A - b_E)\tau^A + (b_B - b_E)\tau^B + (b_C - b_E)\tau^C + (b_D - b_E)\tau^B + b_E\tau^E$$

where the coefficients are all positive. Moreover,

$$\begin{aligned} &h^A = (1, 1, 0, 0, 0),\ h^B = (0, 1, 1, 0, 0),\ h^C = (0, 0, 1, 1, 0), \\ &h^D = (0, 0, 0, 1, 1),\ h^E = (0, 1, 1, 1, 1). \end{aligned}$$

Hence, the contribution of T^0_π to the density is

$$\frac{1}{5!}\frac{1}{\lambda_A + \lambda_B}\frac{1}{\lambda_B + \lambda_C}\frac{1}{\lambda_C + \lambda_D}\frac{1}{\lambda_D + \lambda_E}\frac{1}{\lambda_B + \lambda_C + \lambda_D + \lambda_E}.$$

The cone T^1_π contributes

$$\frac{1}{5!}\frac{1}{\lambda_E + \lambda_D}\frac{1}{\lambda_D + \lambda_C}\frac{1}{\lambda_C + \lambda_B}\frac{1}{\lambda_B + \lambda_A}\frac{1}{\lambda_A + \lambda_B + \lambda_C + \lambda_D},$$

and so the total density is (recall that $|\lambda| = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha$)

$$\frac{1}{5!}\frac{1}{(\lambda_A + \lambda_B)(\lambda_B + \lambda_C)(\lambda_C + \lambda_D)(\lambda_D + \lambda_E)}\left(\frac{1}{|\lambda| - \lambda_A} + \frac{1}{|\lambda| - \lambda_E}\right).$$

4.3 Combinatorial statement

We want to prove that the intersection of every pre-stratum with the set of (π, λ, τ) such that $\operatorname{area}(\pi, \lambda, \tau) \leq 1$ has finite volume. The crucial step is

Proposition 4.11. *Let $(\tau^\beta)_{\beta \in \mathcal{A}}$ be a basis of $\mathbb{R}^{\mathcal{A}}$ contained in the closure of T_π^δ for some $\delta \in \{0,1\}$, and let $h^\beta = -\Omega_\pi(\tau^\beta)$ for $\beta \in \mathcal{A}$. Given any non-empty proper subset $\mathcal{B}$ of $\mathcal{A}$, we have*

$$\#\big\{\beta \in \mathcal{A} : h_\alpha^\beta = 0 \text{ for all } \alpha \in \mathcal{B}\big\} + \#\mathcal{B} \leq d,$$

and the inequality is strict unless $\mathcal{B}$ contains $\alpha(1-\delta)$ but not $\alpha(\delta)$.

Proof. We suppose $\delta = 0$, as the other case is analogous. Let $h = -\Omega_\pi(\tau)$ for some τ in the closure of T_π^0. By (2.12) and (4.6),

$$h_\alpha = b_\alpha^0 - b_\alpha^1 = b_{\alpha_0^-}^0 - b_{\alpha_1^-}^1 \,. \tag{4.13}$$

The symbol α_ε^- is not defined when $\pi_\varepsilon(\alpha) = 1$, but (4.13) remains valid in that case, as long as one interprets $b_{\alpha_\varepsilon^-}^\varepsilon$ to be zero. By the definition of $T_\pi^0 \subset T_\pi^+$ in (2.10) and (4.8), and the assumption that τ is in the closure of T_π^0,

$$b_\alpha^0 \geq 0 \ \text{ for all } \alpha \in \mathcal{A} \quad \text{and} \quad b_\alpha^1 \leq 0 \ \text{ for all } \alpha \in \mathcal{A} \setminus \{\alpha(1)\}. \tag{4.14}$$

Therefore, given any $\alpha \neq \alpha(1)$,

$$h_\alpha = 0 \quad \Rightarrow \quad b_\alpha^0 = b_\alpha^1 = 0 = b_{\alpha_0^-}^0 = b_{\alpha_1^-}^1. \tag{4.15}$$

A part of (4.15) remains valid even when $\alpha = \alpha(1)$:

$$h_\alpha = 0 \quad \Rightarrow \quad b_{\alpha_0^-}^0 = b_{\alpha_1^-}^1 = 0, \tag{4.16}$$

because $\alpha_1^- \neq \alpha(1)$. Finally, adding the relations

$$h_{\alpha(0)} = b_{\alpha(0)}^0 - b_{\alpha(0)}^1 \quad \text{and} \quad h_{\alpha(1)} = b_{\alpha(1)}^0 - b_{\alpha(1)}^1,$$

and recalling that $b_{\alpha(0)}^0 = \sum_{\beta \in \mathcal{A}} \tau_\alpha = b_{\alpha(1)}^1$, we get that

$$h_\alpha = 0 \text{ for both } \alpha \in \{\alpha(0), \alpha(1)\} \quad \Rightarrow \quad b_{\alpha(1)}^0 = b_{\alpha(0)}^1 = 0. \tag{4.17}$$

Now let $\mathcal{B}$ be a non-empty proper subset of $\mathcal{A}$, and assume $h_\alpha = 0$ for all $\alpha \in \mathcal{B}$.

Case 1: $\mathcal{B}$ does not contain $\alpha(1)$. Define

$$\mathcal{B}_\varepsilon = \mathcal{B} \cup \{\alpha_\varepsilon^- : \alpha \in \mathcal{B}\} \quad \text{for } \varepsilon \in \{0,1\}. \tag{4.18}$$

Then (4.15) gives that

$$b_\beta^\varepsilon = 0 \ \text{ for all } \beta \in \mathcal{B}_\varepsilon \text{ and } \varepsilon \in \{0,1\}. \tag{4.19}$$

We claim that there exists $\varepsilon \in \{0, 1\}$ such that

$$\#\mathcal{B}_\varepsilon > \#\mathcal{B}. \tag{4.20}$$

Indeed, it follows from the definition (4.18) that $\mathcal{B}$ is contained in $\mathcal{B}_\varepsilon$. Moreover, the two sets coincide only if $\alpha_\varepsilon^- \in \mathcal{B}$ for every $\alpha \in \mathcal{B}$ or, in other words, if

$$\mathcal{B} = \pi_\varepsilon^{-1}(\{1, \ldots, k\}) \quad \text{for some } 1 \le k \le d. \tag{4.21}$$

Note that $k < d$, because $\mathcal{B}$ is a proper subset of $\mathcal{A}$. So, since π is irreducible, (4.21) can not hold simultaneously for both $\varepsilon = 0$ and $\varepsilon = 1$. Hence, there exists ε such that $\mathcal{B}_\varepsilon \neq \mathcal{B}$. This proves the claim. Now fix any such ε. Since the map $\tau \mapsto b^\varepsilon$ is injective, and the $(\tau^\beta)_{\beta\in\mathcal{A}}$ are linearly independent, (4.19) and (4.20) give

$$\#\{\beta \in \mathcal{A} : h_\alpha^\beta = 0 \text{ for all } \alpha \in \mathcal{B}\} \le d - \#\mathcal{B}_\varepsilon < d - \#\mathcal{B}.$$

Case 2: $\mathcal{B}$ contains $\alpha(1)$ but not $\alpha(0)$. Let $\mathcal{B}_1 = \mathcal{B} \setminus \{\alpha(1)\} \cup \{\alpha_1^- : \alpha \in \mathcal{B}\}$. The relations (4.15) and (4.16) imply that

$$b_\beta^1 = 0 \ \text{ for all } \beta \in \mathcal{B}_1.$$

Let $k \ge \pi_1(\alpha(0))$ be maximum such that $\bar\beta = \pi_1^{-1}(k)$ is not in $\mathcal{B}$. The assumption that $\mathcal{B}$ contains $\alpha(1)$ but not $\alpha(0)$ ensures that k is well defined and less than d. Then $\bar\beta = \alpha_1^-$ for some $\alpha \in \mathcal{B}$, and so $\bar\beta \in \mathcal{B}_1$. This shows that

$$\mathcal{B}_1 \supset \mathcal{B} \setminus \{\alpha(1)\} \cup \{\bar\beta\},$$

and so $\#\mathcal{B}_1 \ge \#\mathcal{B}$. Hence, just as before,

$$\#\{\beta \in \mathcal{A} : h_\alpha^\beta = 0 \text{ for all } \alpha \in \mathcal{B}\} \le d - \#\mathcal{B}_1 \le d - \#\mathcal{B}.$$

Case 3: $\mathcal{B}$ contains both $\alpha(0)$ and $\alpha(1)$. Define $\mathcal{B}_0 = \mathcal{B} \cup \{\alpha_0^- : \alpha \in \mathcal{B}\}$. By (4.15), (4.16), (4.17),

$$b_\beta^0 = 0 \quad \text{for all } \beta \in \mathcal{B}_0.$$

It is easy to check that $\mathcal{B}_0$ contains $\mathcal{B}$ strictly. Indeed, the two sets can coincide only if $\alpha_0^- \in \mathcal{B}$ for every $\alpha \in \mathcal{B}$, that is, if $\mathcal{B} = \pi_0^{-1}(\{1, \ldots, k\}$ for some k. Since $\mathcal{B}$ contains $\alpha(0) = \pi_0^{-1}(d)$, this would imply $\mathcal{B} = \mathcal{A}$, contradicting the hypothesis. It follows, just as in the first case, that

$$\#\{\beta \in \mathcal{A} : h_\alpha^\beta = 0 \text{ for all } \alpha \in \mathcal{B}\} \le d - \#\mathcal{B}_0 < d - \#\mathcal{B}.$$

The proof of Proposition 4.11 is complete. □

Remark 4.12. The inequality in Proposition 4.11 is not always strict. Indeed, let $\tau^A, \ldots, \tau^E$ be the generators of T_π^0 in Example 4.10, and let $\mathcal{B} = \{A\}$. Then $\mathcal{B}$ contains $A = \alpha(1)$ but not $E = \alpha(0)$. Note also that

$$\{\beta : h_A^\beta = 0\} = \{B, C, D, E\}$$

has exactly $4 = d - \#\mathcal{B}$ elements. Thus, the equality holds in this case. In fact, if the inequality were strict in all cases, then arguments as in the next section would imply that the measure ν is finite. However, the latter is usually not true, as we have already seen in Example 4.7.

4.4 Finite volume

Let C be a Rauzy class and $\hat{\mathcal{S}} = \hat{\mathcal{S}}(C)$ be the corresponding pre-stratum. Define the normalized pre-stratum to be the subset $\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_1(C)$ of all $(\pi, \lambda, \tau) \in \hat{\mathcal{S}}$ such that $\operatorname{area}(\lambda, \tau) \leq 1$.

Theorem 4.13. *For every Rauzy class C, the normalized pre-stratum $\hat{\mathcal{S}}_1$ has finite volume: $\hat{m}(\hat{\mathcal{S}}_1) < \infty$.*

Proof. Recall that $\hat{\mathcal{S}}_1$ is obtained from the subset of all $(\pi, \lambda, \tau) \in \hat{\mathcal{H}}$ such that $\operatorname{area}(\lambda, \tau) \leq 1$ and

$$\sum_{\alpha \neq \alpha(1-\varepsilon)} \lambda_\alpha \leq 1 \leq \sum_{\alpha \in \mathcal{A}} \lambda_\alpha \tag{4.22}$$

by identifying (π, λ, τ) with $\hat{\mathcal{R}}(\pi, \lambda, \tau)$ when $\sum_{\alpha \neq \alpha(1-\varepsilon)} \lambda_\alpha = 1$. Thus,

$$\operatorname{vol}(\hat{\mathcal{S}}_1) = \sum_{\pi \in C} \int \rho(\pi, \lambda)\, d\lambda, \tag{4.23}$$

where $\rho(\pi, \lambda)$ is the d-dimensional volume of $\{\tau \in T_\pi^+ : \operatorname{area}(\lambda, \tau) \leq 1\}$, and the integral is over the set of $\lambda \in \mathbb{R}_+^{\mathcal{A}}$ satisfying (4.22). Let T^i, $i = 1, \ldots, k$ be a decomposition of T_π^+ (up to a codimension 1 subset) into simplicial cones. Then, by Proposition 4.6,

$$\rho(\pi, \lambda) = \frac{1}{d!} \sum_{i=1}^{k} \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^{i,\beta}} \tag{4.24}$$

where $h^{i,\beta} = -\Omega_\pi(\tau^{i,\beta})$ and $(\tau^{i,\beta})_{\beta \in \mathcal{A}}$ is a basis of generators of T^i. We may assume that each T^i is contained either in T_π^0 or in T_π^1, and we do so in what follows. Let us consider (compare (2.64) also)

$$\Lambda_{\mathcal{A}} \times \mathbb{R} \ni (\lambda, s) \mapsto e^s \lambda \in \mathbb{R}_+^{\mathcal{A}}. \tag{4.25}$$

Recall that $d\lambda = e^{sd} d_1\lambda\, ds$, where $d_1\lambda$ denotes the $(d-1)$-dimensional volume induced on the simplex $\Lambda_{\mathcal{A}}$ by the Riemannian metric of $\mathbb{R}^{\mathcal{A}}$. Notice that, given $(\lambda, s) \in \Lambda_{\mathcal{A}} \times \mathbb{R}$, the vector $e^s\lambda$ satisfies (4.22) if and only if $0 \leq s \leq t_R(\pi, \lambda)$, where t_R is the Rauzy renormalization time defined in (2.65). Recall also that $\lambda \mapsto \rho(\pi, \lambda)$ is homogeneous of degree $-d$. Thus, after change of variables, (4.23) becomes

$$\operatorname{vol}(\hat{\mathcal{S}}_1) = \sum_{\pi \in C} \int_{\Lambda_{\mathcal{A}}} \int_0^{t_R(\pi,\lambda)} \rho(\pi, e^s\lambda) e^{sd}\, ds\, d_1\lambda = \sum_{\pi \in C} \int_{\Lambda_{\mathcal{A}}} \rho(\pi, \lambda) t_R(\pi, \lambda)\, d_1\lambda.$$

Using (4.24) and the definition of $t_R(\pi, \lambda)$, this gives

$$\operatorname{vol}(\hat{\mathcal{S}}_1) = \frac{1}{d!} \sum_{\pi \in C} \sum_{i=1}^{k} \int_{\Lambda_{\mathcal{A}}} -\log\left(1 - \lambda_{\alpha(1-\varepsilon)}\right) \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^{i,\beta}}\, d_1\lambda, \tag{4.26}$$

where ε is the type of (π, λ). Therefore, to prove the theorem we only have to show that the integral is finite, for every fixed $\pi \in C$ and $i = 1, \ldots, k$.

For simplicity, we write $h^\beta = h^{i,\beta}$ in what follows. Also, we assume T^i is contained in T^0_π; the other case is analogous. This implies the corresponding basis of generators $(\tau^{i,\beta})_{\beta \in \mathcal{A}}$ is contained in the closure of T^0_π.

Let $\mathcal{N}$ denote the set of integer vectors $n = (n_\alpha)_{\alpha \in \mathcal{A}}$ such that $n_\alpha \geq 0$ for all $\alpha \in \mathcal{A}$, and the n_α are not all zero. For each $n \in \mathcal{N}$, define

$$\Lambda(n) = \{\lambda \in \Lambda_{\mathcal{A}} : 2^{-n_\alpha} \leq \lambda_\alpha d < 2^{-n_\alpha+1} \text{ for every } \alpha \in \mathcal{A}\}, \tag{4.27}$$

except that for $n_\alpha = 0$ the second inequality is omitted.

Lemma 4.14. *There exists $c_1 > 0$ depending only on the dimension d such that*

$$\operatorname{vol}_{d-1} \Lambda(n) \leq c_1 2^{-\sum_{\mathcal{A}} n_\alpha}$$

for all $n \in \mathcal{N}$. Moreover, the family $\Lambda(n)$, $n \in \mathcal{N}$ covers $\Lambda_{\mathcal{A}}$.

Proof. If $\sum_{\alpha \in \mathcal{A}} \lambda_\alpha = 1$ then $\lambda_\beta \geq 1/d$ for some $\beta \in \mathcal{A}$, and so λ belongs to some $\Lambda(n)$ with $n_\beta = 0$. This shows that these sets $\Lambda(n)$ do cover $\Lambda_{\mathcal{A}}$. To prove the volume estimate, fix n and $\beta \in \mathcal{A}$ such that $n_\beta = 0$. When λ varies in $\Lambda(n)$, the $(d-1)$-dimensional vector $(\lambda_\alpha)_{\alpha \neq \beta}$ varies in some subset $S(n)$ of the product space $\prod_{\alpha \neq \beta} [0, 2^{-n_\alpha+1}]$. The $(d-1)$-dimensional volume of $S(n)$ is bounded above by $2^{d-1} 2^{-\sum_{\alpha \in \mathcal{A}} n_\alpha}$. Then, since $\Lambda(n)$ is a graph over $S(n)$,

$$\operatorname{vol}_{d-1} \Lambda(n) \leq \sqrt{d} \operatorname{vol}_{d-1} S(n) \leq c_1 2^{-\sum_{\alpha \in \mathcal{A}} n_\alpha},$$

where $c_1 = \sqrt{d}\, 2^{d-1}$. The proof is complete. $\square$

It is clear that $\lambda_{\alpha(1-\varepsilon)} < 1/2$, and so

$$-\log\left(1 - \lambda_{\alpha(1-\varepsilon)}\right) \leq 2\lambda_{\alpha(1-\varepsilon)} = 2 \min\{\lambda_{\alpha(0)}, \lambda_{\alpha(1)}\}.$$

Therefore, for each fixed π and i, the integral in (4.26) is bounded above by

$$\sum_{n \in \mathcal{N}} \int_{\Lambda(n)} 2 \min\{\lambda_{\alpha(0)}, \lambda_{\alpha(1)}\} \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^\beta}\, d_1\lambda. \tag{4.28}$$

For each $\beta \in \mathcal{A}$, let $\mathcal{A}(\beta)$ be the subset of $\alpha \in \mathcal{A}$ such that $h^\beta_\alpha > 0$. Let $c_2 > 0$ be the minimum of the non-zero h^β_α, over all α and β. Then

$$\lambda \cdot h^\beta = \sum_{\mathcal{A}(\beta)} h^\beta_\alpha \lambda_\alpha \geq \sum_{\mathcal{A}(\beta)} c_2 d^{-1} 2^{-n_\alpha} \geq c_2 d^{-1} 2^{-\min_{\mathcal{A}(\beta)} n_\alpha} \tag{4.29}$$

for every $\lambda \in \Lambda(n)$ and $\beta \in \mathcal{A}$. Using Lemma 4.14 we deduce that

$$\begin{aligned}\int_{\Lambda(n)} 2\min\{\lambda_{\alpha(0)},\lambda_{\alpha(1)}\} \prod_{\beta\in\mathcal{A}} \frac{1}{\lambda\cdot h^\beta}\, d_1\lambda \\ \le K\, 2^{-\max_\varepsilon n_{\alpha(\varepsilon)}+\sum_\beta \min_{\mathcal{A}(\beta)} n_\alpha - \sum_\alpha n_\alpha},\end{aligned} \tag{4.30}$$

where the constant $K = (2c_1)(d/c_2)^d$. Using Proposition 4.11, we obtain

Lemma 4.15.

$$\max_{\varepsilon\in\{0,1\}} n_{\alpha(\varepsilon)} - \sum_{\beta\in\mathcal{A}} \min_{\alpha\in\mathcal{A}(\beta)} n_\alpha + \sum_{\alpha\in\mathcal{A}} n_\alpha \ge \max_{\alpha\in\mathcal{A}} n_\alpha.$$

Proof. Let $0 = n^0 < n^1 < \cdots$ be the different values taken by n_α, and $\mathcal{B}^i$, $i \ge 0$ be the set of values of $\alpha \in \mathcal{A}$ such that $n_\alpha \ge n^i$. On the one hand,

$$\sum_{\alpha\in\mathcal{A}} n_\alpha = \sum_{i\ge1} n^i\big(\#\mathcal{B}^i - \#\mathcal{B}^{i+1}\big) = \sum_{i\ge1} \#\mathcal{B}^i(n^i - n^{i-1}). \tag{4.31}$$

On the other hand, $\min_{\mathcal{A}(\beta)} n_\alpha \ge n^i$ if and only if $\mathcal{A}(\beta) \subset \mathcal{B}^i$. Consequently,

$$\begin{aligned}\sum_{\beta\in\mathcal{A}} \min_{\mathcal{A}(\beta)} n_\alpha &= \sum_{i\ge1} n^i\big(\#\{\beta : \mathcal{A}(\beta)\subset\mathcal{B}^i\} - \#\{\beta : \mathcal{A}(\beta)\subset\mathcal{B}^{i+1}\}\big) \\ &= \sum_{i\ge1} \#\{\beta : \mathcal{A}(\beta)\subset\mathcal{B}^i\}(n^i - n^{i-1}).\end{aligned} \tag{4.32}$$

Observe that $\mathcal{A}(\beta) \subset \mathcal{B}^i$ if and only if $h_\alpha^\beta = 0$ for all $\alpha \in \mathcal{A} \setminus \mathcal{B}^i$. So, by Proposition 4.11 (with $\mathcal{B} = \mathcal{A} \setminus \mathcal{B}^i$),

$$\#\{\beta : \mathcal{A}(\beta)\subset\mathcal{B}^i\} < \#\mathcal{B}^i \tag{4.33}$$

except, possibly, if $\mathcal{B}^i$ contains $\alpha(0)$ but not $\alpha(1)$. On the one hand, if (4.33) does hold then

$$\#\mathcal{B}^i(n^i - n^{i-1}) - \#\{\beta : \mathcal{A}(\beta)\subset\mathcal{B}^i\}(n^i - n^{i-1}) \ge (n^i - n^{i-1}). \tag{4.34}$$

On the other hand, if $\mathcal{B}^i$ contains $\alpha(0)$ but not $\alpha(1)$ then $n_{\alpha(1)} < n^i \le n_{\alpha(0)}$. Let i_1 be the smallest and i_2 be the largest value of i for which this happens. Then

$$\#\mathcal{B}^i - \#\{\beta : \mathcal{A}(\beta)\subset\mathcal{B}^i\} \ge 0 \quad \text{for} \quad i_1 \le i \le i_2 \tag{4.35}$$

$$\text{and} \quad \max\{n_{\alpha(0)}, n_{\alpha(1)}\} = n_{\alpha(0)} \ge n^{i_2} - n^{i_1-1} = \sum_{i=i_1}^{i_2} (n^i - n^{i-1}).$$

Putting (4.34) and (4.35) together, we find that

$$\max_{\varepsilon\in\{0,1\}} n_{\alpha(\varepsilon)} - \sum_{\beta\in\mathcal{A}} \min_{\mathcal{A}(\beta)} n_\alpha + \sum_{\alpha\in\mathcal{A}} n_\alpha \ge \sum_{i\ge1}^{k} (n^i - n^{i-1}) = \max_{\alpha\in\mathcal{A}} n_\alpha.$$

This proves the lemma. ◻

Replacing the conclusion of the lemma in (4.30) we obtain, for every $\in \mathcal{N}$,

$$\int_{\Lambda(n)} 2\min\{\lambda_{\alpha(0)}, \lambda_{\alpha(1)}\} \prod_{\beta\in\mathcal{A}} \frac{1}{\lambda \cdot h^\beta}\, d_1\lambda \le K 2^{-\max_{\mathcal{A}} n_\alpha}. \tag{4.36}$$

For each $m \ge 0$ there are at most $(m+1)^d$ choices of $n \in \mathcal{N}$ with $\max_{\mathcal{A}} n_\alpha = m$. So, (4.36) implies that the integral in (4.26) is bounded above by

$$\sum_{m=0}^{\infty} K(m+1)^d 2^{-m} < \infty$$

for every $\pi \in C$ and every $1 \le i \le k$. The proof of Theorem 4.13 is complete. □

4.5 Recurrence and inducing

Given a measurable map $f : M \to M$ and a measure μ on M, we call (f, μ) *recurrent* if for any positive measure set $E \subset M$ and μ-almost every $x \in E$ there exists $n \ge 1$ such that $f^n(x) \in E$. The classical Poincaré recurrence theorem asserts that if μ is invariant and *finite* then (f, μ) is recurrent. Similar observations hold for flows as well.

Lemma 4.16. *The Teichmüller flow $\mathcal{T}^t : \hat{\mathcal{S}} \to \hat{\mathcal{S}}$ and semi-flow $T^t : \hat{S} \to \hat{S}$ are recurrent, for the corresponding invariant measures $\hat{m}$ and $\hat{\nu}$. The Rauzy-Veech renormalization maps $\mathcal{R} : \mathcal{H} \to \mathcal{H}$ and $R : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$ are also recurrent, for the corresponding invariant measures m and ν.*

Proof. Since $\hat{m}$ is a finite measure, by Theorem 4.13, the claim that $(\mathcal{T}^t, \hat{m})$ is recurrent is a direct consequence of the Poincaré recurrence theorem. The claim for $(T^t, \hat{\nu})$ follows immediately, because $\hat{\nu} = P_*(\hat{m})$ and $T^t \circ P = P \circ \mathcal{T}^t$: given any positive measure set $D \subset \hat{S}$, the fact that $\hat{m}$-almost every point of $P^{-1}(D)$ returns to $P^{-1}(D)$ under $\mathcal{T}^t$ implies that $\hat{\nu}$-almost every point of D returns to D under T^t. Similarly, the statement for $(\mathcal{R}, m)$ follows immediately from the fact that $(\mathcal{T}^t, \hat{m})$ is recurrent, $\mathcal{R}$ is the return map of $\mathcal{T}^t$ to the cross-section $\mathcal{S}$, and a subset of the cross-section as positive m-measure if and only the set of flow orbits has positive $\hat{m}$-measure. For the same reasons, the fact that $(T^t, \hat{\nu})$ is recurrent implies that (R, ν) is recurrent. □

If (f, μ) is recurrent then, given any positive measure $D \subset M$ there is a first-return map $f_D : D \to D$ of f to D, defined by

$$f_D(x) = f^n(x), \quad n = \min\{k \ge 1 : f^k(x) \in D\}$$

at almost every point $x \in D$. We call f_D the map *induced* by f on D.

Lemma 4.17. *The induced map f_D preserves the restriction of μ to D.*

Proof. Suppose first that f is invertible. Then, given any measurable set $E \subset D$, the pre-image $f_D^{-1}(E)$ is the disjoint union of all $f^{-k}(E_k)$, $k \geq 1$ where E_k is the set of points $x \in E$ such that $f^{-k}(x) \in D$ but $f^{-j}(x) \notin D$ for $0 < j < k$. Since these E_k are pairwise disjoint, we get

$$\mu(f_D^{-1}(E)) = \sum_{k \geq 1} \mu(f^{-k}(E_k)) = \sum_{k \geq 1} \mu(E_k) = \mu(E).$$

To treat the general, possibly non-invertible, case, consider the natural extension $(\tilde{f}, \tilde{\mu})$ of the system (f, μ). This is defined by

$$\tilde{f} : \tilde{M} \to \tilde{M}, \quad \tilde{f}(\ldots, x_n, \ldots, x_0) = (\ldots, x_n, \ldots, x_0, f(x_0))$$

where $\tilde{M}$ is the space of all sequences $(x_n)_n$ on M such that $f(x_n) = x_{n-1}$ for all $n \geq 1$. Moreover, $\tilde{\mu}$ is the unique $\tilde{f}$-invariant measure such that $\pi_*(\tilde{\mu}) = \mu$, where $\pi : \tilde{M} \to M$ is the projection $(x_n)_n \mapsto x_0$. Clearly, $\pi \circ \tilde{f} = f \circ \pi$. Moreover,

$$\pi \circ \tilde{f}_{\tilde{D}} = f_D \circ \pi$$

where $\tilde{f}_{\tilde{D}}$ denotes the map induced by $\tilde{f}$ on $\tilde{D} = \pi^{-1}(D)$. Then, using the previous paragraph,

$$\mu(f_D^{-1}(E)) = \tilde{\mu}(\pi^{-1}(f_D^{-1}(E)) = \tilde{\mu}(\tilde{f}_{\tilde{D}}^{-1}(\pi^{-1}(E))) = \tilde{\mu}(\pi^{-1}(E)) = \mu(E).$$

This completes the proof. $\square$

Remark 4.18. It is clear that if f is ergodic for μ then f_D is ergodic for the restricted measure $\mu \mid D$. Indeed, given any $E \subset D$, let $F = \cup_{n=0}^{\infty} F_n$, where $F_0 = E$ and

$$F_n = \{x \in M : f^n(x) \in E \text{ but } f^k(x) \notin E \text{ for al } 0 \leq j < n\} \quad \text{for } n \geq 1.$$

If E is f_D-invariant then F is f-invariant. Suppose $\mu(E) > 0$. Then $\mu(F) > 0$ and so, by hypothesis, $\mu(F) = 1$. Consequently, $\mu(E) = \mu(F \cap D) = \mu(D)$. This shows that f_D is ergodic if f is. We are going to prove a partial converse to this fact.

We say that (f, μ) is a *Markov system* if the measure μ is f-invariant and there exists a countable partition $(M_j)_j$ of a full measure subset of M, such that each M_j is mapped bijectively to a full measure subset of M. Such systems always admit a Jacobian. Indeed, let μ_j be the measure defined on each M_j by $\mu_j(E) = \mu(f(E))$. Since μ is invariant, $\mu \leq \mu_j$ and, in particular, μ is absolutely continuous with respect to μ_j. The set where the Radon-Nikodym derivative vanishes has zero μ-measure:

$$\mu\big(\{x : \frac{d\mu}{d\mu_j}(x) = 0\}\big) = \int_{\{x : \frac{d\mu}{d\mu_j}(x) = 0\}} \frac{d\mu}{d\mu_j} \, d\mu_j = 0.$$

Hence, $J_\mu f(x) = (d\mu/d\mu_j)^{-1}(x)$ is well-defined at μ-almost every point in each M_j, and it is a Jacobian of f relative to μ:

$$\int_E J_\mu f \, d\mu = \int_E \left(\frac{d\mu}{d\mu_j}\right)^{-1} d\mu = \int_E d\mu_j = \mu_j(E) = \mu(f(E))$$

for every measurable set $E \subset M_j$ and every $j \geq 1$.

Lemma 4.19. *Assume (f, μ) is a Markov system. If the map induced by f on some of the Markov domains M_j is ergodic for the restriction of μ to M_j, then (f, μ) itself is ergodic.*

Proof. Let $F \subset M$ be f-invariant. Then $E = F \cap M_j$ is f_{M_j}-invariant and so, either $\mu(E) = 0$ or $\mu(M_j \setminus E) = 0$. In the first case, the existence of a Jacobian implies that $\mu(f(E)) = 0$. Notice that $f(E) = F$, up to a zero measure set, because $f : M_j \to M$ is essentially surjective and F is an invariant set. It follows that $\mu(F) = 0$. In the second case, a similar argument shows that $\mu(M \setminus F) = 0$. This proves that f is ergodic. $\square$

We are going to apply these observations to the Rauzy-Veech renormalization map R, and the R-invariant measure ν constructed in Section 4.1. Recall that R maps each $\{\pi\} \times \Lambda_{\pi,\varepsilon}$ bijectively to $\{\pi'\} \times \Lambda_{\mathcal{A}}$, where π' is the type ε successor of π and

$$\Lambda_{\pi,\varepsilon} = \{\lambda \in \Lambda_{\mathcal{A}} : (\pi, \lambda) \text{ has type } \varepsilon\}.$$

For each $n \geq 1$ and $\underline{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_{n-1}) \in \{0, 1\}^n$, define

$$\Lambda_{\pi,\underline{\varepsilon},n} = \{\lambda \in \Lambda_{\mathcal{A}} : R^k(\pi, \lambda) \text{ has type } \varepsilon_k \text{ for } k = 0, 1, \ldots, n-1\}. \tag{4.37}$$

Then R^n maps every $\{\pi\} \times \Lambda_{\pi,\underline{\varepsilon},n}$ bijectively to $\{\pi^n\} \times \Lambda_{\mathcal{A}}$. As a consequence of (1.11), the simplex $\Lambda_{\pi,\underline{\varepsilon},n}$ is the image of $\Lambda_{\mathcal{A}}$ under the projectivization of Θ^{n*}, where $\Theta^{n*} = \Theta^{n*}_{\pi,\lambda}$ for any $(\pi, \lambda) \in \Lambda_{\pi,\underline{\varepsilon},n}$. By Corollary 1.21, one may find π, $N \geq 1$, and $\underline{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_{N-1})$ such that $\Lambda_* = \Lambda_{\pi,\underline{\varepsilon},N}$ is relatively compact in $\Lambda_{\mathcal{A}}$. Let Λ_* be fixed from now on and denote by $R_* : \Lambda_* \to \Lambda_*$ the map induced by R^N on $\Lambda_* \approx \{\pi\} \times \Lambda_*$. For x in a full measure subset of Λ_*, let $k \geq 1$ be the smallest positive integer such that $R^{kN}(x) \in \Lambda_*$. Then the set $\Lambda_{\pi,\underline{\theta},(k+1)N}$ that contains x satisfies

$$R^{kN}(\Lambda_{\pi,\underline{\theta},(k+1)N}) = \Lambda_*.$$

In particular, $R_* = R^{kN}$ on the set $\Lambda_{\pi,\underline{\theta},(k+1)N}$. This proves that $\big(R_*, (\nu \mid \Lambda_*)\big)$ is a Markov system.

Proposition 4.20. *The Markov system $\big(R_*, (\nu \mid \Lambda_*)\big)$ is ergodic.*

The proof of this proposition appears in Section 4.7. It uses the notion of projective metric, that we recall in Section 4.6. This notion will be useful again later. Also in Section 4.7, we deduce from the proposition that the renormalization maps R and $\mathcal{R}$, and the Teichmüller flow $\mathcal{T}^t$ are ergodic, relative to their invariant measures ν, $\hat{\nu}$, and $\hat{m}$.

4.6 Projective metrics

Birkhoff [6] introduced the notion of projective metric associated to a general convex cone C in any vector space. Here we only need the case $C = \mathbb{R}_+^{\mathcal{A}}$.

Given any $u, v \in C$, define

$$a(u,v) = \inf\{\frac{v_\alpha}{u_\alpha} : \alpha \in \mathcal{A}\} \quad \text{and} \quad b(u,v) = \sup\{\frac{v_\beta}{u_\beta} : \beta \in \mathcal{A}\}. \tag{4.38}$$

Notice that

$$v - tu \in C \Leftrightarrow t < a(u,v) \quad \text{and} \quad su - v \in C \Leftrightarrow s > b(u,v). \tag{4.39}$$

We call *projective metric* associated to $C = \mathbb{R}_+^{\mathcal{A}}$ the function $\mathrm{d_p}(\cdot\,,\cdot)$ defined by

$$\mathrm{d_p}(u,v) = \log \frac{b(u,v)}{a(u,v)} = \log \sup\{\frac{u_\alpha}{v_\alpha}\frac{v_\beta}{u_\beta} : \alpha, \beta \in \mathcal{A}\} \tag{4.40}$$

for each $u, v \in C$. This terminology is justified by the next lemma, which says that $\mathrm{d_p}(\cdot\,,\cdot)$ induces a distance in the projective quotient of C. The lemma is an easy consequence of the definition (4.40).

Lemma 4.21. *For all $u, v, w \in C$,*

(a) $\mathrm{d_p}(u,v) = \mathrm{d_p}(v,u)$

(b) $\mathrm{d_p}(u,v) + \mathrm{d_p}(v,w) \geq \mathrm{d_p}(u,w)$

(c) $\mathrm{d_p}(u,v) \geq 0$

(d) $\mathrm{d_p}(u,v) = 0$ *if and only if there exists $t > 0$ such that $u = tv$.*

Let $G\colon \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ be a linear operator such that $G(C) \subset C$ or, equivalently, such that all the entries $G_{\alpha,\beta}$ of the matrix of G are non-negative. Then

$$t < a(u,v) \Leftrightarrow v - t\,u \in C \ \Rightarrow G(v) - tG(u) \in C \Leftrightarrow t < a(G(u), G(v)).$$

This means that $a(u,v) \leq a(G(u), G(v))$ and a similar argument proves that $b(u,v) \geq b(G(u), G(v))$. Therefore,

$$\mathrm{d_p}(G(u), G(v)) \leq \mathrm{d_p}(u,v) \quad \text{for all } u, v \in C. \tag{4.41}$$

It follows from Lemma 4.21 that, restricted to the simplex $\Lambda_{\mathcal{A}}$, the function d_p is a genuine metric. We call $g : \Lambda_{\mathcal{A}} \to \Lambda_{\mathcal{A}}$ a *projective map* if there exists a linear isomorphism $G : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ such that $G(\mathbb{R}_+^{\mathcal{A}}) \subset \mathbb{R}_+^{\mathcal{A}}$ and

$$g(\lambda) = \frac{G(\lambda)}{\sum_{\alpha\in\mathcal{A}} G(\lambda)_\alpha} = \frac{G(\lambda)}{\sum_{\alpha,\beta\in\mathcal{A}} G_{\alpha,\beta}\lambda_\beta}\,. \tag{4.42}$$

We say g is the projectivization of G. The relation (4.41) means that projective maps never expand the projective metric on the simplex.

A set $K \subset \Lambda_{\mathcal{A}}$ is relatively compact in $\Lambda_{\mathcal{A}}$ if and only if the coordinates of its points are all larger than some positive constant. So, it follows directly from the definition (4.40) that if K is relatively compact in $\Lambda_{\mathcal{A}}$ then it has finite diameter relative to the projective metric:

$$\sup_{x,y\in K} \mathrm{d_p}(x,y) < \infty.$$

We shall see in Proposition 4.23 that if the entries of G are strictly positive or, equivalently, if the image of g is relatively compact in $\Lambda_{\mathcal{A}}$, then the inequality in (4.41) is strict. Thus, in that case the maps G and g are uniform contractions relative to the projective metrics in $\mathbb{R}_+^{\mathcal{A}}$ and $\Lambda_{\mathcal{A}}$, respectively.

Lemma 4.22. *Let $g : \Lambda_{\mathcal{A}} \to \Lambda_{\mathcal{A}}$ be a projective map and Dg be its derivative. Then $\log|\det Dg|$ is $(d+1)$-Lipschitz continuous for the projective distance.*

Proof. We use the following observation: if a functional $h(\lambda) = \sum_\beta h_\beta \lambda_\beta$ has non-negative coefficients, $h_\beta \geq 0$, then $\log h(\lambda)$ is 1-Lipschitz relative to the projective distance. Indeed,

$$\log h(\sigma) - \log h(\lambda) = \log \frac{\sum_\beta h_\beta \sigma_\beta}{\sum_\beta h_\beta \lambda_\beta} \leq \log \sup_\beta \frac{\sigma_\beta}{\lambda_\beta} = \log b(\lambda,\sigma).$$

Recall the definition (4.38). Since $\sum_\beta \lambda_\beta = 1 = \sum_\beta \sigma_\beta$, we also have $a(\lambda,\sigma) \leq 1$. It follows that $\log b(\lambda,\sigma) \leq \mathrm{d_p}(\lambda,\sigma)$. This justifies our observation.

Now let g be the projectivization of some linear isomorphism G. We begin by expressing Dg in terms of G. Let $\dot{\Lambda}_{\mathcal{A}}$ represent the hyperplane tangent to the simplex $\Lambda_{\mathcal{A}}$. From (4.42) we find that, for any $\dot{\lambda} \in \dot{\Lambda}_{\mathcal{A}}$,

$$Dg(\lambda)\dot{\lambda} = \frac{G(\dot{\lambda})}{s(\lambda)} - \frac{G(\lambda)}{s(\lambda)} \frac{\sum_\alpha G(\dot{\lambda})_\alpha}{s(\lambda)}, \qquad s(\lambda) = \sum_{\alpha,\beta} G_{\alpha,\beta}\lambda_\beta.$$

This may be rewritten as $Dg(\lambda) = P_\lambda \circ s(\lambda)^{-1} \circ G$, where $G : \dot{\Lambda}_{\mathcal{A}} \to G(\dot{\Lambda}_{\mathcal{A}})$, we use $s(\lambda)^{-1}$ to mean division by the scalar $s(\lambda)$ on the vector hyperplane $G(\dot{\Lambda}_{\mathcal{A}})$, and $P_\lambda : G(\dot{\Lambda}_{\mathcal{A}}) \to \dot{\Lambda}_{\mathcal{A}}$ is the projection along the direction of $G(\lambda)$. Consequently,

$$\log\det Dg(\lambda) = \log\det P_\lambda - (d-1)\log s(\lambda) + \log\det G$$

We are going to show that each of the three terms on the right hand side is Lipschitz relative to the projective metric. Indeed, $\log\det G$ is constant. By the observation in the first paragraph, $\log s(\lambda)$ is 1-Lipschitz. Finally,

$$\log\det P_\lambda = \log(n_0 \cdot G(\lambda)) - \log(n_1 \cdot G(\lambda))$$

where n_0 and n_1 are unit vectors orthogonal to the hyperplanes $\dot{\Lambda}_{\mathcal{A}}$ and $G(\dot{\Lambda}_{\mathcal{A}})$, respectively. Both n_i have non-negative coefficients: on the one hand, n_0 is collinear to $(1,\ldots,1)$; on the other, n_1 is collinear to $G^*(1,\ldots,1)$, and the adjoint operator G^* has non-negative coefficients since G does. Using the observation in the first paragraph once more, it follows that each $\log(n_i \cdot G(\lambda))$ is a 1-Lipschitz function. Altogether, $\log\det Dg(\lambda)$ is $(d+1)$-Lipschitz relative to the projective metric, as claimed. $\square$

For proving Proposition 4.20, this is all we need to know about projective metrics. In the remainder of the present section we prove a few other properties that will be useful at latter occasions. The first one means that if $g : \Lambda_{\mathcal{A}} \to \Lambda_{\mathcal{A}}$ is a projective map such that $g(\Lambda_{\mathcal{A}})$ has finite $\mathrm{d_p}$-diameter in $\Lambda_{\mathcal{A}}$ then g is a uniform contraction relative to the projective metric:

Proposition 4.23. *For any $\Delta > 0$ there is $\theta < 1$ such that if the diameter of $G(C)$ relative to $\mathrm{d_p}$ is less than Δ then*

$$\mathrm{d_p}(G(u), G(v)) \le \theta\, \mathrm{d_p}(u, v) \quad \textit{for all } u, v \in C.$$

Proof. Given any $u, v \in C$ and $n \ge 1$, consider any $t_n < a(u, v)$ and $s_n > b(u, v)$. By (4.39),

$$z_n = v - t_n\, u \in C \quad \text{and} \quad w_n = s_n\, u - v \in C.$$

Next, consider $T_n < a(z_n, w_n)$ and $S_n > b(z_n, w_n)$. Then, as before,

$$G(s_n u - v) - T_n\, G(v - t_n u) \in C \quad \text{and} \quad S_n\, G(v - t_n u) - G(s_n u - v) \in C.$$

This may be rewritten as

$$(s_n + t_n T_n) G(u) - (1 + T_n) G(v) \in C \quad \text{and} \quad (1 + S_n) G(v) - (s_n + t_n S_n) G(u) \in C.$$

By (4.39), this is the same as

$$b(G(u), G(v)) < \frac{s_n + t_n T_n}{1 + T_n} \quad \text{and} \quad a(G(u), G(v)) > \frac{s_n + t_n S_n}{1 + S_n}.$$

Combining these two inequalities we see that $\mathrm{d_p}(G(u), G(v))$ can not exceed

$$\log\left(\frac{s_n + t_n T_n}{1 + T_n} \cdot \frac{1 + S_n}{s_n + t_n S_n}\right) = \log\left(\frac{s_n/t_n + T_n}{1 + T_n} \cdot \frac{1 + S_n}{s_n/t_n + S_n}\right).$$

The last term can be rewritten as

$$\log\left(\frac{s_n}{t_n} + T_n\right) - \log(1 + T_n) - \log\left(\frac{s_n}{t_n} + S_n\right) + \log(1 + S_n) =$$
$$= \int_0^{\log(s_n/t_n)} \left(\frac{e^x\, dx}{e^x + T_n} - \frac{e^x\, dx}{e^x + S_n}\right),$$

and this is not larger than

$$\sup_{x>0} \frac{e^x (S_n - T_n)}{(e^x + T_n)(e^x + S_n)} \log\left(\frac{s_n}{t_n}\right).$$

Now we use the following elementary fact:

$$\sup_{y>0} \frac{y(S_n - T_n)}{(y+T_n)(y+S_n)} = \frac{1-(T_n/S_n)^{1/2}}{1+(T_n/S_n)^{1/2}} \tag{4.43}$$

(the supremum is attained at $y = (S_n T_n)^{1/2}$). Now, we may choose T_n and S_n arbitrarily close to $a(G(z_n), G(w_n))$ and $b(G(z_n), G(w_n))$, respectively. In particular, since

$$\mathrm{d_p}(G(z_n), G(w_n)) = \log \frac{b(G(z_n), G(w_n))}{a(G(z_n), G(w_n))} < \Delta,$$

we may assume that $\log(S_n/T_n) < \Delta$. Then the right hand side of (4.43)

$$\frac{1-(T_n/S_n)^{1/2}}{1+(T_n/S_n)^{1/2}} < \frac{1-e^{-\Delta/2}}{1+e^{-\Delta/2}} = \tanh\big(\frac{\Delta}{4}\big).$$

Take $\theta = \tanh(\Delta/4)$. It follows that

$$\mathrm{d_p}(G(u), G(v)) < \theta \log\left(\frac{s_n}{t_n}\right).$$

Finally, we may assume that $t_n \to a(u,v)$ and $s_n \to b(u,v)$. Then the last factor converges to $\mathrm{d_p}(u,v)$, and so the conclusion of the proposition follows. □

Remark 4.24. The hypothesis that the $\mathrm{d_p}$-diameter is finite holds if the coefficients of the matrix of G are all positive. Indeed, in that case we may fix $\delta > 0$ such that $\delta \le G_{\alpha,\beta} \le \delta^{-1}$ for all $\alpha, \beta \in \mathcal{A}$. Then,

$$a(G(z), G(w)) = \inf_{\alpha} \frac{\sum_\beta G_{\alpha,\beta}\, w_\beta}{\sum_\beta G_{\alpha,\beta}\, z_\beta} \ge c\delta^2 \quad \text{and} \quad b(G(z), G(w)) \le c\delta^{-2}$$

with $c = \sum_\beta w_\beta / \sum_\beta z_\beta$. Hence, $\mathrm{d_p}(G(z), G(w)) \le -4\log\delta$ for all $z, w \in C$.

Next, we observe that this projective metric is complete:

Proposition 4.25. *Any* $\mathrm{d_p}$-*Cauchy sequence* $(\lambda^n)_n$ *is* $\mathrm{d_p}$-*convergent. Moreover, the normalization* $(\lambda^n/|\lambda^n|)_n$ *is norm-convergent.*

Proof. Let $(\lambda^n)_n$ be a $\mathrm{d_p}$-Cauchy sequence in C: given any $\varepsilon > 0$, there exists $N \ge 1$ such that $\mathrm{d_p}(\lambda^m, \lambda^n) \le \varepsilon$ for all $m, n \ge N$. Up to dropping a finite number of terms, we may suppose that $\mathrm{d_p}(\lambda^m, \lambda^n) \le 1$ for all $m, n \ge 1$. Then,

$$\frac{1}{e} \le \frac{\lambda^m_\alpha \lambda^n_\beta}{\lambda^n_\alpha \lambda^m_\beta} \le e \quad \text{for all } \alpha, \beta \in \mathcal{A} \text{ and } m, n \ge 1. \tag{4.44}$$

As a consequence, writing $R = e \sup\{\lambda^1_\alpha / \lambda^1_\beta : \alpha, \beta \in \mathcal{A}\}$ we get

$$\frac{1}{R} \le \frac{\lambda^n_\alpha}{\lambda^n_\beta} \le R \quad \text{for all } \alpha, \beta \in \mathcal{A} \text{ and } n \ge 1. \tag{4.45}$$

It is no restriction to suppose that $|\lambda^n| = 1$ for all $n \geq 1$. Then

$$\inf_{\alpha\in\mathcal{A}} \lambda_\alpha^n \leq 1 \leq \sup_{\beta\in\mathcal{A}} \lambda_\beta^n \quad \text{and} \quad \inf_{\alpha\in\mathcal{A}} \frac{\lambda_\alpha^n}{\lambda_\alpha^m} \leq 1 \leq \sup_{\beta\in\mathcal{A}} \frac{\lambda_\beta^n}{\lambda_\beta^m} \tag{4.46}$$

for all $m, n \geq 1$. The first part of (4.46) together with (4.45) imply

$$\frac{1}{R} \leq \lambda_\alpha^n \leq R \quad \text{for all } \alpha \in \mathcal{A} \text{ and } n \geq 1. \tag{4.47}$$

The second part of (4.46) together with $\mathrm{d_p}(\lambda^m, \lambda^n) \leq \varepsilon$ give

$$e^{-\varepsilon} \leq \inf_{\alpha\in\mathcal{A}} \frac{\lambda_\alpha^n}{\lambda_\alpha^m} \leq 1 \leq \sup_{\beta\in\mathcal{A}} \frac{\lambda_\beta^n}{\lambda_\beta^m} \leq e^{\varepsilon} \tag{4.48}$$

for all $m, n \geq N$. It follows that

$$\sup_{\alpha\in\mathcal{A}} |\lambda_\alpha^m - \lambda_\alpha^n| \leq \sup_{\alpha\in\mathcal{A}} \lambda^m \cdot \sup_{\alpha\in\mathcal{A}} \left| \frac{\lambda_\alpha^n}{\lambda_\alpha^m} - 1 \right| \leq R(e^{\varepsilon} - 1).$$

This shows that $(\lambda^n)_n$ is a Cauchy sequence with respect to the usual norm in $\mathbb{R}^{\mathcal{A}}$. It follows that the sequence converges to some $\lambda \in \mathbb{R}^{\mathcal{A}}$. Passing to the limit in (4.47) we find that $R^{-1} \leq \lambda_\alpha \leq R$ for all $\alpha \in \mathcal{A}$ and, in particular, $\lambda \in C$. Passing to the limit in (4.48), we get

$$e^{-\varepsilon} \leq \inf_{\alpha\in\mathcal{A}} \frac{\lambda_\alpha^n}{\lambda_\alpha} \leq 1 \leq \sup_{\beta\in\mathcal{A}} \frac{\lambda_\beta^n}{\lambda_\beta} \leq e^{\varepsilon}$$

for all $n \geq N$. This means that $a(\lambda, \lambda^n) \geq e^{-\varepsilon}$ and $b(\lambda, \lambda^n) \leq e^{\varepsilon}$, and so $\mathrm{d_p}(\lambda, \lambda^n) \leq 2\varepsilon$ for all $n \geq N$. Therefore, $(\lambda^n)_n$ is $\mathrm{d_p}$-convergent to λ. □

4.7 Ergodicity theorem

Applying the conclusions in the first half of the previous section to the inverse branches of the map $R_* : \Lambda_* \to \Lambda_*$ introduced in Section 4.5, we can give the

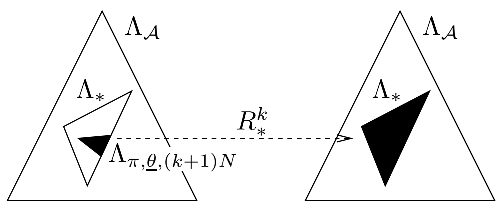


Figure 4.4:

Proof of Proposition 4.20. The domain Λ_* has finite diameter $D_* > 0$ for the projective metric $\mathrm{d_p}$, because it is relatively compact in $\Lambda_{\mathcal{A}}$. By Lemma 4.22, $\log|\det R_*^{-k}|$ is $(d+1)$-Lipschitz continuous relative to $\mathrm{d_p}$, for every inverse branch $R_*^{-k} : \Lambda_* \to \Lambda_{\pi,\underline{\theta},(k+1)N}$ of any iterate R_*^k of the map R_*. See Figure 4.4. Consequently,

$$\log \frac{|\det R_*^{-k}|(x)}{|\det R_*^{-k}|(y)} \le (d+1)D_* \tag{4.49}$$

for any $x, y \in \Lambda_*$ and every inverse branch. Now let $E \subset \Lambda_{\mathcal{A}}$ be any R_*-invariant set with $\nu(E) > 0$. Then E has positive Lebesgue measure as well. Then, for any $\delta > 0$ there exists $k \ge 1$ such that

$$d_1\lambda(\Lambda_{\pi,\underline{\theta},(k+1)N} \setminus E) < \delta d_1\lambda(\Lambda_{\pi,\underline{\theta},(k+1)N}).$$

Taking the images under R_*^k and using (4.49), we find that

$$d_1\lambda(\Lambda_{\mathcal{A}} \setminus E) < \delta e^{(d+1)D_*} d_1\lambda(\Lambda_{\mathcal{A}}).$$

Since δ is arbitrary, we conclude that E has full Lebesgue measure in $\Lambda_{\mathcal{A}}$. It follows that it also has full ν measure in $\Lambda_{\mathcal{A}}$. This proves ergodicity. □

Remark 4.26. Each inverse branch $R_*^{-k} : \Lambda_* \to \Lambda_{\pi,\underline{\theta},(k+1)N}$ is the projectivization of a linear map Θ^{kN*}, and so it extends to a bijection from the whole simplex $\Lambda_{\mathcal{A}}$ to the set $\Lambda_{\pi,\underline{\tau},kN}$ that contains $\Lambda_{\pi,\underline{\theta},(k+1)N}$. Notice that $\Lambda_{\pi,\underline{\tau},kN}$ is contained in Λ_*, which is relatively compact in $\Lambda_{\mathcal{A}}$. Using Proposition 4.23, we get that all these inverse branches contract the projective metric, with contraction rate uniformly bounded from 1. Thus, $R_* : \Lambda_* \to \Lambda_*$ is a uniformly expanding map. Although we do not use this fact, it could be combined with Lemma 4.22 to give an alternative proof that R_* and R admit invariant measures absolutely continuous with respect to Lebesgue measure.

Corollary 4.27. *The Rauzy-Veech renormalization map R is ergodic relative to the invariant measure ν. Moreover, every R-invariant measure absolutely continuous with respect to Lebesgue measure coincides with a multiple of ν.*

Proof. We have seen in Proposition 4.20 that the map R_* induced by R^N is ergodic relative to the restriction of ν to Λ_*. Using Lemma 4.19, we conclude that (R^N, ν) is ergodic. This implies that (R, ν) is ergodic. Uniqueness is a
* consequence of ergodicity. □

Together with Proposition 4.6, Corollary 4.27 completes the proof of Theorem 4.1. From the previous arguments we also get

Corollary 4.28. *The invertible Rauzy-renormalization map $\mathcal{R}$ is ergodic, for the invariant measure m, and the Teichmüller flow $\mathcal{T}$ is ergodic, for the invariant measure $\hat{m}$, restricted to the subset*
* $\{(\pi, \lambda, \tau) : \mathrm{area}(\lambda, \tau) = 1\}$.

4.8 Zorich measure

Here we prove Theorem 4.2. Recall that the invertible Zorich renormalization $\mathcal{Z} : Z_* \to Z_*$ was defined in Section 2.10 as the first return map of the Rauzy-Veech renormalization $\mathcal{R} : \mathcal{H} \to \mathcal{H}$ to the domain $Z_* = Z_0 \cup Z_1$, where

$$Z_0 = \{(\pi, \lambda, \tau) \in \mathcal{H} : \lambda_{\alpha(0)} > \lambda_{\alpha(1)} \quad \text{and} \quad \sum_{\alpha \in \mathcal{A}} \tau_\alpha > 0\}$$

and

$$Z_1 = \{(\pi, \lambda, \tau) \in \mathcal{H} : \lambda_{\alpha(0)} < \lambda_{\alpha(1)} \quad \text{and} \quad \sum_{\alpha \in \mathcal{A}} \tau_\alpha < 0\}.$$

It follows from the definition (and Lemma 4.17) that $\mathcal{Z}$ preserves the restriction of the measure m to Z_*. Moreover, $\mathcal{Z}$ preserves the restriction of m to the domain $\{\text{area}\,(\lambda, \tau) \le c\} \cap Z_*$, for any $c > 0$. In this regard, observe that $\mathcal{Z}$ preserves the area (2.30), since $\mathcal{R}$ does.

Also by construction, $P \circ \mathcal{Z} = Z \circ P$, where $P : \mathcal{H} \to C \times \Lambda_{\mathcal{A}}$ denotes the canonical projection and Z is the Zorich renormalization map introduced in Section 1.8. Therefore, Z preserves the measure

$$\mu = P_*\big(m \mid Z_* \cap \{\text{area}\,(\lambda, \tau) \le 1\}\big).$$

Arguing in just the same way as in Section 4.1, we see that the measure μ is absolutely continuous with respect to Lebesgue measure, with density

$$\frac{d\mu}{d_1\lambda}(\pi, \lambda) = \text{vol}\,(\{\tau \in T_\pi^\varepsilon : \text{area}\,(\lambda, \tau) \le 1\}) = \frac{1}{d!} \sum_{i=1}^{k} \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^{i,\beta}}, \qquad (4.50)$$

where ε is the type of (π, λ). Here the notations are as before:

$$T_\pi^0 = \{\tau \in T_\pi^+ : \sum_{\alpha \in \mathcal{A}} \tau_\alpha > 0\} \quad \text{and} \quad T_\pi^1 = \{\tau \in T_\pi^+ : \sum_{\alpha \in \mathcal{A}} \tau_\alpha < 0\},$$

$T^1, \ldots, T^k$ are pairwise disjoint simplicial cones covering the polyhedral cone T_π^ε up to a positive codimension subset, $(\tau^{i,\beta})_{\beta \in \mathcal{A}}$ is a basis of generators of each T^i, and $h^{i,\beta} = -\Omega_\pi(\tau^{i,\beta})$ for each i and β.

The relation (4.50) shows that the density of the absolutely continuous Z-invariant measure μ is given by a rational function with degree $-d$ and bounded from zero. The next step is to show that this measure μ is finite.

Example 4.29. Let us give an explicit expression for the density of μ when π is the pair defined in (4.5). We consider first the case when $(\pi, \lambda) \in C \times \Lambda_{\mathcal{A}}$ has type 0. We have seen in Section 4.2 that T_π^0 is a simplicial cone, and admits $\tau^\alpha = (\tau^\alpha_\beta)_{\beta \in \mathcal{A}}$ defined by

$$\tau^\alpha_\beta = \begin{cases} 1 & \text{if } \beta = \alpha \\ -1 & \text{if } \beta = \alpha_0^+ \\ 0 & \text{in all other cases} \end{cases} \qquad \text{if } \alpha \ne \alpha(0)$$

$$\tau^\alpha_\beta = \begin{cases} 1 & \text{if } \pi_0(\beta) = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{if } \alpha = \alpha(0).$$

as a volume 1 basis of generators. Then $h^\alpha = -\Omega_\pi(\tau^\alpha)$ is given by

$$h^\alpha_\beta = \begin{cases} 1 & \text{if } \beta = \alpha \text{ or } \beta = \alpha_0^+ \\ 0 & \text{otherwise} \end{cases} \quad \text{if } \alpha \neq \alpha(0)$$

$$h^\alpha_\beta = \begin{cases} 0 & \text{if } \pi_0(\beta) = 1 \text{ or } \beta = \alpha(1) \\ 1 & \text{otherwise} \end{cases} \quad \text{if } \alpha = \alpha(0).$$

It follows that

$$\frac{d\mu}{d_1\lambda}(\pi,\lambda) = \frac{1}{d!} \prod_{\alpha \neq \alpha(0)} \left(\frac{1}{\lambda_\alpha + \lambda_{\alpha_0^+}} \right) \cdot \frac{1}{\sum_{\beta \neq \alpha(1)} \lambda_\beta}.$$

The case when (π, λ) has type 1 is analogous, and one gets

$$\frac{d\mu}{d_1\lambda}(\pi,\lambda) = \frac{1}{d!} \prod_{\alpha \neq \alpha(1)} \left(\frac{1}{\lambda_\alpha + \lambda_{\alpha_1^+}} \right) \cdot \frac{1}{\sum_{\beta \neq \alpha(0)} \lambda_\beta}.$$

In particular, for $d = 2$ this gives

$$\frac{d\mu}{d_1\lambda}(\lambda) = \begin{cases} 1/(2\lambda_B) & \text{if } \lambda_A < \lambda_B \\ 1/(2\lambda_A) & \text{if } \lambda_B < \lambda_A. \end{cases}$$

Notice that the density is bounded on $\Lambda_\mathcal{A}$, and so the measure μ is finite. While boundedness is specific to the case $d = 2$, finiteness holds in general, as we are going to see.

Proposition 4.30. *The measure $\mu(C \times \Lambda_\mathcal{A})$ is finite.*

Proof. Given $\pi \in C$, let Λ_ε be the subset of $\lambda \in \mathbb{R}_+^\mathcal{A}$ such that $\lambda_{\alpha(\varepsilon)} > \lambda_{\alpha(1-\varepsilon)}$. Then

$$\mu(C \times \Lambda_\mathcal{A}) = \sum_{\pi \in C} \sum_{\varepsilon = 0,1} \int_{\Lambda_\varepsilon} \operatorname{vol}\left(\{\tau \in T_\pi^\varepsilon : \operatorname{area}(\lambda, \tau) \leq 1\}\right) d_1\lambda.$$

Using (4.50) we deduce that

$$\mu(C \times \Lambda_\mathcal{A}) = \sum_{\pi \in C} \sum_{\varepsilon = 0,1} \int_{\Lambda_\varepsilon} \sum_{i=1}^{k} \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^{i,\beta}} \, d_1\lambda. \tag{4.51}$$

To prove the proposition we only have to show that the integral is finite, for every fixed $\pi \in C$, $\varepsilon \in \{0, 1\}$, and $i = 1, \dots, k$. Let us consider $\varepsilon = 0$; the case $\varepsilon = 1$ is analogous. Then the basis of generators $(\tau^{i,\beta})_{\beta \in \mathcal{A}}$ is contained in the closure of T_π^0. For simplicity, we write $h^\beta = h^{i,\beta}$ in what follows.

Let $\mathcal{N}$ denote the set of integer vectors $n = (n_\alpha)_{\alpha \in \mathcal{A}}$ such that $n_\alpha \geq 0$ for all $\alpha \in \mathcal{A}$, and the n_α are not all zero. As in (4.27), define

$$\Lambda(n) = \{\lambda \in \Lambda_\mathcal{A} : 2^{-n_\alpha} \leq \lambda_\alpha d < 2^{-n_\alpha + 1} \text{ for every } \alpha \in \mathcal{A}\},$$

except that for $n_\alpha = 0$ the second inequality is omitted. By Lemma 4.14, the $\Lambda(n)$ cover $\Lambda_{\mathcal{A}}$ and satisfy

$$c_1 2^{-\sum_{\mathcal{A}} n_\alpha} \le \operatorname{vol}_{d-1} \Lambda(n) \le c_1^{-1} 2^{-\sum_{\mathcal{A}} n_\alpha} \tag{4.52}$$

for some $c_1 > 0$. In what follows we consider $n_{\alpha(0)} \le n_{\alpha(1)}$, for the corresponding $\Lambda(n)$ suffice to cover $\Lambda_0 = \{\lambda_{\alpha(0)} > \lambda_{\alpha(1)}\}$. For each $\beta \in \mathcal{A}$, let $\mathcal{A}(\beta)$ be the subset of $\alpha \in \mathcal{A}$ such that $h_\alpha^\beta > 0$. Let $c_2 > 0$ be the minimum of the non-zero h_α^β, over all α and β. Then

$$\lambda \cdot h^\beta = \sum_{\mathcal{A}(\beta)} h_\alpha^\beta \lambda_\alpha \ge \sum_{\mathcal{A}(\beta)} c_2 d^{-1} 2^{-n_\alpha} \ge c_2 d^{-1} 2^{-\min_{\mathcal{A}(\beta)} n_\alpha} \tag{4.53}$$

for every $\beta \in \mathcal{A}$. Using (4.52) we deduce that

$$\int_{\Lambda(n)} \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^\beta} \, d_1\lambda \le K \, 2^{\sum_{\beta \in \mathcal{A}} \min_{\mathcal{A}(\beta)} n_\alpha - \sum_{\alpha \in \mathcal{A}} n_\alpha}, \tag{4.54}$$

where the constant $K = (2/c_1)(d/c_2)^d$.

Lemma 4.31. *Assuming $n_{\alpha(0)} \le n_{\alpha(1)}$, we have*

$$-\sum_{\beta \in \mathcal{A}} \min_{\alpha \in \mathcal{A}(\beta)} n_\alpha + \sum_{\alpha \in \mathcal{A}} n_\alpha \ge \max_{\alpha \in \mathcal{A}} n_\alpha.$$

Proof. Let $0 = n^0 < n^1 < \cdots$ be the different values taken by n_α, and $\mathcal{B}^i$, $i \ge 0$ be the set of values of $\alpha \in \mathcal{A}$ such that $n_\alpha \ge n^i$. On the one hand,

$$\sum_{\alpha \in \mathcal{A}} n_\alpha = \sum_{i \ge 1} n_i \big(\#\mathcal{B}^i - \#\mathcal{B}^{i+1}\big) = \sum_{i \ge 1} \#\mathcal{B}^i (n^i - n^{i-1}). \tag{4.55}$$

On the other hand, $\min_{\mathcal{A}(\beta)} n_\alpha \ge n_i$ if and only if $\mathcal{A}(\beta) \subset \mathcal{B}^i$. Consequently,

$$\begin{aligned} \sum_{\beta \in \mathcal{A}} \min_{\mathcal{A}(\beta)} n_\alpha &= \sum_{i \ge 1} n_i \big(\#\{\beta : \mathcal{A}(\beta) \subset \mathcal{B}^i\} - \#\{\beta : \mathcal{A}(\beta) \subset \mathcal{B}^{i+1}\}\big) \\ &= \sum_{i \ge 1} \#\{\beta : \mathcal{A}(\beta) \subset \mathcal{B}^i\}(n^i - n^{i-1}). \end{aligned} \tag{4.56}$$

Observe that $\mathcal{A}(\beta) \subset \mathcal{B}^i$ if and only if $h_\alpha^\beta = 0$ for all $\alpha \in \mathcal{A} \setminus \mathcal{B}^i$. Observe also that the assumption $n_{\alpha(0)} \le n_{\alpha(1)}$ means that if $\alpha(1) \in \mathcal{B}^i$ then $\alpha(0) \in \mathcal{B}^i$. Using Proposition 4.11 (with $\mathcal{B} = \mathcal{A} \setminus \mathcal{B}^i$), we obtain

$$\#\{\beta : \mathcal{A}(\beta) \subset \mathcal{B}^i\} < \#\mathcal{B}^i. \tag{4.57}$$

Putting (4.55)-(4.57) together, we find that

$$-\sum_{\beta \in \mathcal{A}} \min_{\mathcal{A}(\beta)} n_\alpha + \sum_{\alpha \in \mathcal{A}} n_\alpha \ge \sum_{i \ge 1}^{k} (n^i - n^{i-1}) = \max_{\alpha \in \mathcal{A}} n_\alpha.$$

This proves the lemma. □

Replacing the conclusion of the lemma in (4.54) we obtain, for every $n \in \mathcal{N}$,

$$\int_{\Lambda(n)} \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^\beta} \, d_1\lambda \leq K 2^{-\max_{\mathcal{A}} n_\alpha}. \tag{4.58}$$

For each $m \geq 0$ there are at most $(m+1)^d$ vectors $n \in \mathcal{N}$ with $\max_{\mathcal{A}} n_\alpha = m$. So, (4.58) implies that the integral in (4.51) is bounded above by

$$\sum_{m=0}^{\infty} K(m+1)^d 2^{-m} < \infty$$

for every $\pi \in C$ and $1 \leq i \leq k$. The proof of Proposition 4.30 is complete. $\square$

To finish the proof of Theorem 4.2 we only have to observe that the system (Z, μ) is ergodic. This can be shown in the same way we proved, in Corollary 4.27, that (R, ν) is ergodic. We just outline the arguments. As noted before, (Z, μ) is a Markov system. Since μ is invariant and finite, (Z, μ) is a recurrent system. Consider any relatively compact subsimplex $\{\pi\} \times \Lambda_*$ which is mapped to a whole $\{\pi_0\} \times \Lambda_{\mathcal{A}}$ by some iterate Z^N. The map induced by Z on Λ_* has a bounded distortion property as in Lemma 4.22. For the same reason as in Proposition 4.20, that implies the induced map is ergodic relative to μ restricted to Λ_*. It follows, using Lemma 4.19, that Z itself is ergodic relative to μ. This proves the claim.

Notes

Theorem 4.1 is due to Masur [41] and Veech [54], independently. Our presentation is closer to the latter paper, but several steps have been simplified with the help of Marmi, Moussa, Yoccoz [40, 61]. For alternative proofs, see Rees [48], Kerckhoff [29], and Boshernitzan [8]. Theorem 4.2 is due to Zorich [63].

Chapter 5

Lyapunov Exponents

In this chapter we study the dynamical properties of Teichmüller flows and renormalization operators in greater depth, and use the conclusions to analyze the quantitative behavior of geodesics on typical translation surfaces. We have already seen in Section 5.7 that almost every Abelian differential α admits an asymptotic cycle $c_1 \in H_1(M, \mathbb{R})$, such that

$$\frac{1}{l}[\gamma(p,l)] \to c_1 \quad \text{uniformly as the length } l \to \infty,$$

where $[\gamma(p,l)] \in H_1(M, \mathbb{R})$ is the homology class defined by the vertical geodesic segment of length l starting at any point p. Here we obtain a much more precise description of the asymptotic behavior of long geodesic segments, also for almost all Abelian differentials.

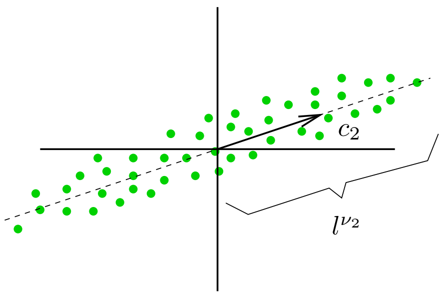


Figure 5.1:

Indeed, we are going to see that the component of $[\gamma(p,l)]$ orthogonal to the line $L_1 = \mathbb{R}\, c_1$ is asymptotic to some $c_2 \in H_1(M, \mathbb{R})$ and its norm

$$\operatorname{dist}([\gamma(p,l)], L_1) \lesssim l^{\nu_2}$$

(meaning ν_2 is the smallest exponent ν such that the left hand side is less than l^{ν} for every large l) for some constant $\nu_2 < 1$. Figure 7.1 illustrates possible values of this orthogonal component for different values of l. More generally, for any $j = 2, \ldots, g$, the component of $[\gamma(p,l)]$ orthogonal to the subspace $L_{j-1} = \mathbb{R}\, c_1 \oplus \cdots \oplus \mathbb{R}\, c_{j-1}$ is asymptotic to some $c_j \in H_1(M,\mathbb{R})$ and its norm

$$\operatorname{dist}([\gamma(p,l)], L_{j-1}) \lesssim l^{\nu_j}$$

for some constant $\nu_j < \nu_{j-1}$. Finally, the component of $[\gamma(p,l)]$ orthogonal to $L_g = \mathbb{R}\, c_1 \oplus \cdots \oplus \mathbb{R}\, c_g$ is uniformly bounded in norm. Let us state these facts more precisely:

Theorem 5.1. *For any connected component $\mathcal{C}$ of a stratum $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ there exist numbers $1 > \nu_2 > \cdots > \nu_g > 0$ and for almost every Abelian differential $\alpha \in \mathcal{C}$ there exist subspaces $L_1 \subset L_2 \subset \cdots \subset L_g$ of the homology $H_1(M,\mathbb{R})$ such that $\dim L_i = i$ for $i = 1, 2, \ldots, g$ and*

1. *for all $i < g$ the deviation of $[\gamma(p,l)]$ from L_i has amplitude $l^{\nu_{i+1}}$:*

$$\limsup_{l\to\infty} \frac{\log \operatorname{dist}([\gamma(p,l)], L_i)}{\log l} = \nu_{i+1} \quad \textit{uniformly in } p \in M$$

2. *the deviation $\operatorname{dist}([\gamma(p,l)], L_g)$ of $[\gamma(p,l)]$ from L_g is uniformly bounded (the bound depends only on α and the choice of the norm).*

This remarkable statement was discovered by Zorich in the early nineties, from computer calculations of $[\gamma(p,l)]$ for various translation surfaces. An explanation was provided by Zorich and Kontsevich, in terms of the Lyapunov spectrum of the Teichmüller flow in the connected component $\mathcal{C}$. As we observed in Chapter 4, the natural volume on each connected component of the strata of the moduli space $\mathcal{A}_g$ is invariant and ergodic under the Teichmüller flow, restricted to any hypersurface of constant area. Then, we may use the Oseledets theorem (Oseledets [46] and Section 7.1) to conclude that the Teichmüller flow has a well-defined Lyapunov spectrum with respect to this measure. It is not difficult to show (see Section 7.4) that this spectrum has the form

$$\begin{gathered} 2 \geq 1+\nu_2 \geq \cdots \geq 1+\nu_g \geq 1 = \cdots = 1 \geq 1-\nu_g \geq \cdots \geq 1-\nu_2 \geq 0 \geq \\ -1+\nu_2 \geq \cdots \geq -1+\nu_g \geq -1 = \cdots = -1 \geq -1-\nu_g \geq \cdots \geq -1-\nu_2 \geq -2, \end{gathered}$$

where the so-called *trivial exponents* ± 1 have multiplicity $\kappa - 1$. It was observed by Veech [56, 58] that the Teichmüller flow is non-uniformly hyperbolic, which amounts to saying that $\nu_2 < 1$. A short proof is given in Sections 7.3 and 7.4. Zorich and Kontsevich [32, 62, 64, 65] conjectured that all the inequalities in the previous formula are strict, and proved that Theorem 7.1 would follow from this conjecture. The argument is presented in Sections 7.5 and 7.6.

In this direction, Forni [14] proved that one always has $\nu_g > 0$. This result implies the Zorich-Kontsevich conjecture in genus 2 and has also been used to obtain other properties of geodesic flows on translation surfaces, like the weak mixing theorem of Avila, Forni [3]. The full statement of the Zorich-Kontsevich conjecture was proved by Avila, Viana [4]:

Theorem 5.2. *For each connected component $\mathcal{C}$ of any stratum $\mathcal{A}_g(m_1,\ldots,m_\kappa)$ the non-trivial Lyapunov exponents of the Teichmüller flow are all distinct:*

$$\begin{aligned}2 >&1+\nu_2 > \cdots > 1+\nu_g > 1-\nu_g > \cdots > 1-\nu_2 > \\ &> -1+\nu_2 > \cdots > -1+\nu_g > -1-\nu_g > \cdots > -1-\nu_2 > -2.\end{aligned}$$

The connection between the Teichmüller flow on the connected component $\mathcal{C}$ and the geodesic flow of typical Abelian differentials $\alpha \in \mathcal{C}$ is made through another object, the Zorich linear cocycle, that we introduce and analyze in Sections 7.2 and 7.3. We make the connection precise in Sections 7.4 through 7.6, where we also use it to prove Theorem 7.1 from Theorem 7.2. In the rest of the chapter we describe the main ingredients in the proof of Theorem 7.2. There are two parts. In Theorem 7.50 (Sections 7.7 and 7.8) we give general sufficient conditions for simplicity of the Lyapunov spectrum. In Theorem 7.64 (Sections 7.9 and 7.10) we check that these conditions are fulfilled by the Zorich cocycles.

5.1 Oseledets theorem

We begin by recalling some basic facts and terminology relative to linear cocycles and the multiplicative ergodic theorem of Oseledets [46].

Cocycles over maps. Let μ be a probability measure on some space M and $f: M \to M$ be a measurable transformation that preserves μ. Let $\pi: \mathcal{E} \to M$ be a finite-dimensional vector bundle endowed with a Riemannian norm $\|\cdot\|$. A *linear cocycle* (or vector bundle morphism) over f is a map $F: \mathcal{E} \to \mathcal{E}$ such that

$$\pi \circ F = f \circ \pi$$

and the action $A(x): \mathcal{E}_x \to \mathcal{E}_{f(x)}$ of F on each fiber is a linear isomorphism [1]. The action of the nth iterate is given by $A^n(x) = A(f^{n-1}(x)) \cdots A(f(x)) \cdot A(x)$, for every $n \geq 1$. Given any $y > 0$, we denote $\log^+ y = \max\{\log y, 0\}$.

Theorem 5.3. *Assume the function $\log^+ \|A(x)\|$ is μ-integrable. Then, for μ-almost every $x \in M$, there exists $k = k(x)$, numbers $\lambda_1(x) > \cdots > \lambda_k(x)$, and a filtration $\mathcal{E}_x = F_x^1 > \cdots > F_x^k > \{0\} = F_x^{k+1}$ of the fiber, such that*

1. *$k(f(x)) = k(x)$ and $\lambda_i(f(x)) = \lambda_i(x)$ and $A(x) \cdot F_x^i = F_{f(x)}^i$ and*
2. *$\displaystyle\lim_{n\to+\infty} \frac{1}{n} \log \|A^n(x)v\| = \lambda_i(x)$ for all $v \in F_x^i \setminus F_x^{i+1}$ and all $i = 1, \ldots, k$.*

The *Lyapunov exponents* λ_i and the subspaces F^i depend in a measurable (but usually not continuous) fashion on the base point. The statement of the

[1] It is often possible to assume that the vector bundle is trivial, meaning that $\mathcal{E} = M \times \mathbb{R}^d$, restricting to some full μ-measure subset of M if necessary. Then $A(\cdot)$ takes values in the linear group $\mathrm{GL}(d,\mathbb{R})$ of invertible $d \times d$ matrices.

theorem, including the values of $k(x)$, the $\lambda_i(x)$, and the $F^i(x)$, is not affected if one replaces $\|\cdot\|$ by any other Riemann norm $|||\cdot|||$ equivalent to it in the sense that there exists some μ-integrable function $c(\cdot)$ such that

$$e^{-c(x)}\|v\| \leq |||v||| \leq e^{c(x)}\|v\| \quad \text{for all } v \in T_x M. \tag{5.1}$$

When the measure μ is ergodic, the values of $k(x)$ and of each of the $\lambda_i(x)$ are constant on a full measure subset, and so are the dimensions of the subspaces F_x^i. We call $\dim F_x^i - \dim F_x^{i+1}$ the *multiplicity* of the corresponding Lyapunov exponent $\lambda_i(x)$. The *Lyapunov spectrum* of F is the set of all Lyapunov exponents, each counted with multiplicity. The Lyapunov spectrum is *simple* if all Lyapunov exponents have multiplicity 1.

The invertible case. If the transformation f is invertible then so is the cocycle F. Applying Theorem 7.3 also to the inverse F^{-1} and combining the invariant filtrations of the two cocycles, one gets a stronger conclusion than in the general non-invertible case:

Theorem 5.4. *Let $f : M \to M$ be invertible and both functions $\log^+\|A(x)\|$ and $\log^+\|A^{-1}(x)\|$ be μ-integrable. Then, for μ-almost every point $x \in M$, there exists $k = k(x)$, numbers $\lambda_1(x) > \cdots > \lambda_k(x)$, and a decomposition $\mathcal{E}_x = E_x^1 \oplus \cdots \oplus E_x^k$ of the fiber, such that*

1. *$A(x) \cdot E_x^i = E_{f(x)}^i$ and $F_x^i = \oplus_{j=i}^k E_x^j$ and*

2. *$\displaystyle\lim_{n\to\pm\infty} \frac{1}{n} \log \|A^n(x)v\| = \lambda_i(x)$ for all non-zero $v \in E_x^i$ and*

3. *$\displaystyle\lim_{n\to\pm\infty} \frac{1}{n} \log \angle\Big(\bigoplus_{i\in I} E_{f^n(x)}^i, \bigoplus_{j\in J} E_{f^n(x)}^j\Big) = 0$ for all I and J with $I \cap J = \emptyset$.*

Note that the multiplicity of each Lyapunov exponent λ_i coincides with the dimension $\dim E_x^i = \dim F_x^i - \dim F_x^{i+1}$ of the associated *Oseledets subspace* E_x^i. From the conclusion of the theorem one easily gets that

$$\lim_{n\to\pm\infty} \frac{1}{n} \log |\det A^n(x)| = \sum_{i=1}^k \lambda_i(x) \dim E_x^i. \tag{5.2}$$

In most cases we deal with, the determinant is constant equal to 1. Then the sum of all Lyapunov exponents, counted with multiplicity, is identically zero.

Remark 5.5. Any sum $F_x^i = \oplus_{j=i}^k E_x^j$ of Oseledets subspaces corresponding to the smallest Lyapunov exponents depends only on the forward iterates of the cocycle. Analogously, any sum of Oseledets subspaces corresponding to the largest Lyapunov exponents depends only on the backward iterates.

Remark 5.6. The *natural extension* of a (non-invertible) map $f : M \to M$ is defined on the space $\hat{M}$ of sequences $(x_n)_{n\leq 0}$ with $f(x_n) = x_{n+1}$ for $n < 0$, by

$$\hat{f} : \hat{M} \to \hat{M}, \quad (\ldots, x_n, \ldots, x_0) \mapsto (\ldots, x_n, \ldots, x_0, f(x_0)).$$

Let $P : \hat{M} \to M$ be the canonical projection assigning to each sequence $(x_n)_{n\le 0}$ the term x_0. It is clear that $\hat{f}$ is invertible and $P \circ \hat{f} = f \circ P$. Every f-invariant probability μ lifts to a unique $\hat{f}$-invariant probability $\hat{\mu}$ such that $P_*\hat{\mu} = \mu$. Every cocycle $F : \mathcal{E} \to \mathcal{E}$ over f extends to a cocycle $\hat{F} : \hat{\mathcal{E}} \to \hat{\mathcal{E}}$ over $\hat{f}$, as follows: $\hat{\mathcal{E}}_{\hat{x}} = \mathcal{E}_{P(\hat{x})}$ and $\hat{A}(\hat{x}) = A(P(\hat{x}))$, where $\hat{A}(\hat{x})$ denotes the action of $\hat{F}$ on the fiber $\hat{\mathcal{E}}_{\hat{x}}$. Clearly, $\int \log^+ \|\hat{A}\|\, d\hat{\mu} = \int \log^+ \|A\|\, d\mu$ and, assuming the integrals are finite, the two cocycles F and $\hat{F}$ have the same Lyapunov spectrum and the same Oseledets filtration. Moreover, $\int \log^+ \|\hat{A}^{-1}\|\, d\hat{\mu} = \int \log^+ \|A^{-1}\|\, d\mu$ and when the integrals are finite we may apply Theorem 7.4 to the cocycle $\hat{F}$.

Symplectic cocycles. Suppose there exists some symplectic form, that is, some non-degenerate alternate 2-form ω_x on each fiber $\mathcal{E}_x$, which is preserved by the linear cocycle F:

$$\omega_{f(x)}(A(x)u, A(x)v) = \omega_x(u,v) \quad \text{for all } x \in M \text{ and } u, v \in \mathcal{E}_x.$$

Assume the symplectic form is integrable, in the sense that there exists a μ-integrable function $x \mapsto c(x)$ such that

$$|\omega_x(u,v)| \le e^{c(x)}\|u\|\,\|v\| \quad \text{for all } x \in M \text{ and } u, v \in \mathcal{E}_x.$$

Remark 5.7. We are going to use the following easy observation. Let μ be an invariant ergodic probability for a transformation $f : M \to M$, and let $\phi : M \to \mathbb{R}$ be a μ-integrable function. Then

$$\lim_{n\to\infty} \frac{1}{n}\phi(f^n(x)) = 0 \quad \mu\text{-almost everywhere.}$$

This follows from the Birkhoff ergodic theorem applied to $\psi(x) = \phi(f(x)) - \phi(x)$. Note that the argument remains valid under the weaker hypothesis that the function ψ be integrable.

Proposition 5.8. *If F preserves an integrable symplectic form then its Lyapunov spectrum is symmetric: if λ is a Lyapunov exponent at some point x then so is $-\lambda$, with the same multiplicity.*

This statement can be justified as follows. Consider any i and j such that $\lambda_i(x) + \lambda_j(x) \ne 0$. For all $v^i \in E^i_x$ and $v^j \in E^j_x$,

$$|\omega_x(v^i, v^j)| = |\omega_{f^n(x)}(A^n(x)v^i, A^n(x)v^j)| \le e^{c(f^n(x))}\|A^n(x)v^i\|\|A^n(x)v^j\|$$

for all $n \in \mathbb{Z}$. Since $c(x)$ is integrable the first factor has no exponential growth: by Remark 7.7,

$$\lim_{n\to\pm\infty} \frac{1}{n} c(f^n(x)) = 0 \quad \text{almost everywhere.}$$

The assumption implies that $\|A^n(x)v^i\|\|A^n(x)v^j\|$ goes to zero exponentially fast, either when $n \to +\infty$ or when $n \to -\infty$. So, the right hand side of the previous inequality goes to zero either when $n \to +\infty$ or when $n \to -\infty$. Therefore, in either case, the left hand side must vanish. This proves that

$$\lambda_i(x) + \lambda_j(x) \neq 0 \quad \Rightarrow \quad \omega_x(v^i, v^j) = 0 \ \text{ for all } v^i \in E_x^i \text{ and } v^j \in E_x^j.$$

Since the symplectic form is non-degenerate, it follows that for every i there exists j such that $\lambda_i(x) + \lambda_j(x) = 0$. We are left to check that the multiplicities of such symmetric exponents coincide. We may suppose $\lambda_i(x) \neq 0$, of course. Let s be the dimension of E_x^i. Using a Gram-Schmidt argument, one constructs a basis $v_1^i, \ldots, v_s^i$ of E_x^i and a family of vectors $v_1^j, \ldots, v_s^j$ in E_x^j such that

$$\omega_x(v_p^i, v_q^j) = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

Notice that $\omega_x(v_p^i, v_q^i) = 0 = \omega_x(v_p^j, v_q^j)$ for all p and q, since $\lambda_i(x) = -\lambda_j(x)$ is non-zero. The relations (7.3) imply that the $v_1^j, \ldots, v_s^j$ are linearly independent, and so $\dim E_x^j \geq \dim E_x^i$. The converse inequality is proved in the same way.

Adjoint linear cocycle. Let $\pi^* : \mathcal{E}^* \to M$ be another vector bundle which is *dual* to $\pi : \mathcal{E} \to M$, in the sense that there exists a nondegenerate bilinear form

$$\mathcal{E}_x^* \times \mathcal{E}_x \ni (u, v) \mapsto u \cdot v \in \mathbb{R}, \quad \text{for each } x \in M.$$

The *annihilator* of a subspace $E^* \subset \mathcal{E}_x^*$ is the subspace $E \subset \mathcal{E}_x$ of all $v \in \mathcal{E}_x$ such that $u \cdot v = 0$ for all $u \in E^*$. We also say that E^* is the annihilator of E. Notice that $\dim E + \dim E^* = \dim \mathcal{E}_x = \dim \mathcal{E}_x^*$. The norm $\|\cdot\|$ may be transported from $\mathcal{E}$ to $\mathcal{E}^*$ through the duality:

$$\|u\| = \sup\{|u \cdot v| : v \in \mathcal{E}_x \text{ with } \|v\| = 1\} \quad \text{for } u \in \mathcal{E}_x^*. \tag{5.4}$$

For $x \in M$, the *adjoint* of $A(x)$ is the linear map $A^*(x) : \mathcal{E}_{f(x)}^* \to \mathcal{E}_x^*$ defined by

$$A^*(x)u \cdot v = u \cdot A(x)v \quad \text{for every } u \in \mathcal{E}_{f(x)}^* \text{ and } v \in \mathcal{E}_x. \tag{5.5}$$

The inverses $A^{-1*}(x) : \mathcal{E}_x^* \mapsto \mathcal{E}_{f(x)}^*$ define a linear cocycle $F^{-1*} : \mathcal{E}^* \to \mathcal{E}^*$ over f.

Proposition 5.9. *The Lyapunov spectra of F and F^{-1*} are symmetric to one another at each point.*

Indeed, the definitions (7.4) and (7.5) imply $\|A^*(x)\| = \|A(x)\|$ and, analogously, $\|A^{-1*}(x)\| = \|A^{-1}(x)\|$ for any $x \in M$. Thus, F^{-1*} satisfies the integrability condition in Theorem 7.4 if and only if F does. Let $\mathcal{E}_x = \oplus_{j=1}^k E_x^j$ be the Oseledets decomposition of F at each point x. For each $i = 1, \ldots, d$ define

$$E_x^{i*} = \text{ annihilator of } E_x^1 \oplus \cdots \oplus E_x^{i-1} \oplus E_x^{i+1} \oplus \cdots \oplus E_x^k. \tag{5.6}$$

The decomposition $\mathcal{E}_x^* = \oplus_{j=1}^k E_x^{j*}$ is invariant under F^{-1*}. Moreover, given any $u \in E_x^{i*}$ and any $n \geq 1$,

$$\|A^{-n*}(x)u\| = \max_{\|v\|=1} |A^{-n*}(x)u \cdot v| = \max_{\|v\|=1} |u \cdot A^{-n}(x)v|.$$

Fix any $\varepsilon > 0$. Begin by considering $v \in E^i_{f^n(x)}$. Then $A^{-n}(x)v \in E^i_x$, and so

$$|u \cdot A^{-n}(x)v| \geq c\,\|u\|\,\|A^{-n}(x)v\| \geq c\,\|u\|\,e^{-(\lambda_i(x)+\varepsilon)n}$$

for every n sufficiently large, where $c = c(E^i_x, E^{i*}_x) > 0$. Consequently,

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^{-n*}(x)u\| \geq -(\lambda_i(x) + \varepsilon). \tag{5.7}$$

Next, observe that a general unit vector $v \in \mathcal{E}_{f^n(x)}$ may be written

$$v = \sum_{j=1}^{k} v^j \quad \text{with } v^j \in E^j_{f^n(x)}.$$

Using part 3 of Theorem 7.4, we see that every $\|v^j\| \leq e^{\varepsilon n}$ if n is sufficiently large. Therefore, given any $u \in E^{i*}_x$,

$$|u \cdot A^{-n}(x)v| = |u \cdot A^{-n}(x)v^i| \leq \|u\|\,e^{-(\lambda_i(x)-\varepsilon)n}\,\|v^i\| \leq \|u\|\,e^{-(\lambda_i(x)-2\varepsilon)n}$$

for every unit vector $v \in \mathcal{E}_{f^n(x)}$, and so

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^{-n*}(x)u\| \leq -(\lambda_i(x) - 2\varepsilon). \tag{5.8}$$

Since $\varepsilon > 0$ is arbitrary, the relations (7.7) and (7.8) show that the Lyapunov exponent of F^{-1*} along E^{i*}_x is precisely $-\lambda_i(x)$, for every $i = 1, \ldots, k$. Thus, $\mathcal{E}^*_x = \oplus_{j=1}^k E^{j*}_x$ must be the Oseledets decomposition of F^{-1*} at x. Observe, in addition, that $\dim E^{i*}_x = \dim E^i_x$ for all $i = 1, \ldots, k$.

Cocycles over flows. We call *linear cocycle over a flow* $f^t : M \to M$, $t \in \mathbb{R}$ a flow extension $F^t : \mathcal{E} \to \mathcal{E}$, $\quad t \in \mathbb{R}$ such that $\pi \circ F^t = f^t \circ \pi$ and the action $A^t(x) : \mathcal{E}_x \to \mathcal{E}_{f^t(x)}$ of F^t on every fiber is a linear isomorphism. Notice that $A^{t+s}(x) = A^s(f^t(x)) \cdot A^t(x)$ for all $t, s \in \mathbb{R}$.

Theorem 5.10. *Assume* $\log^+ \|A^t(x)\|$ *is* μ*-integrable for all* $t \in \mathbb{R}$. *Then, for* μ*-almost every* $x \in M$, *there exists* $k = k(x) \leq d$, *numbers* $\lambda_1(x) > \cdots > \lambda_k(x)$, *and a decomposition* $\mathcal{E}_x = E^0_x \oplus E^1_x \oplus \cdots \oplus E^k_x$ *of the fiber, such that*

1. $A^t(x) \cdot E^i_x = E^i_{f^t(x)}$ *and* E^0_x *is tangent to the flow lines*
2. $\displaystyle\lim_{t\to\pm\infty} \frac{1}{t} \log \|A^t(x)v\| = \lambda_i(x)$ *for all non-zero* $v \in E^i_x$
3. $\displaystyle\lim_{t\to\pm\infty} \frac{1}{t} \log \angle\big(\bigoplus_{i\in I} E^i_{f^t(x)}, \bigoplus_{j\in J} E^j_{f^t(x)}\big) = 0$ *for all* I *and* J *with* $I \cap J = \emptyset$.

As a consequence, the relation (7.2) also extends to the continuous time case, as do the observations made in the previous sections for discrete time cocycles.

An important special case is the *derivative cocycle* $Df^t : TM \to TM$ over a smooth flow $f^t : M \to M$. We call Lyapunov exponents and Oseledets subspaces of the flow the corresponding objects for this cocycle Df^t, $t \in \mathbb{R}$.

Cocycle induced over a return map. The following construction will be useful later. Let $f : M \to M$ be a transformation, not necessarily invertible, μ be an invariant probability measure, and D be some positive measure subset of M. Let $\rho(x) \geq 1$ be the first return time to D, defined for almost every $x \in D$. Given any cocycle $F = (f, A)$ over f, there exists a corresponding cocycle $G = (g, B)$ over the first return map $g(x) = f^{\rho(x)}(x)$, defined by $B(x)v = A^{\rho(x)}(x)v$.

Proposition 5.11. *1. The normalized restriction μ_D of the measure μ to the domain D is invariant under the first return map g.*

2. $\log^+ \|B^{\pm 1}\|$ are integrable for μ_D if $\log^+ \|A^{\pm 1}\|$ are integrable for μ.

3. For μ-almost every $x \in D$, the Lyapunov exponents of G at x are obtained multiplying the Lyapunov exponents of F at x by some constant $c(x) \geq 1$.

Proof. First, we treat the case when the transformation f is invertible. For each $j \geq 1$, let D_j be the subset of points $x \in D$ such that $\rho(x) = j$. The $\{D_j : j \geq 1\}$ is a partition of a full measure subset of D, and so is the $\{f^j(D_j) : j \geq 1\}$. Notice also that $g \mid D_j = f^j \mid D_j$ for all $j \geq 1$. For any measurable set $E \subset D$ and any $j \geq 1$,

$$\mu\big(g^{-1}(E \cap f^j(D_j))\big) = \mu\big(f^{-j}(E \cap f^j(D_j))\big) = \mu\big(E \cap D_j\big),$$

because μ is invariant under f. It follows that

$$\mu\big(g^{-1}(E)\big) = \sum_{j=1}^{\infty} \mu\big(g^{-1}(E \cap f^j(D_j))\big) = \sum_{j=1}^{\infty} \mu\big(E \cap D_j\big) = \mu(E).$$

This implies that μ_D is invariant under g, as claimed in part (1). Next, from the definition $B(x) = A^{\rho(x)}(x)$ we conclude that

$$\int_D \log^+ \|B\| \, d\mu = \sum_{j=1}^{\infty} \int_{D_j} \log^+ \|A^j\| \, d\mu \leq \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} \int_{D_j} \log^+ \|A \circ f^i\| \, d\mu.$$

Since μ is invariant under f and the domains $f^i(D_j)$ are pairwise disjoint for all $0 \leq i \leq j-1$, it follows that

$$\int_D \log^+ \|B\| \, d\mu \leq \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} \int_{f^i(D_j)} \log^+ \|A\| \, d\mu \leq \int \log^+ \|A\| \, d\mu.$$

The corresponding bound for the norm of the inverse is obtained in the same way. This implies part (2) of the proposition. To prove part (3), define

$$c(x) = \lim_{k \to \infty} \frac{1}{k} \sum_{j=0}^{k-1} \rho(f^j(x)).$$

Notice that ρ is integrable relative to μ_D:

$$\int_D \rho\, d\mu = \sum_{j=1}^{\infty} j\mu(D_j) = \sum_{j=1}^{\infty}\sum_{i=0}^{j-1} \mu(f^i(D_j)) \le 1.$$

Thus, by the ergodic theorem, $c(x)$ is well defined at μ_D-almost every x. It is clear from the definition that $c(x) \ge 1$. Now, given any vector $v \in \mathcal{E}_x \setminus \{0\}$ and a generic point $x \in D$,

$$\lim_{k\to\infty} \frac{1}{k} \log \|B^k(x)v\| = c(x) \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\|$$

(we are assuming $\log^+ \|A\|$ is μ-integrable and so Theorem 7.3 ensures that both limits exist). This proves part (3) of the proposition, when f is invertible.

Finally, we extend the proposition to the non-invertible case. Let $\hat{f}$ be the natural extension of f and $\hat{\mu}$ be the lift of μ (Remark 7.6). Denote $\hat{D} = P^{-1}(D)$. It is clear that the $\hat{f}$-orbit of a point $\hat{x} \in \hat{D}$ returns to $\hat{D}$ at some time n if and only if the f-orbit of $x = P(\hat{x})$ returns to D at time n. Thus, the first return map of $\hat{f}$ to the domain $\hat{D}$ is

$$\hat{g}(x) = \hat{f}^{\rho(x)}(\hat{x}), \quad x = P(\hat{x}),$$

and so it satisfies $P \circ \hat{g} = g \circ P$. It is also clear that the normalized restriction $\hat{\mu}_D$ of $\hat{\mu}$ to the domain $\hat{D}$ satisfies $P_*\hat{\mu}_D = \mu_D$. By the invertible case, $\hat{\mu}_D$ is invariant under $\hat{g}$. It follows that μ_D is invariant under g:

$$\mu_D(g^{-1}(E)) = \hat{\mu}_D(P^{-1}g^{-1}(E)) = \hat{\mu}_D(\hat{g}^{-1}P^{-1}(E)) = \hat{\mu}_D(P^{-1}(E)) = \mu_D(E),$$

for every measurable set $E \subset D$. This settles part (1). Now let $\hat{F} = (\hat{f}, \hat{A})$ be the natural extension of the cocycle F (Remark 7.6) and $\hat{G}$ be the cocycle it induces over $\hat{g}$:

$$\hat{G}(\hat{x}, v) = (\hat{g}(\hat{x}), \hat{B}(\hat{x})v), \quad \hat{B}(\hat{x}) = \hat{A}^{\rho(x)}(\hat{x}).$$

By definition, $\hat{A}(\hat{x}) = A(x)$, and so $\hat{B}(\hat{x}) = B(x)$. Consequently,

$$\int \log^+ \|A\|\, d\mu = \int \log^+ \|\hat{A}\|\, d\hat{\mu} \quad \text{and} \quad \int \log^+ \|B\|\, d\mu_D = \int \log \|\hat{B}\|\, d\hat{\mu}_D.$$

By the invertible case, $\log^+ \|\hat{B}\|$ is $\hat{\mu}_D$-integrable if $\log^+ \|\hat{A}\|$ is $\hat{\mu}$-integrable. It follows that $\log^+ \|B\|$ is μ_D-integrable if $\log^+ \|A\|$ is μ-integrable. The same argument applies to the inverses. This settles part (2) of the proposition. Part (3) also extends easily to the non-invertible case: as observed in Remark 7.6, the Lyapunov exponents of $\hat{F}$ at $\hat{x}$ coincide with the Lyapunov exponents of F at x. For the same reasons, the Lyapunov exponents of $\hat{G}$ at $\hat{x}$ coincide with the Lyapunov exponents of G at x. By the invertible case, the exponents of $\hat{G}$ at $\hat{x}$ are obtained multiplying the exponents of $\hat{F}$ at $\hat{x}$ by some positive factor. Consequently, the exponents of G at x are obtained multiplying the exponents of F at x by that same factor. This concludes the proof of the proposition. $\square$

5.2 Rauzy-Veech-Zorich cocycles

Let C be the extended Rauzy class associated to a given connected component of stratum $\mathcal{C}$ (Section 6.6). Consider $(\pi, \lambda) \in C \times \mathbb{R}_+^{\mathcal{A}}$ and let $\varepsilon \in \{0, 1\}$ be its type. In Section 1.2 we introduced a linear isomorphism $\Theta = \Theta_{\pi,\lambda}$ defined by

$$\Theta_{\alpha,\beta} = \begin{cases} 1 & \text{if either } \alpha = \beta \text{ or } (\alpha, \beta) = (\alpha(1-\varepsilon), \alpha(\varepsilon)) \\ 0 & \text{in all other cases.} \end{cases}$$

In other words, all the entries $\Theta_{\alpha,\beta}$ of the matrix of Θ are zero, except for those on the diagonal and the one where α is the loser and β is the winner of (π, λ).

We also defined the Rauzy-Veech induction $\hat{R}(f)$ of the interval exchange transformation f defined by (π, λ) to be another interval exchange transformation, corresponding to a certain partition $(I'_\alpha)_{\alpha\in\mathcal{A}}$ of the interval

$$I' = I \setminus f(I_{\alpha(1)}) \ \text{ if } \varepsilon = 0 \quad \text{and} \quad I' = I \setminus I_{\alpha(0)} \ \text{ if } \varepsilon = 1.$$

In either case, $\hat{R}(f)(x) = f^{r(x)}(x)$ where $r = r_{\pi,\lambda}$ is the first return time to I' under f, given by $r(x) = 2$ on the (loser) interval $I'_{\alpha(1-\varepsilon)}$ and $r(x) = 1$ on all the other I'_α. By construction, $f(I'_{\alpha(1-\varepsilon)}) \subset I_{\alpha(\varepsilon)}$ if $\varepsilon = 0$ and $I'_{\alpha(1-\varepsilon)} \subset I_{\alpha(\varepsilon)}$ if $\varepsilon = 1$. Thus, in either case,

$$\Theta_{\alpha,\beta} = \#\{0 \le i < r(I'_\alpha) : f^i(I'_\alpha) \subset I_\beta\} \quad \text{for all } \alpha, \beta \in \mathcal{A}. \tag{5.9}$$

For interval exchange maps. The *Rauzy-Veech cocycle* associated to the extended Rauzy class C is the linear cocycle over the Rauzy-Veech renormalization $R : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}$ defined by

$$F_R : C \times \Lambda_{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}}, \quad (\pi, \lambda, v) \mapsto (R(\pi, \lambda), \Theta_{\pi,\lambda}(v)). \tag{5.10}$$

Note that $F_R^n(\pi, \lambda, v) = (R^n(\pi, \lambda), \Theta_{\pi,\lambda}^n(v))$ for all $n \ge 1$, where

$$\Theta^n = \Theta^n_{\pi,\lambda} = \Theta_{\pi^{n-1},\lambda^{n-1}} \cdots \Theta_{\pi',\lambda'} \Theta_{\pi,\lambda} \quad \text{and} \quad (\pi^i, \lambda^i) = R^i(\pi, \lambda).$$

In Proposition 7.14 below we obtain an important interpretation of this linear cocycle. For each $n \ge 1$, let I^n be the domain of definition of $\hat{R}^n(f)$ and $(I^n_\alpha)_{\alpha\in\mathcal{A}}$ the corresponding partition into subintervals. The proposition asserts that *each entry $\Theta^n_{\alpha,\beta}$ of the matrix of Θ^n counts the number of visits of I^n_α to the interval I_β during the induction time.* Before giving the precise statement, we need to collect a few basic facts. Notice that

$$\hat{R}^{n+1}(f)(x) = \hat{R}\big[\hat{R}^n(f)\big](x) = \begin{cases} \hat{R}^n(f)(x) & \text{if } r_{\pi^n,\lambda^n}(x) = 1 \\ \hat{R}^n(f) \circ \hat{R}^n(f)\,(x) & \text{if } r_{\pi^n,\lambda^n}(x) = 2. \end{cases}$$

Consequently, $\hat{R}^n(f)(x) = f^{r^n(x)}(x)$ where the nth Rauzy-Veech *induction time* $r^n = r^n_{\pi,\lambda}$ is defined by

$$r^1_{\pi,\lambda} = r_{\pi,\lambda} \quad \text{and} \quad r^{n+1}_{\pi,\lambda}(x) = \begin{cases} r^n_{\pi,\lambda}(x) & \text{if } r_{\pi^n,\lambda^n}(x) = 1 \\ r^n_{\pi,\lambda}(x) + r^n_{\pi,\lambda}\big(\hat{R}^n(f)(x)\big) & \text{if } r_{\pi^n,\lambda^n}(x) = 2. \end{cases}$$

We shall write $r^n(I^n_\alpha) = r^n_{\pi,\lambda}(I^n_\alpha)$ to mean $r^n(x) = r^n_{\pi,\lambda}(x)$ for any $x \in I^n_\alpha$.

Remark 5.12. If (π, λ) satisfies the Keane condition then $\min\{r^n(x) : x \in I^n\}$ goes to infinity as $n \to \infty$. Indeed, recall that $r^n(x)$ is the (first) return time of x to the interval I^n. Recall also that I^n approaches the origin as $n \to \infty$, by Corollary 1.20. According to Lemma 1.16, the origin can not be a periodic point. Thus, the return times must go to infinity, as claimed.

Lemma 5.13. *The function $r^n_{\pi,\lambda}$ is constant on I^n_α for any $n \geq 1$ and $\alpha \in \mathcal{A}$. Moreover, given any $0 \leq j < r^n(I^n_\alpha)$ there exists $\beta \in \mathcal{A}$ such that $f^j(I^n_\alpha) \subset I_\beta$.*

Proof. The case $n = 1$ is clear from the definition of the Rauzy-Veech induction. The proof proceeds by induction. Suppose first that α is not the loser of (π^n, λ^n). Then $I^{n+1}_\alpha \subset I^n_\alpha$ (they coincide unless α is the winner) and $r^{n+1}(x) = r^n(x)$ for every $x \in I^{n+1}_\alpha$. So, both claims in the lemma follow immediately from the induction hypothesis.

Figure 5.2:

Now take α to be the loser of (π^n, λ^n). Let $w \in \mathcal{A}$ be the winner. Suppose first that (π^n, λ^n) has type 0. Then $I^{n+1}_\alpha = I^n_\alpha$ and $\hat{R}^n(f)(I^{n+1}_\alpha) \subset I^n_w$, as shown on the left hand side of Figure 7.2. Hence,

$$r^{n+1}(x) = r^n(I^n_\alpha) + r^n(I^n_w) \quad \text{for all } x \in I^{n+1}_\alpha,$$

which proves the first claim. Moreover, $f^j(I^{n+1}_\alpha) = f^j(I^n_\alpha)$ for $0 \leq j < r^n(I^n_\alpha)$ and $f^j(\hat{R}^n(f)(I^{n+1}_\alpha)) \subset f^j(I^n_w)$ for $0 \leq j < r^n(I^n_w)$. Hence, the second claim in the lemma follows directly from the induction hypothesis as well. Now suppose that (π^n, λ^n) has type 1. Then $I^{n+1}_\alpha \subset I^n_w$ and $\hat{R}^n(f)(I^{n+1}_\alpha) = I^n_\alpha$, as shown on the right hand side of Figure 7.2. Hence,

$$r^{n+1}(x) = r^n(I^n_w) + r^n(I^n_\alpha) \quad \text{for all } x \in I^{n+1}_\alpha,$$

which proves the first claim. Moreover, $f^j(I^{n+1}_\alpha) \subset f^j(I^n_w)$ for $0 \leq j < r^n(I^n_w)$ and $f^j(\hat{R}^n(f)(I^{n+1}_\alpha)) \subset f^j(I^n_\alpha)$ for $0 \leq j < r^n(I^n_\alpha)$. In view of the induction hypothesis, this proves the second claim in the lemma. □

Lemma 7.13 will be used in the proof of the next proposition, through the following immediate consequence: for any $0 \leq j < r^n_{\pi,\lambda}(I^n_\alpha)$, any $J \subset I^n_\alpha$, and any $\beta \in \mathcal{A}$, we have

$$f^j(J) \subset I_\beta \quad \text{if and only if} \quad f^j(I^n_\alpha) \subset I_\beta. \tag{5.11}$$

Proposition 5.14. *For every $\alpha, \beta \in \mathcal{A}$ and every $n \geq 1$,*

$$\Theta^n_{\alpha,\beta} = \#\{0 \leq j < r^n_{\pi,\lambda}(I^n_\alpha) : f^j(I^n_\alpha) \subset I_\beta\}.$$

Proof. The case $n = 1$ is precisely (7.9). The proof proceeds by induction. Let $l, w \in \mathcal{A}$ be, respectively, the loser and the winner of (π^n, λ^n). We have

$$\Theta^{n+1}_{\alpha,\beta} = \sum_{\gamma\in\mathcal{A}} (\Theta_{\pi^n,\lambda^n})_{\alpha,\gamma}\Theta^n_{\gamma,\beta} = \begin{cases} \Theta^n_{\alpha,\beta} & \text{if } \alpha \neq l \\ \Theta^n_{\alpha,\beta} + \Theta^n_{w,\beta} & \text{if } \alpha = l. \end{cases}$$

Suppose first that $\alpha \neq l$. Then $I^{n+1}_\alpha \subset I^n_\alpha$ and $r^{n+1}(I^{n+1}_\alpha) = r^n(I^n_\alpha)$. Using (7.11) we get that $f^j(I^{n+1}_\alpha) \subset I_\beta$ if and only if $f^j(I^n_\alpha) \subset I_\beta$, for any $0 \le j < r^n(I^n_\alpha)$. These observations show that

$$\#\{0 \le j < r^{n+1}(I^{n+1}_\alpha) : f^j(I^{n+1}_\alpha) \subset I_\beta\} = \#\{0 \le j < r^n(I^n_\alpha) : f^j(I^n_\alpha) \subset I_\beta\}.$$

By the induction hypothesis, the expression on the right hand side is equal to $\Theta^n_{\alpha,\beta} = \Theta^{n+1}_{\alpha,\beta}$ and so the statement follows in this case.

Now we treat the case $\alpha = l$. Suppose first that (π^n, λ^n) has type 0. Then $I^{n+1}_\alpha = I^n_\alpha$ and $\hat{R}^n(f)(I^{n+1}_\alpha) \subset I^n_w$ (left hand side of Figure 7.2). Hence,

$$r^{n+1}(I^{n+1}_\alpha) = r^n(I^n_\alpha) + r^n(I^n_w).$$

Using (7.11) we find that $f^j(\hat{R}^n(f)(I^{n+1}_\alpha)) \subset I_\beta$ if and only if $f^j(I^n_w) \subset I_\beta$, for any $0 \le j < r^n(I^n_w)$. Thus, the number of $0 \le j < r^{n+1}(I^{n+1}_\alpha)$ such that $f^j(I^{n+1}_\alpha) \subset I_\beta$ is equal to

$$\#\{0 \le j < r^n(I^n_\alpha) : f^j(I^n_\alpha) \subset I_\beta\} + \#\{0 \le j < r^n(I^n_w) : f^j(I^n_w) \subset I_\beta\}.$$

By the induction hypothesis, this sum is equal to $\Theta^n_{\alpha,\beta} + \Theta^n_{w,\beta} = \Theta^{n+1}_{\alpha,\beta}$. This settles the type 0 case. Now suppose (π^n, λ^n) has type 1. Then $I^{n+1}_\alpha \subset I^n_w$ and $\hat{R}^n(f)(I^{n+1}_\alpha) = I^n_\alpha$ (right hand side of Figure 7.2). Hence,

$$r^{n+1}(I^{n+1}_\alpha) = r^n(I^n_w) + r^n(I^n_\alpha).$$

Using (7.11) once more, we find that $f^j(I^{n+1}_\alpha) \subset I_\beta$ if and only if $f^j(I^n_w) \subset I_\beta$, for $0 \le j < r^n(I^n_w)$. Moreover, $f^j(\hat{R}^n(f)(I^{n+1}_\alpha)) \subset I_\beta$ if and only if $f^j(I^n_\alpha) \subset I_\beta$, for $0 \le j < r^n(I^n_\alpha)$. Thus, the number of $0 \le j < r^{n+1}(I^{n+1}_\alpha)$ such that $f^j(I^{n+1}_\alpha) \subset I_\beta$ is equal to

$$\#\{0 \le j < r^n(I^n_w) : f^j(I^n_w) \subset I_\beta\} + \#\{0 \le j < r^n(I^n_\alpha) : f^j(I^n_\alpha) \subset I_\beta\}.$$

This sum is equal to $\Theta^n_{w,\beta} + \Theta^n_{\alpha,\beta} = \Theta^{n+1}_{\alpha,\beta}$ and so the proof is complete. □

From Proposition 7.14 we also get an alternative proof of Corollary 1.21:

Corollary 5.15. *Suppose the interval exchange transformation f defined by (π, λ) is minimal. Then there is $N \ge 1$ such that $\Theta^N_{\alpha,\beta} \ge 1$ for all α, $\beta \in \mathcal{A}$.*

Proof. We use the following equivalent formulation of minimality: given any compact set $K \subset I$ and any open set $A \subset I$, there exists $N_1 \ge 1$ such that for any $x \in K$ we have $f^j(x) \in A$ for some $0 \le j < N_1$. Fix K to be the closure of the domain I' of $\hat{R}(f)$. Then there exists $N_2 \ge 1$ such that for any $x \in K$

and any $\beta \in \mathcal{A}$ we have $f^j(x) \in I_\beta$ for some $0 \le j < N_2$. By Remark 7.12, we may fix $N \ge 1$ such that $r^N(x) \ge N_2$ for all $x \in I^N$. Since $I^N \subset K$, we get that for every $\alpha, \beta \in \mathcal{A}$ and every $x \in I_\alpha^N$ there exists $0 \le j < r^N(x)$ such that $f^j(x) \in I_\beta$. Using (7.11) we conclude that $f^j(I_\alpha^N) \subset I_\beta$ for any such j. In view of Proposition 7.14, this means that $\Theta_{\alpha,\beta}^N \ge 1$ for every $\alpha, \beta \in \mathcal{A}$. □

For translation surfaces. The *invertible Rauzy-Veech cocycle* associated to an extended Rauzy class C is the linear cocycle over the invertible Rauzy-Veech renormalization $\mathcal{R} : \mathcal{H} \to \mathcal{H}$ defined by

$$F_{\mathcal{R}} : \mathcal{H} \times \mathbb{R}^{\mathcal{A}} \to \mathcal{H} \times \mathbb{R}^{\mathcal{A}}, \quad (\pi, \lambda, \tau, v) \mapsto (\mathcal{R}(\pi, \lambda, \tau), \Theta_{\pi,\lambda}(v)). \tag{5.12}$$

Recall $\mathcal{H} = \mathcal{H}(C)$ is the set of all (π, λ, τ) such that $\pi \in C$, $\lambda \in \Lambda_{\mathcal{A}}$, and $\tau \in T_\pi^+$.

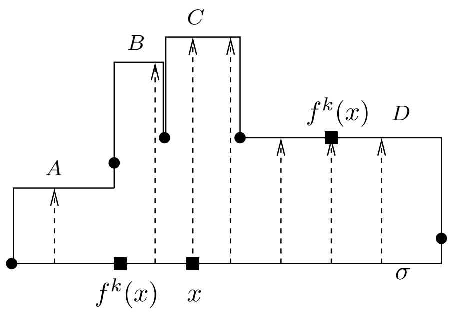


Figure 5.3:

Proposition 7.14 may be reinterpreted in terms of the suspension of the interval exchange map f defined by (π, λ), as follows. Take the suspension surface M to be represented in the form of zippered rectangles, corresponding to data (π, λ, τ, h). Let us recall some notation from Section 5.7. For each x in the basis horizontal segment σ and for each $k \ge 1$, let $[\gamma(x, k)] \in H_1(M, \mathbb{R})$ be the homology class represented by the vertical geodesic segment from x to $f^k(x)$, with the endpoints joined by the horizontal segment they determine inside σ. See Figure 7.3. Moreover, for each $\beta \in \mathcal{A}$, let $[v_\beta]$ be the homology class represented by a vertical segment crossing from bottom to top the rectangle labeled by β, with its endpoints joined by a horizontal segment inside σ.

Corollary 5.16. *For any $n \ge 1$, $\alpha \in \mathcal{A}$, and $x \in I_\alpha^n$,*

$$[\gamma(x, r^n(I_\alpha^n))] = \sum_{\beta \in \mathcal{A}} \Theta_{\alpha,\beta}^n [v_\beta].$$

Proof. Proposition 7.14 means that the vertical geodesic segment $\gamma(x, r^n(x))$ intersects each horizontal segment $I_\beta \times \{0\} \subset \sigma$ exactly $\Theta_{\alpha,\beta}^n$ times. Equivalently, $\gamma(x, r^n(x))$ crosses $\Theta_{\alpha,\beta}^n$ times the rectangle labeled by each $\beta \in \mathcal{A}$. The claim follows immediately. □

Zorich cocycles. Recall that $n(\pi,\lambda) \geq 1$ is the smallest integer for which the type of $R^n(\pi,\lambda)$ is different from the type of (π,λ). In Sections 1.8 and 2.10 we introduced the Zorich renormalization

$$Z : C \times \Lambda_{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}}, \quad Z(\pi,\lambda) = R^{n(\pi,\lambda)}(\pi,\lambda)$$

and its invertible version

$$\mathcal{Z} : Z_* \to Z_*, \quad \mathcal{Z}(\pi,\lambda,\tau) = \mathcal{R}^{n(\pi,\lambda)}(\pi,\lambda,\tau),$$

on the set Z_* of all data $(\pi,\lambda,\tau) \in \mathcal{H}(C)$ such that the type of (π,λ) is different from the type of τ. Now we let

$$\Gamma = \Gamma_{\pi,\lambda} = \Theta_{\pi,\lambda}^{n(\pi,\lambda)} \tag{5.13}$$

and introduce the *Zorich cocycle* over the map Z

$$F_Z : C \times \Lambda_{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}} \to C \times \Lambda_{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}}, \quad F_Z(\pi,\lambda,v) = (Z(\pi,\lambda), \Gamma_{\pi,\lambda}(v)),$$

and the *invertible Zorich cocycle* over the map $\mathcal{Z}$

$$F_{\mathcal{Z}} : Z_* \times \mathbb{R}^{\mathcal{A}} \to Z_* \times \mathbb{R}^{\mathcal{A}}, \quad F_{\mathcal{Z}}(\pi,\lambda,\tau,v) = (\mathcal{Z}(\pi,\lambda,\tau), \Gamma_{\pi,\lambda}(v)).$$

In Section 4.8 we found an ergodic Z-invariant probability measure μ absolutely continuous with respect Lebesgue measure along $\Lambda_{\mathcal{A}}$. Moreover, we have seen in Corollary 5.17 that $\mathcal{Z}$ is equivalent to the natural extension of Z, up to zero measure sets. Hence, $F_{\mathcal{Z}}$ may be seen as the natural extension of F_Z, in the sense of Remark 7.6, and so the two cocycles have the same Lyapunov spectrum. We are going to see, in Proposition 7.18, that this Lyapunov spectrum is indeed well defined. Before that, let us translate to this setting of Zorich cocycles the properties of Rauzy-Veech cocycles we have just obtained.

By definition, the Zorich induction $\hat{Z}(f)(x) = \hat{R}^{n(\pi,\lambda)}(f)(x) = f^{z(x)}(x)$, where

$$z(x) = z_{\pi,\lambda}(x) = r_{\pi,\lambda}^{n(\pi,\lambda)}(x).$$

More generally, $\hat{Z}^m(f)(x) = \hat{R}^{n^m(\pi,\lambda)}(f)(x) = f^{z^m(x)}(x)$ for all $m \geq 1$, where

$$n^m(\pi,\lambda) = \sum_{j=0}^{m-1} n(Z^j(\pi,\lambda)) \quad \text{and} \quad z^m(x) = z_{\pi,\lambda}^m(x) = r_{\pi,\lambda}^{n^m(\pi,\lambda)}(x). \tag{5.14}$$

Denote by $J_\alpha^m = I_\alpha^{n^m(\pi,\lambda)}$, $\alpha \in \mathcal{A}$ the partition subintervals corresponding to the interval exchange map $Z^m(f) = R^{n^m(\pi,\lambda)}(f)$.

We shall write $z^m(J_\alpha^m) = z_{\pi,\lambda}^m(J_\alpha^m)$ to mean $z^m(x) = z_{\pi,\lambda}^m(x)$ for any $x \in J_\alpha^m$.

Corollary 5.17. *For every $\alpha,\beta \in \mathcal{A}$ and every $m \geq 1$,*

1. $\Gamma_{\alpha,\beta}^m = \#\{0 \leq j < z_{\pi,\lambda}^m(J_\alpha^m) : f^j(J_\alpha^m) \subset I_\beta\}$ *and*
2. $[\gamma(x, z_{\pi,\lambda}^m(J_\alpha^m))] = \sum_{\beta\in\mathcal{A}} \Gamma_{\alpha,\beta}^m [v_\beta]$ *and*

3. $z^m_{\pi,\lambda}(J^m_\alpha) = \sum_{\beta\in\mathcal{A}} \Gamma^m_{\alpha,\beta}$.

Proof. Parts 1 and 2 follow directly from Proposition 7.14 and Corollary 7.16, respectively, simply by restricting the conclusions to appropriate subsequences. Part 3 is obtained summing the equality in part 1 over all $\beta \in \mathcal{A}$. $\square$

Now we check that the Zorich cocycles satisfy the integrability condition in the Oseledets Theorem 7.4. For convenience, in what follows we take the norm of a vector or a matrix to be given by the largest absolute value of the coefficients.

Proposition 5.18. *The functions $(\pi, \lambda) \mapsto \log^+ \|\Gamma^{\pm 1}_{\pi,\lambda}\|$ are integrable relative to the invariant probability measure μ of the Zorich renormalization Z.*

Proof. Notice that $\det \Gamma_{\pi,\lambda} = 1$ for all $(\pi, \lambda) \in C \times \Lambda_{\mathcal{A}}$. So, in view of our choice of the norm, $\|\Gamma_{\pi,\lambda}\| = \|\Gamma^{-1}_{\pi,\lambda}\| \geq 1$ for all (π, λ). Thus, we only have to prove that $(\pi, \lambda) \mapsto \log \|\Gamma_{\pi,\lambda}\|$ is integrable. To that end we use *

Lemma 5.19. *Let $w = \alpha(\varepsilon)$ be the winner of (π, λ) and $s = \pi_{1-\varepsilon}(w)$ be its position in the other line of the pair π. For any integer $L \geq 1$,*

$$\max_{\alpha,\beta\in\mathcal{A}} \Gamma_{\alpha,\beta} > L \quad \Rightarrow \quad \lambda_w > L \sum_{\pi_{1-\varepsilon}(\gamma)>s} \lambda_\gamma.$$

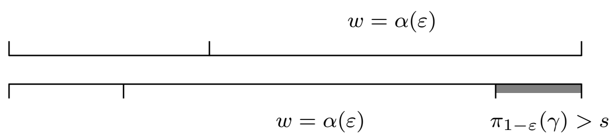


Figure 5.4:

Proof. During the $n(\pi, \lambda)$ iterates that define $\Gamma_{\pi,\lambda}$ the winner does not change. Consequently, for all the matrices $\Theta_{R^i(\pi,\lambda)}$, $0 \leq i < n(\pi, \lambda)$ involved, we have

$$\Theta_{\alpha,\beta} = \begin{cases} 1 & \text{if either } \alpha = \beta \text{ or } \alpha = \text{loser and } \beta = w \\ 0 & \text{in all other cases.} \end{cases}$$

It follows, by induction on the iterate, that

$$\Gamma_{\alpha,\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{if } \alpha \neq \beta \neq w \\ \text{number of times } \alpha \text{ is the loser} & \text{if } \alpha \neq \beta = w. \end{cases} \tag{5.15}$$

The losers during those iterates are taken from $\pi^{-1}_{1-\varepsilon}(d), \ldots, \pi^{-1}_{1-\varepsilon}(s+1)$, in cyclic order. See Figure 7.4. Therefore,

$$\min_{\pi_{1-\varepsilon}(\gamma)>s} \Gamma_{\gamma,w} \sum_{\pi_{1-\varepsilon}(\gamma)>s} \lambda_\gamma < \lambda_w,$$

and the difference between the maximum and the minimum is at most 1. As a direct consequence, we get that for any integer $L \geq 1$,

$$\max_{\alpha,\beta \in \mathcal{A}} \Gamma_{\alpha,\beta} > L \quad \Rightarrow \quad \min_{\pi_{1-\varepsilon}(\gamma)>s} \Gamma_{\gamma,w} \geq L \quad \Rightarrow \quad \lambda_w > L \sum_{\pi_{1-\varepsilon}(\gamma)>s} \lambda_\gamma,$$

just as we claimed. The proof of Lemma 7.19 is complete. □

Let us proceed with the proof of Proposition 7.18. Let $\mathcal{N}$ denote the set of integer vectors $n = (n_\alpha)_{\alpha \in \mathcal{A}}$ such that $n_\alpha \geq 0$ for all $\alpha \in \mathcal{A}$, and the n_α are not all zero. For each $n \in \mathcal{N}$, define

$$\Lambda(n) = \{\lambda \in \Lambda_{\mathcal{A}} : 2^{-n_\alpha} \leq \lambda_\alpha d < 2^{-n_\alpha+1} \text{ for every } \alpha \in \mathcal{A}\},$$

except that for $n_\alpha = 0$ the second inequality is omitted. We have seen in (4.58) that there exists a constant $K > 0$ such that

$$\mu\big(\{\pi\} \times \Lambda(n)\big) = \int_{\Lambda(n)} \prod_{\beta \in \mathcal{A}} \frac{1}{\lambda \cdot h^\beta}\, d_1\lambda \leq K 2^{-\max_{\mathcal{A}} n_\alpha}$$

Lemma 7.19 implies that, for any integer $L \geq 1$,

$$\|F_Z\| > L \quad \Rightarrow \quad \lambda_{\alpha(\varepsilon)} > L\lambda_{\alpha(1-\varepsilon)} \quad \Rightarrow \quad \lambda_{\alpha(1-\varepsilon)} < L^{-1}.$$

Taking $L = 2^k d$, with $k \geq 0$, we find that

$$\|F_Z\| > 2^k d \quad \Rightarrow \quad \lambda_{\alpha(1-\varepsilon)} d < 2^{-k} \quad \Rightarrow \quad \lambda \in \bigcup_{\max n_\alpha \geq k} \Lambda(n).$$

For each $k \geq 0$ there are at most $(k+1)^d$ vectors $n \in \mathcal{N}$ with $\max_{\mathcal{A}} n_\alpha = k$. So, the previous observations yield

$$\mu\big(\{\|F_Z\| > 2^k d\}\big) \leq \sum_{l=k}^{\infty} \sum_{\max n_\alpha = l} \mu(\Lambda(n)) \leq \sum_{l=k}^{\infty} K(l+1)^d 2^{-l} \leq K'(k+1)^d 2^{-k}$$

for some constant K'. This inequality implies that $\|F_Z\|^\theta$ is μ-integrable for all $\theta < 1$. In particular, $\log \|F_Z\|$ is μ-integrable. □

This proposition ensures that Zorich cocycles have well defined Lyapunov exponents which, since the measure μ is ergodic, are constant on a full μ-measure set. Next, we analyze the corresponding Lyapunov spectra.

5.3 Lyapunov spectra of Zorich cocycles

Symmetry. First, we prove that these Lyapunov spectra have a symmetric structure:

Proposition 5.20. *The Lyapunov spectrum of the Zorich cocycle F_Z corresponding to any connected component of a stratum $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ has the form*

$$\theta_1 \geq \theta_2 \geq \cdots \geq \theta_g \geq 0 = \cdots = 0 \geq -\theta_g \geq \cdots \geq -\theta_2 \geq -\theta_1$$

where 0 *occurs with multiplicity $\kappa - 1$.*

Proof. The proof has three main steps, corresponding to Lemmas 7.21 to 7.23. First, we exhibit a $2g$-dimensional subbundle H which is invariant under F_Z. Next, we prove that the Lyapunov exponents corresponding to Oseledets subspaces transverse to H are all zero. Then, we check that the restriction of F_Z to the invariant subbundle preserves a symplectic form, and so its Lyapunov spectrum is symmetric around zero. Let us detail each of these steps.

Let $H = \{(\pi, \lambda) \times H_\pi\}$ be the subbundle of $C \times \Lambda_\mathcal{A} \times \mathbb{R}^\mathcal{A}$ whose fiber over each $(\pi, \lambda) \in C \times \Lambda^\mathcal{A}$ is the subspace $H_\pi = \Omega_\pi(\mathbb{R}^\mathcal{A})$ introduced in Section 1.9. Since Ω_π is anti-symmetric, H_π is the orthogonal complement of $\ker \Omega_\pi$. We have seen in Proposition 2.8 and Lemma 2.10 that $\dim \ker \Omega_\pi = \kappa - 1$ and $\dim H_\pi = 2g$, where κ is the number of singularities and g is the genus.

Lemma 5.21. *H_π is invariant under the linear cocycles F_R and F_Z.*

Proof. Let $(\pi', \lambda') = R(\pi, \lambda)$. By Lemma 1.29, we have $\Theta\, \Omega_\pi\, \Theta^* = \Omega_{\pi'}$. It follows that $\Theta^{-1*}(\ker \Omega_\pi) = \ker \Omega_{\pi'}$. In other words, the subbundle whose fiber over each $(\pi, \lambda) \in C \times \Lambda^\mathcal{A}$ is the subspace $\ker \Omega_\pi$ is invariant under the adjoint Rauzy-Veech cocycle

$$F_R^{-1*} : (\pi, \lambda, v) \mapsto (\pi', \lambda', \Theta^{-1*}(v)).$$

Since H_π is the orthogonal complement of the kernel, it follows that the subbundle H is invariant under the Rauzy-Veech cocycle F_R. Consequently, H is invariant under the Zorich cocycle $F_Z(\pi, \lambda, v) = F_R^{n(\pi,\lambda)}(\pi, \lambda, v)$ as well. □

Let us denote $(\pi^n, \lambda^n) = R^n(\pi, \lambda)$, for generic $(\pi, \lambda) \in C \times \Lambda_\mathcal{A}$ and $n \geq 1$.

Lemma 5.22. *There exists $C_0 > 0$ such that the component of every $\Theta^n_{\pi,\lambda}(v)$ orthogonal to H_{π^n} is bounded by $C_0\|v\|$ for any (π, λ, v) and any $n \geq 1$.*

Proof. Let σ be the permutation of $\{0, 1, \ldots, d\}$ associated to π, as defined in Section 2.4. More generally, let σ^n be the permutation associated to π^n, for $n \geq 1$. Consider the basis $\{\lambda(\mathcal{O}) : \mathcal{O}$ is an orbit of σ not containing $0\}$ of $\ker \Omega_\pi$ we introduced in Section 2.6. We have seen in Lemma 2.16 that the dynamics of Θ^{-1*} on the invariant subbundle $\{(\pi, \lambda) \times \ker \Omega_\pi\}$ is quite trivial: the image of every $\lambda(\mathcal{O})$ coincides with some element $\lambda(\mathcal{O}')$ of the basis of $\ker \Omega_{\pi'}$. It follows that for every $n \geq 1$ there exists a bijection $\mathcal{O} \mapsto \mathcal{O}^n$ between the set of orbits of σ not containing 0 and the set of orbits of σ^n not containing 0, such that $\Theta^{-n*}(\lambda(\mathcal{O})) = \lambda(\mathcal{O}^n)$ for all $\mathcal{O}$. Then,

$$\lambda(\mathcal{O}^n) \cdot \Theta^n(v) = \Theta^{n*}(\lambda(\mathcal{O}^n)) \cdot v = \lambda(\mathcal{O}) \cdot v \quad \text{for every } v \in \ker \Omega_\pi \text{ and every } \mathcal{O}.$$

This implies that the component of $\Theta^n(v)$ in the direction of $\ker \Omega_{\pi^n}$ is bounded in norm by $C_0\|v\|$, for some constant C_0 that depends only on the choice of the norm ($C_0 = 1$ if the bases $\{\lambda(\mathcal{O})\}$ are orthonormal). $\square$

Recall that F_Z and the invertible Zorich cocycle $F_{\mathcal{Z}}$ have the same Lyapunov spectrum. Let $E^i_{\pi,\lambda,\tau}$ be any Oseledets subspace of $F_{\mathcal{Z}}$ transverse to H. Given any non-zero $v \in E^i_{\pi,\lambda}$ and $n \geq 1$, denote $v_n = \Gamma^n_{\pi,\lambda}(v)$. Write $v_n = v_n^H + v_n^K$, where v_n^H is the projection to H and v_n^K is the projection to the orthogonal complement of H. According to Lemma 7.22, $\|v_n^K\| \leq C_0\|v\|$ for all n. Moreover, given any $\varepsilon > 0$,

$$\|v_n^K\| \geq e^{-\varepsilon n}\|v_n^H\| \quad \text{for all large } n,$$

because the angles between the iterates of $E^i_{\pi,\lambda}$ and the subbundle H decay at most sub-exponentially (part 3 of Theorem 7.4). This implies

$$e^{-2\varepsilon n}\|v\| \leq \|v_n\| \leq e^{2\varepsilon n}\|v\| \quad \text{for all large } n.$$

Thus, the Lyapunov exponent corresponding to $E^i_{\pi,\lambda}$ is smaller than 2ε in absolute value. Since ε is arbitrary, the exponent must vanish, as we claimed.

In Section 1.9 we introduced a symplectic form on each $H_\pi = \Omega_\pi(\mathbb{R}^{\mathcal{A}})$, given by

$$\omega_\pi : H_\pi \times H_\pi \to \mathbb{R}, \quad \omega_\pi(\Omega_\pi(u), \Omega_\pi(v)) = -u \cdot \Omega_\pi(v).$$

This defines a symplectic form ω on the invariant subbundle H. Moreover, we checked that $\Theta_{\pi,\lambda} : H_\pi \to H_{\pi'}$ is symplectic relative to the forms ω_π and $\omega_{\pi'}$. In other words,

Lemma 5.23. *The symplectic form ω is invariant under both F_R and F_Z.*

According to Proposition 7.8, this implies that the Lyapunov spectrum of F_Z restricted to H is symmetric around zero. This ends the proof of Proposition 7.20. $\square$

Extremal Lyapunov exponents. The final step in Proposition 7.24 is to show that the extremal exponents have multiplicity 1:

Proposition 5.24. *The largest Lyapunov exponent θ_1 and the smallest Lyapunov exponent $-\theta_1$ of every Zorich cocycle F_Z are simple, and the same is true for the adjoint cocycle F_Z^{-1*}.*

A much stronger fact will be obtained later, in Theorem 7.32: all Lyapunov exponents $\pm\theta_j$ are distinct and non-zero.

Proof. By Proposition 7.20 the spectra of F_Z and F_Z^{-1*} are symmetric with respect to the origin. By Proposition 7.9 they are symmetric to one another. Thus, it suffices to prove that the smallest exponent of the adjoint cocycle F_Z^{-1*} is simple. This is done as follows.

By Corollary 7.15, for almost every (π, λ) we may find $N \geq 1$ such that $\Theta^N_{\alpha,\beta} \geq 1$ for all $\alpha, \beta \in \mathcal{A}$. Then the same is true for every Θ^k, $k \geq N$ and, in particular, $\Gamma^k_{\alpha,\beta} \geq 1$ for all $\alpha, \beta \in \mathcal{A}$ and $k \geq N$. Let (π, λ) and N be fixed. In view of the Markov structure of the Zorich renormalization (recall Section 1.8), λ is contained in some subsimplex D of $\Lambda_{\mathcal{A}}$ such that $Z^N \mid \{\pi\} \times D$ coincides with the projective map defined by Γ^{-N*} and maps the domain bijectively to $\{\pi^N\} \times \Lambda_{\pi^N,1-\varepsilon}$. Since the coefficients of Γ^{N*} are all positive, D is relatively compact in $\Lambda_{\mathcal{A}}$. See Figure 7.5.

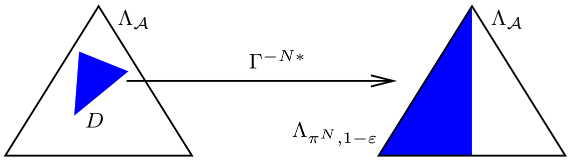


Figure 5.5:

By Poincaré recurrence, for μ-almost every $(\pi, \lambda) \in \{\pi\} \times D$ there exists a first return time $\rho(\pi, \lambda) \geq N$ to the domain $\{\pi\} \times D$ under the map Z. Note that $\mu(\{\pi\} \times D) > 0$, since μ is positive on open sets. The normalized restriction μ_D of the measure μ to the domain $\{\pi\} \times D$ is invariant and ergodic under the return map (Proposition 7.11)

$$\tilde{Z} : \{\pi\} \times D \to \{\pi\} \times D, \quad \tilde{Z}(\pi, \lambda) = Z^{\rho(\pi,\lambda)}(\pi, \lambda).$$

The adjoint Zorich cocycle induces a linear cocycle $\tilde{F}_Z$ over $\tilde{Z}$, given by

$$\tilde{F}_Z(\pi, \lambda, v) = (\tilde{Z}(\pi, \lambda), \tilde{\Gamma}_{\pi,\lambda}(v)), \qquad \tilde{\Gamma} = \tilde{\Gamma}_{\pi,\lambda} = \Gamma^{-\rho(\pi,\lambda)*}_{\pi,\lambda}.$$

Corollary 5.25. *The functions $\log^+ \|\tilde{\Gamma}^{\pm 1}\|$ are μ_D-integrable, and the smallest Lyapunov exponent of the adjoint Zorich cocycle F_Z^{-1*} for μ is simple if and only if the smallest Lyapunov exponent of $\tilde{F}_Z$ is simple at μ_D-almost every point.*

Proof. Proposition 7.18 gives that the functions $\log^+ \|\Gamma^{\pm 1}\|$ are μ-integrable. So, the first statement in the corollary follows immediately from part (2) of Proposition 7.11, applied to $F = F_Z^{-1*}$. Moreover, part (3) of Proposition 7.11 gives that the Lyapunov exponents of $\tilde{F}_Z$ at a generic point x are the products of the Lyapunov exponents of F_Z^{-1*} by some constant $c(x)$. The last statement in the corollary is a direct consequence. $\square$

Thus, to prove Proposition 7.24 it suffices to show that the smallest exponent of $\tilde{F}_Z$ is simple. Let $\mathcal{C} = \{v \in \mathbb{R}^{\mathcal{A}}_+ : v/|v| \in D\}$ be the cone associated to D. The definition of $\tilde{\Gamma}$ implies that

$$\tilde{\Gamma}^{-1}_{\pi,\lambda}(\mathbb{R}^{\mathcal{A}}_+) = \Gamma^{\rho*}_{\pi,\lambda}(\mathbb{R}^{\mathcal{A}}_+) = \Gamma^{N*}_{\pi,\lambda}\big(\Gamma^{(\rho-N)*}_{\pi^N,\lambda^N}(\mathbb{R}^{\mathcal{A}}_+)\big) \subset \Gamma^{N*}_{\pi,\lambda}(\mathbb{R}^{\mathcal{A}}_+) \subset \mathcal{C}$$

for every (π, λ). Thus, the cocycle $\tilde{F}_Z$ admits a backward invariant cone which is relatively compact, in the sense that its intersection with the simplex $\Lambda_{\mathcal{A}}$ is relatively compact inside the simplex. So, at this point the proposition is a direct consequence of the following Perron-Fröbenius type result:

Lemma 5.26. *Let $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$, $F(x,v) = (f(x), A(x)v)$ be a linear cocycle over a transformation $f : M \to M$, such that $\log^+ \|A^{\pm 1}\|$ are integrable with respect to some f-invariant probability ν. Assume there exists some relatively compact cone $\mathcal{C} \subset \mathbb{R}_+^{\mathcal{A}}$ such that $A(x)^{-1}(\mathbb{R}_+^{\mathcal{A}}) \subset \mathcal{C}$ for all $x \in M$. Then the smallest exponent of F with respect to ν has multiplicity 1 at almost ν-every point.*

Proof. There are two main parts. First, we identify the invariant line bundle associated to the smallest Lyapunov exponent $\lambda(x)$. Then, we prove that any vector outside this invariant subbundle grows (or decays) at exponential rate strictly larger than $\lambda(x)$, under positive iteration.

Since $\mathcal{C}$ is relatively compact, it has finite diameter relative to the projective metric on the cone $\mathbb{R}_+^d$. Thus, by Proposition 4.23, every $A(x)^{-1} : \mathbb{R}_+^d \to \mathcal{C}$ is a contraction with respect to the projective metric, with uniform contraction rate (depending only on $\mathcal{C}$). It follows that the width of $A^n(x)^{-1}(\mathbb{R}_+^{\mathcal{A}})$ is bounded by $C_1 e^{-an}$, for some $C_1 > 0$ and $a > 0$ that depend only on $\mathcal{C}$. In particular, the intersection of all these cones reduces to a half-line at every $x \in M$:

$$\bigcap_{n=1}^{\infty} A^n(x)^{-1}(\mathbb{R}_+^{\mathcal{A}}) = \mathbb{R}_+ \xi(x) \tag{5.16}$$

for some vector $\xi(x) \in \mathcal{C}$ which we may choose with norm 1. It is clear from (7.16) that the line bundle $\mathbb{R}\xi(x)$ is invariant under the cocycle F. Let $\lambda(x)$ be the corresponding Lyapunov exponent. We claim that any vector v which is not in $\mathbb{R}\xi(x)$ grows, under positive iteration, at exponential rate larger or equal than $\lambda(x) + a$. This implies that all the other Lyapunov exponents are at least $\lambda(x) + a$, which proves the lemma. Thus, we are left to proving this claim.

Let v be any unit vector outside $\mathbb{R}\xi(x)$. It follows from the definition (7.16) that some iterate of v is outside the cone $\mathbb{R}_+^{\mathcal{A}}$. Thus, it is no restriction to assume right from the start that $v \notin \mathbb{R}_+^{\mathcal{A}}$. Since $\xi(y) \in \mathcal{C}$ for every $y \in M$, the coefficients of any $\xi(y)$ are uniformly bounded from zero. Hence, there exists $c_1 > 0$, depending only on this bound, such that

$$\frac{A^n(x)v}{\|A^n(x)v\|} + c_1 \frac{A^n(x)\xi(x)}{\|A^n(x)\xi(x)\|} = \frac{A^n(x)v}{\|A^n(x)v\|} + c_1\, \xi(f^n(x)) \in \mathbb{R}_+^{\mathcal{A}} \quad \text{for every } n \geq 1.$$

Then, by the considerations in the previous paragraph, the angle between $\xi(x)$ and

$$\frac{v}{\|A^n(x)v\|} + c_1 \frac{\xi(x)}{\|A^n(x)\xi(x)\|} = A^n(x)^{-1}\Big(\frac{A^n(x)v}{\|A^n(x)v\|} + c_1 \frac{A^n(x)\xi(x)}{\|A^n(x)\xi(x)\|}\Big)$$

is bounded by $C_1 e^{-an}$. Since $v \notin \mathbb{R}_+^{\mathcal{A}}$ and $\xi(x) \in \mathcal{C}$, the angle between $\xi(x)$ and v is bounded below by some constant $c_2 > 0$ that depends only on the cone $\mathcal{C}$. Thus, the previous property implies that

$$\frac{\|A^n(x)\xi(x)\|}{\|A^n(x)v\|} \leq C_2 e^{-an},$$

where the constant C_2 depends only on c_1, c_2, and C_1, and so is determined by the cone $\mathcal{C}$. This implies that

$$\|A^n(x)v\| \geq C_2^{-1} e^{an} \|A^n(x)\xi(x)\| \geq c(x) e^{(\lambda(x)+a)n},$$

for every $n \geq 1$ and $v \notin \mathbb{R}_+\xi(x)$, as claimed. □

At this point the proof of Proposition 7.24 is complete. □

Extremal Oseledets subspaces. Recall that F_Z and the invertible Zorich cocycle $F_{\mathcal{Z}}$ have the same Lyapunov spectrum. Besides, the same is true for the adjoint cocycle $F_{\mathcal{Z}}^{-1*}$, as a consequence of Propositions 7.9 and 7.20. For either of these invertible cocycles, we are going to give an explicit description of the Oseledets subspaces associated to the extremal Lyapunov exponents $\pm\theta_1$.

Let us start with the cocycle $F_{\mathcal{Z}}$. For each $x = (\pi, \lambda, \tau) \in \mathcal{H}$, consider the following subspaces of $\mathbb{R}^{\mathcal{A}}$:

- E_x^s = line spanned by $w = \Omega_\pi(\lambda)$ and E_x^u = line spanned by $h = -\Omega_\pi(\tau)$
- E_x^c = ω_π-symplectic orthogonal to $E_x^u \oplus E_x^s$, that is,

$$E_x^c = \{v \in \mathbb{R}^{\mathcal{A}} : \omega_\pi(v, w) = 0 \text{ for all } w \in E_x^u \oplus E_x^s\}.$$

E_x^u is not symplectic orthogonal to E_x^s: indeed, $\omega_\pi(\Omega_\pi(\lambda), \Omega_\pi(\tau)) = -\lambda \cdot \Omega_\pi(\tau)$ is strictly positive, since both λ and $h = -\Omega_\pi(\tau)$ have only positive coordinates. Thus, E_x^c has codimension 2 and $\mathbb{R}^{\mathcal{A}} = E_x^u \oplus E_x^c \oplus E_x^s$.

Lemma 5.27. *The splitting $E^u \oplus E^c \oplus E^s$ is invariant under the invertible Zorich cocycle $F_{\mathcal{Z}}$. Moreover, E^u corresponds to the largest Lyapunov exponent θ_1, E^s corresponds to the smallest Lyapunov exponent $-\theta_1$, and E^c corresponds to the remaining Lyapunov exponents.*

Proof. Let $(\pi', \lambda', \tau') = \mathcal{R}(\pi, \lambda, \tau)$. Then, by (1.11) and Lemma 1.29, we have $\Theta(\Omega_\pi(\lambda)) = \Omega_{\pi'}(\Theta^{-1*}(\lambda)) = \Omega_{\pi'}(\lambda')$. Analogously, $\Theta(\Omega_\pi(\tau)) = \Omega_{\pi'}(\tau')$, by (2.33) and Lemma 1.29. This proves that E^u and E^s are invariant under the invertible Rauzy-Veech cocycle $F_{\mathcal{R}}$. Then the symplectic orthogonal E^c is also invariant, since $F_{\mathcal{R}}$ preserves the symplectic form ω_π (Corollary 1.30). It follows that all three subbundles are invariant under the Zorich cocycle $F_{\mathcal{Z}}$ as well.

Since $h = -\Omega_\pi(\tau)$ lies in the positive cone, and the matrices of $F_{\mathcal{Z}}$ have non-negative coefficients, E^u must be contained in the Oseledets subspace corresponding to the largest Lyapunov exponent. As this exponent is simple (Proposition 7.24), it follows that E^u coincides with that Oseledets subspace. This implies that E^s corresponds to the smallest Lyapunov exponent, since it is an invariant direction which is not contained in the symplectic orthogonal to E^u (recall the arguments following Proposition 7.8). Then the complementary invariant subbundle E^c must coincide with the sum of the Oseledets subspaces corresponding to the remaining Lyapunov exponents. □

Now we deal with the cocycle $F_{\mathcal{Z}}^{-1*}$. For each $x = (\pi, \lambda, \tau) \in \mathcal{H}$, define

- E^{s*}_x = line spanned by λ and E^{u*}_x = line spanned by τ
- $E^{c*}_x = \omega'_\pi$-symplectic orthogonal to $E^{u*}_x \oplus E^{s*}_x$.

E^{u*}_x is not symplectic orthogonal to E^{s*}_x since $\omega'_\pi(\lambda,\tau) = -\lambda \cdot \Omega_\pi(\tau)$ is strictly positive. Thus, E^{c*}_x has codimension 2 and $\mathbb{R}^{\mathcal{A}} = E^{u*}_x \oplus E^{c*}_x \oplus E^{s*}_x$.

Lemma 5.28. *The splitting $E^{u*} \oplus E^{c*} \oplus E^{s*}$ is invariant under the adjoint cocycle $F_{\mathcal{Z}}^{-1*}$. The subspace E^{u*} corresponds to the largest Lyapunov exponent θ_1, the subspace E^{s*} corresponds to the smallest Lyapunov exponent $-\theta_1$, and E^{c*} corresponds to the remaining Lyapunov exponents.*

Proof. The relations (1.11) and (2.33) imply that E^{u*} and E^{s*} are invariant under $F_{\mathcal{Z}}^{-1*}$. From Lemma 1.29 and (7.13) we get that

$$\Omega_{\pi^n}\Gamma^{-n*}(v) = \Gamma^n\Omega_\pi(v) \quad \text{for every } n \in \mathbb{Z} \text{ and } v \in \mathbb{R}^{\mathcal{A}}. \tag{5.17}$$

Since the set of combinatorial data π^n is finite, the norms of the Ω_{π^n} are uniformly bounded. Thus, taking $v = \tau$ and $n > 0$ in (7.17), and using Lemma 7.27,

$$\lim_{n\to+\infty} \frac{1}{n} \log \|\Gamma^{-n*}(\tau)\| \geq \lim_{n\to+\infty} \frac{1}{n} \log \|\Gamma^n\Omega_\pi(\tau)\| = \theta_1 \,,$$

and so the Lyapunov exponent of the cocycle $F_{\mathcal{Z}}^{-1*}$ along the invariant direction $E^{u*}_x = \mathbb{R}\tau$ is equal to θ_1. Analogously, taking $v = \lambda$ and $n < 0$ in (7.17), and then using Lemma 7.27 once more,

$$\lim_{n\to-\infty} \frac{1}{n} \log \|\Gamma^{-n*}(\lambda)\| \leq \lim_{n\to-\infty} \frac{1}{n} \log \|\Gamma^n\Omega_\pi(\lambda)\| = -\theta_1 \,.$$

and so the Lyapunov exponent of the cocycle $F_{\mathcal{Z}}^{-1*}$ along the invariant direction $E^{s*}_x = \mathbb{R}\lambda$ is equal to $-\theta_1$. It follows that E^{c*} is also invariant under $F_{\mathcal{Z}}^{-1*}$ and coincides with the sum of the remaining Oseledets subspaces.

□

5.4 Zorich cocycles and Teichmüller flows

In this section we relate the Lyapunov spectrum of the Teichmüller flow, on each connected component of stratum, to the Lyapunov spectrum of the corresponding Zorich cocycle:

Proposition 5.29. *The Lyapunov spectrum of the Teichmüller flow on any connected component $\mathcal{C}$ of a stratum $\mathcal{A}_g(m_1,\ldots,m_\kappa)$ has the form*

$$\{\pm 1 \pm \nu_i : i = 1,\ldots,g\} \cup \{1,\ldots,1\} \cup \{-1,\ldots,-1\}$$

where ± 1 appear with multiplicity $\kappa - 1$, and $\nu_i = \theta_i/\theta_1$ for $i = 1,\ldots,g$.

Proof. Let us begin by recalling the construction in Section 2.10. The pre-stratum $\hat{\mathcal{S}} = \hat{\mathcal{S}}(C)$ associated to a Rauzy class C is the quotient of the space

$$\hat{\mathcal{H}} = \{(\pi, \lambda, \tau) : \pi \in C, \lambda \in \mathbb{R}_+^{\mathcal{A}}, \tau \in T_\pi^+\}$$

by the equivalence relation generated by

$$(\pi, e^{t_R}\lambda, e^{-t_R}\tau) \sim \mathcal{R}(\pi, \lambda, \tau) = (\pi', \Theta^{-1*}(e^{t_R}\lambda), \Theta^{-1*}(e^{-t_R}\tau)), \tag{5.18}$$

where the Rauzy renormalization time $t_R = t_R(\pi, \lambda)$ is characterized by

$$|\Theta^{-1*}(e^{t_R}\lambda)| = |\lambda|\,. \tag{5.19}$$

By definition, the Teichmüller flow $\mathcal{T}^t$, $t \in \mathbb{R}$ on $\hat{\mathcal{S}}$ is the projection under the quotient map of the flow defined on $\hat{\mathcal{H}}$ by

$$(\pi, \lambda, \tau) \mapsto (\pi, e^t\lambda, e^{-t}\tau). \tag{5.20}$$

The image $\mathcal{S} \subset \hat{\mathcal{S}}$ of the subset $\mathcal{H} = \{(\pi, \lambda, \tau) \in \hat{\mathcal{H}} : |\lambda| = 1\}$ under the quotient map is a cross section for the Teichmüller flow: the return time coincides with the Rauzy renormalization time and the Poincaré return map is identified with the Rauzy-Veech renormalization $\mathcal{R} : \mathcal{H} \to \mathcal{H}$. From (7.18) and (7.20) we see that the derivative of the time-t_R map of the Teichmüller flow has the form

$$D\mathcal{T}^{t_R}(\pi, \lambda, \tau) = \begin{pmatrix} e^{t_R}\Theta^{-1*} & 0 \\ 0 & e^{-t_R}\Theta^{-1*} \end{pmatrix} : \mathbb{R}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}}. \tag{5.21}$$

We also consider a smaller cross-section $\mathcal{S}_* \subset \mathcal{S}$ which is the image under the quotient map of $Z_* = \{(\pi, \lambda, \tau) \in Z_0 \cup Z_1 : |\lambda| = 1\}$. Recall $Z_\varepsilon \subset \mathcal{H}$ is the set of all (π, λ, τ) such that (π, λ) has type ε and τ has type ε, for $\varepsilon = 0, 1$. The corresponding Poincaré map coincides with the first return map of $\mathcal{R}$ to the cross-section, which is the Zorich renormalization $\mathcal{Z}$, and the first return time is the *Zorich renormalization time*

$$t_Z = t_Z(\pi, \lambda) = \sum_{j=0}^{n(\pi,\lambda)-1} t_R(R^j(\pi, \lambda)), \tag{5.22}$$

which is also characterized by (recall (7.13) and (7.19))

$$|\Gamma^{-1*}(e^{t_Z}\lambda)| = |\lambda|\,. \tag{5.23}$$

Using (7.13) and (7.22), one immediately gets an analogue of (7.21) for the derivative of the time-t_Z map of the Teichmüller flow:

$$D\mathcal{T}^{t_Z}(\pi, \lambda, \tau) = \begin{pmatrix} e^{t_Z}\Gamma^{-1*} & 0 \\ 0 & e^{-t_Z}\Gamma^{-1*} \end{pmatrix} : \mathbb{R}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}}. \tag{5.24}$$

These matrices $P(\pi, \lambda, \tau) = D\mathcal{T}^{t_Z}(\pi, \lambda, \tau)$ define a linear cocycle over the Zorich renormalization $\mathcal{Z}$, that we denote by F_P. The nth iterate is described by

$$P^n(\pi, \lambda, \tau) = D\mathcal{T}^{t_Z^n}(\pi, \lambda, \tau) = \begin{pmatrix} e^{t_Z^n}\Gamma^{-n*} & 0 \\ 0 & e^{-t_Z^n}\Gamma^{-n*} \end{pmatrix} \tag{5.25}$$

where

$$t_Z^n = t_Z^n(\pi, \lambda) = \sum_{j=0}^{n-1} t_Z(Z^j(\pi, \lambda)).$$

We are going to relate the Lyapunov spectra of the Teichmüller flow and of the Zorich cocycle through the Lyapunov spectrum of this cocycle F_P. For this we need

Lemma 5.30. *For μ-almost every (π, λ),*

$$\lim_{n\to\infty} \frac{1}{n} t_Z^n(\pi, \lambda) = \theta_1.$$

Proof. From the definition (7.23),

$$0 \le t_Z(\pi, \lambda) = -\log |\Gamma_{\pi,\lambda}^{-1*}(\lambda)| = \log \frac{|\lambda|}{|\Gamma_{\pi,\lambda}^{-1*}(\lambda)|} \le \log \left(d \, \|\Gamma_{\pi,\lambda}^*\|\right) \tag{5.26}$$

(take the norm of a matrix to be given by the largest absolute value of the coefficients). Then, since $\|\Gamma_{\pi,\lambda}^*\| = \|\Gamma_{\pi,\lambda}\|$, Proposition 7.18 immediately implies that the function $(\pi, \lambda) \mapsto t_Z(\pi, \lambda)$ is μ-integrable. Thus, we may use the ergodic theorem to conclude that

$$\lim_{n\to\infty} \frac{1}{n} t_Z^n(\pi, \lambda) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} t_Z(Z^j(\pi, \lambda))$$

exists, at almost every point. Moreover, analogously to (7.26),

$$t_Z^n(\pi, \lambda) = -\log |\Gamma_{\pi,\lambda}^{-n*}(\lambda)| = \log \frac{|\lambda|}{|\Gamma_{\pi,\lambda}^{-n*}(\lambda)|} \le \log \left(d \, \|\Gamma_{\pi,\lambda}^{n*}\|\right) = \log \left(d \, \|\Gamma_{\pi,\lambda}^{n}\|\right)$$

for every $n \ge 1$. Consequently, applying Theorem 7.3 to the Zorich cocycle F_Z,

$$\lim_{n\to\infty} \frac{1}{n} t_Z^n(\pi, \lambda) \le \lim_{n\to\infty} \frac{1}{n} \log \|\Gamma_{\pi,\lambda}^n\| = \theta_1 \, . \tag{5.27}$$

Next, fix some compact subset K of the simplex $\Lambda_{\mathcal{A}}$ and some positive constant $c = c(K)$ such that every vector $v \in K$ satisfies $v_\alpha \ge c$ for every $\alpha \in \mathcal{A}$. By ergodicity of the Zorich renormalization (Theorem 4.2), there exist $n_j \to \infty$ for which $\lambda^{n_j}/|\lambda^{n_j}| \in K$ and so $\lambda_\alpha^{n_j} \ge c\,|\lambda^{n_j}|$ for all $\alpha \in \mathcal{A}$ and all $j \ge 1$. For these iterates,

$$t_Z^{n_j}(\pi, \lambda) = \log \frac{|\lambda|}{|\Gamma_{\pi,\lambda}^{-n_j*}(\lambda)|} \ge \log \left(c \, \|\Gamma_{\pi,\lambda}^{n_j*}\|\right) = \log \left(c \, \|\Gamma_{\pi,\lambda}^{n_j}\|\right).$$

In view of the previous observations, this implies that

$$\lim_{n\to\infty} \frac{1}{n} t_Z^n(\pi, \lambda) \ge \lim_{n\to\infty} \frac{1}{n} \log \|\Gamma_{\pi,\lambda}^n\| = \theta_1 \, . \tag{5.28}$$

The relations (7.27) and (7.28) prove the claim of the lemma. ∎

From (7.25) we immediately see that if E_x is an Oseledets subspace for F_Z^{-1*}, corresponding to a Lyapunov exponent θ, then $E_x \times \{0\}$ and $\{0\} \times E_x$ are Oseledets subspaces for F_P, corresponding to exponents

$$\theta + \lim_{n\to\infty} \frac{1}{n} t_Z^n(\pi, \lambda) \quad \text{and} \quad \theta - \lim_{n\to\infty} \frac{1}{n} t_Z^n(\pi, \lambda),$$

respectively. Therefore, using Lemma 7.30,

$$\operatorname{Lyap\,spec}(F_P) = \big(\operatorname{Lyap\,spec}(F_{\mathcal{Z}}^{-1*}) + \theta_1\big) \cup \big(\operatorname{Lyap\,spec}(F_{\mathcal{Z}}^{-1*}) - \theta_1\big)$$

(for any ergodic $\mathcal{Z}$-invariant probability). From Propositions 7.9 and 7.20 we *
get that $\operatorname{Lyap\,spec}(F_{\mathcal{Z}}^{-1*}) = \operatorname{Lyap\,spec}(F_{\mathcal{Z}})$. Thus,

$$\begin{aligned} \operatorname{Lyap\,spec}(F_P) &= (\operatorname{Lyap\,spec}(F_{\mathcal{Z}}) + \theta_1) \cup (\operatorname{Lyap\,spec}(F_{\mathcal{Z}}) - \theta_1)\,. \\ &= \{\pm\theta_1 \pm \theta_i : i = 1, \dots, g\} \cup \{\pm\theta_1, \dots, \pm\theta_1\} \end{aligned} \tag{5.29}$$

where the exponents $\pm\theta_1$ appear with multiplicity $\kappa - 1$. The definition (7.25) also gives that if E_x is an Oseledets subspace for the derivative $D\mathcal{T}^t$ of the flow, corresponding to Lyapunov exponent θ, then it is also an Oseledets subspace for F_P, corresponding to the exponent

$$\theta \lim_{n\to\infty} \frac{1}{n} t_Z^n(\pi, \lambda).$$

Using Lemma 7.30 once more, we conclude that *

$$\operatorname{Lyap\,spec}(F_P) = \theta_1 \operatorname{Lyap\,spec}(\mathcal{T}). \tag{5.30}$$

The statement of the proposition follows by combining (7.29) and (7.30). □

Notice that $1 - \nu_1 = 0 = -1 + \nu_1$ and, in view of Proposition 7.24, these are the only vanishing Lyapunov exponents for the Teichmüller flow. The corresponding Oseledets subspace may be described explicitly:

Corollary 5.31. *The vanishing Lyapunov exponents of the Teichmüller flow are associated to the invariant 2-dimensional subbundle*

$$E_x^{00} = (\mathbb{R}\lambda, 0) \oplus (0, \mathbb{R}\tau) \subset \mathbb{R}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}}, \quad x = (\pi, \lambda, \tau).$$

The dynamics on this subbundle is trivial: up to an appropriate choice of bases, $D\mathcal{T}^t \mid E_x^{00} = \mathrm{id}$ *for every* $x \in \hat{\mathcal{S}}$ *and* $t \in \mathbb{R}$. *The intersection of* E_x^{00} *with the tangent space to the hypersurfaces of constant area coincides with the flow direction.*

Proof. From the proof of Proposition 7.29 we see that the Lyapunov exponents $1-\nu_1 = 0$ and $-1+\nu_1 = 0$ arise from $(E_x^{s*}, 0)$ and $(0, E_x^{u*})$, where E_x^{s*} and E_x^{u*} are the Oseledets subbundles of $F_{\mathcal{Z}}^{-1*}$ associated, respectively, to the smallest Lyapunov exponent $-\theta_1$ and the largest Lyapunov exponent θ_1. By Lemma 7.28, $E_x^{s*} = \mathbb{R}\lambda$ and $E_x^{u*} = \mathbb{R}\tau$. This proves the first claim in the lemma.

To prove the second one, consider the basis $\{(\lambda, 0), (0, \tau)\}$ of the plane E^{00}_x, defined for each $x = (\pi, \lambda, \tau) \in \mathcal{S}$. From (7.21) we get that

$$D\mathcal{T}^{t_R}(x)(\lambda, 0) = (\lambda', 0) \quad \text{and} \quad D\mathcal{T}^{t_R}(x)(0, \tau) = (0, \tau'),$$

where $(\pi', \lambda', \tau') = \mathcal{R}(\pi, \lambda, \tau) = \mathcal{T}^{t_R}(\pi, \lambda, \tau)$. This means that $D\mathcal{T}^{t_R}(x) = \mathrm{id}$ for every $x \in \mathcal{S}$, relative to these bases. Then, since $\mathcal{R}$ is the first return map to the cross-section $\mathcal{S}$, there exists a unique extension of the basis of E^{00}_x to every x in the pre-stratum $\hat{\mathcal{S}}$, relative to which $D\mathcal{T}^t(x) = \mathrm{id}$ in all cases.

From the definition (2.30) we see that the tangent space to the hypersurfaces of constant area at each $x = (\pi, \lambda, \tau)$ is the hyperplane of all $(\dot{\lambda}, \dot{\tau}) \in \mathbb{R}^{\mathcal{A}} \times \mathbb{R}^{\mathcal{A}}$ such that

$$\dot{\lambda} \cdot \Omega_\pi(\tau) + \lambda \cdot \Omega_\pi(\dot{\tau}) = 0.$$

So, its intersection with E^{00}_x is the space of all $(a\lambda, b\tau)$, $a, b \in \mathbb{R}$ such that

$$a\lambda \cdot \Omega_\pi(\tau) + \lambda \cdot \Omega_\pi(b\tau) = 0, \quad \text{that is} \quad a + b = 0.$$

In other words, the intersection is the line $\mathbb{R}(\lambda, -\tau)$. It is clear from the form of the Teichmüller flow, that the tangent vector field is $(\pi, \lambda, \tau) \mapsto (\lambda, -\tau)$. This completes the proof. $\square$

5.5 Asymptotic flag theorem: preliminaries

We call *restricted* (respectively, *invertible restricted*) Zorich cocycle the restriction of F_Z (respectively, $F_{\mathcal{Z}}$) to the invariant subbundle $H = \{(\pi, \lambda) \times H_\pi\}$. Recall Lemma 7.21. For simplicity, we also denote these restrictions by F_Z and $F_{\mathcal{Z}}$. According to (2.53), we may consider them to act on the trivial fiber bundles

$$C \times \Lambda_{\mathcal{A}} \times H^1(M, \mathbb{R}) \quad \text{and} \quad Z_* \times H^1(M, \mathbb{R}),$$

respectively, with their adjoints F_Z^{-1*} and $F_{\mathcal{Z}}^{-1*}$ acting on

$$C \times \Lambda_{\mathcal{A}} \times H_1(M, \mathbb{R}) \quad \text{and} \quad Z_* \times H_1(M, \mathbb{R}),$$

respectively. Consider on H_π and $\mathbb{R}^{\mathcal{A}}/\ker \Omega_\pi$ the Riemann metrics induced by the canonical metric in $\mathbb{R}^{\mathcal{A}}$. Then endow $H^1(M, \mathbb{R})$ and $H_1(M, \mathbb{R})$ with the metrics transported through the identifications in (2.46) and (2.52). In view of Proposition 7.29, Theorem 7.2 may be restated as

Theorem 5.32. *The Lyapunov spectrum of every restricted Zorich cocycle is simple:* $\theta_1 > \theta_2 > \cdots > \theta_g > -\theta_g > \cdots > -\theta_2 > -\theta_1$.

An outline of the proof of this theorem will be given later in Sections 7.7 through 7.10. Here and in the next section we are going to prove Theorem 7.1 from Theorem 7.32. For this, we need to recall some terminology.

Let $\sigma = I \times \{0\}$ be the basis horizontal segment in a representation of the surface M as zippered rectangles. We have seen in Section 5.7 that to each

$\gamma(p,l)$ we may associate a vertical segment $\gamma(x,k)$ with endpoints in σ, such that the difference $[\gamma(p,l)] - [\gamma(x,k)]$ is uniformly bounded in the homology. Moreover, k and l are comparable, up to product by the area of the surface. Recall the relations (5.16) and (5.17). Up to identifying $H_1(M,\mathbb{R}) \approx \mathbb{R}^{\mathcal{A}}/\ker\Omega_\pi$ through (2.46), part 2 of Corollary 7.17 may be written as

$$[\gamma(x, z^m(x))] = \sum_{\beta\in\mathcal{A}} [v_\beta]\Gamma^m_{\alpha,\beta} = \sum_{\beta\in\mathcal{A}} [v_\beta]\Gamma^{m*}_{\beta,\alpha} = \Gamma^{m*}_{\pi,\lambda}([v_\alpha]) \tag{5.31}$$

for every $x \in J^m_\alpha$. The *annihilator* of a subspace $L \subset H_1(M,\mathbb{R})$ is the subspace $L^\perp$ of all cohomology classes $\phi \in H^1(M,\mathbb{R})$ such that $c\cdot\phi = \int_c \phi$ vanishes for every $c \in L$. Recall (2.51). Then, for any $c \in H_1(M,\mathbb{R})$ and any subspace $L \subset H_1(M,\mathbb{R})$, *

$$\operatorname{dist}(c,L) = \max\{|c\cdot\phi| : \phi \in L^\perp, \|\phi\| = 1\}.$$

Let us choose the exponents ν_i and the subspaces L_i in Theorem 7.1 as follows:

- $\nu_i = \theta_i/\theta_1$ and
- $L_i \subset H_1(M,\mathbb{R})$ is the sum of the Oseledets subspaces corresponding to the Lyapunov exponents $\theta_1,\dots,\theta_i$ of the linear cocycle $F_{\mathcal{Z}}^{-1*}$.

In view of (7.6), this means that the annihilator of L_i is the sum of the Oseledets subspaces associated to the Lyapunov exponents $\theta_{i+1},\dots,\theta_g,-\theta_g,\dots,-\theta_1$ of the linear cocycle $F_{\mathcal{Z}}$. Then, Theorem 7.1 is an immediate consequence of

Theorem 5.33. *For every $1 \le i < g$ and any $\phi \in L_i^\perp \setminus L_{i+1}^\perp$,*

$$\limsup_{k\to\infty} \frac{\log\big|[\gamma(x,k)]\cdot\phi\big|}{\log k} = \nu_{i+1} \quad \text{uniformly in } x \in \sigma. \tag{5.32}$$

Moreover, $\big|[\gamma(x,k)]\cdot\phi\big|$ is uniformly bounded, for every $\phi \in L_g^\perp$.

The proof of Theorem 7.33 occupies both this and the next section. All the arguments in the two sections are for μ-almost every (π,λ). In particular, we assume from the start that the associated interval exchange transformation is uniquely ergodic.

Preparatory results. Recall that we represent by $\{e_\alpha : \alpha\in\mathcal{A}\}$ the canonical basis of $\mathbb{R}^{\mathcal{A}}$. For each $\alpha\in\mathcal{A}$ and $m \ge 1$,

$$\Gamma^{m*}_{\pi,\lambda}(e_\alpha) = \sum_{\beta\in\mathcal{A}} \Gamma^m_{\alpha,\beta} e_\beta$$

is the α-line vector of the matrix of $\Gamma^m_{\pi,\lambda}$.

Proposition 5.34. *For every $\alpha \in \mathcal{A}$,*

$$\lim_{m\to\infty} \frac{1}{m} \log \|\Gamma^{m*}_{\pi,\lambda}(e_\alpha)\| = \theta_1.$$

Proof. The conclusion is independent of the choice of the norm: during the proof we take it to be given by the largest absolute value of the coefficients. From Theorem 7.3 we immediately get that, μ-almost everywhere,

$$\limsup_{m\to\infty} \frac{1}{m} \log \|\Gamma^{m*}_{\pi,\lambda}(e_\alpha)\| \le \lim_{m\to\infty} \frac{1}{m} \log \|\Gamma^{m}_{\pi,\lambda}\| = \theta_1. \tag{5.33}$$

So, we only have to prove the lower bound: for every $\delta > 0$ and $\alpha \in \mathcal{A}$,

$$\liminf_{m\to\infty} \frac{1}{m} \log \|\Gamma^{m*}_{\pi,\lambda}(e_\alpha)\| \ge \theta_1 - \delta, \tag{5.34}$$

μ-almost everywhere. To this end notice that, as a consequence of Corollary 7.15, the entries $\Gamma^j_{\alpha,\beta}$ of the matrix of $\Gamma^j_{\pi,\lambda}$ are eventually positive, for μ-almost every (π, λ). In particular, $\mu(V_l) \to 1$ when $l \to \infty$, where

$$V_l = \{(\pi,\lambda) : \Gamma^j_{\alpha,\beta} \ge 1 \text{ for all } \alpha, \beta \in \mathcal{A} \text{ and all } j \ge l\}.$$

Fix ε small enough so that $(\theta_1 - \varepsilon)(1 - 2\varepsilon) \ge \theta_1 - \delta$, and then let $l \ge 1$ be fixed such that $\mu(V_l) > 1 - \varepsilon$. By ergodicity,

$$\lim_{m\to\infty} \frac{1}{m} \#\{0 \le j < m : Z^j(\pi,\lambda) \in V_l\} = \mu(V_l),$$

for μ-almost all (π, λ). Thus, on a full μ-measure set we may find $N = N(\pi, \lambda)$ such that for every $m \ge N$ we have

$$\frac{1}{m} \#\{0 \le j < m : Z^j(\pi,\lambda) \in V_l\} \ge \mu(V_l) - \varepsilon \ge 1 - 2\varepsilon \tag{5.35}$$

and also, recalling (7.33),

$$\frac{1}{m} \log \|\Gamma^{m}_{\pi,\lambda}\| \in (\theta_1 - \varepsilon, \theta_1 + \varepsilon). \tag{5.36}$$

Let $n \ge 2N + l$. Taking $m = n - l$ in (7.35), we get that there exists $j(n)$ such that $(n - l)(1 - 2\varepsilon) \le j(n) \le (n - l)$ and

$$Z^{j(n)}(\pi,\lambda) \in V_l. \tag{5.37}$$

In particular, $j(n) \ge 2N(1 - 2\varepsilon) \ge N$ (assume $\varepsilon < 1/4$ from the start), and so we may take $m = j(n)$ in (7.36):

$$\log \|\Gamma^{j(n)}_{\pi,\lambda}\| \ge j(n)(\theta_1 - \varepsilon). \tag{5.38}$$

From (7.37) and $n - j(n) \ge l$ we see that the entries of $\Gamma^{n-j(n)}_{Z^{j(n)}(\pi,\lambda)}$ are all positive. Therefore, for any $\alpha \in \mathcal{A}$,

$$\|\Gamma^{n*}_{\pi,\lambda}(e_\alpha)\| = \|\Gamma^{j(n)*}_{\pi,\lambda}\big(\Gamma^{(n-j(n))*}_{Z^{j(n)}(\pi,\lambda)}(e_\alpha)\big)\| \ge \|\Gamma^{j(n)*}_{\pi,\lambda}\| = \|\Gamma^{j(n)}_{\pi,\lambda}\|\,.$$

Using (7.38) we conclude that

$$\log \|\Gamma^{n*}_{\pi,\lambda}(e_\alpha)\| \geq j(n)(\theta_1 - \varepsilon) \geq (n-l)(1-2\varepsilon)(\theta_1 - \varepsilon),$$

and so

$$\liminf_{n\to\infty} \frac{1}{n} \log \|\Gamma^{n*}_{\pi,\lambda}(e_\alpha)\| \geq (1-2\varepsilon)(\theta_1 - \varepsilon) \geq \theta_1 - \delta$$

for every $\alpha \in \mathcal{A}$. This completes the proof of (7.34) which, together with (7.33), implies the proposition. □

Proposition 5.35. *For any $\alpha \in \mathcal{A}$, there exist $0 \leq l_1 < \cdots < l_d$ such that $\{\Gamma^{l_s*}_{\pi,\lambda}(e_\alpha) : s = 1, \ldots, d\}$ is a basis of $\mathbb{R}^{\mathcal{A}}$.*

Proof. By Theorem 5.1 and Remark 5.16, for almost every (π, λ) the intersection of all $\Theta^{l*}(\mathbb{R}^{\mathcal{A}}_+)$, $l \geq 0$ coincides with the half-line spanned by λ. Since this intersection is decreasing and, by the definition (7.13), the Γ^l are a subsequence of the Θ^l, we also have that the intersection of all $\Gamma^{l*}(\mathbb{R}^{\mathcal{A}}_+)$, $l \geq 0$ coincides with $\mathbb{R}_+\lambda$. This implies that $\Gamma^{l*}(e_\alpha)$ converges to the direction of λ in the projective space, as $l \to \infty$. Let $E \subset \mathbb{R}^\alpha$ be the subspace generated by the $\Gamma^{l*}(e_\alpha)$, $l \geq 0$. Since E is a closed subset, the previous observation implies that $\lambda \in E$. Suppose the conclusion of the proposition is false, that is, the subspace E has positive codimension. Then the subspace spanned by the (integer) vectors $\Gamma^{l*}(e_\alpha)$, $l \geq 0$ inside $\mathbb{Q}^{\mathcal{A}}$ also has positive codimension. Let $q \in \mathbb{Q}^{\mathcal{A}}$ be a non-zero vector in the orthogonal complement to this subspace. Then every vector in E is rationally dependent:

$$\sum_{\alpha\in\mathcal{A}} q_\alpha v_\alpha = 0 \quad \text{for every } v \in E.$$

This is a contradiction, because λ is rationally independent (for almost every λ). This contradiction proves that the $\Gamma^{l*}(e_\alpha)$, $l \geq 0$ span the whole $\mathbb{R}^{\mathcal{A}}$. Thus, we may choose $l_1 < \cdots < l_d$ as in the statement. □

Special subsequence. The proof of Theorem 7.33 is long and combinatorially subtle. In order to motivate the strategy and help the reader keep track of what is going on, we begin by stating and proving a special case where k runs only over the subsequence of Zorich induction times $z^m(x)$: the arguments are much more direct in this setting, while having the same flavor as the actual proof. This special case is not really used in the sequel, so the reader may also choose to proceed immediately to the next section.

We represent by J^m the domain of the mth Zorich induction $\hat{Z}^m(f)$, for any $m \geq 1$, and by $\{J^m_\alpha \colon \alpha \in \mathcal{A}\}$ the corresponding partition into subintervals. Corresponding to the case $m = 0$, we let $J = I$ and $J_\alpha = I_\alpha$ for any $\alpha \in \mathcal{A}$.

Proposition 5.36. *For every $1 \leq i < g$ and $\phi \in L_i^\perp \setminus L_{i+1}^\perp$,*

$$\limsup_{m\to\infty} \frac{\log \left|[\gamma(x, z^m(x))] \cdot \phi\right|}{\log z^m(x)} = \nu_{i+1} \quad \textit{uniformly in } x \in J^m.$$

This is an immediate consequence of Lemmas 7.37 through 7.39 that follow.

Lemma 5.37.

$$\lim_{m\to\infty} \frac{1}{m} \log z^m(x) = \theta_1 \quad \textit{uniformly in } x \in J^m.$$

Proof. By Corollary 7.17, $z^m(x) = \sum_{\beta\in\mathcal{A}} \Gamma^m_{\alpha,\beta}$ for every $\alpha \in \mathcal{A}$ and $x \in J^m_\alpha$. Consequently,

$$\min_{\alpha\in\mathcal{A}} \|\Gamma^{m*}(e_\alpha)\| \leq z^m(x) \leq d \max_{\alpha\in\mathcal{A}} \|\Gamma^{m*}(e_\alpha)\|$$

for every $x \in J^m$ (take the norm of a vector to be given by the largest absolute value of its coefficients). Proposition 7.34 asserts that $m^{-1} \log \|\Gamma^{m*}(e_\alpha)\|$ converges to θ_1 for every $\alpha \in \mathcal{A}$. Using this fact on the left hand side and on the right hand side of the previous formula we get that $m^{-1} \log z^m(x)$ converges uniformly to θ_1, as claimed. □

Lemma 5.38. *For every $1 \leq i < g$ and $\phi \in H^1(M,\mathbb{R}) \setminus L^\perp_{i+1}$,*

$$\limsup_{m\to\infty} \frac{1}{m} \log \left|[\gamma(x, z^m(x))] \cdot \phi\right| \geq \theta_{i+1} \quad \textit{uniformly in } x \in J^m.$$

Proof. By (7.31), we have $[\gamma(x, z^m(x)] = \Gamma^{m*}([v_\alpha])$ for every $x \in J^m_\alpha$. Take $l_1 < \ldots < l_d$ as in Proposition 7.35, such that $\Gamma^{l_s*}(e_\alpha)$, $s = 1, \ldots, d$ generate $\mathbb{R}^\mathcal{A}$. Since the Zorich cocycles are locally constant (recall the proof of Proposition 7.24), we may find a simplex $D \subset \Lambda_\mathcal{A}$ such that all $(\pi, \lambda') \in \{\pi\} \times D$ share the same Γ^{l_s*} for $s = 1, \ldots, d$. By Poincaré recurrence, there exist infinitely many iterates $0 < k_1 < k_2 < \cdots$ such that $Z^{k_j}(\pi, \lambda) \in \{\pi\} \times D$. Since

$$\Gamma^{l_s*}([v_\alpha]), \quad s = 1, \ldots, d$$

generate $H_1(M,\mathbb{R})$, there exists $c_1 = c_1(\alpha) > 0$ and for each k_j we may find l_s, $s = s(j)$ such that

$$\left|\Gamma^{k_j}(\phi) \cdot \Gamma^{l_s*}([v_\alpha])\right| \geq c_1 \|\Gamma^{k_j}(\phi)\|.$$

This relation may be rewritten as

$$\left|[\gamma(x, z^{m_j}(x))] \cdot \phi\right| = \left|\Gamma^{m_j*}([v_\alpha]) \cdot \phi\right| = \left|\Gamma^{l_s*}([v_\alpha]) \cdot \Gamma^{k_j}(\phi)\right| \geq c_1 \|\Gamma^{k_j}(\phi)\|,$$

where $m_j = k_j + l_s$. The condition $\phi \notin L^\perp_{i+1}$ means that ϕ is outside the sum of the Oseledets subspaces associated to the exponents $\theta_{i+2}, \ldots, \theta_g, -\theta_g, \ldots, -\theta_1$ of the cocycle $F_\mathcal{Z}$. So, for any $\varepsilon > 0$ we may find $c_2 = c_2(\varepsilon) > 0$ such that

$$\|\Gamma^k(\phi)\| \geq c_2 e^{(\theta_{i+1}-\varepsilon)k} \|\phi\| \quad \text{for all } k \geq 0.$$

Combining the last inequalities we obtain $\left|[\gamma(x, z^{m_j}(x))] \cdot \phi\right| \geq c e^{(\theta_{i+1}-\varepsilon)k_j}$, where $c = c_1 c_2 \|\phi\|$ depends only on α, ε, and ϕ. It is clear that $m_j/k_j \to 1$ as $j \to \infty$, since l_s takes only finitely many values. So this last inequality implies the conclusion of the lemma. □

Lemma 5.39. *For every* $1 \le i < g$ *and* $\phi \in L_i^\perp$,

$$\limsup_{m\to\infty} \frac{1}{m} \log \bigl|[\gamma(x, z^m(x))] \cdot \phi\bigr| \le \theta_{i+1} \quad \textit{uniformly in } x \in J^m.$$

Proof. From the relation (7.31) we get that, for any $x \in J_\alpha^m$,

$$[\gamma(x, z^m(x))] \cdot \phi = \Gamma^{m*}([v_\alpha]) \cdot \phi = [v_\alpha] \cdot \Gamma^m(\phi).$$

The condition $\phi \in L_i^\perp$ means ϕ belongs to the sum of the Oseledets subspaces associated to the exponents $\theta_{i+1}, \ldots, \theta_g, -\theta_g, \ldots, -\theta_1$ of the cocycle $F_{\mathcal{Z}}$. Hence, given any $\varepsilon > 0$, there exists $c_3 = c_3(\pi, \lambda, \varepsilon) > 0$ such that

$$\|\Gamma^m(\phi)\| \le c_3 e^{(\theta_{i+1}+\varepsilon)m} \|\phi\| \quad \text{for all } m \ge 1.$$

Combining these observations we find that

$$|[\gamma(x, z^m(x))] \cdot \phi| \le C e^{(\theta_{i+1}+\varepsilon)m} \|\phi\| \quad \text{for all } m \ge 1, \tag{5.39}$$

where C is the product of c_3 by an upper bound for the norm of every $[v_\alpha]$. This proves the lemma. $\square$

Remark 5.40. The arguments in Lemma 7.39 remain valid for $i = g$: in the place of (7.39), one obtains

$$|[\gamma(x, z^m(x))] \cdot \phi| \le C e^{(-\theta_g+\varepsilon)m} \|\phi\| \quad \text{for all } m \ge 1.$$

Since $-\theta_g < 0$, this implies that $|[\gamma(x, z^m(x))] \cdot \phi|$ converges to zero as $m \to \infty$, uniformly in $x \in J^m$.

5.6 Asymptotic flag theorem: proof

In this section we prove the full statement of Theorem 7.33. For the reader's convenience, we split the arguments into three main steps, that are presented in the next three subsections:

Preparation. Given $x \in \sigma$ and $k \ge 1$, let $m = m(x, k)$ be the largest integer such that the orbit segment $f^j(x)$, $0 \le j \le k$ hits the interval J^m at least twice. That is,

$$m = m(x, k) = \max\bigl\{l \ge 0 : \#\{0 \le j \le k : f^j(x) \in J^l\} \ge 2\bigr\}. \tag{5.40}$$

Note that if $x \in J^n$ then $m(x, z^n(x)) \ge n$, because $z^n(x)$ is the first return time to J^n. Thus, the next result is a kind of extension of Lemma 7.37:

Lemma 5.41.

$$\lim_{k\to\infty} \frac{\log k}{m(x, k)} = \theta_1 \quad \textit{uniformly in } x \in \sigma.$$

Proof. Let $x_j = f^j(x)$, where $j \geq 0$ is the first time x hits J^m. By the definition of m, the orbit of x_j returns to J^m before time $k - j$. So, using part 3 of Corollary 7.17,

$$k \geq k - j \geq z^m(x_j) \geq \min_{\alpha \in \mathcal{A}} \sum_{\beta \in \mathcal{A}} \Gamma^m_{\alpha,\beta} \geq \min_{\alpha \in \mathcal{A}} \|\Gamma^{m*}(e_\alpha)\| . \tag{5.41}$$

By the definition of m, the orbit segment $f^j(x)$, $0 \leq j \leq k$ intersects J^{m+1} at most once. Suppose for a while that, in fact, there is no intersection. Since we take the interval exchange f to be minimal, there are iterates $-r < 0 < k < s$ such that $f^{-r}(x)$ and $f^s(x)$ belong to J^{m+1}. Take r and s smallest and denote $x_{-r} = f^{-r}(x)$. Then, using once more part 3 of Corollary 7.17,

$$k \leq r + s = z^{m+1}(x_{-r}) \leq \max_{\alpha \in \mathcal{A}} \sum_{\beta \in \mathcal{A}} \Gamma^{m+1}_{\alpha,\beta} \leq d \max_{\alpha \in \mathcal{A}} \|\Gamma^{(m+1)*}(e_\alpha)\| .$$

In general, if the orbit segment $f^j(x)$, $0 \leq j \leq k$ does intersect J^{m+1}, we may apply the previous argument to the subsegments before and after the intersection. In this way we find that

$$k \leq 2 \max_{\alpha \in \mathcal{A}} \sum_{\beta \in \mathcal{A}} \Gamma^{m+1}_{\alpha,\beta} \leq 2d \max_{\alpha \in \mathcal{A}} \|\Gamma^{(m+1)*}(e_\alpha)\| . \tag{5.42}$$

This relation also ensures that m goes to infinity, uniformly in x, when k goes to infinity. Now let $\varepsilon > 0$. By Proposition 7.34 there is $n_\varepsilon > 0$ such that

$$\frac{1}{n} \log \|\Gamma^{n*}(e_\alpha)\| \in \big(\theta_1 - \varepsilon, \theta_1 + \varepsilon\big) \quad \text{for all } n \geq n_\varepsilon \text{ and } \alpha \in \mathcal{A}.$$

Assume k is large enough to ensure $m \geq n_\varepsilon$. Then (7.41) and (7.42) yield

$$m(\theta_1 - \varepsilon) \leq \log k \leq \log(2d) + (m+1)(\theta_1 + \varepsilon).$$

Dividing by m and passing to the limit as $k \to \infty$, we obtain

$$\theta_1 - \varepsilon \leq \liminf_{k \to \infty} \frac{\log k}{m} \leq \limsup_{k \to \infty} \frac{\log k}{m} \leq \theta_1 + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves the claim in the lemma. □

To complete the proof of Theorem 7.33 we only need the following three propositions (compare Lemmas 7.38 and 7.39 and Remark 7.40).

Proposition 5.42. *For every $\phi \in H^1(M, \mathbb{R}) \setminus L^\perp_{i+1}$ and $1 \leq i < g$,*

$$\limsup_{k \to \infty} \frac{1}{m(x,k)} \log \big|[\gamma(x,k)] \cdot \phi\big| \geq \theta_{i+1} \quad \textit{uniformly in } x \in \sigma.$$

Proposition 5.43. *For every $\phi \in L^\perp_i$ and $1 \leq i < g$,*

$$\limsup_{k \to \infty} \frac{1}{m(x,k)} \log \big|[\gamma(x,k)] \cdot \phi\big| \leq \theta_{i+1} \quad \textit{uniformly in } x \in \sigma.$$

Proposition 5.44. *There exists $C > 0$ such that $\big|[\gamma(x,k)] \cdot \phi\big| \le C\|\phi\|$ for any $x \in \sigma$, any $k \ge 1$, and any $\phi \in L_g^{\perp}$.*

The proofs are given in the next two subsections. Before that, we need to introduce some terminology. Given any $x \in \sigma$ and $k \ge 1$, define

$$s(x,k;\pi,\lambda) = \sum_{\beta \in \mathcal{A}} s_\beta(x,k;\pi,\lambda) e_\beta$$

where $\{e_\beta : \beta \in \mathcal{A}\}$ is the canonical basis of $\mathbb{R}^{\mathcal{A}}$ and each $s_\beta(x,k;\pi,\lambda)$ is the number of visits of x to the subinterval I_β before time k:

$$s_\beta(x,k;\pi,\lambda) = \#\{0 \le j < k : f^j(x) \in I_\beta\}.$$

Observe that, whenever $x \in J_\alpha^m$, part 1 of Corollary 7.17 gives

$$s_\beta(x,z^m(x);\pi,\lambda) = \#\{0 \le j < z^m(x) : f^j(x) \in I_\beta\} = \Gamma^m_{\alpha,\beta},$$

for all $\beta \in \mathcal{A}$. Equivalently,

$$s(x,z^m(x);\pi,\lambda) = \Gamma^{m*}(e_\alpha). \tag{5.43}$$

Observe, in addition, that $s(x,k;\pi,\lambda) \in \mathbb{R}^{\mathcal{A}}$ corresponds to the homology class $[\gamma(x,k)]$ under the identification (2.46). In what follows, $v \in H_\pi$ is the vector corresponding to $\phi \in H^1(M,\mathbb{R})$ under the identification (2.52). Then,

$$[\gamma(x,k)] \cdot \phi = s(x,k;\pi,\lambda) \cdot v. \tag{5.44}$$

Lower bound. For the proof of Proposition 7.42 we need the following auxiliary result:

Lemma 5.45. *Let $\alpha \in \mathcal{A}$ be the first symbol on the top line of π. Then there exists $r \ge 1$ and a sequence $(n_j)_j \to \infty$ such that*

$$\liminf_{j\to\infty} \frac{1}{n_j} \log |\Gamma^{n_j *}_{\pi,\lambda}(e_\alpha) \cdot v| \ge \theta_{i+1} \quad \textit{and} \quad J^{n_j+r} \subset J^{n_j}_\alpha \textit{ for all } j \ge 1.$$

Proof. The condition $\phi \notin L_{i+1}^{\perp}$ means that ϕ (thus, v) is outside the sum of the Oseledets subspaces associated to the Lyapunov exponents θ_{i+1}, …, θ_g, $-\theta_g$, …, $-\theta_1$ of the cocycle $F_{\mathcal{Z}}$. So, for any $\varepsilon > 0$, there exists $c_0 = c_0(\pi,\lambda,\varepsilon) > 0$ such that

$$\|\Gamma^l_{\pi,\lambda}(v)\| \ge c_0 e^{(\theta_{i+1}-\varepsilon)l}\|v\| \quad \text{for every } l \ge 1. \tag{5.45}$$

By Proposition 7.35, there exist $l_1 < \cdots < l_d$ such that

$$\Gamma^{l_s *}_{\pi,\lambda}(e_\alpha), \quad s = 1,\ldots,d \quad \text{is a basis of } \mathbb{R}^{\mathcal{A}}. \tag{5.46}$$

It follows from the definition of the induction operator, in (1.4)–(1.6), that the first symbol on the top (or bottom) line of π^n is always α, for all n. Thus, the left endpoint of J^n_α coincides with $\partial J^n = 0$ for every n. By Corollary 1.20, the diameter of J^n goes to zero as $n \to \infty$. Then, there exists $r \geq 1$ such that

$$J^{l_s+r} \subsetneqq J^{l_s}_\alpha \quad \text{for every } s = 1, \ldots, d. \tag{5.47}$$

By continuity, (7.47) remains valid for any $(\hat{\pi}, \hat{\lambda})$ in a small neighborhood U of (π, λ). Reducing U if necessary, we may suppose that

$$\Gamma^{l_s *}_{\hat{\pi},\hat{\lambda}}(e_\alpha) = \Gamma^{l_s *}_{\pi,\lambda}(e_\alpha), \quad \text{for every } (\hat{\pi}, \hat{\lambda}) \in U \text{ and } s = 1, \ldots, d. \tag{5.48}$$

By Poincaré recurrence, there exist infinitely many iterates $t_1 < \cdots < t_j < \cdots$ such that

$$Z^{t_j}(\pi, \lambda) \in U.$$

In view of (7.46), there exists $c_1 > 0$ and for each j there exists some $s = s(j)$ such that

$$\|\Gamma^{l_s *}_{\pi,\lambda}(e_\alpha) \cdot \Gamma^{t_j}_{\pi,\lambda}(v)\| \geq c_1 \|\Gamma^{t_j}_{\pi,\lambda}(v)\| .$$

Take $n_j = t_j + l_s$. In view of (7.48), the previous inequality leads to

$$|\Gamma^{n_j *}_{\pi,\lambda}(e_\alpha) \cdot v| = |\Gamma^{l_s *}_{Z^{t_j}(\pi,\lambda)}(e_\alpha) \cdot \Gamma^{t_j}_{\pi,\lambda}(v)| = |\Gamma^{l_s *}_{\pi,\lambda}(e_\alpha) \cdot \Gamma^{t_j}_{\pi,\lambda}(v)| \geq c_1 \|\Gamma^{t_j}_{\pi,\lambda}(v)\| .$$

Combined with (7.45), this gives that

$$|\Gamma^{n_j *}_{\pi,\lambda}(e_\alpha) \cdot v| \geq c_0 c_1 e^{(\theta_{i+1} - \varepsilon) t_j} \|v\| \quad \text{for every } j \geq 1.$$

Clearly, $|n_j - t_j| \leq \max\{l_s : s = 1, \ldots, d\}$ for all $j \geq 1$, and so t_j / n_j converges to 1 as $j \to \infty$. So, the previous inequality implies that there exists $c_2 = c_2(\varepsilon)$ such that

$$|\Gamma^{n_j *}_{\pi,\lambda}(e_\alpha) \cdot v| \geq c_2 e^{(\theta_{i+1} - \varepsilon) t_j} \|v\| \quad \text{for every } j \geq 1.$$

This proves the first claim in the lemma. To prove the second one, observe that

$$J^n = [0, |\hat{\lambda}^n|) \quad \text{and} \quad J^n_\alpha = [0, \hat{\lambda}^n_\alpha) \quad \text{for all } n \geq 1,$$

where $(\pi^n, \hat{\lambda}^n) = \hat{Z}^n(\pi, \lambda)$. Keep in mind that $Z^n(\pi, \lambda) = (\pi^n, \lambda^n)$ for all n, where $\lambda^n = \hat{\lambda}^n / |\hat{\lambda}^n|$. Denote $(\pi^{n,l}, \hat{\lambda}^{n,l}) = \hat{Z}^l(\pi^n, \lambda^n)$, for every $n \geq 1$ and $l \geq 0$. Then

$$\hat{\lambda}^{n,l} = \frac{\hat{\lambda}^{n+l}}{|\hat{\lambda}^n|} \quad \text{for every } n \geq 1 \text{ and } l \geq 0.$$

The relation (7.47), applied to the points $Z^{t_j}(\pi, \lambda) \in U$, means that

$$|\hat{\lambda}^{t_j, l_s + r}| < \hat{\lambda}^{t_j, l_s}_\alpha \quad \text{for all } j \geq 1.$$

Multiplying both sides by $|\hat{\lambda}^{t_j}|$ we obtain that

$$|\hat{\lambda}^{t_j + l_s + r}| < \hat{\lambda}^{t_j + l_s}_\alpha$$

and this implies that $J^{n_j + r} \subset J^{n_j}_\alpha$, for all $j \geq 1$. □

Proof of Proposition 7.42. Given $r \geq 1$ and $(n_j)_j$ as in Lemma 7.45, let us define $p_j = p_j(x)$ to be the first time the orbit of x hits the interval J^{n_j+r}, that is,

$$p_j = \min\{n \geq 0 : f^n(x) \in J^{n_j+r}\}.$$

It is clear from the definition (7.40) that $m(x, p_j) \leq n_j + r$, and so

$$\begin{aligned}\limsup_{j\to\infty} \frac{1}{m(x,p_j)} \log |s(x, p_j; \pi, \lambda) \cdot v| &\geq \limsup_{j\to\infty} \frac{1}{n_j + r} \log |s(x, p_j; \pi, \lambda) \cdot v| \\ &= \limsup_{j\to\infty} \frac{1}{n_j} \log |s(x, p_j; \pi, \lambda) \cdot v| .\end{aligned}$$

If the limit on the right hand side is greater or equal than θ_{i+1} then the same is true for the limit on the left hand side which, in view of (7.44), implies that the conclusion of the proposition holds. So, we may assume that the limit is strictly less than θ_{i+1}: there exist $a > 0$ and $c_3 > 0$ such that

$$|s(x, p_j; \pi, \lambda) \cdot v| \leq c_3 e^{n_j(\theta_{i+1}-a)} \|v\| \quad \text{for all } j \geq 1. \tag{5.49}$$

Then let $q_j = q_j(x)$ be the first time the orbit of x returns to J^{n_j} after time p_j:

$$q_j = \min\{n > p_j : f^n(x) \in J^{n_j}\}.$$

In other words, $q_j = p_j + z^{n_j}(x_j)$, where $x_j = f^{p_j}(x)$. Clearly,

$$s(x, q_j; \pi, \lambda) = s(x, p_j; \pi, \lambda) + s(x_j, z^{n_j}(x_j); \pi, \lambda).$$

By construction, $x_j \in J^{n_j+r} \subset J_\alpha^{n_j}$. Thus, using (7.43), this relation may be rewritten as

$$s(x, q_j; \pi, \lambda) = s(x, p_j; \pi, \lambda) + \Gamma_{\pi,\lambda}^{n_j *}(e_\alpha).$$

It follows, using (7.45) and (7.49), that

$$\begin{aligned}|s(x, q_j; \pi, \lambda) \cdot v| &\geq |\Gamma_{\pi,\lambda}^{n_j *}(e_\alpha) \cdot v| - |s(x, p_j; \pi, \lambda) \cdot v| \\ &\geq c_0 e^{n_j(\theta_{i+1}-\varepsilon)} \|v\| - c_3 e^{n_j(\theta_{i+1}-a)} \|v\| .\end{aligned}$$

Taking $\varepsilon < a$, this implies that there exists $c_4 = c_4(\pi, \lambda, \varepsilon) > 0$ such that

$$|s(x, q_j; \pi, \lambda) \cdot v| \geq c_4 e^{n_j(\theta_{i+1}-\varepsilon)} \|v\|$$

for all $j \geq 1$. In view of (7.44), this implies that

$$\limsup_{k\to\infty} \frac{1}{m} \log \left|[\gamma(x, k)] \cdot \phi\right| = \limsup_{k\to\infty} \frac{1}{m} \log \left|s(x, k; \pi, \lambda) \cdot v\right| \geq \theta_{i+1} - \varepsilon,$$

uniformly. Proposition 7.43 follows, since $\varepsilon > 0$ is arbitrary. $\square$

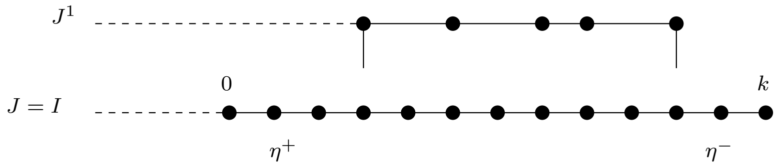


Figure 5.6:

Upper bound. The strategy to prove Proposition 7.43 is to stratify the orbit segment $f^j(x)$, $0 \le j \le k$ according to increasing renormalization depth, relating each stratification level to some subsegment that starts and ends at returns to a domain J^l. Let us explain this in more detail, with the help of Figure 7.6.

Recall J^1 denotes the domain of the Zorich induction $\hat{Z}(f)$ of the transformation f. Given $x \in \sigma$ and $k \ge 1$, define

$$\begin{aligned} \eta^+ &= \eta^+(x, k; \pi, \lambda) = \min\{j \ge 0 : f^j(x) \in J^1\} \quad \text{and} \\ \eta^- &= \eta^-(x, k; \pi, \lambda) = \min\{j \ge 0 : f^{k-j}(x) \in J^1\}. \end{aligned} \tag{5.50}$$

In other words, η^+ is the first time and $k - \eta^-$ is the last time the orbit segment hits the interval J^1. Denote $x_1 = f^{\eta^+}(x)$. Then, time $k - \eta^+ - \eta^-$ is a return of the point x_1 to the interval J^1 under the map f, and so

$$f^{k-\eta^+-\eta^-}(x_1) = \hat{Z}(f)^{k_1}(x_1) \tag{5.51}$$

for some $k_1 \ge 1$. It is clear that

$$\begin{aligned} s(x, k; \pi, \lambda) = s(x_1, k - \eta^+ - \eta^-; \pi, \lambda) \\ + s(x, \eta^+; \pi, \lambda) + s(f^{k-\eta^-}(x), \eta^-; \pi, \lambda). \end{aligned} \tag{5.52}$$

Compare Figure 7.6. The first term on the right hand side will be estimated through the following recurrence relation:

Lemma 5.46. *For every $x \in \sigma$ and $k \ge 1$,*

$$s(x_1, k - \eta^+ - \eta^-; \pi, \lambda) = \Gamma^*_{\pi,\lambda}\left(s(x_1, k_1; \pi^1, \lambda^1)\right)$$

* *where $(\pi^1, \lambda^1) = Z(\pi, \lambda)$ and the number $k_1 \ge 1$ is defined by (7.51).*

Proof. Denote $g = Z(f)$. By (7.51), we have $f^{k-\eta^+-\eta^-}(x_1) = g^{k_1}(x_1)$. Clearly,

$$s_\beta(x_1, k - \eta^+ - \eta^-; \pi, \lambda) = \sum_{i=0}^{k_1-1} s_\beta(g^i(x_1), z^1(g^i(x_1)); \pi, \lambda),$$

for every $\beta \in \mathcal{A}$. By part 1 of Corollary 7.17,

$$s_\beta(g^i(x_1), z^1(g^i(x_1)); \pi, \lambda) = \#\{0 \le j < z^1(g^i(x_1)) : f^j(g^i(x_1)) \in I_\beta\} = \Gamma_{\alpha,\beta}$$

whenever $g^i(x_1) \in J^1_\alpha$. Replacing in the previous relation,

$$\begin{aligned} s_\beta(x_1, k - \eta^+ - \eta^-; \pi, \lambda) &= \sum_{\alpha \in \mathcal{A}} \#\{0 \le i < k_1 : g^i(x_1) \in J^1_\alpha\} \, \Gamma_{\alpha,\beta}. \\ &= \sum_{\alpha \in \mathcal{A}} \Gamma_{\alpha,\beta} s_\alpha(x_1, k_1; \pi^1, \lambda^1). \end{aligned}$$

This means that $s(x_1, k - \eta^+ - \eta^-; \pi, \lambda) = \Gamma^* \left(s(x_1, k_1; \pi^1, \lambda^1)\right)$, as claimed. □

The sum of the last two terms in (7.52) will be bounded using the next lemma. Recall we take the norm of a vector to be given by the largest absolute value of its coefficients.

Lemma 5.47. *For every $x \in \sigma$, $k \ge 1$, and $l \ge 1$,*

$$\|s(x, \eta^+; \pi, \lambda) + s(f^{k-\eta^-}(x), \eta^-; \pi, \lambda)\| \le 2\|\Gamma_{\pi,\lambda}\| \, .$$

Proof. Take $r \ge 0$ minimum such that $\bar{x} = f^{-r}(x) \in J^1$. This is well defined, since the interval exchange f is minimal. Then $r + \eta^+$ is the first return time of $\bar{x}$ to J^1, that is, $r + \eta^+ = z(\bar{x})$. Clearly, $s_\beta(x, \eta^+; \pi, \lambda) \le s_\beta(\bar{x}, z(\bar{x}); \pi, \lambda)$ for every $\beta \in \mathcal{A}$. From part 1 of Corollary 7.17 we get that

$$s_\beta(\bar{x}, z(\bar{x}); \pi, \lambda) = \#\{0 \le j < z(\bar{x}) : f^j(\bar{x}) \in I_\beta\} \le \max_{\alpha \in \mathcal{A}} \Gamma_{\alpha,\beta}$$

for every $\beta \in \mathcal{A}$. Therefore,

$$\|s(x, \eta^+; \pi, \lambda)\| \le \|s(\bar{x}, z(\bar{x}); \pi, \lambda)\| \le \max_{\alpha,\beta \in \mathcal{A}} \Gamma_{\alpha,\beta} = \|\Gamma\|.$$

Analogously, $\|s(f^{k-\eta^-}(x), \eta^-; \pi, \lambda)\| \le \|\Gamma\|$. The lemma follows. □

Replacing Lemmas 7.46 and 7.47 in (7.52), we obtain that

$$s(x, k; \pi, \lambda) = \Gamma^*_{\pi,\lambda}\big(s(x_1, k_1; \pi^1, \lambda^1)\big) + r(x, k; \pi, \lambda) \tag{5.53}$$

with $\|r(x, k; \pi, \lambda)\| \le 2\|\Gamma_{\pi,\lambda}\|$, for every $x \in \sigma$ and $k \ge 1$.

Applying this relation to the orbit segment $Z(f)^i(x_1)$, $0 \le i < k_1$, we obtain

$$s(x_1, k_1; \pi^1, \lambda^1) = \Gamma^*_{\pi^1,\lambda^1}\big(s(x_2, k_2; \pi^2, \lambda^2)\big) + r(x_1, k_1; \pi^1, \lambda^1),$$

where $(\pi^2, \lambda^2) = Z^2(\pi, \lambda)$ and $\|r(x_1, k_1; \pi^1, \lambda^1)| \le 2\|\Gamma_{\pi^1,\lambda^1}\|$. Thus,

$$\begin{aligned} s(x, k; \pi, \lambda) &= \Gamma^*_{\pi,\lambda} \cdot \left[\Gamma^*_{\pi^1,\lambda^1}\big(s(x_2, k_2; \pi^2, \lambda^2)\big) + r(x_1, k_1; \pi^1, \lambda^1)\right] + r(x, k; \pi, \lambda) \\ &= \Gamma^{2*}_{\pi,\lambda}\big(s(x_2, k_2; \pi^2, \lambda^2)\big) + \Gamma^*_{\pi,\lambda}\big(r(x_1, k_1; \pi^1, \lambda^1)\big) + r(x, k; \pi, \lambda). \end{aligned}$$

Write $(\pi^j, \lambda^j) = Z^j(\pi, \lambda)$ for $j \ge 0$. Repeating this procedure m times, we obtain (compare Figure 7.7)

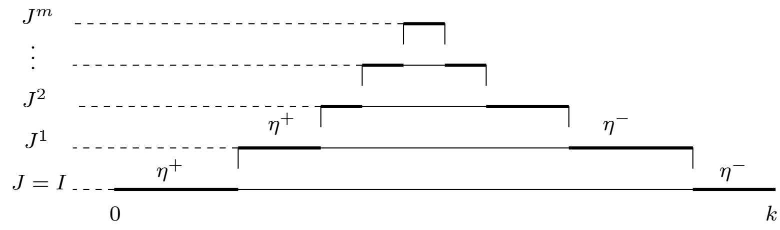


Figure 5.7:

Lemma 5.48. *For every $x \in \sigma$ and $k \geq 1$,*

$$s(x, k; \pi, \lambda) = \Gamma^{m*}_{\pi,\lambda}\big(s(x_m, k_m; \pi^m, \lambda^m)\big) + \sum_{j=0}^{m-1} \Gamma^{j*}_{\pi,\lambda}\big(r(x_j, k_j; \pi^j, \lambda^j)\big)$$

with $\|s(x_m, k_m; \pi^m, \lambda^m)\| \leq 2\|\Gamma_{\pi^m,\lambda^m}\|$ *and* $\|r(x_j, k_j; \pi^j, \lambda^j)\| \leq 2\|\Gamma_{\pi^j,\lambda^j}\|$ *for every* $0 \leq j \leq m-1$.

Proof. All that is left to prove is the bound on the norm of $s(x_m, k_m; \pi^m, \lambda^m)$. Let $\hat{\Gamma} = \Gamma_{\pi^m,\lambda^m}$ and let $\hat{\Gamma}_{\alpha,\beta}$, $\alpha, \beta \in \mathcal{A}$ be its coefficients. Denote $g = Z^m(f)$. The definition of m implies that the orbit segment $g^j(x_m)$, $0 \leq j \leq k_m$ intersects J^{m+1} at most once. Suppose first that there is no intersection. Since g is minimal, there exist $-r < 0 \leq k_m < s$ such that both $x_{-r} = g^{-r}(x_m)$ and $x_s = g^s(x_m)$ are in J^{m+1}. Take r and s minimum. Then $r+s$ coincides with the first Zorich inducing time $z_{\pi^m,\lambda^m}(x_{-r})$ of the point x_{-r} for the transformation g. So, using part 1 of Corollary 7.17,

$$s_\beta(x_m, k_m; \pi^m, \lambda^m) \leq s_\beta(x_{-r}, r+s; \pi^m, \lambda^m) \leq \max_{\alpha \in \mathcal{A}} \hat{\Gamma}_{\alpha,\beta}$$

for every $\beta \in \mathcal{A}$. It follows that

$$\|s(x_m, k_m; \pi^m, \lambda^m)\| = \max_{\beta \in \mathcal{A}} s_\beta(x_m, k_m; \pi^m, \lambda^m) \leq \max_{\alpha,\beta \in \mathcal{A}} \hat{\Gamma}_{\alpha,\beta} = \|\hat{\Gamma}\|.$$

If $g^j(x)$, $0 \leq j \leq k_p$ does intersect J^{m+1}, we may apply the same argument as before to the subsegments before and after the intersection. Then, adding the two bounds, we find that $\|s(x_m, k_m; \pi^m, \lambda^m)\| \leq 2\|\hat{\Gamma}\|$, as claimed. ☐

Proof of Proposition 7.43. The condition $\phi \in L_i^\perp$ means that ϕ (thus, v) belongs to the sum of the Oseledets subspaces associated to the exponents $\theta_{i+1}, \ldots, \theta_g, -\theta_g, \ldots, -\theta_1$ of the cocycle $F_{\mathcal{Z}}$. Hence, given any $\varepsilon > 0$, there exists $c_0 = c_0(\pi, \lambda, \varepsilon)$ such that

$$\|\Gamma^j_{\pi,\lambda}(v)\| \leq c_0 e^{(\theta_{i+1}+\varepsilon)j} \|v\| \quad \text{for every } j \geq 0. \tag{5.54}$$

By Proposition 7.18, the function $\phi(\tilde{\pi}, \tilde{\lambda}) = \log \|\Gamma_{\tilde{\pi},\tilde{\lambda}}\|$ is μ-integrable. So, we may apply Remark 7.7 to conclude that, for any $\varepsilon > 0$ there is $c_1 = c_1(\pi, \lambda, \varepsilon)$

such that

$$\|r(x_j, k_j; \pi^j, \lambda^j)\| \leq 2\|\Gamma_{\pi^j,\lambda^j}\| \leq c_1 e^{\varepsilon j} \quad \text{for every } j \geq 0. \tag{5.55}$$

Using Lemma 7.48, we find that

$$s(x, k; \pi, \lambda) \cdot v = s(x_m, k_m; \pi^m, \lambda^m) \cdot \Gamma^m_{\pi,\lambda}(v) + \sum_{j=0}^{m-1} r(x_j, k_j; \pi^j, \lambda^j) \cdot \Gamma^j_{\pi,\lambda}(v).$$

and so, using also (7.54) and (7.55),

$$\big|s(x, k; \pi, \lambda) \cdot v\big| \leq \sum_{j=0}^{m} c_0 e^{(\theta_{i+1}+\varepsilon)j} \|v\|\, c_1 e^{\varepsilon j} = c_0 c_1 \|v\| \sum_{j=0}^{m} e^{(\theta_{i+1}+2\varepsilon)j}. \tag{5.56}$$

Assuming $\varepsilon > 0$ is small enough, the exponent $\theta_{i+1} + 2\varepsilon$ is positive, and so the sum is bounded by a multiple of the last term. Thus, there exists $c_2 = c_2(\pi, \lambda, \varepsilon)$ such that

$$\big|s(x, k; \pi, \lambda) \cdot v\big| \leq c_2 e^{(\theta_{i+1}+2\varepsilon)m} \tag{5.57}$$

for every $x \in \sigma$ and $k \geq 1$. In view of (7.44), this implies that

$$\limsup_{k\to\infty} \frac{1}{m} \log \big|[\gamma(x, k) \cdot \phi\big| = \limsup_{k\to\infty} \frac{1}{m} \log \big|s(x, k; \pi, \lambda) \cdot v\big| \leq \theta_{i+1} + 2\varepsilon,$$

uniformly. As $\varepsilon > 0$ is arbitrary, the conclusion of Proposition 7.43 follows. □

Proof of Proposition 7.44. This is similar to the proof of Proposition 7.43. The condition $\phi \in L_g^\perp$ means that ϕ (thus, v) belongs to the sum of the Oseledets subspaces associated to the exponents $-\theta_g, \ldots, -\theta_1$ of the cocycle $F_{\mathcal{Z}}$. Fix $0 < 2a < \theta_g$. Then there exists $c_3 = c_3(\pi, \lambda) > 0$ such that

$$\|\Gamma^j_{\pi,\lambda}(v)\| \leq c_3 e^{-2aj} \|v\| \quad \text{for every } j \geq 0. \tag{5.58}$$

Just as in (7.55), there is also $c_4 = c_4(\pi, \lambda) > 0$ such that

$$\|r(x_j, k_j; \pi^j, \lambda^j)\| \leq 2\|\Gamma_{\pi^j,\lambda^j}\| \leq c_4 e^{aj} \quad \text{for every } j \geq 0. \tag{5.59}$$

Then, analogously to (7.56),

$$\big|s(x, k; \pi, \lambda) \cdot v\big| \leq c_3 c_4 \|v\| \sum_{j=0}^{m} e^{-aj} \tag{5.60}$$

and this is bounded by $c_5\|v\|$ for some constant $c_5 = c_5(\pi, \lambda) > 0$. This proves Proposition 7.44. □

5.7 Simplicity criterium

In these last four sections we outline the proof of Theorem 7.32. Here we state an abstract sufficient condition for the Lyapunov spectra of a certain class of linear cocycles to be simple. The main steps in the proof are presented in the next section. Then, we explain how this criterium may be used to obtain the theorem.

We consider cocycles $F : \Sigma \times \mathbb{R}^d \to \Sigma \times \mathbb{R}^d$, $F(x,v) = (f(x), A(x)v)$ over a transformation $f : \Sigma \to \Sigma$ together with an invariant ergodic probability measure μ, satisfying the following conditions:

(c1) $f : \Sigma \to \Sigma$ is the shift map on $\Sigma = \mathcal{I}^{\mathbb{Z}}$, where the alphabet $\mathcal{I}$ is either finite or countable

(c2) μ has *bounded distortion*, meaning that it is positive on cylinders and there exists $C = C(\mu) > 0$ such that

$$\frac{1}{C} \le \frac{\mu([i_m, \ldots, i_{-1} : i_0, i_1, \ldots, i_n])}{\mu([i_m, \ldots, i_{-1}])\mu([i_0, i_1, \ldots, i_n])} \le C$$

for every $i_m, \ldots, i_0, \ldots, i_n$ and $m \le n$ with $m \le 0$ and $n \ge -1$.

(c3) $A : \Sigma \to \mathrm{GL}(d, \mathbb{R})$ is locally constant: $A(\ldots, i_{-1}, i_0, i_1, \ldots) = A(i_0)$.

By *cylinder* we mean any set $[i_m, \ldots, i_{-1} : i_0, \ldots, i_n]$ of sequences $x \in \Sigma$ such that $x_j = i_j$ for all $j = m, \ldots, -1, 0, 1, \ldots, n$ (the colon locates the zeroth term; it is omitted when either $m = 0$ or $n = -1$). We also denote

$$\Sigma^+ = \mathcal{I}^{\{n \ge 0\}} \qquad W^s_{loc}(x) = \{y \in \Sigma : y_n = x_n \text{ for all } n \ge 0\}$$

$$\Sigma^- = \mathcal{I}^{\{n < 0\}} \qquad W^u_{loc}(x) = \{y \in \Sigma : y_n = x_n \text{ for all } n < 0\}$$

Condition (c3) above may be relaxed: the theory we are presenting extends to certain continuous cocycles not necessarily locally constant. See [5, 7].

Our simplicity criterium is formulated in terms of the monoid associated to the cocycle. In this context, a *monoid* is just a subset of $\mathrm{GL}(d, \mathbb{R})$ closed under multiplication and containing the identity. The *associated monoid* $\mathcal{B} = \mathcal{B}(F)$ is the smallest monoid that contains all $A(i)$, $i \in \mathcal{I}$. Let $\mathrm{Gr}(\ell, \mathbb{R}^d)$ be the Grassmannian manifold of ℓ-dimensional subspaces of $\mathbb{R}^d$, for any $1 \le \ell < d$.

We need the notion of eccentricity of a linear isomorphism $B : \mathbb{R}^d \to \mathbb{R}^d$, which is defined as follows. Let $\sigma_1^2 \ge \cdots \ge \sigma_d^2$ be the eigenvalues of the operator B^*B, in non-increasing order. The eigenvalues are indeed real and positive: if $B^*B(v) = \lambda v$ then $B(v) \cdot B(v) = \lambda(v \cdot v)$. Geometrically, their positive square roots $\sigma_1 \ge \cdots \ge \sigma_d > 0$ measure the semi-axes of the ellipsoid $\{B(v) : \|v\| = 1\}$. The *eccentricity* of B is

$$\mathrm{Ecc}(B) = \min_{1 \le \ell < d} \mathrm{Ecc}(\ell, B),$$

where $\mathrm{Ecc}(\ell, B) = \sigma_\ell / \sigma_{\ell+1}$ is called the *ℓ-eccentricity*. See Figure 7.8. That is, B has large eccentricity if the ratios of any two semi-axes are far from 1.

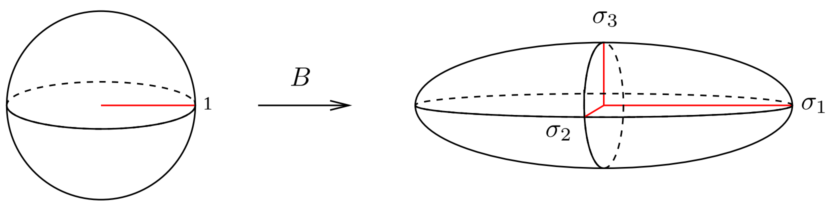


Figure 5.8:

Definition 5.49. We say that the cocycle F (and the associated monoid $\mathcal{B}$) is cocycle cocycle

- *pinching* if it contains elements with arbitrarily large eccentricity $\mathrm{Ecc}(B)$
- *twisting* if for any $E \in \mathrm{Gr}(\ell, \mathbb{R}^d)$ and any finite family $G_1, \ldots, G_N$ of elements of $\mathrm{Gr}(d-\ell, \mathbb{R}^d)$ there exists $B \in \mathcal{B}$ such that $B(E) \cap G_i = \{0\}$ for all $i = 1, \ldots, N$.

It is evident from the definition that any monoid that contains a pinching submonoid is also pinching, and analogously for twisting.

Theorem 5.50. *Assume f, μ, F satisfy conditions (c1), (c2), (c3) above. If F is pinching and twisting then its Lyapunov spectrum relative to (f, μ) is simple.*

Remark 5.51. Pinching and twisting are often easy to establish. For instance, suppose a (general) monoid $\mathcal{B}$ contains some element B_1 whose eigenvalues all have distinct norms. Then $\mathcal{B}$ is pinching, since the powers B_1^n have arbitrarily large eccentricity as $n \to \infty$. Suppose, in addition, that the monoid contains some element B_2 satisfying $B_2(V) \cap W = \{0\}$ for any pair of subspaces V and W which are sums of eigenspaces of B_1 and have complementary dimensions. Then $\mathcal{B}$ is twisting. Indeed, given any E, G_1, ..., G_n as in the definition, we have that $B_1^n(E)$ is close to some sum V of ℓ eigenspaces of B_1, and every $B_1^{-n}(G_i)$ is close to some sum W_i of $d-\ell$ eigenspaces of B_1, as long as n is large enough. It follows that

$$B_2(B_1^n(E)) \cap B_1^{-n}(G_i) = \{0\}, \quad \text{that is,} \quad B_1^n B_2 B_1^n(E) \cap G_i = \{0\}.$$

A converse to these observations is given in [4, Lemma A.5].

Example 5.52. Suppose there are symbols t and b in the alphabet $\mathcal{I}$ such that

$$A(t) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad A(b) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then the associated monoid is pinching and twisting. Indeed,

$$B = A(t)A(b) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

is hyperbolic and so its powers have arbitrarily large eccentricity. This proves pinching. To prove twisting, consider E and $G_1, \ldots, G_N \in \mathrm{Gr}(1, \mathbb{R}^2)$. Fix k large enough so that no $A(t)^{-k}(G_i)$ coincides with any of the eigenspaces E^u and E^s of B. Then $B^n(E) \cap A(t)^{-k}(G_i) = \{0\}$, that is, $A(t)^k B^n(E) \cap G_i = \{0\}$ for all i and any sufficiently large n. See Figure 7.9: the dotted lines express the fact that $A(t)$ and $A(b)$ act by sheer along the horizontal axis and the vertical axis, respectively.

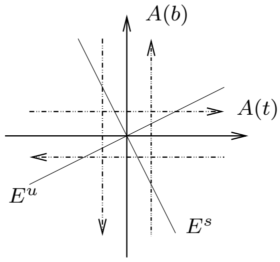


Figure 5.9:

In Section 7.8 we outline the proof of Theorem 7.50. The strategy is inspired by the following observations. Suppose a cocycle does have $\ell \in \{1, \ldots, d-1\}$ Lyapunov exponents, counted with multiplicity, which are strictly larger than all the other ones. Then the sum $\xi^+(x)$ of the corresponding Oseledets subspaces defines an invariant section of $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$ which is an "attractor" for the action of F on the Grassmannian bundle: one may find $\xi^+(x)$ as a limit for $A^n(f^{-n}(x))$, $n \geq 1$ acting on the Grassmannian $\mathrm{Gr}(\ell, \mathbb{R}^d)$, as illustrated in Figure 7.10. Observe also that $\xi^+(x)$ is constant on local unstable sets $W^u_{loc}(x)$ because, as observed in Remark 7.5, it is determined by the backward iterates of the cocycle alone and, clearly, the sequence of backward iterates is constant on local unstable sets.

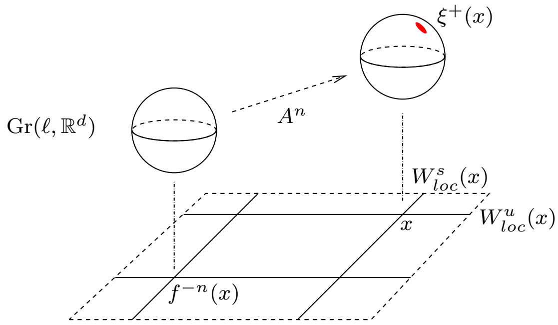


Figure 5.10:

The first main step in the proof of Theorem 7.50 is to show that such an invariant section does exist under the assumptions of the theorem. This is stated in Proposition 7.53 and then we explain how the theorem can be deduced from it. The way we actually construct the invariant section to prove the proposition is as the limit of the iterates under $A^n(f^{-n}(x))$, $n \geq 1$ of certain measures in $\mathrm{Gr}(\ell, \mathbb{R}^d)$. These measures are obtained projecting invariant measures of the cocycle of a special class, that we call *u-states*. The statement is given in Proposition 7.58 and then we explain how Proposition 7.53 may be obtained from it.

The role of u-states is to provide some dynamically meaningful relation between fibers of the Grassmannian bundle $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$ at different points, especially points in the same local unstable set. Indeed, these are probability measures on the Grassmannian bundle whose conditional probabilities on the fibers of points in the same local unstable set are all equivalent. For instance, a measure on $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$ whose conditional probabilities are Dirac masses, that is, a measure of the form

$$m(X \times Y) = \int_X \delta_{\xi(x)}(Y)\, d\mu(x),$$

is a u-state if and only if the function $\xi : \Sigma \to \mathrm{Gr}(\ell, \mathbb{R}^d)$ is constant on local unstable sets. These observations are important for the proof of Proposition 7.58, that we briefly sketch in the last part of Section 7.8.

5.8 Proof of the simplicity criterium

In this section we outline the proof of Theorem 7.50. The presentation is in successive layers, so as to allow the reader to choose an appropriate level of detail. The complete arguments can be found in [5, Appendix] and [4].

Invariant section. First, we explain how Theorem 7.50 can be obtained from the following proposition (see Figure 7.11):

Proposition 5.53. *Fix $\ell \in \{1, \ldots, d-1\}$. Assume F is pinching and twisting. Then there is a measurable section $\xi^+ : \Sigma \to \mathrm{Gr}(\ell, \mathbb{R}^d)$ such that*

1. *ξ^+ is constant on local unstable sets and F-invariant, that is, it satisfies $A(x)\xi^+(x) = \xi^+(f(x))$ at μ-almost every point*
2. *the ℓ-eccentricity $\mathrm{Ecc}(\ell, A^n(f^{-n}(x))) \to \infty$ and the image $E^+(x, n)$ of the ℓ-subspace most expanded under $A^n(f^{-n}(x))$ converges to $\xi^+(x)$ as $n \to \infty$*
3. *for any $V \in \mathrm{Gr}(d - \ell, \mathbb{R}^d)$, the subspace $\xi^+(x)$ is transverse to V at μ-almost every point.*

We want to show that ξ^+ is precisely the sum of the Oseledets subbundles corresponding to the ℓ (strictly) largest Lyapunov exponents. There are three main steps. First, we find a candidate ξ^- for being the sum of the remaining

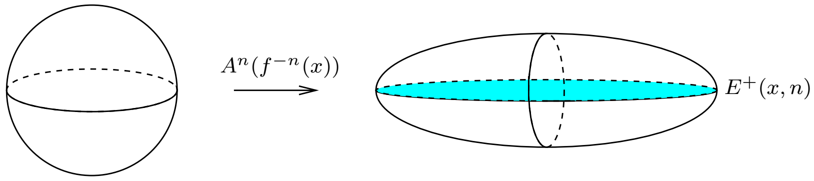


Figure 5.11:

Oseledets subbundles. Next, we check that ξ^+ and ξ^- are transverse to each other at almost every point. Finally, we prove that the Lyapunov exponents of the cocycle along ξ^+ are indeed strictly larger than the exponents along ξ^-. Let us detail each of these steps a bit more.

To begin with, observe that Proposition 7.53 may be applied to the inverse cocycle F^{-1}, since conditions (c1), (c2), (c3) are invariant under time reversal, and we also have

Lemma 5.54. *A monoid $\mathcal{B}$ is pinching and twisting if and only if the inverse $\mathcal{B}^{-1} = \{B^{-1} : B \in \mathcal{B}\}$ is pinching and twisting.*

Considering the action of F^{-1} on the Grassmannian $\mathrm{Gr}(d-\ell, \mathbb{R}^d)$ of complementary dimension, we find an invariant section $\xi^- : \Sigma \to \mathrm{Gr}(d-\ell, \mathbb{R}^d)$ satisfying the analogues of properties (1), (2), (3) in the proposition. In particular, ξ^- is constant on local *stable* sets of f. Next, we need to show that ξ^+ and ξ^- are transverse to each other:

Lemma 5.55. $\xi^+(x) \oplus \xi^-(x) = \mathbb{R}^d$ *for μ-almost every $x \in \Sigma$.*

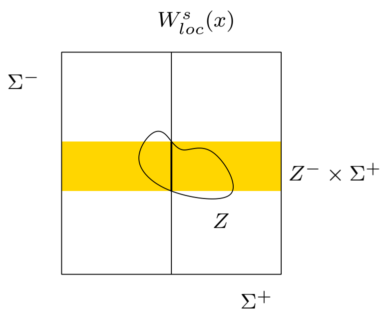


Figure 5.12:

This is easy to see, with the help of Figure 7.12. Indeed, suppose the claim fails on a set $Z \subset \Sigma$ with $\mu(Z) > 0$. Using the bounded distortion property (c2), one can see that there exist points $x \in \Sigma$ such that $Z^- \times \Sigma^+$ has positive μ-measure, where $Z^- = W^s_{loc}(x) \cap Z$. Define $V = \xi^-(x)$. Then $\xi^-(y) = V$ for all $y \in W^s_{loc}(x)$ and ξ^+ is not transverse to V on $Z^- \times \Sigma^+$. This contradicts the last part of Proposition 7.53.

The third and last step in deducing Theorem 7.50 from Proposition 7.53 is

Lemma 5.56. *The Lyapunov exponents of $F \mid \xi^+$ are strictly larger than the Lyapunov exponents of $F \mid \xi^-$.*

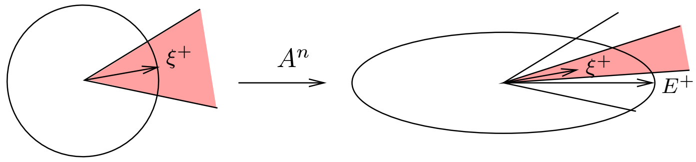


Figure 5.13:

This lemma is deduced along the following lines. See also Figure 7.13. Let us consider a cone field C^+ around the invariant section ξ^+. Fix some compact subset K with positive measure and some large number $N \geq 1$, such that

- $E^+(x,n) \subset C^+(x)$ for every $x \in K$ and $n \geq N$. This is possible, because part 2 of Proposition 7.53 asserts that $E^+(x,n)$ is close to $\xi^+(x)$ when n is large.
- $A^n(x)C^+(x) \subset C^+(f^n(x))$ for every $x \in K$ and $n \geq N$ such that $f^n(x) \in K$. This is possible because the iterates of the cone $C^+(x)$ approach the image $E^+(f^n(x),n)$ of the most expanded subspace as n goes to infinity.

Reducing K if necessary, we may also assume that no point of K returns to it in less than N iterates. Then the previous property means that the cone field is invariant under the cocycle $\tilde{F}$ induced by F over the first return map. This implies, by a variation of the argument in Lemma 7.26, that there is a gap between the first ℓ Lyapunov exponents of $\tilde{F}$ and the remaining ones. Consequently, by Corollary 7.25, the same is true for the original cocycle F.

This finishes our outline of the proof of Theorem 7.50 from the invariant section Proposition 7.53. In what follows we comment on the proof of the proposition.

Invariant u-states. We are going to explain how Proposition 7.53 can be obtained from a statement about iterations of certain probability measures on the Grassmannian given in Proposition 7.58.

A probability m on $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$ is a *u-state* if it projects down to μ and there is $C > 0$ such that

$$\frac{m([i_s,\ldots,i_{-1}:i_0,\ldots,i_p]\times X)}{\mu([i_0,\ldots,i_p])} \leq C\frac{m([i_s,\ldots,i_{-1}:j_0,\ldots,j_q]\times X)}{\mu([j_0,\ldots,j_q])}$$

for every $i_s, \dots, i_0, \dots, i_p, j_0, \dots, j_q$ and $X \subset \mathrm{Gr}(\ell, \mathbb{R}^d)$. Notice that, since μ has bounded distortion, this is the same as saying there is $C' > 0$ such that

$$\frac{m([i_s, \dots, i_{-1} : i_0, \dots, i_p] \times X)}{\mu([i_s, \dots, i_{-1} : i_0, \dots, i_p])} \le C' \frac{m([i_s, \dots, i_{-1} : j_0, \dots, j_q] \times X)}{\mu([i_s, \dots, i_{-1} : j_0, \dots, j_q])}.$$

In other words, up to a uniform factor, the m-measures of any two "parallelepipeds" $[i_s, \dots, i_{-1} : i_0, \dots, i_p] \times X$ along the same $[i_s, \dots, i_{-1}] \subset \Sigma^-$ are comparable to the μ-measures of their "bases" $[i_s, \dots, i_{-1} : i_0, \dots, i_p]$. See Figure 7.14.

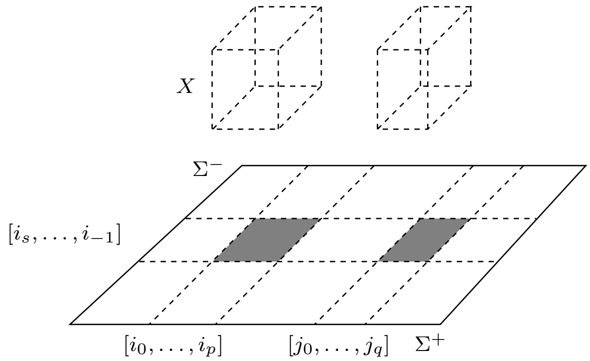


Figure 5.14:

Yet another equivalent formulation is that m is a u-state if it admits a disintegration

$$m = \int_\Sigma m_x \, d\mu(x), \qquad m_x \text{ a probability on } \mathrm{Gr}(\ell, \mathbb{R}^d),$$

where m_x is equivalent to m_y whenever $x \in W^u_{loc}(y)$, with derivative uniformly bounded by C.

It is easy to see that u-states always exist: for instance, $m = \mu \times \nu$ for any probability ν in the Grassmannian. Even more,

Lemma 5.57. *There exist u-states on $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$ which are invariant under the action of the cocycle on $\mathrm{Gr}(\ell, \mathbb{R}^d)$.*

The arguments are quite standard. The iterates of any u-state under the cocycle are also u-states, with uniform distortion constant C. It follows that the iterates form a relatively compact set, for the weak topology in the space of probability measures in $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$, and every measure in the closure is still a u-state. Hence, any Cesaro weak limit of the iterates is an invariant u-state.

One calls *hyperplane section* of $\mathrm{Gr}(\ell, \mathbb{R}^d)$ associated to any $G \in \mathrm{Gr}(d-\ell, \mathbb{R}^d)$ the subset of all $E \in \mathrm{Gr}(\ell, \mathbb{R}^d)$ such that $E \cap G \neq \{0\}$.

Proposition 5.58. *Let m be an invariant u-state in $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$ and ν be its projection to $\mathrm{Gr}(\ell, \mathbb{R}^d)$. Then*

1. *the support of ν is not contained in any hyperplane section of the Grassmannian*

2. *for μ-almost every $x \in M$, the push-forwards $\nu^n(x)$ of ν under $A^n(f^{-n}(x))$ converge to a Dirac measure at some point $\xi^+(x) \in \mathrm{Gr}(\ell, \mathbb{R}^d)$.*

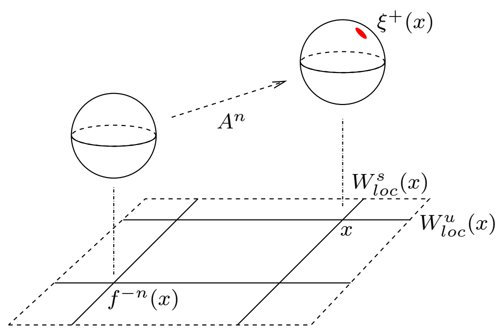


Figure 5.15:

See Figure 7.15. To deduce Proposition 7.53 from Proposition 7.58 it suffices to use the following linear algebra statement:

Lemma 5.59. *Let $L_n : \mathbb{R}^d \to \mathbb{R}^d$ be a sequence of linear isomorphisms and ρ be a probability measure on $\mathrm{Gr}(\ell, \mathbb{R}^d)$ which is not supported in any hyperplane section of $\mathrm{Gr}(\ell, \mathbb{R}^d)$. If the push-forwards $(L_n)_*\rho$ converge to a Dirac measure δ_ξ then the eccentricity $\mathrm{Ecc}(\ell, L_n) \to \infty$ and the images $E^+(L_n)$ of the most expanded ℓ-subspace converge to ξ.*

Convergence to a Dirac mass. Finally, we comment on the proof of Proposition 7.58. Part 1 of the proposition corresponds to

Lemma 5.60. *If F is twisting then the projection ν of any invariant u-state m is not supported inside any hyperplane section of $\mathrm{Gr}(\ell, \mathbb{R}^d)$.*

Proof. We claim that $B(\operatorname{supp}\nu) \subset \operatorname{supp}\nu$ for every $B \in \mathcal{B}$. The lemma is an easy consequence. Indeed, consider any subspace $E \in \operatorname{supp}\nu$ and suppose the support was contained in a hyperplane section S, associated to some $G \in \mathrm{Gr}(d-\ell, \mathbb{R}^d)$. Then $B(E) \in S$ or, equivalently, $B(E) \cap G \neq \{0\}$ for all $B \in \mathcal{B}$, which would contradict the twisting assumption. Therefore, we only have to prove the claim. Moreover, it suffices to consider the case when $B = A(j_0)$ for some $j_0 \in \mathcal{I}$. Let j_0 be fixed and $\xi \in \mathrm{Gr}(\ell, \mathbb{R}^d)$ be any point in $\operatorname{supp}\nu$. By definition, $m(\Sigma \times V) > 0$ for any neighborhood V of ξ. Equivalently, there exists some $i_0 \in \mathcal{I}$ such that $m([i_0] \times V) > 0$. Since m is a u-state, the measure of any $[i_0] \times V$ is positive if and only if the measure of $[j_0] \times V$ is positive. Hence, $m([j_0] \times V) > 0$ for any neighborhood V of ξ. Since $F([j_0] \times V) \subset \Sigma \times B(V)$ and m is F-invariant, it follows that $m(\Sigma \times B(V))$ is also positive, for any neighborhood V of ξ. This implies that $B(\xi)$ is also in the support of ν, as we wanted to prove. □

Now let us discuss part 2 of Proposition 7.58. There are three main steps. The first, and most delicate, is to show that some subsequence converges to a Dirac measure:

Lemma 5.61. *For almost every x there exist $n_j \to \infty$ such that $\nu^{n_j}(x)$ converges to a Dirac measure.*

Let us give some heuristic explanation of the construction of such a subsequence. See also Figure 7.16. By hypothesis, there exist elements

$$B_1^p = A(i_{p-1}) \cdots A(i_1)A(i_0)$$

of the associated monoid $\mathcal{B}$ with arbitrarily strong eccentricity. By ergodicity, for μ-almost every $x \in \Sigma$ there exist $m_j \to \infty$ such that $f^{-m_j}(x) \in [i_0, \ldots, i_{p-1}]$, and so

$$A^{m_j}(f^{-m_j}(x)) = C_j B_1^p$$

for some $C_j \in \mathrm{GL}(d, \mathbb{R})$. We want to argue that $C_j B_1^p$ has strong eccentricity, because B_1^p does, and so, using that ν is not supported in a hyperplane section, the measure

$$A^{m_j}(f^{-m_j}(x))_* \nu = \left(C_j B_1^p\right)_* \nu$$

is strongly concentrated near the image of the most expanded ℓ-subspace. In order to justify this kind of assertion, one would need to ensure that, somehow, the strongly pinching behavior of B_1^p is not destroyed by C_j. The following observation, by Furstenberg [16], that the space of projective maps on the Grassmannian has a natural compactification, gives some hope this might be possible.

We call *projective map* on $\mathrm{Gr}(\ell, \mathbb{R}^d)$ any transformation induced on the Grassmannian by a linear isomorphism of $\mathbb{R}^d$. More generally, we call *quasi-projective map* on $\mathrm{Gr}(\ell, \mathbb{R}^d)$ induced by a linear map $L : \mathbb{R}^d \to \mathbb{R}^d$, the transformation [2] $L_\#$ that assigns to every $E \in \mathrm{Gr}(\ell, \mathbb{R}^d)$ with $E \cap \ker L = \{0\}$ its image $L(E) \in \mathrm{Gr}(\ell, \mathbb{R}^d)$. This is defined on the complement of the *kernel* of the quasi-projective map, defined by

$$\ker L_\# = \left\{E \in \mathrm{Gr}(\ell, \mathbb{R}^d) : E \cap \ker L \neq \{0\}\right\}.$$

We assume L is not identically zero. Then, clearly, $\ker L_\#$ is contained in some hyperplane section of the Grassmannian. We may always consider $\|L\| = 1$, since multiplying L by any constant does not change the definition. Thus, the space of quasi-projective maps inherits a compact topology from the unit ball of linear operators in $\mathbb{R}^d$.

[2] In most places we use the same notation for the linear map and the quasi-projective map induced by it.

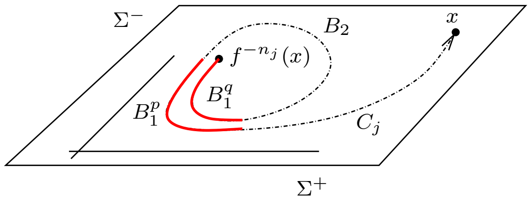


Figure 5.16:

Therefore, the family of all C_j one obtains in the previous construction is contained in some compact set of quasi-projective maps. Of course, this does not yet mean that the effect of the C_j on eccentricity is bounded (which would ensure the strongly pinching behavior of the factor B_1^p prevails). The problem is that the image E_p^+ of the ℓ-dimensional subspace most expanded by B_1^p may be contained in the kernel of any accumulation point $C_\#$ of the sequence C_j in the space of quasi-projective maps: in that case the maps C_j are strongly distorting near $\ker C_\#$ and so they might indeed cancel out the eccentricity of B_1^p. To make the previous arguments work one needs to avoid this situation, that is, one needs to ensure that $C_\#$ may always be chosen so that its kernel does not contain E_p^+. More precisely, one can argue as follows. See Figure 7.16.

Let B_1^p and $C_\#$ be fixed, as before. Consider another arbitrarily eccentric element

$$B_1^q = A(j_{q-1}) \cdots A(j_1) A(j_0) \in \mathcal{B}$$

and let E_q^+ be the image of its most expanded ℓ-dimensional subspace. By the twisting condition, there exists some

$$B_2 = A(k_s) \cdots A(k_1)$$

that maps E_q^+ outside the kernel of $C_\# B_1^p$. Moreover, by ergodicity, there exists some sequence $n_l = m_{j_l} + q + s \to \infty$ such that

$$f^{-n_l}(x) \in [j_0, \ldots, j_{q-1}, k_1, \ldots, k_s, i_0, \ldots, i_{p-1}]$$

and so $A^{n_l}(f^{-n_l}(x)) = C_j B_1^p B_2 B_1^q$. By construction, E_q^+ is outside the kernel of $C'_\# = C_\# B_1^p B_2$. Thus, the previous arguments now make sense. *

Now we move on with the arguments. Let $m^{(n)}(x)$ denote the projection to the Grassmannian of the normalized restriction of m to the cylinder $[i_{-n}, \ldots, i_{-1}]$ that contains x. The second step in the proof of part 2 of Proposition 7.58 is

Lemma 5.62. *The sequence $m^{(n)}(x)$ converges almost surely to some probability $m(x)$ on $\mathrm{Gr}(\ell, \mathbb{R}^d)$, and $m(\cdot)$ is almost everywhere constant on local unstable sets.*

The first claim follows from a simple martingale argument. From the construction we easily see that $\{m(x)\}$ is a disintegration of m relative to the partition of $\Sigma \times \mathrm{Gr}(\ell, \mathbb{R}^d)$ into the sets $W^u_{loc}(x) \times \mathrm{Gr}(\ell, \mathbb{R}^d)$, and that gives the second claim.

The final step in the proof of Proposition 7.58 is the following lemma, which is a consequence of the definition of u-state:

Lemma 5.63. *There exists $C = C(m) > 0$ such that*

$$\frac{1}{C} \leq \frac{\nu^n(x)}{m^{(n)}(x)} \leq C \quad \textit{for all } x.$$

From Lemmas 7.62 and 7.63 we get that, given any $x \in \Sigma$ and any accumulation point $\nu(x)$ of the sequence $\nu^n(x)$,

$$\frac{1}{C} \leq \frac{\nu(x)}{m(x)} \leq C$$

In particular, any two accumulation points are equivalent. Now, by Lemma 7.61, some accumulation point is a Dirac measure $\delta_{\xi(x)}$, at almost every point. Clearly, this implies the accumulation point is unique, and the sequence $\nu^n(x)$ does converge to a Dirac measure, as we claimed. This finishes our sketch of the proofs of Proposition 7.58 and, thus, Theorem 7.50.

5.9 Zorich cocycles are pinching and twisting

Now, to deduce Theorem 7.32 we only have to check that Theorem 7.50 may be applied to the restricted Zorich cocycles. Let us begin by verifying the hypotheses (c1), (c2), (c3).

We have seen in Section 1.8 that Z is a Markov map. More precisely, there exists a finite partition $\{\Lambda_{\pi,\varepsilon} : \pi \in C \text{ and } \varepsilon = 0, 1\}$ and a countable refinement

$$\Lambda^*_{\pi,\varepsilon,n} = \{\lambda \in \Lambda_{\pi,\varepsilon} : \varepsilon^1 = \cdots = \varepsilon^{n-1} = \varepsilon \neq \varepsilon^n\}.$$

such that Z maps every $\{\pi\} \times \Lambda^*_{\pi,\varepsilon,n}$ bijectively onto $\{\pi^n\} \times \Lambda_{\pi^n,1-\varepsilon}$. This is not exactly a full shift, but it is easy to extend the criterium to this slightly more general version of condition (c1).

In Section 4.8 we constructed a Z-invariant probability μ which is ergodic and equivalent to volume $d\lambda$. This measure μ has bounded distortion, and so
* condition (c2) is met. Finally, the cocycle F_Z is constant on each atom $\Lambda^*_{\pi,\varepsilon,n}$, because

$$\Gamma_{\pi,\lambda} = \Theta^{n(\pi,\lambda)}_{\pi,\lambda}$$

depends only on π and the types of all $R^j(\pi, \lambda)$ with $0 \leq j < n(\pi, \lambda)$. In other words, $\Gamma_{\pi,\lambda}$ depends only on π and ε^j for $1 \leq j < n(\pi, \lambda)$, and so it is constant on every atom $\Lambda^*_{\pi,\varepsilon,n}$ of the Markov partition. This gives condition (c3).

Thus, now we only have to check that

Theorem 5.64. *Every restricted Zorich cocycle is pinching and twisting.*

The proof of this theorem will be outlined in the next section. The strategy is to argue by induction on the complexity of the stratum, that is, on the genus g and the number κ of singularities. Indeed, we look for orbits of $\mathcal{T}^t$ that spend a long time close to the boundary of each stratum and, hence, pick up the behavior of the flow on "simpler" strata. Figure 7.17 illustrates this idea: think of the upper hemisphere as a stratum, whose boundary is a simpler stratum, represented by the equator (the actual geometry of strata near the boundary is much more complex than the figure suggests, and is still poorly understood).

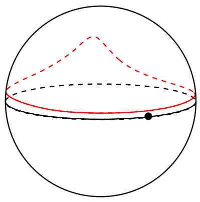


Figure 5.17:

Before we explain in more detail how this strategy is implemented to give the inductive step of the proof of Theorem 7.64, let us note that the initial step of the induction, corresponding to the torus case $g = 1$, $\kappa = 0$, $d = 2$ is easy. Indeed, in this case there is only one permutation pair

$$\pi = \begin{pmatrix} A & B \\ B & A \end{pmatrix}.$$

The top case of the renormalization corresponds to $\lambda_A < \lambda_B$, and the bottom case corresponds to $\lambda_B < \lambda_A$. In every case, the cocycle is given by

$$\Theta_{top} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Theta_{bot} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then, arguing just as in Example 7.52, we get that F is pinching and twisting.

5.10 Relating to simpler strata

Here we outline the inductive step in the proof of Theorem 7.64. Fix any permutation pair $\pi \in \mathcal{C}$ and denote by $\mathcal{B}_\pi$ the submonoid of $\mathcal{B}$ corresponding to orbit segments $(\pi^0, \lambda^0), \ldots, (\pi^k, \lambda^k)$ such that $\pi^0 = \pi = \pi^k$. It suffices to prove that the action of $\mathcal{B}_\pi$ on the space H_π is pinching and twisting.

The proof is by induction on the complexity of the stratum, that is, the genus and the number of singularities. Recall that Abelian differentials in simpler strata, contained in the boundary of $\mathcal{A}_g(m_1, \ldots, m_\kappa)$, may be obtained by collapsing two or more singularities of some Abelian differential in $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ together, as illustrated in Figure 7.18. The multiplicity of the new singularity is the sum of the multiplicities of the original ones.

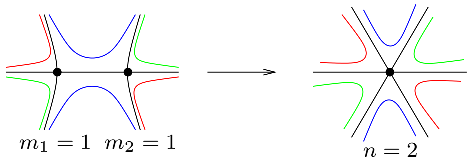


Figure 5.18:

This strategy is more easily implemented at the level of interval exchange transformations. In that setting, *approaching the boundary* corresponds to *making some coefficient λ_α very small.* Then it remains small for a long time under iteration by the renormalization operator.

Simple reductions and simple extensions. We consider two operations on the combinatorics, that we call simple reduction and simple extension. *Simple reduction* $\pi \mapsto \pi'$ corresponds to removing one letter from both top and bottom lines of the permutation pair. *Simple extension* $\pi' \mapsto \pi$ corresponds to inserting one letter at appropriate positions of both top and bottom lines. See the formula:

$$\pi = \begin{pmatrix} a_1 & \cdots & a_{i-1} & \mathbf{c} & a_{i+1} & \cdots & \cdots & \cdots & \cdots & \cdots & a_d \\ b_1 & \cdots & & \cdots\cdots & \cdots & \cdots & b_{j-1} & \mathbf{c} & b_{j+1} & \cdots & b_d \end{pmatrix}$$

$$\updownarrow$$

$$\pi' = \begin{pmatrix} a_1 & \cdots & a_{i-1} & a_{i+1} & \cdots & \cdots & \cdots & \cdots & a_d \\ b_1 & \cdots & \cdots & \cdots & \cdots & b_{j-1} & b_{j+1} & \cdots & b_d \end{pmatrix}$$

The two operations are not exactly inverse to each other, because there are some restrictions on the insertion locations in the simple extension: the inserted letter can not be last in either line and can not be first in both rows simultaneously.

Lemma 5.65. *Given any π there exists π' such that π is a simple extension of π'. Moreover, either $g(\pi) = g(\pi')$ or $g(\pi) = g(\pi') + 1$.*

We also take advantage of the symplectic structure preserved by the Zorich cocycles. A subspace V of a symplectic space (H, ω) is called *isotropic* if

$$\omega(v_1, v_2) = 0 \quad \text{for any } v_1, v_2 \in V.$$

Let $\mathrm{Iso}(\ell, H) \subset \mathrm{Gr}(\ell, H)$ denote the submanifold of isotropic subspaces with dimension ℓ. The *symplectic reduction* of H by some $v \in \mathbb{P}(H)$ is the quotient H^v by the direction of v of the symplectic orthogonal of v. Note $\dim H^v = \dim H - 2$. The *stabilizer* of v is the submonoid $\mathcal{B}^v$ of elements of $\mathcal{B}$ that preserve v. The *induced action* of the cocycle on the symplectic reduction is the natural action of the stabilizer $\mathcal{B}^v$ on H^v.

* **Lemma 5.66.** *In the context of Lemma 7.65,*

1. *If $g(\pi) = g(\pi')$ then there is a symplectic isomorphism $H_{\pi'} \to H_\pi$ that conjugates the action of $\mathcal{B}_{\pi'}$ on $H_{\pi'}$ to the action of some submonoid of $\mathcal{B}_\pi$ on H_π.*
2. *If $g(\pi) = g(\pi')+1$, there is some symplectic reduction H^v_π of H_π and some symplectic isomorphism $H_{\pi'} \to H^v_\pi$ that conjugates the action of $\mathcal{B}_{\pi'}$ on $H_{\pi'}$ to the action induced by some submonoid of $\mathcal{B}_\pi$ on H^v_π.*

The proof of Theorem 7.64 may be split into proving two propositions that we state in the sequel. Towards establishing the twisting property, we prove isotropic

Proposition 5.67. *The action of $\mathcal{B}_\pi$ on $\mathrm{Iso}(\ell, H_\pi)$ is minimal: any closed invariant set is either empty or the whole ambient space.*

It follows, in particular, that $\mathcal{B}_\pi$ *twists isotropic subspaces* of H_π: given any $E \in \mathrm{Iso}(\ell, H_\pi)$ and any finite family $G_1, \dots, G_N$ of elements of $\mathrm{Gr}(d-\ell, \mathbb{R}^d)$, there exists $B \in \mathcal{B}_\pi$ such that $B(E) \cap G_i = \{0\}$ for all $j = 1, \dots, N$. This is a direct consequence of the proposition, and the observation that hyperplane sections

$$\{W : W \cap G_i \neq \{0\}\}$$

have empty interior in $\mathrm{Iso}(\ell, H_\pi)$.

To compensate for this weaker twisting statement, we prove a stronger form of pinching:

Proposition 5.68. *The action of $\mathcal{B}_\pi$ on H_π is strongly pinching: given any $C > 0$ there exist $B \in \mathcal{B}_\pi$ for which*

$$\log \sigma_g > C \quad \textit{and} \quad \frac{\log \sigma_j}{\log \sigma_{j+1}} > C \quad \textit{for all } 1 \leq j < g.$$

Clearly, for symplectic actions in dimension $d = 2$, twisting is equivalent to isotropic twisting and it is also equivalent to minimality. Moreover, pinching is the same as strong pinching. In any dimension,

Lemma 5.69. *Let a monoid $\mathcal{B}$ act symplectically on a symplectic space (H, ω). If $\mathcal{B}$ twists isotropic subspaces and is strongly pinching then it is twisting and pinching.*

This shows that Theorem 7.64 does follow from Propositions 7.67 and 7.68.

Proof of minimality. Here we outline the proof of Proposition 7.67. Given any π, take π' such that π is a simple extension of π'. In the first case of Lemma 7.66 we immediately get, by induction, that the action of $\mathcal{B}_\pi$ on $\mathrm{Iso}(\ell, H_\pi)$ is minimal. In the second case, the starting point of the proof of Proposition 7.67 is the observation that the action $\mathcal{B}_\pi$ on $\mathbb{P}(H)$ is minimal: any closed invariant set is either empty or the whole projective space. Then the proof of the proposition proceeds by induction on the dimension, using the following lemma:

Lemma 5.70. *If the action of $\mathcal{B}$ on $\mathbb{P}(H)$ is minimal and there is $v \in \mathbb{P}(H)$ such that the induced action of $\mathcal{B}^v$ on $\operatorname{Iso}(\ell-1, H^v)$ is minimal, then the action of $\mathcal{B}$ on $\operatorname{Iso}(\ell, H)$ is minimal.*

The proof of the lemma goes as follows. Consider the fibration

$$\mathcal{I}(H) = \bigcup_{E \in \operatorname{Iso}(\ell,H)} \{E\} \times \mathbb{P}(E) \to \mathbb{P}(H), \quad (E, \lambda) \mapsto \lambda.$$

The fiber over each $\lambda \in \mathbb{P}(H)$ is precisely $\operatorname{Iso}(\ell-1, H^\lambda)$. There is a natural action of $\mathcal{B}$ on $\mathcal{I}(H)$, and we are going to see that this action is minimal. Indeed, let $C \subset \mathcal{I}(H)$ be a closed invariant set and C_λ denote its intersection with the fiber of each $\lambda \in \mathbb{P}(H)$. The hypothesis implies that C_λ is either empty or the whole $\operatorname{Iso}(\ell-1, H^\lambda)$. In the first case, let Λ be the set of $\lambda \in \mathbb{P}(H)$ for which C_λ is empty. In the second case, let Λ be the set of $\lambda \in \mathbb{P}(H)$ for which C_λ is the whole fiber of λ. In either case, Λ is a closed, non-empty, invariant subset of $\mathbb{P}(H)$, and so it must be the whole projective space. This proves that $C = \emptyset$ in the first case and $C = \mathcal{I}(H)$ in the second case. Thus, the action of $\mathcal{B}$ on $\mathcal{I}(H)$ is minimal, as we claimed. Using the natural projection $\mathcal{I}(H) \to \operatorname{Iso}(\ell, H)$, $(E, \lambda) \mapsto E$ one immediately deduces that the action of $\mathcal{B}$ on the isotropic submanifold is minimal.

Proof of strong pinching. Finally, we outline the proof of Proposition 7.68. We denote by $\theta_1(B) \geq \cdots \geq \theta_g(B)$ the non-negative Lyapunov exponents (i.e. logarithms of the norms of the eigenvalues) of a symplectic isomorphism B. We are going to use the following criterium for strong pinching:

Lemma 5.71. *Let $\mathcal{B}$ be a monoid acting symplectically on H, $\dim H = 2g$. Assume for every $C > 0$ there exists some $B \in \mathcal{B}$ such that*

1. *1 is an eigenvalue of B with 1-dimensional eigenspace*
2. *$\theta_{g-1}(B) > 0$*
3. *$\theta_j(B) > C\theta_{j+1}(B)$ for every $1 \leq j \leq g-2$.*

Then $\mathcal{B}$ is strongly pinching.

Notice that the eigenvalue 1 must have even algebraic multiplicity, because B is symplectic. Moreover, the second condition ensures the multiplicity is at most two. Thus, B contains a unipotent block

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

In terms of the singular values of the powers B^n, this implies that

$$\sigma_g(B^n) \approx n \quad \text{and} \quad \sigma_j(B^n) \approx e^{n\theta_j(B)} \quad \text{for } j = 1, \ldots, g-1$$

and so $\mathcal{B}$ is indeed strongly pinching.

Another useful observation is that the property of being (or not) strongly pinching is not affected if one replaces the permutation pair π by any other one $\tilde{\pi}$ in the same Rauzy class. That is because one can find monoid elements γ_1 and γ_2 such that

$$\mathcal{B}_\pi = \gamma_1 \mathcal{B}_{\tilde{\pi}} \gamma_2$$

and then it is not difficult to deduce that the action of $\mathcal{B}_\pi$ on H_π is strongly pinching if and only if the action of $\mathcal{B}_{\tilde{\pi}}$ on $H_{\tilde{\pi}}$ is strongly pinching.

The next step is to reduce the general statement to the case when the Rauzy class is *minimal*, meaning that the number of symbols $d = 2g$. In general, $d = 2g + \kappa - 1$ where κ is the number of singularities. Thus, in terms of the Teichmüller flow, this corresponds to reducing the problem to the *minimal stratum* $\mathcal{A}_g(2g-2)$ of Abelian differentials having a unique singularity. It is implemented through the following refinement of Lemma 7.65:

Lemma 5.72. *Let C be a non-minimal Rauzy class, that is, such that $d > 2g$. Then there exists $\pi \in C$ and there exists π' such that π is a simple extension of π' and $g(\pi) = g(\pi')$.*

Then, by Lemma 7.66, the action of $\mathcal{B}_\pi$ on H_π admits some sub-action conjugate to the action of $\mathcal{B}_{\pi'}$ on $H_{\pi'}$, and so the former is strongly pinching if the latter is. Iterating this procedure, one must eventually reach a permutation pair in a minimal component.

The minimal case is more delicate, because we need to relate the minimal stratum of $\mathcal{A}_g$ with some stratum of a *different* moduli space $\mathcal{A}_{g'}$. The crucial ingredient is

Lemma 5.73. *Any minimal Rauzy class contains some permutation pair*

$$\pi = \begin{pmatrix} A & \alpha_2^0 & \cdots & \alpha_{d-1}^0 & Z \\ Z & \alpha_2^1 & \cdots & \alpha_{d-1}^1 & A \end{pmatrix}$$

such that the following reduction is irreducible:

$$\pi' = \begin{pmatrix} \alpha_2^0 & \cdots & \alpha_{d-1}^0 \\ \alpha_2^1 & \cdots & \alpha_{d-1}^1 \end{pmatrix}.$$

Moreover, $g(\pi') = g(\pi) - 1$ and the Rauzy class of π' is also minimal.

This fact is a consequence of the Kontsevich-Zorich Lemma 6.13, which expresses at the combinatorial level the surgery procedure they called "bubbling a handle" (or, more precisely, its inverse). See Section 6.8.

The final step is to use the inductive assumption that the action of $\mathcal{B}_{\pi'}$ on $H_{\pi'}$ is strongly pinching to construct a parabolic element $B \in \mathcal{B}_\pi$ in the way described in Lemma 7.71.

Notes

Theorem 7.1 was proved by Zorich [63, 65], conditioned to the Zorich-Kontsevich conjecture (Theorem 7.2). The latter was proved by Forni [14] in the genus 2 case, and by Avila, Viana [4] in full generality. The notion of asymptotic cycle was introduced by Schwartzman [50], in a context of flows on metric spaces. Most of Sections 7.2 through 7.6 is taken from Zorich [63, 64, 65]. Sections 7.7 and 7.8 are based on Avila, Viana [5, 4] and Bonatti, Viana [7]. A simplicity criterium for Lyapunov exponents of independent random matrices was first given by Guivarc'h, Raugi [19], and their condition was improved by Gol'dsheid, Margulis [18]. Theorem 7.50 is due to Bonatti, Viana [7] and Avila, Viana [4, Appendix]. An extension to non-locally constant cocycles was given by Avila, Viana [5]. Theorem 7.64 and Sections 7.9 and 7.10 are based on Avila, Viana [4].

Appendix A

Teichmüller Theory

In this appendix we briefly review some main ideas in Teichmüller theory, to help motivate and put in perspective the problems dealt with in the text. For simplicity, we restrict ourselves to the context of compact surfaces. For detailed expositions of the theory, the reader is referred to Ahlfors [1], Gardiner [17], and Lehto [37].

Riemann surfaces. A *Riemann surface* is a smooth surface endowed with an atlas whose coordinate changes are holomorphic maps of the plane. A homeomorphism $f : R \to S$ between two Riemann surfaces is *conformal* if its representations in local charts are holomorphic maps of the plane. This may be expressed geometrically, as follows. The conformal structure defines a field of infinitesimal circles on each of the two Riemann surfaces, transported from the field of infinitesimal circles of the plane through the local charts. Then f is conformal if and only if it is differentiable and the derivative maps the field of infinitesimal circles on R to the one on S. One can also define conformality in analytic terms, using local coordinates as follows. Write $z = x + iy$ and $f(z) = u + iv$, and define

$$\partial_z f = (\partial_x f - i\partial_y f)/2 \quad \text{and} \quad \partial_{\bar{z}} f = (\partial_x f + i\partial_y f)/2$$

where $\partial_x f = \partial_x u + i\partial_x v$ and $\partial_y f = \partial_y u + i\partial_y v$. Then f is conformal if and if

$$\partial_{\bar{z}} f = 0,$$

which is just another way of writing the Cauchy-Riemann conditions. Two Riemann surfaces are *conformally equivalent* if there exists a conformal homeomorphism between them.

Quasi-conformal mappings. Let $f : R \to S$ be a diffeomorphism between two Riemann surfaces. Then f maps the infinitesimal circle at each point $z \in R$ to an infinitesimal ellipse at $f(z) \in S$. This ellipse may be described, with the aid of local coordinates, as follows. Firstly, the ratio of the lengths of the major axis and the minor axis is

$$\kappa_f(z) = \frac{|\partial_z f| + |\partial_{\bar z} f|}{|\partial_z f| - |\partial_{\bar z} f|}(z). \tag{A.1}$$

Note that $|\partial_z f| > |\partial_{\bar z} f|$ since we take f to preserve orientation. Secondly, the directions of the major and minor axes are the images, under the derivative, of the directions

$$\frac{\partial_{\bar z} f\, d\bar z}{\partial_z f\, dz}(z) > 0 \quad \text{and} \quad \frac{\partial_{\bar z} f\, d\bar z}{\partial_z f\, dz}(z) < 0, \tag{A.2}$$

respectively, where $dz = dx + idy$ and $d\bar z = dx - idy$ are the isomorphisms of the tangent space $T_z R$ to the complex plane $\mathbb{C}$ induced by the local coordinate z. Equivalently, the infinitesimal circle at $f(z)$ is the image of the infinitesimal ellipse at z whose ratio of the lengths of the axes is (A.1) and whose minor and major axes are in the directions (A.2), respectively. This equivalent point of view is illustrated in Figure A.1. The *dilatation* $\kappa_f(z)$ and the *Beltrami coefficient* of f

$$\mu_f(z) = \frac{\partial_{\bar z} f}{\partial_z f}$$

are related by

$$\kappa_f(z) = \frac{1 + |\mu_f(z)|}{1 - |\mu_f(z)|}.$$

By compactness, we have $\sup_z \kappa_f(z) < \infty$ and $\sup_z |\mu_f(z)| < 1$.

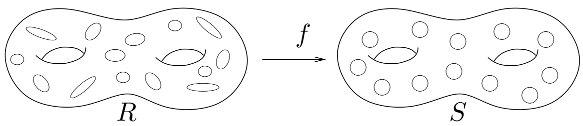


Figure A.1:

More generally, let $f : R \to S$ be a homeomorphism which is differentiable at almost every point [1]. The previous notions make sense, almost everywhere. The homeomorphism is called *quasi-conformal* if there exists $K < \infty$ such that the dilatation $\kappa_f(z) \le K$ for almost every z. An equivalent formulation is that there exists $k < 1$ such that $|\mu_f(z)| \le k$ for almost every point. We denote by $\kappa(f)$ and $\mu(f)$ the smallest possible values of K and k. They are related by

$$\kappa(f) = \frac{1 + \mu(f)}{1 - \mu(f)} \quad \Leftrightarrow \quad \mu(f) = \frac{\kappa(f) - 1}{\kappa(f) + 1}\,.$$

Geometrically, quasi-conformality means that the (measurable) field of infinitesimal ellipses defined by f on R, by pull-back from the field of infinitesimal circles on S as described in Figure A.1, has uniformly bounded eccentricity. Notice also that the inverse f^{-1} is quasi-conformal if f is, with $\kappa(f) = \kappa(f^{-1})$.

[1] The precise condition is that $\partial_z f$ and $\partial_{\bar z} f$ exist as distributions and are locally integrable.

Beltrami differentials. A *Beltrami differential* $\mu\, d\bar{z}/dz$ on a Riemann surface R assigns to each local chart z a measurable complex function $\mu(z)$ such that, if w is any other local chart, then the corresponding function $\mu'(w)$ satisfies

$$\mu'(w) = \mu(z)\frac{dw}{dz} \Big/ \overline{\frac{dw}{dz}}$$

on the intersection of the domains. This implies $|\mu'(w)| = |\mu(z)|$, and so the L^∞ norm $\|\mu\|_\infty$ is well defined, independently of coordinates.

Given a quasi-conformal homeomorphism $f : R \to S$ from R to any other Riemann surface, we call $\mu_f(z)(d\bar{z}/dz)$ the Beltrami differential of f. Quasi-conformality means that $\|\mu_f\|_\infty < 1$. The Ahlfors-Bers (or measurable Riemann mapping) theorem says that, conversely, every Beltrami differential in the unit L^∞-ball corresponds to some quasi-conformal homeomorphism:

Theorem A.1. *Every Beltrami differential with L^∞-norm less than* 1 *is the Beltrami differential of some quasi-conformal homeomorphism $f : R \to S$. Moreover, f is unique up to post-composition with a conformal map.*

Thus, every Beltrami differential in the unit L^∞-ball defines a conformal structure on the surface, obtained by pull-back from S under any quasi-conformal map $f : R \to S$ as in the theorem. There is also an associated (measurable) field of infinitesimal ellipses in R, given by the pull-back of the field of infinitesimal circles in S under any such $f : R \to S$, and characterized by (A.1) and (A.2).

Moduli space of Riemann surfaces. Given any $g \geq 1$, we denote by $\mathcal{M}_g$ the *moduli space* of conformal structures of genus g, that is, the set of Riemann surfaces of genus g modulo conformal equivalence. As we are going to explain, $\mathcal{M}_g$ may be viewed as the quotient of a nicer space, the Teichmüller space $\mathcal{T}_g$, which is a complex manifold of complex dimension $3g - 3$ (respectively, 1 if $g = 1$), by some discrete group, the so-called modular group. It follows that $\mathcal{M}_g$ is a complex variety, also with dimension $3g - 3$ (respectively, 1 if $g = 1$). The quotient map $\mathcal{T}_g \to \mathcal{M}_g$ has branching points, where $\mathcal{M}_g$ fails to be a manifold.

Let us begin by reviewing alternative constructions of the moduli space. The following theorem simplifies the theory somewhat in the compact case:

Theorem A.2. *Any homotopy class of orientation preserving homeomorphisms between compact Riemann surfaces contains some quasi-conformal map.*

Any quasi-conformal homeomorphism $f : R \to S$ defines a conformal structure on the smooth surface underlying R, by pull-back from S. Conversely, every conformal structure can be obtained in this way since, as a consequence of the previous theorem, all conformal structures on the same compact smooth surface are *quasi-conformally equivalent*, meaning there exists quasi-conformal homeomorphisms between them. Thus, the moduli space $\mathcal{M}_g$ may be viewed [2]

[2]This remains valid in the non-compact case, because the moduli space is *defined* as the set of equivalence classes of conformal structures quasi-conformally equivalent to the conformal structure of R.

as the space $\mathcal{M}(R)$ of quasi-conformal maps from R to any other Riemann surface, modulo the following equivalence relation: $f_1 : R \to S_1$ and $f_2 : R \to S_2$ are equivalent if there exists some conformal homeomorphism $h : S_1 \to S_2$.

Then, through the measurable Riemann mapping theorem, $\mathcal{M}(R)$ is also naturally identified with the quotient of the unit L^∞-ball of Beltrami differentials on R by the equivalence relation that identifies two Beltrami differentials μ_1 and μ_2 if the corresponding quasi-conformal maps $f_1 : R \to S_1$ and $f_2 : R \to S_2$ are equivalent. Geometrically, given any conformal homeomorphism $h : S_1 \to S_2$, the fields of infinitesimal ellipses on R associated to μ_1 and μ_2 are mapped to one another by the derivatives of $h' = f_2^{-1} \circ h \circ f_1$ and its inverse.

Teichmüller space. There is a stronger equivalence relation in the space of quasi-conformal maps on R, where one requires that $f_2 \circ f_1^{-1} : S_1 \to S_2$ be homotopic to some conformal map $h : S_1 \to S_2$. We denote by $[f]$ the equivalence class of a quasi-conformal map f for this relation. The set of such equivalence classes is denoted $\mathcal{T}(R)$ and called the *Teichmüller space*.

If $h : S_1 \to S_2$ is conformal map then $h' = f_2^{-1} \circ h \circ f_1$ is conformal relative to the conformal structures obtained by pull-back under f_1 and f_2. Clearly, h' is homotopic to the identity if and only if h is homotopic to $f_2 \circ f_1^{-1}$. Thus, one may also define $\mathcal{T}(R)$ as the set of all conformal structures [3] on the smooth surface underlying R, modulo the equivalence relation that identifies conformal structures that are mapped to one another by some homeomorphism homotopic to the identity. Yet another equivalent definition of $\mathcal{T}(R)$ is as the unit L^∞-ball of Beltrami differentials, modulo the equivalence relation that identifies all Beltrami differentials when the corresponding quasi-conformal maps belong to the same Teichmüller equivalence class.

There is a natural distance in $\mathcal{T}(R)$, the *Teichmüller metric*, defined by

$$\mathrm{d_T}([f_1],[f_2]) = \frac{1}{2} \inf \big\{ \log \kappa(g) : g \in [f_2 \circ f_1^{-1}] \big\}.$$

Symmetry corresponds to $\kappa(g) = \kappa(g^{-1})$, and the triangle inequality follows from $\kappa(g' \circ g'') \le \kappa(g')\kappa(g'')$. The infimum in the definition is always attained. Indeed, the Teichmüller theorem states that, *in every Teichmüller equivalence class there exists a homeomorphism that minimizes the dilatation.* In addition, this homeomorphism is unique, up to composition with conformal maps, and has a very special product structure. We shall explain the last statement later, after introducing the notion of quadratic differential.

The metric space $(\mathcal{T}(R), \mathrm{d_T})$ is complete and homeomorphic to a cell in $\mathbb{R}^{6g-6}$ (in $\mathbb{R}^2$, if $g = 1$). Furthermore, $\mathcal{T}(R)$ has the structure of a complex manifold with complex dimension $3g - 3$ (complex dimension 1, if $g = 1$), and the Teichmüller distance is determined by that complex structure. The Teichmüller distance is also a Finsler metric, given by the integration of a norm on the tangent bundle of $\mathcal{T}(R)$, but not a Riemannian metric (the norm does not come from an inner product).

[3] In the non-compact case take structures quasi-conformally equivalent to the one of R.

Mapping class group. The *modular group*, or *mapping class group*, is the group $\mathrm{Mod}(R)$ of homotopy classes of quasi-conformal maps $g : R \to R$. Recall (Theorem A.2) that in the compact case every homotopy class of orientation preserving homeomorphisms contains some quasi-conformal map. The modular group acts on $\mathcal{T}(R)$ through

$$\langle g \rangle [f] = [f \circ g^{-1}],$$

where $\langle g \rangle$ is the homotopy class of h, and $[f]$ and $[f \circ g]$ are the Teichmüller equivalence classes of $f : R \to S$ and $f \circ g : R \to S$, respectively. This action is by isometries of the Teichmüller metric:

$$\mathrm{d_T}([f_1 \circ g], [f_2 \circ g]) = \frac{1}{2} \inf\{\log \kappa(h) : h \in [f_2 \circ f_1^{-1}]\} = \mathrm{d_T}([f_1], [f_2]).$$

It is easy to see that the quotient space $\mathcal{T}(R)/\mathrm{Mod}(R)$ coincides with the moduli space $\mathcal{M}(R)$ of complex structures on R. Indeed, suppose $f_1 : R \to S_1$ and $f_2 : R \to S_2$ represent the same point on $\mathcal{M}(R)$, that is, there exists a conformal homeomorphism $h : S_1 \to S_2$. Then $g = f_2^{-1} \circ h \circ f_1$ is a quasi-conformal homeomorphism of R and $[f_2] = \langle g \rangle [h \circ f_1] = \langle g \rangle [f_1]$. In the converse direction, if $[f_2] = \langle g \rangle [f_1] = [f_1 \circ g^{-1}]$ for some quasi-conformal homeomorphism g, then there exists some conformal homeomorphism $h : S_1 \to S_2$. In particular, f_1 and f_2 represent the same point in the moduli space $\mathcal{M}(R)$.

Quadratic differentials. A *quadratic differential* $\phi\, dz^2$ on a Riemann surface assigns to each point a complex quadratic form on the corresponding tangent space, depending meromorphically on the point. See Strebel [51]. In other words, given any local coordinate z on the Riemann surface, a quadratic differential corresponds to an expression $\phi(z)dz^2$ where the coefficient $\phi(z)$ is a meromorphic function. Moreover, the expression $\phi'(w)dw^2$ with respect to any other local coordinate w must satisfy

$$\phi'(w) = \phi(z) \left(\frac{dz}{dw} \right)^2$$

on the intersection of the domains. We call the quadratic differential *holomorphic* if the coefficient $\phi(z)$ is a holomorphic function, relative to any local coordinate.

A non-zero quadratic differential comes with a pair of transverse foliations, $\mathcal{F}^h$ and $\mathcal{F}^v$, defined as follows. The horizontal direction and the vertical direction at a point z are defined by

$$\phi(z)\, dz^2 > 0 \quad \text{and} \quad \phi(z)\, dz^2 < 0$$

respectively. By integrating these directions, one obtains the *horizontal foliation* $\mathcal{F}^h$ and the *vertical foliation* $\mathcal{F}^v$, respectively. The definition does not make sense at the zeros and poles of the differential, where the two foliations may exhibit singularities, as illustrated in Figure A.2.

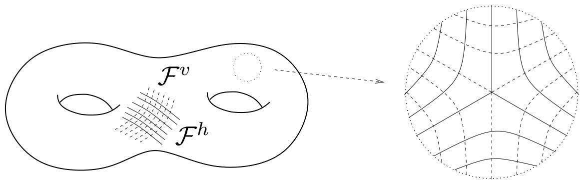


Figure A.2:

The *norm* of a quadratic differential $\phi\, dz^2$ is defined by

$$\|\phi\| := \int |\phi(z)|\, dz\, d\bar{z}$$

(decompose the surface into domains of local charts and integrate the absolute value of the corresponding coefficient on each of the domains). We call the quadratic differential *integrable* if its norm is finite. Then all its poles, if any, are simple. On compact Riemann surfaces the converse is also true: quadratic differentials with no poles of order larger 1 are integrable.

Abelian differentials. Let $\phi\, dz^2$ be an integrable quadratic differential. Near any point p which is neither a zero nor a pole, one may choose *adapted* local coordinates

$$\zeta = \int_p^z \sqrt{\phi(w)}\, dw,$$

so that $\phi(z)\, dz^2 = d\zeta^2$. Near a zero of order (or multiplicity) $m \geq 1$ or near a pole (case $m = -1$) one may take

$$\zeta = \left(\int_p^z \sqrt{\phi(w)}\, dw \right)^{2/m+2}$$

and then $\phi(z)\, dz^2 = (m/2+1)^2 \zeta^m d\zeta^2 = d\left(\zeta^{\frac{m}{2}+1}\right)^2$. Thus, the quadratic differential defines a flat Riemann metric on the surface, transported from $\mathbb{C}$ by such adapted coordinates, with conical singularities at the zeros and poles of the coefficient. The total area of this Riemann metric coincides with the norm of $\phi\, dz^2$.

The horizontal and vertical fields of directions, $d\zeta^2 > 0$ and $d\zeta^2 < 0$, are constant in adapted coordinates, which means that they are parallel for the flat metric. Any other adapted coordinate ζ' near a regular point, obtained from a different choice of the starting local coordinate z or of the sign of the square root, satisfies

$$\zeta' = \pm\zeta + \text{const}$$

on the intersection of their domains. This structure falls short of a translation surface only because of the $\pm$ sign: it is sometimes called a *half-translation* *surface.* Near a singularity, changes of adapted coordinates may also be involve multiplication by a $(m+2)$nd root of unity.

In the case when $\phi\, dz^2$ is the square of a holomorphic 1-form $\alpha\, dz$, that is when $\phi(z)dz^2 = \big(\alpha(z)dz\big)^2$, one may define adapted coordinates near any regular point p by

$$\zeta = \int_p^z \alpha(z)\, dz$$

so that $\alpha(z)\, dz = d\zeta$. Changes between such coordinates, near regular points, are always of the form

$$\zeta' = \zeta + \text{const}\,.$$

A holomorphic 1-form is called an *Abelian differential* and the observation we just made means that each Abelian differential determines a translation structure on the surface. The geodesics in the associated flat metric are the straight lines $\arg d\zeta = \text{const}$. In this case, the horizontal direction field is naturally oriented by $d\zeta > 0$, and so it gives rise to a parallel *vector* field. The same is true about the vertical direction field, of course.

Notice that the square of an Abelian differential is always a quadratic differential. Such quadratic differentials are called *orientable.*

Teichmüller theorem. The Teichmüller theorem states that every equivalence class $[f] \in \mathcal{T}(R)$ includes a map $f : R \to S$ that minimizes the dilatation in the class: $\kappa(f) \leq \kappa(f')$ for any $f' \in [g]$. This extremal homeomorphism f is unique, up to post-composition with conformal maps, and it admits a very precise description, as follows (see Figure A.3).

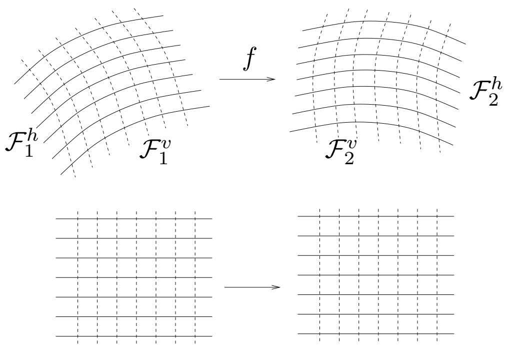


Figure A.3:

There exist integrable quadratic differentials $\phi = \phi\, dz^2$ on R and $\psi = \psi\, dz^2$ on S, such that f maps the horizontal foliation of ϕ to the horizontal foliation of ψ, and analogously for the vertical foliations. Moreover, in (horizontal,vertical) coordinates, the map f has the form

$$f(x,y) = \big(K^{1/2}x, K^{-1/2}y\big)$$

for some $K < \infty$. Let us state this more precisely. Firstly, f maps each zero of ϕ to a zero of ψ with the same order, and analogously for the (simple) poles. Secondly, near regular points the Beltrami coefficient $\mu_f(\zeta)$ is constant equal to some $k < 1$, if one considers adapted local coordinates $\zeta = x + iy$ in R and $\eta = u + iv$ in S, such that $\phi(z)dz^2 = d\zeta^2$ and $\psi(w)dw^2 = d\eta^2$. This means, in other words, that

$$f(x+iy) = K^{1/2}u + iK^{-1/2}v, \qquad \text{with } K = \frac{1+k}{1-k}\,.$$

In particular, $\kappa(f)$ is constant equal to K and the map f is real analytic, outside the singularities (zeros and poles) of the quadratic differentials.

We call ϕ the *initial* differential and ψ the *final* differential of f. By definition, the pull-back of ψ under f is an (integrable) quadratic differential for the conformal structure $[f]$. Thus, by varying either ϕ or k one obtains deformations of the initial conformal structure, together with quadratic differentials for the deformed conformal structures. This observation leads to the following important definition.

Teichmüller flow. Consider $[f] \in \mathcal{T}(R)$ with $f : R \to S$, and let ψ be an integrable quadratic differential on the Riemann surface S, and ϕ be its pull-back by f. For each $t \in \mathbb{R}$, define

$$K_t = e^{2t}, \quad k_t = \frac{K_t - 1}{K_t + 1}\,, \quad \mu_t = k_t \frac{|\psi|}{\psi}\,.$$

Let $g_t : S \to S_t$ be a quasi-conformal map with Beltrami differential μ_t. Then g_t is an minimizing map in the sense of the Teichmüller theorem, with ψ as initial differential and some ψ_t as final differential on S_t. Then $[g_t \circ f]$ is a curve in the Teichmüller space $\mathcal{T}(R)$, and the pull-back ϕ_t of each ψ_t under $f \circ g_t$ is a quadratic differential for the conformal structure defined by $[g_t \circ f]$. Then

$$\big([f], \phi\big) \mapsto \big([g_t \circ f], \phi_t\big)$$

defines a flow in the space $\mathcal{Q}(R)$ of pairs $([f], \phi)$ such that $[f] \in \mathcal{T}(R)$ and ϕ is an integrable quadratic differential for the conformal structure defined by $[f]$. This is called the Teichmüller flow. See Figure A.4.

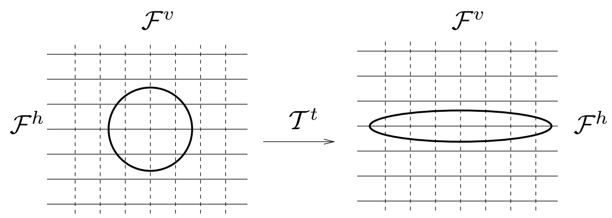


Figure A.4:

The space $\mathcal{Q}(R)$ is a fiber bundle over the Teichmüller space, where each fiber is the space of integrable quadratic differentials for the conformal structure associated to the base point. The unit subbundle $\mathcal{Q}_1(R)$ is the subset of pairs for which the norm

$$\|\phi\| := \int |\phi| \, dz \, d\bar{z} = 1.$$

From the definition we have that the Teichmüller flow preserves the norm and, in particular, leaves this unit subbundle invariant.

Given any quadratic differential $\phi \, dz^2$ and Beltrami differential $\mu \, d\bar{z}/dz$, their product is a well-defined area form $\mu\phi \, dz \, d\bar{z}$. This observation provides a duality between the two spaces

$$\langle \mu, \phi \rangle := \int \mu(z)\phi(z) \, dz \, d\bar{z}$$

through which the a convenient quotient of space of bounded Beltrami differentials may be identified with the space of integrable quadratic differentials.

The measurable Riemann mapping theorem allows us to view Beltrami differentials as directions of deformation of the conformal structure, that is, as a sort of tangent vectors to the moduli space or the Teichmüller space. Then, in view of the previous observations, the space $\mathcal{Q}(R)$ may be thought of as a cotangent bundle to $\mathcal{T}(R)$ or $\mathcal{M}(R)$. Observe also that the projection of each flow trajectory down to $\mathcal{T}(R)$ is a geodesic relative to the Teichmüller metric: since g_t is an extremal map,

$$\mathrm{d_T}([f], [g_t \circ f]) = \frac{1}{2}\kappa(g_t) = \frac{1}{2} \log K_t = t$$

for all $t \in \mathbb{R}$. For these reasons, one often speaks of the Teichmüller *geodesic* flow on the cotangent bundle of $\mathcal{T}(R)$ or $\mathcal{M}(R)$.

Moduli spaces of Abelian and quadratic differentials. Two Abelian differentials on surfaces of genus g are conformally equivalent if they are mapped to one another by some conformal homeomorphism. The moduli space $\mathcal{A}_g$ is the space of conformally equivalence classes. $\mathcal{A}_g$ is a complex orbifold of (complex) dimension $d = 4g - 3$. It is naturally stratified by the subsets $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ of Abelian differentials whose zeros have multiplicities $m_1, \ldots, m_\kappa \geq 1$. Here $\kappa \geq 0$ is the number of zeros and the multiplicities must satisfy the Gauss-Bonnet (or Euler-Poincaré) formula

$$m_1 + \cdots + m_\kappa = 2g - 2.$$

Each stratum $\mathcal{A}_g(m_1, \ldots, m_\kappa)$ is a complex orbifold of the moduli space with (complex) dimension $2g + \kappa - 1$. This dimension is highest when all singularities have multiplicity 1 (and so $\kappa = 2g - 2$), in which case it coincides with the dimension of the whole $\mathcal{A}_g$. It is minimum when there is only one singularity, with the maximum multiplicity $2g-2$. We call $\mathcal{A}_g(1,\ldots,1)$ the *principal stratum* and $\mathcal{A}_g(2g-2)$ the *minimal stratum* of $\mathcal{A}_g$.

Similarly, the moduli space $\mathcal{Q}_g$ of quadratic differentials on surfaces of genus g is naturally stratified by the subsets $\mathcal{Q}_g(n_1,\ldots,n_\sigma)$ having $\sigma \geq 0$ singularities, with orders $n_1,\ldots,n_\sigma \in \{-1\} \cup \mathbb{N}$. The cases $n_i = -1$ correspond to simple poles, whereas the singularity with $n_i \geq 1$ are zeros of the differential. These orders must satisfy

$$n_1 + \cdots + n_\sigma = 4g-4.$$

However, unlike the Abelian case, a few strata of quadratic differentials turn out to be empty. See Masur, Smillie [42]. Each non-empty stratum $\mathcal{Q}_g(n_1,\ldots,n_\sigma)$ is a complex orbifold with (complex) dimension $2g+\sigma-2$. In particular, the complex dimension of the *principal stratum* $\mathcal{Q}_g(1,\ldots,1)$ is $6g-6$.

One may identify $\mathcal{A}_g$ with the suborbifold of orientable quadratic differentials in $\mathcal{Q}_g$, via the map $\alpha \mapsto q = \alpha^2$. This maps sends each $\mathcal{A}_g(m_1,\ldots,m_\kappa)$ inside $\mathcal{Q}_g(n_1,\ldots,n_\sigma)$ with $\sigma = \kappa$ and $n_i = 2m_i$ for all i. In the converse direction (see Lanneau [35, Section 2]), every quadratic differential $q \in \mathcal{Q}_g(n_1,\ldots,n_\sigma)$ may be lifted to an Abelian differential on a convenient branched cover [4] of the surface. Assuming the n_i have been ordered in such a way that the first $\theta \in \{0,\ldots,\sigma\}$ are odd, and the remaining $\sigma-\theta$ are even, then this corresponds to an embedding of $\mathcal{Q}_g(n_1,\ldots,n_\theta,n_{\theta+1},\ldots,n_\sigma)$ inside

$$\mathcal{A}_{\hat{g}}(n_1+1,\ldots,n_\theta+1,\frac{1}{2}n_{\theta+1},\frac{1}{2}n_{\theta+1},\ldots,\frac{1}{2}n_\sigma,\frac{1}{2}n_\sigma).$$

Thus, the number of zeros is $\kappa = \theta + 2(\sigma-\theta)$ zeros, and the two genera are related by

$$\theta + 4g - 4 = 2\hat{g} - 2.$$

Veech [59] and Arnoux discovered that strata of Abelian differentials need not be connected. Recently, Kontsevich, Zorich [33] gave a complete classification of the connected components: while most strata turn out to be connected, some may have two or even three components. In the non-connected case, the components are distinguished by two invariants called hyperellipticity and spin parity. The maximum of three connected components is attained by the minimal stratum $\mathcal{A}_g(2g-2)$ for $g \geq 4$. Lanneau [35, 34] carried out a similar classification in the case of quadratic differentials: in this case, may have at most two connected components.

Measured foliations. Let us also recall some basic ingredients in Thurston's theory of measured foliations. For detailed presentations, see Thurston [52], Fathi, Laudenbach, Poenaru [13], and Casson, Bleiber [10].

Vertical and horizontal foliations of a quadratic differential ϕ come with additional structure, namely, certain transverse length measures, defined by

$$\ell(\gamma) = \int_\gamma \sqrt{|\phi(z)|}\,|dz|$$

[4] The cover is connected precisely if the quadratic differential is *not* the square of an Abelian differential.

for any curve γ transverse to the foliation. If the quadratic differential is orientable, that is, the square of an Abelian differential $\alpha(z)\,dz$, this length measure just takes the form

$$\ell(\gamma) = \int_\gamma |\alpha(z)|\,|dz|.$$

Observe that the vertical foliation and the horizontal foliation are characterized by being tangent, at every regular point, to the kernel of the closed real 1-forms $\Im\alpha$ and $\Re\alpha$. Their leaves are geodesics for the associated flat metric and, conversely, any geodesic is a vertical (or horizontal) leaf for the product of the Abelian differential by some norm 1 complex constant c.

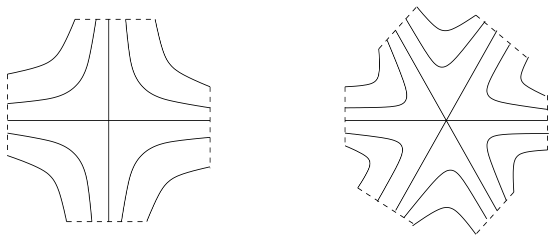


Figure A.5:

In general, one calls *measured foliation* defined by a real closed 1-form β, the foliation whose leaves are the curves tangent at every regular point to the kernel of β. It is assumed that β vanishes only at finitely many points, and these are saddle-type singularities of the foliation (possibly degenerate). See Figure A.5. The 1-form defines a transverse measure to the foliation

$$\ell(\gamma) = \int_\gamma |\beta|,$$

and this measure is invariant under all holonomy maps (projections from one cross-section to another along the leaves of the foliation). In particular, any Poincaré return map of the measured foliation to a given a cross-section γ preserves the transverse measure. One can always parametrize γ in such a way that this invariant measure correspond to the usual Lebesgue measure in the parameter. Then, assuming the return map is piecewise continuous, it must be an interval exchange transformation.

The global structure of measured foliations is well described by the theorem of Maier [38]: the surface splits into open components of two types, *periodic* or *minimal*, separated by saddle-connections; periodic components consist of closed leaves, whereas all leaves in a minimal component are dense in the component. By *saddle-connection* one means a leaf that connects two singularities; the saddle-connection is called a *homoclinic loop* if the two singularities coincide. If there are no saddle-connections then there is only one component, and it must be of minimal type (at least for genus larger than 1).

Most measured foliations can be realized as the horizontal or vertical foliation of a translation surface. Indeed, a measured foliation is realizable in this way if and only if any two regular points can be joined by an increasing cross-section to the foliation, that is, a curve on whose tangent vectors the 1-form β is always positive. This fact was proved, independently and in different contexts by Calabi [9] and Hubbard, Masur [22]. Also independently, Katok [24] proved that measured foliations without saddle-connections are always realizable.

Bibliography

[1] L. Ahlfors. *Lectures on quasiconformal mappings.* Van Nostrand Mathematical Studies, No. 10. D. Van Nostrand Co., 1966.

[2] M. Atiyah. Riemann surfaces and spin structures. *Annales Sci. E.N.S.*, 4:47–62, 1971.

[3] A. Avila and G. Forni. Weak mixing for interval exchange transformations and translation flows. *Annals of Math.*, 165:637–664, 2007.

[4] A. Avila and M. Viana. Simplicity of Lyapunov spectra: proof of the Zorich-Kontsevich conjecture. *Acta Math.*, 198:1–56, 2007.

[5] A. Avila and M. Viana. Symplicity of Lyapunov spectra: A sufficient criterion. *Portugaliae Mathematica*, 64:311–376, 2007.

[6] G. Birkhoff. *Lattice theory*, volume 25. A.M.S. Colloq. Publ., 1967.

[7] C. Bonatti and M. Viana. Lyapunov exponents with multiplicity 1 for deterministic products of matrices. *Ergod. Th. & Dynam. Sys*, 24:1295–1330, 2004.

[8] M. Boshernitzan. A condition for minimal interval exchange maps to be uniquely ergodic. *Duke Math. J.*, 52:723–752, 1985.

[9] E. Calabi. An intrinsic characterization of harmonic one-forms. In *Global Analysis (Papers in Honor of K. Kodaira)*, pages 101–117. Univ. Tokyo Press, Tokyo, 1969.

[10] A. Casson and S. Bleiler. *Automorphisms of surfaces after Nielsen and Thurston*, volume 9 of *London Mathematical Society Student Texts.* Cambridge University Press, 1988.

[11] A. Eskin, H. Masur, and A. Zorich. Moduli spaces of abelian differentials: the principal boundary, counting problems, and the Siegel-Veech constants. *Publ. Math. Inst. Hautes Études Sci.*, 97:61–179, 2003.

[12] H. Farkas and I. Kra. *Riemann surfaces*, volume 71 of *Graduate Texts in Mathematics.* Springer-Verlag, second edition, 1992.

[13] A. Fathi, F. Laudenbach, and V. Poenaru. *Travaux de Thurston sur les surfaces*, volume 66 of *Astérisque*. Société Mathématique de France, 1979. Séminaire Orsay, With an English summary.

[14] G. Forni. Deviation of ergodic averages for area-preserving flows on surfaces of higher genus. *Ann. of Math.*, 155:1–103, 2002.

[15] W. Fulton. *Algebraic topology.* Springer Verlag, 1995.

[16] H. Furstenberg. Boundary theory and stochastic processes on homogeneous spaces. In *Harmonic analysis in homogeneous spaces*, volume XXVI of *Proc. Sympos. Pure Math. (Williamstown MA, 1972)*, pages 193–229. Amer. Math. Soc., 1973.

[17] F. Gardiner. *Teichmüller theory and quadratic differentials.* Pure and Applied Mathematics. John Wiley & Sons Inc., 1987.

[18] I. Ya. Gol'dsheid and G. A. Margulis. Lyapunov indices of a product of random matrices. *Uspekhi Mat. Nauk.*, 44:13–60, 1989.

[19] Y. Guivarc'h and A. Raugi. Products of random matrices : convergence theorems. *Contemp. Math.*, 50:31–54, 1986.

[20] S. Gutkin and C. Judge. The geometry and arithmetic of translation surfaces with applications to polygonal billiards. *Math. Res. Lett.*, 3:391–403, 1996.

[21] S. Gutkin and C. Judge. Affine mappings of translation surfaces: geometry and arithmetic. *Duke Math. J.*, 103:191–213, 2000.

[22] J. Hubbard and H. Masur. Quadratic differentials and foliations. *Acta Math.*, 142:221–274, 1979.

[23] D. Johnson. Spin structures and quadratic forms on surfaces. *J. London. Math. Soc.*, 22:365–373, 1980.

[24] A. Katok. Invariant measures of flows on orientable surfaces. *Dokl. Akad. Nauk SSSR*, 211:775–778, 1973.

[25] A. Katok. Lyapunov exponents, entropy and periodic points of diffeomorphisms. *Publ. Math. IHES*, 51:137–173, 1980.

[26] M. Keane. Interval exchange transformations. *Math. Zeit.*, 141:25–31, 1975.

[27] M. Keane. Non-ergodic interval exchange transformations. *Israel J. Math.*, 26:188–196, 1977.

[28] M. Keane and G. Rauzy. Stricte ergodicité des échanges d'intervalles. *Math. Zeit.*, 174:203–212, 1980.

[29] S. Kerckhoff. Simplicial systems for interval exchange maps and measured foliations. *Ergod. Th. & Dynam. Syst.*, 5:257–271, 1985.

[30] S. Kerckhoff, H. Masur, and J. Smillie. Ergodicity of billiard flows and quadratic differentials. *Ann. of Math.*, 124:293–311, 1986.

[31] H. Keynes and D. Newton. A minimal non-uniquely ergodic interval exchange transformation. *Math. Zeit.*, 148:101–105, 1976.

[32] M. Kontsevich. Lyapunov exponents and Hodge theory. In *The mathematical beauty of physics (Saclay, 1996)*, volume 24 of *Adv. Ser. Math. Phys.*, pages 318–332. World Sci. Publishing, 1997.

[33] M. Kontsevich and A. Zorich. Connected components of the moduli spaces of Abelian differentials with prescribed singularities. *Invent. Math.*, 153:631–678, 2003.

[34] E. Lanneau. Connected components of the strata of the moduli space of quadratic differentials. Preprint www.arxiv.org 2005.

[35] E. Lanneau. Hyperelliptic components of the moduli spaces of quadratic differentials with prescribed singularities. *Comment. Math. Helv.*, 79:471–501, 2004.

[36] E. Lanneau. Parity of the spin structure defined by a quadratic differential. *Geom. Topol.*, 8:511–538 (electronic), 2004.

[37] O. Lehto. *Univalent functions and Teichmüller spaces*, volume 109 of *Graduate Texts in Mathematics.* Springer-Verlag, 1987.

[38] A. G. Maier. Trajectories on closable orientable surfaces. *Sb. Math.*, 12:71–84, 1943. (in Russian).

[39] R. Mañé. *Ergodic theory and differentiable dynamics.* Springer Verlag, 1987.

[40] S. Marmi, P. Moussa, and J.-C. Yoccoz. The cohomological equation for Roth type interval exchange transformations. *Journal of the Amer. Math. Soc.*, 18:823–872, 2005.

[41] H. Masur. Interval exchange transformations and measured foliations. *Ann. of Math*, 115:169–200, 1982.

[42] H. Masur and J. Smillie. Quadratic differentials with prescribed singularities and pseudo-Anosov diffeomorphisms. *Comment. Math. Helv.*, 68:289–307, 1993.

[43] J. Milnor. *Topology from the differentiable viewpoint.* Princeton Landmarks in Mathematics. Princeton University Press, 1997. Revised reprint of the 1965 original.

[44] D. Mumford. Theta characteristics of an algebraic curve. *Annales Sci. E.N.S.*, 4:181–191, 1971.

[45] A. Nogueira and D. Rudolph. Topological weak-mixing of interval exchange maps. *Ergod. Th. & Dynam. Sys.*, 17:1183–1209, 1997.

[46] V. I. Oseledets. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19:197–231, 1968.

[47] G. Rauzy. Échanges d'intervalles et transformations induites. *Acta Arith.*, 34:315–328, 1979.

[48] M. Rees. An alternative approach to the ergodic theory of measured foliations on surfaces. *Ergod. Th. & Dynam. Sys.*, 1:461–488, 1981.

[49] R. Schwartz. Billiards obtuse and irrational. Preprint 2005.

[50] S. Schwartzman. Asymptotic cycles. *Ann. of Math.*, 66:270–284, 1957.

[51] K. Strebel. *Quadratic differentials.* Ergebnisse der Mathematik. Springer Verlag, 1984.

[52] W. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.*, 19:417–431, 1988.

[53] W. Veech. Interval exchange transformations. *J. Analyse Math.*, 33:222–272, 1978.

[54] W. Veech. Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math.*, 115:201–242, 1982.

[55] W. Veech. The metric theory of interval exchange transformations. I. Generic spectral properties. *Amer. J. Math.*, 106:1331–1359, 1984.

[56] W. Veech. The metric theory of interval exchange transformations. II. Approximation by primitive interval exchanges. *Amer. J. Math.*, 106:1361–1387, 1984.

[57] W. Veech. The metric theory of interval exchange transformations. III. The Sah-Arnoux-Fathi invariant. *Amer. J. Math.*, 106:1389–1422, 1984.

[58] W. Veech. The Teichmüller geodesic flow. *Ann. of Math.*, 124:441–530, 1986.

[59] W. Veech. Moduli spaces of quadratic differentials. *J. Analyse Math.*, 55:117–171, 1990.

[60] W. Veech. Geometric realizations of hyperelliptic curves. In *Algorithm, fractals, and dynamics (Okayama/Kyoto 1992)*, pages 217–226. Plenum, 1995.

[61] J.-C. Yoccoz. Continued fraction algorithms for interval exchange maps: an introduction. In *Frontiers in Number Theory, Physics and Geometry*, volume 1 : On random matrices, zeta functions and dynamical systems. Springer Verlag, 2006. École de Physique des Houches, 2003.

[62] A. Zorich. Asymptotic flag of an orientable measured foliation on a surface. In *Geometric study of foliations (Tokyo, 1993)*, pages 479–498. World Sci. Publishing, 1994.

[63] A. Zorich. Finite Gauss measure on the space of interval exchange transformations. Lyapunov exponents. *Ann. Inst. Fourier (Grenoble)*, 46:325–370, 1996.

[64] A. Zorich. Deviation for interval exchange transformations. *Ergodic Theory Dynam. Systems*, 17:1477–1499, 1997.

[65] A. Zorich. How do the leaves of a closed 1-form wind around a surface? In *Pseudoperiodic Topology*, volume 197 of *Amer. Math. Soc. Transl. Ser. 2*, pages 135–178. Amer. Math. Soc., 1999.

[66] A. Zorich. Flat surfaces. In *Frontiers in Number Theory, Physics and Geometry*, volume 1 : On random matrices, zeta functions and dynamical systems. Springer Verlag, 2006. École de Physique des Houches, 2003.

Index