Modular symbols for Teichmüller curves

Curtis T. McMullen

1 April 2019

Abstract

This paper introduces a space of nonabelian modular symbols $\mathcal{S}(V)$ attached to any hyperbolic Riemann surface V, and applies it to obtain new results on polygonal billiards and holomorphic 1-forms. In particular, it shows the scarring behavior of periodic trajectories for billiards in a regular polygon is governed by a countable set of measures homeomorphic to $\omega^\omega + 1$.

Contents

Research supported in part by the NSF. Typeset 2021-09-07 13:01.

1 Introduction

The purpose of this paper is to investigate the following related phenomena:

- the behavior of long, periodic trajectories for billiards in a regular polygon;
- the distribution of the closed geodesics on a singular flat surface $(X, |\omega|)$, when $\mathrm{SL}(X, \omega)$ is a lattice;
- the different ways of describing a Teichmüller curve $V \to \mathcal{M}_g$ by purely topological data; and
- the closure of the projective monodromy group for such a curve.

In all four cases, a countable compact set homeomorphic to the ordinal $\omega^\omega + 1$ emerges.[1]

We will show that these sets are all reflections of the space of nonabelian modular symbols $\mathcal{S}(V)$, which is itself homeomorphic to ω^ω. This new structure is a natural enhancement of the classical modular symbols for $\mathrm{SL}_2(\mathbb{Z})$.

Billiards. Here is a special case of the phenomena we will describe. Consider billiards in a regular n–sided polygon P_n, $n \geq 3$. It is well–known that P_n has optimal dynamics: every billiard trajectory in P_n is either periodic, or uniformly distributed [V1].

Somewhat paradoxically, one can still find long periodic billiard trajectories that are not equidistributed. To describe their behavior, let M_s denote the set of probability measures on P_n that arise from limits of periodic trajectories C_n with slopes $s_n \to s$ and lengths $L(C_n) \to \infty$. We will show:

Theorem 1.1 *The space of limit measures M_s is homeomorphic to $\omega^\omega + 1$ whenever s is a periodic slope and $n \neq 3, 4$ or 6. Its derived set $D^\infty M_s$ consists of a single point, namely normalized area measure on P_n.*

(We note that periodic slopes are dense in $\widehat{\mathbb{R}}$ and include the slopes of the sides of P_n.)

The failure of periodic trajectories to distribute evenly was first observed in the regular pentagon, by Davis and Lelièvre; see [DL, §4] and Figure 1.

[1] The space $E = \omega^\omega + 1$ is characterized by the property that its derived set $D^\infty(E)$ consists of a single point. The first derived set $D(E)$ consists of the limit points of E; $D^{n+1}(E) = D(D^n(E))$; and $D^\infty(E) = \bigcap D^n(E)$.

Theorem 1.1 shows the scarring behavior of such trajectories is governed by a countable set of measures clustered around the uniform distribution.

Equidistribution of C_n can be recovered by requiring that the lengths of the 'continued fractions' of the slopes s_n tend to infinity (see [Mc5, Theorem 3] and Theorem 1.4 below). Equidistribution is automatic if the slope s is aperiodic; in this case M_s is a single point.

Ergodic measures. Theorem 1.1 completes the description of the *closure* of the ergodic invariant measures for billiards in a regular polygon. The closure consists of the measures on periodic orbits, together with $\bigcup M_s$. As we will see in detail below, when s is a periodic slope, *none* of the measures in M_s are ergodic. By comparison, for unipotent flows, limits of ergodic measures remain ergodic [MS].

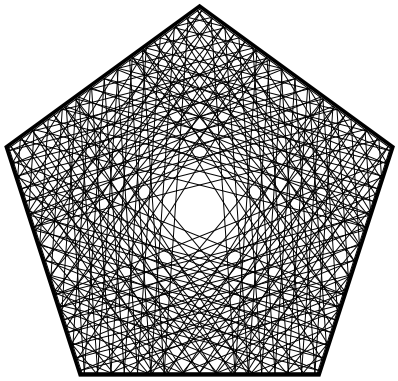

Figure 1. Long periodic billiard trajectories need not be uniformly distributed.

Teichmüller curves. The theory of billiards in rational polygons can be related to the natural action of $\mathrm{SL}_2(\mathbb{R})$ on the bundle $\Omega\mathcal{M}_g \to \mathcal{M}_g$ of holomorphic 1–forms (X, ω) over the moduli space of Riemann surfaces X of genus g.[2] Via this connection, Theorem 1.1 will be an immediate consequence of Theorem 1.2 below.

Let (X, ω) be a holomorphic 1–form of genus g, normalized so that $\int_X |\omega|^2 = 1$. In local coordinates where $\omega = dz$, the geodesics on the singular flat surface $(X, |\omega|)$ are simply straight lines in the complex plane with constant slope s.

[2]The notation ω^ω refers to an ordinal; otherwise, ω will always denote a holomorphic 1–form on a compact Riemann surface $X \in \mathcal{M}_g$.

For any such form we have an exact sequence

$$1 \to \mathrm{Aut}(X,\omega) \to \mathrm{Aff}^+(X,\omega) \stackrel{D}{\to} \mathrm{SL}(X,\omega) \to 1,$$

where the three main terms are (i) the group of holomorphic automorphisms of X preserving ω; (ii) the group of orientation–preserving real affine automorphisms ϕ of (X,ω); and (iii) the stabilizer of (X,ω) in $\mathrm{SL}_2(\mathbb{R})$. We remark that the real affine automorphism are simply the homeomorphisms that send geodesics to geodesics.

Let us focus on the case where $\mathrm{SL}(X,\omega)$ is a *lattice* in $\mathrm{SL}_2(\mathbb{R})$. In this case, the $\mathrm{SL}_2(\mathbb{R})$ orbit of (X,ω) projects to give an algebraic, isometrically immersed *Teichmüller curve*

$$f : V = \mathbb{H}/\mathrm{SL}(X,\omega) \to \mathcal{M}_g.$$

Globally, our assumption that $\mathrm{SL}(X,\omega)$ is a lattice implies the geodesic flow has *optimal dynamics*: for any $s \in \widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, either:

(i) every geodesic with slope s is closed, or
(ii) every geodesic with slope s is dense and equidistributed.

See [V1]. In case (i) we say s is *periodic*; otherwise it is *aperiodic*. In case (ii), we have equidistribution with respect to the probability measure $|\omega|^2$ on X.

Cylinders and cusps. The periodic slopes are dense in $\widehat{\mathbb{R}}$, and correspond to the cusps of $\mathrm{SL}(X,\omega)$. In fact, the smooth periodic geodesics with slope s sweep out a collection of open cylinders

$$A = C(s) = \{A_1, \ldots, A_n\},$$

whose closures cover X. The *modulus* of A_i is defined in terms of its height and circumference, in the metric $|\omega|$, by $m_i = \mathrm{mod}(A_i) = h(A_i)/c(A_i)$. It is known that the cylinders with a given slope have rational ratios of moduli; thus (m_i) is proportional to a unique vector of relatively prime integers,

$$m(s) = (a_1, \ldots, a_n) = (m_1, \ldots, m_n)/m, \tag{1.1}$$

where $m = \gcd(m_1, \ldots, m_n)$. The associated *fundamental twist* is the element of $\mathrm{Aff}^+(X,\omega)$ given by

$$\tau_A = \tau_1^{a_1} \cdots \tau_n^{a_n}, \tag{1.2}$$

where τ_i is an affine right Dehn twist supported in A_i; and $D\tau_A$ is a parabolic element of $\mathrm{SL}(X,\omega)$, fixing the unique cusp on $\partial\mathbb{H}$ corresponding to s.

Limit measures. Consider a sequence of closed geodesics C_n with slopes $s_n \to s$ and lengths

$$L(C_n) = \int_{C_n} |\omega| \to \infty.$$

We say μ is a *limit measure* for slope s if the C_n can be chosen so that

$$\frac{1}{L(C_n)} \int_{C_n} f|\omega| \to \int_X f\,\mu$$

for all $f \in C(X)$. Let M_s denote the compact set of all such probability measures. By (ii) above, if s is aperiodic then $M_s = \{|\omega|^2\}$.

Our main result on limit measures treats the periodic case.

Theorem 1.2 *Let s be a periodic slope for (X, ω). Then either:*

- *The space of limit measures M_s is homeomorphic to $\omega^\omega+1$, and $D^\infty M_s$ consists of a single point, namely the uniform measure $|\omega|^2$ on X; or*
- *The trace field of $\mathrm{SL}(X, \omega)$ is $\mathbb{Q}$, and $M_s = \{|\omega|^2\}$.*

We also obtain a description of the measures in M_s. Given a homology class $C \in H_1(X, \mathbb{R})$, let $\widehat{\mu}(C)$ denote the unique probability measure proportional to

$$\mu(C) = \sum_i \frac{|\langle A_i, C\rangle|}{c(A_i)} \chi_{A_i} |\omega|^2, \tag{1.3}$$

provided this is nonzero. Here $\langle A_i, C\rangle$ denotes the intersection pairing between C and a closed geodesic contained in A_i.

Theorem 1.3 *The measures $\widehat{\mu}(C)$ coming from closed geodesics C on X form a dense subset of M_s.*

Every other measure in M_s, with the possible exception of $|\omega|^2$, is given by $\widehat{\mu}(C)$ for some rational *homology class C.*

Square–tiled surfaces. The first case of Theorem 1.2 can also occur when the trace field is $\mathbb{Q}$. Indeed, there are square–tiled surfaces of genus two where M_s is a single point for some periodic slopes, and $M_s \cong \omega^\omega + 1$ for others; see §9.

Continued fractions. As a complement to Theorem 1.2, we have the following criterion for equidistribution.

Fix a compact set $K \subset V$, e.g. the complement of standard horoball neighborhoods of its cusps. Given a periodic slope s, let $x = -1/s \in \partial\mathbb{H}$

be the corresponding cusp of $\mathrm{SL}(X,\omega)$, let $\gamma(s)$ be the projection to $V = \mathbb{H}/\mathrm{SL}(X,\omega)$ of the hyperbolic geodesic from $z = i$ to x, and let $T(s)$ be the hyperbolic length of $\gamma(s) \cap K$. (Intuitively, $T(s)$ measures the length of the 'continued fraction' for s.)

Theorem 1.4 *If $T(s_n) \to \infty$, then any sequence of closed geodesics C_n with slopes s_n is uniformly distributed on X. The converse holds provided the trace field of $\mathrm{SL}(X,\omega)$ is irrational.*

Theorems 1.1, 1.2, 1.3 and 1.4 are proved in §8.

Twists and the failure of equidistribution. We now turn to a more detailed discussion of Theorem 1.2.

Here is one mechanism for producing badly distributed long geodesics on $(X, |\omega|)$. Let τ_A be the fundamental twist (1.2) at slope s, and let C be a closed geodesic at a different slope t. Then the slopes of the geodesics $C_n = \tau_A^n(C)$ tend to s and the corresponding measures converge to $\widehat{\mu}(C)$, defined using (1.3). This measure is typically distinct from the uniform measure $|\omega|^2$, so equidistribution fails.

Modular symbols. This twist construction, however, does not account for all limit measures, or even a closed subset thereof. To describe the full space of limit measures, we are led to iterate the twist construction along sequences of periodic slopes $(s_1, \ldots, s_n)$, which are encoded by *modular symbols* σ for the Teichmüller curve V.

The space of modular symbols $\mathcal{S}(V)$ is in turn homeomorphic to ω^ω, giving a natural explanation for the appearance of this ordinal in the statement Theorem 1.2. We will see that $\mathcal{S}(V)$ naturally parameterizes the limit measures for all slopes at once, describes how different slopes interact, and explains why they form a countable set rather than, say, a Cantor set.

We will also see that modular symbols label the Thurston multicurve systems (A_i, B_j) presenting a given 1–form (X,ω), their intersection matrices $i(A_i, B_j)$, and, most importantly, *limits and products* of these matrices.

It is essential to study limits to obtain the a closed set of measures. Remarkably, we will find the closure can also be formed by taking all finite *products* of Thurston matrices. See Theorems 1.7, 1.8 and 1.9 below.

In a case like the regular pentagon, where V has just one cusp, the space of invariant measures is naturally a *semigroup*. The category of modular symbols highlights this unexpected multiplicative structure, which also appears in the matrix entries of the $(2, 5, \infty)$ triangle group and will be developed in a sequel.

Hodge theory. The proof of Theorem 1.2 is completed using a Hodge theory argument (§5), based on contraction of the relative period mapping [Mc5], to show that the only limit measure that might not be accounted for by a modular symbol is the uniform measure on X. This argument also shows that Thurston's intersection matrices $i(A_i, B_j)$ decouple in the limit (see equation (1.11) and Theorem 5.1 below).

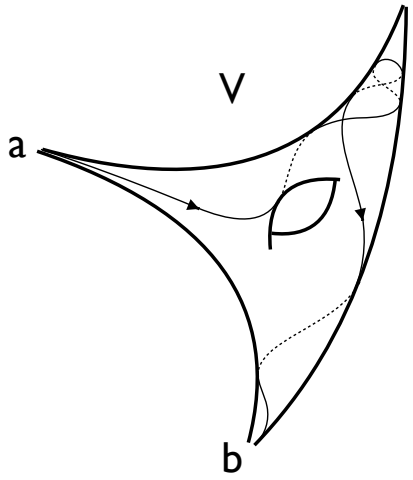

Figure 2. A modular symbol of degree one is a geodesic γ connecting a pair of cusps (a, b) of V.

Definitions. To give a more detailed discussion and formulate these additional results, we begin by defining modular symbols and their associated intersection matrices.

Let $V = \mathbb{H}/\Gamma$ be a hyperbolic Riemann surface. A *modular symbol* of degree d for V is a formal product

$$\sigma = \gamma_1 * \cdots * \gamma_d, \tag{1.4}$$

where $a_0, \ldots, a_d$ are cusps of V, and where γ_i is an oriented hyperbolic geodesic joining a_{i-1} to a_i. Let $\mathcal{S}(V)$ denote the space of all such symbols; $\mathcal{S}^d(V)$, those of degree d; and for any pair of cusps (a, b), let

$$\mathcal{S}_{ab}(V) = \{\sigma \in \mathcal{S}(V) \,:\, a_0 = a \text{ and } a_{\deg(\sigma)} = b\}.$$

The space $\mathcal{S}(V)$ can be regarded as the morphisms in a category whose objects are the cusps of V. (The identity maps have degree 0.) More pre-

cisely, in this category we have

$$\mathcal{S}_{ab} = \mathrm{Mor}(b, a);$$

a chain of geodesics running from a to b gives a morphism from b to a. With this convention we have $\sigma_1 * \sigma_2 = \sigma_1 \circ \sigma_2$. See Figure 2.

The space of modular symbols carries a natural *geometric topology* such that composition is continuous and morphisms of degree ≤ 1 are dense. In the geometric topology, the product σ in equation (1.4) is a limit of geodesics γ_n from a_0 to a_d that make longer and longer excursions into the cusps $a_1, \ldots, a_{d-1}$. It is easy to see we have a homeomorphism

$$\mathcal{S}(V) \cong \omega^\omega,$$

provided V has at least one cusp and $\pi_1(V)$ is not cyclic.

Intersection matrices. We now return to the setting of Teichmüller curves. We define the *intersection matrix* between a pair of cylinder systems $A = C(s)$ and $B = C(t)$ by

$$I(A, B)_{ij} = a_i \cdot i(A_i, B_j). \tag{1.5}$$

Here $(a_i) = m(s)$ is the vector proportional to $(\mathrm{mod}(A_i))$ defined by (1.1). We can regard this integral matrix as a linear map

$$I(A, B) : \mathbb{R}^B \to \mathbb{R}^A;$$

the factors (a_i) are included so that, on the level of $D \in H_1(X, \mathbb{R})$, we have

$$[\tau_A(D)] = [D] + [I(A, B) \cdot D]$$

for all $D \in \mathbb{R}^B$.

The intersection functor. A central and unexpected point of the present paper is that, by extending I to a functor on the full space of modular symbols, we obtain matrices that describe all the measures and currents that arise as limits of closed geodesics on $(X, |\omega|)$. We begin by describing the target of this functor.

The vector space of a cusp. For simplicity, we will assume:

Condition 1.5 *Let s be a periodic slope with $C(s) = \{A_1, \ldots, A_n\}$. Then every $\phi \in \mathrm{Aff}^+(X, \omega)^s$ satisfies $\phi(A_i) = A_i$ for all i.*

Here $\mathrm{Aff}^+(X,\omega)^s$ is the subgroup where $D\phi$ fixes s.

Condition 1.5 can always be achieved by passing to a subgroup of finite index in $\mathrm{Aff}^+(X,\omega)$ (see Theorem 3.1), and all the results which follow hold unconditionally once this is done.

Interactions between cusps. Subject to Condition 1.5, one can functorially attach, to each cusp a of V, a finite set $A = \{A_1, \ldots, A_n\}$ which consistently labels the cylinders in $C(s)$ for every slope s corresponding to a. Hence each cusp of V also determines a finite–dimensional vector space

$$\mathcal{L}(a) = \mathbb{R}^A. \tag{1.6}$$

Given a pair of cusps (a,b) with $(\mathcal{L}(a), \mathcal{L}(b)) = (\mathbb{R}^A, \mathbb{R}^B)$, let

$$\mathcal{L}_{ab}(V) = \mathbb{P}\,\mathrm{Hom}(\mathbb{R}^B, \mathbb{R}^A) \cup \{0\}, \tag{1.7}$$

and let $\mathcal{L}(V) = \bigcup_{(a,b)} \mathcal{L}_{ab}(V)$. The elements of $\mathcal{L}_{ab}(V)$ are matrices up to scale. With its natural topology, $\mathcal{L}_{ab}(V)$ is compact and 0 is an isolated point.

Like $\mathcal{S}(V)$, we can regard $\mathcal{L}(V)$ as the set of morphisms in a category with one object $\mathbb{R}^A$ for each cusp a of V, with

$$\mathcal{L}_{ab}(V) = \mathrm{Mor}(\mathbb{R}^B, \mathbb{R}^A),$$

and with composition defined by matrix multiplication. The composition map is continuous except at points where the product is 0.

The matrix of a modular symbol. We can now define the functor

$$I : \mathcal{S}(V) \to \mathcal{L}(V).$$

Consider first the case of a modular symbol of degree one, i.e. a geodesic γ connecting a pair of cusps (a,b) of V. In this case, a lift of γ to the universal cover of V determines a pair of cusps for $\mathrm{SL}(X,\omega)$ and hence a pair of periodic slopes (s,t); and we simply set

$$[I(\gamma)] = [I(A,B)] \in \mathbb{P}\,\mathrm{Hom}(\mathbb{R}^B, \mathbb{R}^A), \tag{1.8}$$

where $A = C(s)$, $B = C(t)$ and where $I(A,B)$ is defined by equation (1.5). This matrix is independent of the choice of lift of γ.

Since $\mathcal{S}(V)$ is freely generated by its morphisms of degree 1, there is a unique functorial extension of I to all modular symbols. Its value on a general symbol σ is given by

$$[I(\sigma)] = [I(\gamma_1 * \cdots * \gamma_d)] = [I(\gamma_1) \cdot I(\gamma_2) \cdots I(\gamma_d)]. \tag{1.9}$$

Properties of I. We will also need the decoupled matrices

$$R(a,b)_{ij} = [h(A_i)\, c(B_j)] \in \mathcal{L}_{ab}(V), \tag{1.10}$$

which are well–defined up to scale. Their union $\mathcal{R}(V) \subset \mathcal{L}(V)$ is closed under composition. Let $\mathcal{R}_{ab}(V) = \{R(a,b)\}$.

The proof of our main result rests on the two topological properties of I.

Theorem 1.6 *The intersection functor*

$$I : \mathcal{S}(V) \to \mathcal{L}(V)$$

is continuous; and if $\gamma_n \to \infty$ in $\mathcal{S}_{ab}(V)$, then

$$I(\gamma_n) \to [R(a,b)] \in \mathcal{L}_{ab}(V). \tag{1.11}$$

(See Theorems 4.1 and 5.1 below).

Geometrically, the second statement means that parallel transport along γ_n effectively randomizes the curves (A_i), so they become uniformly distributed as seen from the perspective of (B_j). Combining these facts, we will show:

Theorem 1.7 *Letting $T = \{I(\gamma) \,:\, \gamma \in \mathcal{S}^1(V)\}$, we have*

$$\overline{T} = \langle T \rangle \cup \mathcal{R}(V) \subset \mathcal{L}(V). \tag{1.12}$$

Here $\langle T \rangle$ denotes the smallest set of morphisms containing T and closed under composition. Equation (1.12) says that the topological and algebraic completions of T agree, up to the finite set $\mathcal{R}(V)$. Since the algebraic completion of T is countable, so is its topological closure. The shape of $\overline{T}$ is described more precisely by:

Theorem 1.8 *For any pair of cusps (a,b) of V, let $T_{ab} = T \cap \mathcal{L}_{ab}(V)$. We then have:*

$$\overline{T_{ab}} \cong \omega^\omega + 1 \quad \textit{or} \quad T_{ab} = \mathcal{R}_{ab}. \tag{1.13}$$

In the first case, we have $D^\infty \overline{T_{ab}} = \mathcal{R}_{ab}$; in the second case, the trace field of $\mathrm{SL}(X,\omega)$ is $\mathbb{Q}$.

Our results on measures and limits of periodic trajectories follow from a statement similar to Theorem 1.8 for the *columns* of the matrices in $\overline{T_{ab}}$ (see Theorem 7.1).

Thurston's multicurve construction. We now turn to Thurston's construction [Th2]. This construction canonically attaches a holomorphic 1–form $(Y, \eta) \in \Omega\mathcal{M}_g$ to a suitable pair of integral measured laminations (α, β) on a topological surface Σ_g.

Every Teichmüller curve V can be encoded by such a topological pair (α, β); however many pairs give the same V. To make this more precise, let $f : V \to \mathcal{M}_g$ be a complex geodesic generated by (X, ω). Choose a marking homeomorphism $\Sigma_g \cong X$. Using the pair of cylinder systems $(A, B) = (C(s), C(t))$ attached to a lift of γ, define

$$\tau : \mathcal{S}^1(V) \to \mathcal{ML}_g(\mathbb{Z}) \times \mathcal{ML}_g(\mathbb{Z})/\operatorname{Mod}_g$$

by

$$\tau(\gamma) = (\alpha, \beta) = \left(\sum a_i \cdot A_i, \sum b_j \cdot B_j\right). \tag{1.14}$$

Here $(a_i) = m(s)$ and $(b_j) = m(t)$ are defined by equation (1.1), and we have identified A_i and B_j with the simple closed curves they represent on Σ_g. In §10 we will show:

Theorem 1.9 *The map τ gives a natural bijection between the modular symbols of degree 1 on the complex geodesic $f : V \to \mathcal{M}_g$ and the multicurves (α, β) which encode it.*

The intersection matrix $I(\gamma)$ arises naturally in Thurston's construction, and records useful information about V.

Closure of the monodromy group. Finally, we note that the same circle of ideas can be used to study the closure of the (projective) monodromy group G of a Teichmüller curve, defined as the image of the natural map

$$\operatorname{Aff}^+(X, \omega) \to \mathbb{P}\operatorname{End}(H^1(X, \mathbb{R})).$$

For example, when $\operatorname{SL}(X, \omega)$ is a lattice and its trace field is not $\mathbb{Q}$, one can show that

$$\overline{G} \cong \omega^\omega \cup S^1 \times S^1.$$

To be more precise, let $W \subset H^1(X, \mathbb{R})$ be the span of $(\operatorname{Re}\omega, \operatorname{Im}\omega)$, and let $Q \cong S^1 \times S^1$ denote the space of rank one endomorphisms ψ of $H^1(X, \mathbb{R})$, up to scale, such that $\operatorname{Im}(\psi) \subset W$ and $\operatorname{Ker}(\psi) \supset W^\perp$. Then we can write $\overline{G} = K \sqcup Q$, where K is homeomorphic to ω^ω and $Q \subset \overline{K}$. The proof is based on a completion $\overline{\pi}_1(X, p) \cong \omega^\omega$ similar to the space of modular symbols, which bounds the complexity of K from above; on the contractivity of the

complementary period map [Mc5, Theorem 4.1], and on Theorem 7.1 which bounds the complexity of K from below.

An application of modular symbols to the Weil–Petersson metric is given in the Appendix.

Notes and references. The space of modular symbols considered here is well adapted to the study of flat bundles on curves with unipotent monodromy at singular points. Classical modular symbols are an abelian version of the same structure, used to represent relative cohomology classes on $(\overline{V}, \partial V)$ and to study the periods of modular forms for $\mathrm{SL}_2(\mathbb{Z})$ and its congruence subgroups $\Gamma(N)$; see e.g. [Bi], [Maz], [Man], [La, Ch. IV.2]. Drinfeld and Manin used modular symbols to show the cusps of $\mathbb{H}/\Gamma(N)$ form a torsion packet in its Jacobian. Applications of modular symbols to non–arithmetic groups are developed further in [Mc6].

The trajectory shown in Figure 1 starts at the midpoint p of the bottom edge, with slope $s = \sqrt{73225 - 4790\sqrt{5}}/209$; it nearly connects p to a vertex of the pentagon. See [DL] for a detailed study of periodic trajectories in the regular pentagon. Theorem 1.1 above establishes a modified version of [DL, Conj. 4.6]. Theorem 1.6 of [Mc6] gives an explicit description of the limit measures for billiards in a regular pentagon, in terms of the matrix entries for the triangle group group $\Delta(2, 5, \infty) \subset \mathrm{SL}_2(\mathbb{R})$. For surveys on the topic of billiards and moduli spaces, see e.g. [Mas], [Mo1] and [Z].

The ordinal ω^ω also occurs in the study of Pisot numbers [BM], hyperbolic volumes [Th1, §6.6], and cascades of bifurcations of interval exchange maps [Mc4]. Figure 3 of the last paper gives a hint of the behavior of closed trajectories in an octagon, and the mechanism underlying Theorem 1.1; see also [DL, Figure 16].

This paper is a sequel to [Mc5], which establishes the equidistribution results underlying Theorems 1.4 and 5.1.

2 Modular symbols

This section gives a self–contained introduction to the theory of nonabelian modular symbols we will use in the sequel. In particular, we show the space of modular symbols $\mathcal{S}(V)$ is typically homeomorphic to ω^ω, and we give an explicit description of the modular symbols for $\mathrm{SL}_2(\mathbb{Z})$.

Background. Let $V = \mathbb{H}/\Gamma$ be a hyperbolic Riemann surface (or orbifold), presented as the quotient of the upper halfplane $\mathbb{H} = \{z : \mathrm{Im}(z) > 0\}$ by

the action of a discrete group

$$\Gamma \subset G = \mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{Isom}^+(\mathbb{H}).$$

Let $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ denote the boundary of $\mathbb{H}$. A point $x \in \widehat{\mathbb{R}}$ is a *cusp* of Γ if it is the fixed point of a parabolic element $g \in \Gamma$. In this case there is a unique generator p_x for its stabilizer Γ^x which is conjugate to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ in G.

We denote the cusps of Γ by $\kappa(\Gamma)$. There is a natural complex structure on

$$\overline{V} = (\mathbb{H} \cup \kappa(\Gamma))/\Gamma,$$

and when V has finite volume, $\overline{V}$ is compact. We refer to

$$\kappa(V) = \overline{V} - V = \kappa(\Gamma)/\Gamma$$

as the set of cusps of V.

The *thin part* of V, where the injectivity radius is small, consists of collar neighborhoods of short geodesics and horoball neighborhoods of cusps (see e.g. [Th3, §4.5]). The union of the latter forms the *cuspidal thin part* V_{cusp} of V.

Modular symbols. As in the Introduction, a *modular symbol* of degree d for V is a formal product

$$\sigma = \gamma_1 * \cdots * \gamma_d$$

of oriented geodesics running between the cusps of V, such that γ_i ends at the same cusp where γ_{i+1} starts. By adding in these cusps, we obtain a path $\overline{\sigma} : [0, 1] \to \overline{V}$, well–defined up to reparameterization.

Category theory. Whenever σ_1 ends at the same cusp where σ_2 begins, we can form their product $\sigma_1 * \sigma_2$.

The modular symbols $\mathcal{S}(V)$ form the morphisms in a category $\mathcal{C}$ whose objects are the cusps of V. *However*, we must regard a modular symbol running from a to b as morphism from b to a. With this convention, the composition map

$$\mathrm{Mor}(b, a) \times \mathrm{Mor}(c, b) \to \mathrm{Mor}(c, a)$$

is given by $(\sigma_1, \sigma_2) \mapsto \sigma_1 * \sigma_2$, as required by the axioms of a category (cf. [Mun, p.159]).

The morphisms in $\mathcal{C}$ have a natural grading by degree,

$$\mathcal{S}(V) = \bigcup_{d \geq 0} \mathcal{S}^d(V),$$

and $\mathcal{C}$ is freely generated by its morphisms of degree one. The identity maps have degree zero; thus $\mathcal{S}^0(V)$ is in bijection with the cusps of V.

Cusps of Γ. For a more discrete perspective, one can regard a modular symbol of degree 1 as an ordered pair $[x, y]$ of distinct cusps of Γ, subject to the relation

$$[x, y] \sim [gx, gy] \quad \forall g \in \Gamma.$$

The unique hyperbolic geodesic running from x to y in $\mathbb{H}$ projects to the corresponding geodesic γ joining a pair of cusps of V. A general modular symbol of degree d can then be expressed as a product

$$\sigma = [x_0, x_1] * [x_1, x_2] * \cdots * [x_{d-1}, x_d],$$

or more briefly as $\sigma = [x_0, x_1, \ldots, x_d]$.

Topology. The set $\mathcal{S}(V)$ carries a natural *geometric topology*, which makes it into a countable, locally compact Hausdorff space. This topology can be defined briefly as follows: a sequence $\sigma_n \in \mathcal{S}(V)$ converges to ρ if and only if $\overline{\sigma}_n \to \overline{\rho}$ as paths in $\overline{V}$.[3] If σ_n comes closer and closer to a cusp a of V, then this cusp will ultimately be visited by ρ.

Here are some basic properties of $\mathcal{S}(V)$ in the geometric topology.

1. The composition map $\mathcal{S}(V) \times \mathcal{S}(V) \to \mathcal{S}(V)$ is continuous.

2. If a product $\alpha_n * \beta_n$ of modular symbols of degrees d and e converges, then so does each factor, and

$$(\lim \alpha_n) * (\lim \beta_n) = \lim(\alpha_n * \beta_n). \tag{2.1}$$

3. The set $\mathcal{S}^0(V) \cup \mathcal{S}^1(V)$ is dense in $\mathcal{S}(V)$. More generally, for $d \geq 1$ we have

$$\overline{\mathcal{S}^d(V)} = \bigcup_{e \geq d} \mathcal{S}^e(V).$$

4. Provided Γ has infinitely many cusps, $\mathcal{S}^d(V) \neq \emptyset$ for all d, and hence (by the preceding observation) $\mathcal{S}(V)$ is homeomorphic to ω^ω.

5. Let K be a compact subset of V. Then the set of modular symbols σ contained in $K \cup V_{\text{cusp}}$, and with length $L(\sigma \cap K) \leq L_0$, is compact.

[3] Since paths are only well-defined up to reparameterization, convergence here means that there exist homeomorphisms $\phi_n : [0, 1] \to [0, 1]$, fixing the endpoints, such that $\overline{\sigma}_n \circ \phi_n$ converges uniformly to $\overline{\rho}$ on $[0, 1]$.

6. When V has finite volume, we can take $K = V - V_{\text{cusp}}$ and conclude:

The set of modular symbols with $L(\sigma \cap K) \leq L_0$ is compact. (2.2)

Iterated parabolics. The geometric topology can also be characterized from an algebraic perspective. Let $p \in G$ be a parabolic transformation fixing $z \in \widehat{\mathbb{R}} - \{x, y\}$. Then

$$\gamma_n = [x, p^n y] \to [x, z, y] = \sigma$$

as $n \to \infty$. Indeed, the representatives $[x, p^n y]$ and $[p^{-n} x, y]$ of γ_n converge *algebraically* to $[x, z]$ and $[z, y]$ respectively, in the sense that $p^n y \to z$ and $p^{-n} x \to z$. These algebraic limits represent the two components of the path $\overline{\sigma}$ in $V - V_{\text{cusp}}$, so their composition gives σ.

More generally, given distinct cusps $(x, z_1, \ldots, z_e, y)$, and parabolics p_i fixing z_i, we have

$$[x, p_1^{n_1} \cdots p_e^{n_e} y] \to [x, z_1, \ldots, z_e, y] \tag{2.3}$$

as $\inf_i |n_i| \to \infty$. This property explains the density of modular symbols of degree ≤ 1, and together with continuity of composition, it suffices to characterize the topology on $\mathcal{S}(V)$. (See e.g. [MT, §7] for the notions of algebraic and geometric convergence of Kleinian groups, which are similar in spirit.)

Example: $\mathbf{SL_2(\mathbb{Z})}$. As a concrete example, let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Then $V = \mathbb{H}/\mathrm{SL}_2(\mathbb{Z})$ is the $(2, 3, \infty)$ orbifold, and Γ acts transitively on its set of cusps $\kappa(\Gamma) = \mathbb{Q} \cup \{\infty\}$. Since the stabilizer of ∞ in Γ is generated by $z \mapsto z + 1$, we have

$$\mathcal{S}^1(V) = \mathbb{Q}/\mathbb{Z} = \{[\infty, p/q] \ : \ p/q \in [0, 1]\}.$$

It is convenient to describe the modular symbols of degree one using the continued fraction expansion

$$p/q = 1/(a_1 + 1/(a_2 + \cdots + 1/a_n)$$

for rationals $p/q \in [0, 1]$; thus, we will write

$$[\infty, p/q] = \langle a_1, \ldots, a_n \rangle.$$

The continued fraction for a given modular symbol is not quite unique; for example, we have $\langle\rangle = \langle 1 \rangle = [\infty, 0] = [\infty, 1]$, and $\langle a_1, \ldots, a_n, 1 \rangle = \langle a_1, \ldots, a_n + 1 \rangle$.

On the other hand, the modular symbols of higher degree can be simply described by allowing $a_i = \infty$ for one or more i. Indeed, if we let $\overline{A} = A \cup \{\infty\}$ be the one–point compactification of the discrete set $A = \{1, 2, 3, \ldots\}$, then there is a unique continuous map

$$\pi : \coprod_{n=0}^{\infty} \overline{A}^n \to \bigcup_{d \geq 1} \mathcal{S}^d(V)$$

such that $\pi(a_1, \ldots, a_n) = \langle a_1, \ldots, a_n \rangle$ on A^n. This map is surjective, with finite fibers. Thus its domain provides a good approximate model for $\mathcal{S}(V)$.

In concrete terms, the image of $(a_1, \ldots, a_n)$ is obtained by replacing each occurrence of $a_i = \infty$ with $*$. For example,

$$\begin{aligned} \langle 3, 4, \infty, 5, 6, \infty, 7 \rangle &= \langle 3, 4 \rangle * \langle 5, 6 \rangle * \langle 7 \rangle; \\ \langle \infty \rangle &= \langle\rangle * \langle\rangle = \langle 1 \rangle * \langle 1 \rangle; \quad \text{and} \\ \langle 3, \infty, \infty, 5, \infty \rangle &= \langle 3 \rangle * \langle 1 \rangle * \langle 5 \rangle * \langle 1 \rangle. \end{aligned}$$

The degree of $\pi(a_1, \ldots, a_n)$ is always one more than the number of i such that $a_i = \infty$. These statements can all be verified using equation (2.3).

Note that the modular symbols $\langle a_1, \ldots, a_n \rangle$ with $a_i \leq M$ for all i do *not* form a compact set, but those with $n \leq M$ do.

Weil–Petersson geodesics. The modular symbols for $\mathrm{SL}_2(\mathbb{Z})$ arise naturally in the study of the moduli space $\mathcal{M}_{1,1}$; see the Appendix.

3 Teichmüller curves

In this section we review the action of $\mathrm{SL}_2(\mathbb{R})$ on the moduli space of holomorphic 1–forms (X, ω), and its role in constructing algebraic curves $V \looparrowright \mathcal{M}_g$. We then discuss the fundamental twist τ_A associated to a cylinder system $A = C(s)$. Finally we show Condition 1.5 can always be achieved by passing to a finite index subgroup of $\mathrm{Aff}^+(X, \omega)$. The precise statement is:

Theorem 3.1 *Suppose* $\mathrm{SL}(X, \omega)$ *is a lattice. Then there is a subgroup* Φ *of finite index in* $\mathrm{Aff}^+(X, \omega)$ *such that the natural map*

$$\Phi \to \pi_1(V) = \mathrm{PSL}_2(X, \omega)$$

is injective, and the stabilizer Φ^s *of each periodic slope is generated by a power of* τ_A, $A = C(s)$.

We note that $\tau_A(A_i) = A_i$ for each $A_i \in A$, and $\tau_A | H_1(X, \mathbb{R})$ is unipotent. Thus the same is true for every element in Φ^s.

Notation. The natural actions of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R}^2$, $\mathbb{P}^1(\mathbb{R})$, and $\overline{\mathbb{H}} = \mathbb{H} \cup \widehat{\mathbb{R}}$ are given, for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, by

$$\begin{aligned} g \cdot \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} && \text{on } \mathbb{R}^2, \\ g \cdot s &= (ds + c)/(bs + a) && \text{on } \mathbb{P}^1(\mathbb{R}), \text{ and} \\ g \cdot t &= (at + b)/(ct + d) && \text{on } \overline{\mathbb{H}}. \end{aligned}$$

The first two actions are compatible under the relation $s = y/x$ between a vector and its slope.

Complex geodesics. (Cf. [Mc1, §3].) Each form $(X, \omega) \in \Omega\mathcal{M}_g$ generates a holomorphic map

$$\widetilde{f} : \mathbb{H} \to \mathcal{M}_g,$$

characterized by the property that the *real* Teichmüller geodesic $p(s) = \widetilde{f}(ie^{2s})$ satisfies $p(0) = X$ and $p'(0) = [-\overline{\omega}/\omega]$. It is given explicitly by

$$\widetilde{f}(t) = \pi(a_t \cdot (X, \omega)) \quad \text{where} \quad a_t = \frac{1}{\sqrt{\operatorname{Im} t}} \begin{pmatrix} 1 & \operatorname{Re} t \\ 0 & \operatorname{Im} t \end{pmatrix}.$$

The map $\widetilde{f}$ descends to give a *complex geodesic*

$$f : V = \mathbb{H}/\Gamma \to \mathcal{M}_g,$$

where

$$\Gamma = \{g \in \mathrm{SL}_2(\mathbb{R}) \ : \ \widetilde{f}(g \cdot t) = \widetilde{f}(t) \ \forall t \in \mathbb{H}\}.$$

The map f is a generically injective immersion. Using the fact that rotations leave the fibers of π invariant, we can also write

$$V \cong \mathrm{SO}_2(\mathbb{R}) \backslash \mathrm{SL}_2(\mathbb{R}) / \mathrm{SL}(X, \omega).$$

One can readily verify that

$$\Gamma = R \cdot \mathrm{SL}(X, \omega) \cdot R, \quad \text{where} \quad R = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right); \tag{3.1}$$

put differently, if we let $g \in \mathrm{SL}(X, \omega)$ act on $\overline{\mathbb{H}}$ by

$$g \cdot t = \frac{at - b}{-ct + d}, \tag{3.2}$$

then the relation

$$s = -1/x$$

between slopes in $\mathbb{P}^1(\mathbb{R})$ and points $x \in \partial\mathbb{H}$ is compatible with the action of g, and we have $V \cong \mathbb{H}/\operatorname{SL}(X,\omega)$.

Teichmüller curves. Now suppose $\operatorname{SL}(X,\omega)$ is lattice in $\operatorname{SL}_2(\mathbb{R})$. Then V is a hyperbolic Riemann surface (or orbifold) of finite volume, and the map $f : V \to \mathcal{M}_g$ gives an algebraic, isometrically immersed *Teichmüller curve* in moduli space.

In this case, the relation $s = -1/x$ gives a natural bijection between the cusps of Γ and the periodic slopes for (X,ω).

Oriented cylinders. Let s be a periodic slope with associated cylinders

$$C(s) = \{A_1, \ldots, A_n\}.$$

It is often convenient to coherently orient the core curves of these cylinders, so they each determine a homology class $[A_i] \in H_1(X,\mathbb{R})$. We say these cylinders are given *parallel* orientations if

$$\arg \int_{A_i} \omega = \arg\langle [A_i], [\omega] \rangle$$

is independent of i. This means the oriented closed geodesics that sweep out these cylinders all run in the same direction in local charts where $\omega = dz$.

Fundamental twists. Recall from §1 that the slope s determines an integral vector $m(s) = (a_1, \ldots, a_n)$ proportional to $(\operatorname{mod}(A_i))$, and abyby fundamental twist $\tau_A = \prod \tau_i^{a_i}$ (see equations (1.1) and (1.2)). Choose parallel orientations for the cylinders (A_i). Then action of τ_A on $H_1(X,\mathbb{R})$ is given by the unipotent transformation

$$[\tau_A(D)] = [D] + \sum_{i=1}^{n} a_i \langle A_i, D \rangle [A_i]. \tag{3.3}$$

This map is actually independent of the choice of orientations.

Finite covers. In general, there may be affine automorphisms of (X,ω) that permute the cylinders (A_i) or act by fractional twists on some of them. Theorem 3.1 shows these extra maps can be eliminated by passing to a finite subgroup Φ. The proof will use:

Lemma 3.2 *Let Z be a finite volume, noncompact hyperbolic surface, let $N \geq 1$ an integer. Then there exists a finite covering map $p : Z' \to Z$ which is branched of order N over each cusp of Z.*

Proof. Passing to a double cover, if necessary, we can assume that Z has at least two cusps. Then the natural map $\pi_1(Z) \to H_1(Z, \mathbb{Z}/N)$ sends each peripheral loop around a cusp to an element of order N, so its kernel defines the desired covering space $Z' \to Z$. ∎

Proof of Theorem 3.1 . Since the orbifold V has at least one cusp, there is a finite cover $V_0 \to V$ with free fundamental group. Thus $\pi_1(V_0)$ lifts (under D^{-1}) to a subgroup Φ_0 of affine automorphisms of X. Each cusp of a of V_0 corresponds to a periodic slope s, stabilized by the cyclic subgroup $\langle \phi \rangle = \Phi_0^s$. Since the fundamental twist τ_A, $A = C(s)$, also stabilizes s, there is an $N(a) > 0$ such that $\phi^{N(a)} \in \langle \tau_A \rangle$. To complete the proof, let N be the least common multiple of all such $N(a)$, let $\widetilde{V} \to V_0$ be the finite covering provided by the preceding lemma, and let Φ be the subgroup of Φ_0 corresponding to $\pi_1(\widetilde{V}) \subset \pi_1(V_0)$. ∎

4 Intersection matrices

In this section we discuss the intersection matrix $I(\sigma)$ associated to a modular symbol, and prove:

Theorem 4.1 . *The functor $I : \mathcal{S}(V) \to \mathcal{L}(V)$ is continuous.*

As in the Introduction, we will assume the simplifying Condition 1.5 holds. The general case can be treated by passing to a finite cover, using Theorem 3.1; see the end of this section.

The vector space of a cusp. We begin with a formal definition of the vector space $\mathcal{L}(a) = \mathbb{R}^A$ associated to a cusp a of V.

Let $V = \mathbb{H}/\operatorname{SL}(X, \omega)$, where $\operatorname{SL}(X, \omega)$ acts on $\mathbb{H}$ by equation (3.2). Let $S \subset \mathbb{P}(\mathbb{R}^2)$ denote the set of periodic slopes for (X, ω). There is a natural bijection

$$\kappa(V) \cong S/\operatorname{SL}(X, \omega)$$

between the cusps of V and the orbits of periodic slopes, compatible with the map $x \mapsto s = -1/x$ between $\partial\mathbb{H}$ and $\mathbb{P}(\mathbb{R}^2)$. Let $S(a)$ denote the slopes associated to a given cusp a. We can then canonically associate to a the finite set

$$A = \left(\bigcup_{s \in S(a)} C(s) \right) \Big/ \operatorname{Aff}^+(X, \omega).$$

Condition 1.5 insures that the stabilizer of s in the affine group does not permute the elements of $C(s)$; consequently we have a natural bijection between A and $C(s)$ for each $s \in S(a)$.

With this definition, it is clear that the vector space $\mathcal{L}(a) = \mathbb{R}^A$ depends only on the cusp a. These vector spaces form the objects of the category whose morphisms $\mathcal{L}(V)$ are defined by equation (1.7).

The intersection matrix. Next we recapitulate the definition of the functor $I : \mathcal{S}(V) \to \mathcal{L}(V)$.

Let $\gamma = [x, y]$ be a modular symbol of degree one, in the notation of §2, joining a pair of cusps (a, b). Then γ determines a pair of periodic slopes $(s, t) = (-1/x, -1/y)$, a pair of cylinder systems $C(s) = \{A_1, \ldots, A_n\}$ and $C(t) = \{B_1, \ldots, B_m\}$, and an integral vector $m(s) = (a_i)$ proportional to $\mathrm{mod}(A_i)$. By the remarks above, if we write $(\mathcal{L}(a), \mathcal{L}(b)) = (\mathbb{R}^A, \mathbb{R}^B)$, then we can naturally identify A with $\{A_1, \ldots, A_n\}$ and B with $\{B_1, \ldots, B_m\}$. The intersection matrix is the map $I(\gamma) : \mathbb{R}^B \to \mathbb{R}^A$ with matrix entries

$$I(\gamma) = a_i \cdot i(A_i, B_j).$$

If we replace $[x, y]$ with $[gx, gy]$, $g = D\phi \in \mathrm{SL}(X, \omega)$, then the matrix remains the same, because $i(\phi(A_i), \phi(B_j)) = i(A_i, B_j)$ and $\mathrm{mod}(A_i) = \lambda \cdot \mathrm{mod}(\phi(A_i))$ for any affine automorphism ϕ.

Regarded as a matrix up to scale, $[I(\gamma)]$ naturally resides in $\mathcal{L}_{ab}(V)$. There is a unique functorial extension of this map from $\mathcal{S}^1(V)$ to $\mathcal{S}(V)$, given by equation (1.9).

Biconnected matrices. We now turn to the proof of Theorem 4.1.

Given an $n \times m$ real matrix $M_{ij} \geq 0$, one can form a bipartite graph $G(M)$ whose vertices are the rows and columns of M, and where row i is connected to column j if and only if $M_{ij} > 0$. We say M is *biconnected* if $G(M)$ is connected. By an easy argument, one can show:

The product of two biconnected matrices is biconnected.

It is also easy to see that $I(\gamma)$ is biconnected for any geodesic γ, using the fact that $(\bigcup A_i) \cup (\bigcup B_j)$ is connected. Since this property is preserved under products, we have:

Proposition 4.2 *For all $\sigma \in \mathcal{S}(V)$, the intersection matrix $I(\sigma)$ is a biconnected matrix of non–negative integers. In particular, $I(\sigma)$ is nonzero.*

Continuity. The main case of continuity to be considered is the following.

Proposition 4.3 *Suppose we have a sequence of geodesics such that $\gamma_n \to \sigma \in \mathcal{S}(V)$. Then $I(\gamma_n) \to I(\sigma)$ in $\mathcal{L}(V)$.*

Proof. Let us begin with a basic example . Let $x, y, z \in \partial\mathbb{H}$ be three distinct cusps of $\mathrm{SL}(X,\omega)$, let $p \in \mathrm{SL}(X,\omega)$ be a parabolic element generating the stabilizer of z, and let

$$\gamma_n = [x, p^n y] \to [x, z] * [z, y] = \delta_1 * \delta_2 = \sigma.$$

Let (s,t,u) be the slopes $(-1/x, -1/y, -1/z)$, and denote the corresponding families of cylinders by $C(s) = A = (A_i)$, $C(t) = B = (B_j)$, and $C(u) = C = (C_k)$. Let (a_i), (b_j) and (c_k) denote the primitive integral vectors proportional to $(\mathrm{mod}(A_i)), (\mathrm{mod}(B_j))$ and $(\mathrm{mod}(C_j))$. Then

$$\begin{aligned} I(\delta_1)_{ik} &= a_i \cdot i(A_i, C_k) \quad \text{and} \\ I(\delta_2)_{kj} &= c_k \cdot i(C_k, B_j). \end{aligned}$$

Let $\phi_C \in \mathrm{Aff}^+(X,\omega)^u$ be an affine automorphism mapping to p under the natural surjection $\mathrm{Aff}^+(X,\omega) \to \mathrm{SL}(X,\omega)$. We then have

$$I(\gamma_n)_{ij} = a_i \cdot i(A_i, \phi_C^n(B_j)). \tag{4.1}$$

First suppose $\phi_C = \tau_C$. With suitable orientations on A_i, B_j, the geometric and homological intersection numbers agree, and by equation (3.3), we have

$$[\tau_C^n(B_j)] = [B_j] + n \sum_k c_k \langle C_k, B_j \rangle [C_k]. \tag{4.2}$$

As $n \to \infty$, the second term dominates, and therefore we have

$$[\langle A_i, \tau_C^n(B_j) \rangle] \to \left[\sum_k c_k \, \langle C_k, B_j \rangle \langle A_i, C_k \rangle \right] \tag{4.3}$$

in the space $\mathbb{P}\,\mathrm{Hom}(\mathbb{R}^B, \mathbb{R}^A)$ of matrices up to scale. The combination of equations (4.1) and (4.3) then implies that

$$\begin{aligned} \lim [I(\gamma_n)] &= \left[\sum_k a_i \, i(A_i, C_k) \cdot c_k \, i(C_k, B_j) \right] \\ &= [I(\delta_1) \cdot I(\delta_2)] = [I(\sigma)], \end{aligned}$$

as desired.

For the general case, using the fact that τ_C generates a subgroup of finite index in $\mathrm{Aff}^+(X,\omega)^u$, we can write

$$\phi_C^n = \psi \circ \tau_C^m$$

where $m \to \infty$ as $n \to \infty$, and ψ ranges in a finite subset of $\mathrm{Aff}^+(X,\omega)^u$. By Condition 1.5, $\psi(C_k) = C_k$ for all k, so ψ preserves the dominant term in equation (4.2), and the same conclusion holds.

This establishes the basic mechanism of continuity; the general case, described by equation (2.3), can be treated similarly. ∎

Proof of Theorem 4.1. Recall that the composition in $\mathcal{L}(V)$ is continuous near any pair of matrices whose product is nonzero. In particular, Proposition 4.2 insures that composition is continuous on the image of I. Since every modular symbol σ of degree $d \geq 1$ is uniquely the product of d modular symbols of degree 1, continuity of I on the whole of $\mathcal{S}(V)$ follows from Proposition 4.3, using equation (2.1). ∎

General Teichmüller curves. The case of a general Teichmüller curve V — where Condition 1.5 may not hold — can be treated similarly by appealing to Theorem 3.1.

In the notation of that result, one can simply replace $\mathrm{Aff}^+(X,\omega)$ with Φ, $\mathrm{SL}(X,\omega)$ with $D\Phi$, and V with the corresponding finite covering space $\widetilde{V} \to V$. Since the properties insured by Condition 1.5 hold after making this adjustment, the constructions above go through to yield a continuous functor

$$I : \mathcal{S}(\widetilde{V}) \to \mathcal{L}(\widetilde{V}),$$

and the proofs of Theorems 4.1, 1.7 and 1.8 go through as well, with $\widetilde{V}$ replacing V. Intuitively, one can simply imagine that (X,ω) acquires some additional rigidity, which reduces its affine automorphism group to Φ.

5 Hodge theory and equidistribution

In this section we begin the study of the space of all intersection matrices:

$$\mathcal{I}(V) = \{[I(\sigma)] \,:\, \sigma \in \mathcal{S}(V) \quad \text{and} \quad \deg(\sigma) > 0\}.$$

Using results on currents and Hodge theory from [Mc5], we will describe the closure of $\mathcal{I}(V)$ in $\mathcal{L}(V)$ and prove Theorem 1.7.

Note that, for convenience, we have excluded the identity matrices $I(\sigma)$ which arise from the degree zero elements of $\mathcal{S}(V)$. This simplifies later arguments, since we can use the fact that $\mathcal{S}^1(V)$ is dense in $\mathcal{S}(V) - \mathcal{S}^0(V)$.

To state the main result, let $\mathcal{I}_{ab}(V) = \mathcal{I}(V) \cap \mathcal{L}_{ab}(V)$. Recall from equation (1.10) that $\mathcal{R}_{ab}(V)$ consists of the single decoupled matrix

$$[R(a,b)] = [h(A_i)c(B_j)] \in \mathcal{L}_{ab}(V),$$

and that $\mathcal{R}(V) = \bigcup \mathcal{R}_{ab}(V)$. In this section we will show:

Theorem 5.1 *The closure of $\mathcal{I}_{ab}(V)$ in $\mathcal{L}_{ab}(V)$ is given by*

$$\mathcal{I}_{ab}(V) \cup \mathcal{R}_{ab}(V).$$

Moreover, if $\sigma_n \to \infty$ in $\mathcal{S}_{ab}(V)$, then $[I(\sigma_n)] \to [R(a,b)]$.

We also note:

Proposition 5.2 *The sets $\mathcal{I}(V)$ and $\mathcal{R}(V)$ are disjoint, provided the trace field K of $\mathrm{SL}(X,\omega)$ is not $\mathbb{Q}$.*

Proof. The matrices in $\mathcal{I}(V)$ are all rational (up to scale), while the matrices in $\mathcal{R}(V)$ are irrational under the assumption above. Indeed, in §6 we will see there is an integral matrix P such that $P \cdot h(A_i) = \lambda_1 h(A_i)$ and $K = \mathbb{Q}(\lambda_1)$. ∎

Currents and equidistribution. To set the stage for the proof, recall that any holomorphic 1–form (X,ω) determines a foliation $\mathcal{F}(\omega)$ of X, given by horizontal lines in local coordinates where $\omega = dz$. Write $\omega = \alpha + i\beta$ as a linear combination of real harmonic forms. Then the space of closed, positive, 1–dimensional currents carried by $\mathcal{F}(\omega)$ is defined by

$$P(\omega) = \{\text{currents } \xi : d\xi = 0, \xi \wedge \beta = 0 \quad \text{and} \quad \alpha \wedge \xi \geq 0\}. \tag{5.1}$$

(The final condition means $\int f\alpha \wedge \xi \geq 0$ whenever $f \geq 0$.) The space $P(\omega)$ is a convex cone in the natural topology on currents, and it contains the ray $\mathbb{R}_+ \cdot \beta$. The narrower the cone is, the closer the foliation $\mathcal{F}(\omega)$ is to being uniquely ergodic.

This narrowness is estimated effectively in Theorem 1.2 of [Mc5, §3], which states:

Theorem 5.3 *Suppose X lies in a compact subset $K \subset \mathcal{M}_g$, and the Teichmüller geodesic ray generated by (X, ω) spends at least time T in K. Then the closed, positive currents carried by $\mathcal{F}(\omega)$ determine a convex cone*

$$[P(\omega)] \subset H^1(X, \mathbb{R})$$

which meets the unit sphere in a set of diameter $O(e^{-\lambda(K)T})$.

Here the unit sphere and diameter are defined using the Hodge norm on $H^1(X, \mathbb{R})$, and $\lambda(K) > 0$ depends only on K.

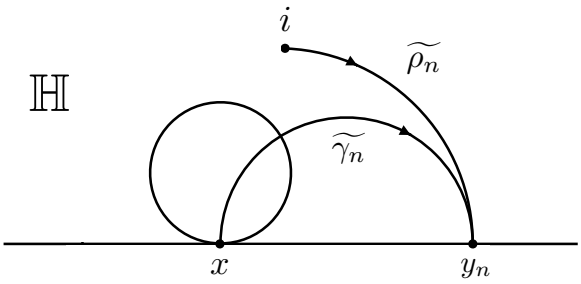

Figure 3. The geodesic $\widetilde{\rho_n}$ is close to $\widetilde{\gamma_n}$ outside a horoball neighborhood of the cusp x.

Proof of Theorem 5.1. Let $\sigma_n \in \mathcal{S}_{ab}(V)$ be a sequence of modular symbols tending to infinity in $\mathcal{S}(V)$ (leaving every compact set). We begin by showing

$$[I(\sigma_n)] \to [R(a, b)]. \tag{5.2}$$

The rest of the Theorem then easily follows.

Since V has finite volume, by removing standard horoball neighborhoods of its cusps we obtain a compact set $K = V - V_{\text{cusp}}$. Then the hyperbolic length of σ_n satisfies

$$L(\sigma_n \cap K) \to \infty \tag{5.3}$$

as $n \to \infty$ (see statement (2.2)). Passing to a subsequence, we can assume that $\lim[I(\sigma_n)]$ exists in $\mathcal{L}_{ab}(V)$. To prove equation (5.2), it suffices to show this limit is $[R(a, b)]$.

Since I is continuous and modular symbols of degree 1 are dense in $\mathcal{S}(V) - \mathcal{S}_0(V)$, we can reduce to the case where $\sigma_n = \gamma_n$ is a sequence of geodesics from a to b. We can then write

$$\gamma_n = [x, y_n],$$

and pass to a subsequence such that $y_n \to y \in \widehat{\mathbb{R}}$. Since $[x, y_n] \sim [x, gy_n]$ for all $g \in \Gamma^x$, we can also arrange that $y \neq x$. Let $s = -1/x$ and $t_n = -1/y_n$ be the corresponding slopes.

Let us normalize so that $\int_X |\omega| = 1$, and so that $\lim t_n = 0$. Then $s \neq 0$, since $y \neq x$. Choose $\theta_n \to 0$ such that $\tan(\theta_n) = t_n$, and let $\omega_n = e^{-i\theta_n}\omega$. Writing $\omega = \alpha + i\beta$ and $\omega_n = \alpha_n + i\beta_n$, we have $\|\beta\| = \|\beta_n\| = 1$ in the Hodge norm on $H^1(X, \mathbb{R})$.

Recall that the complex geodesic $\widetilde{f} : \mathbb{H} \to \mathcal{M}_g$ covering $f : V \to \mathcal{M}_g$ is normalized so that $\widetilde{f}(i) = [X]$ and satisfies $\widetilde{f} = f \circ \pi$, where $\pi : \mathbb{H} \to V$ is the universal covering map. Let $\widetilde{\gamma_n}$ and $\widetilde{\rho_n}$ in $\mathbb{H}$ be the hyperbolic geodesics joining x to y_n and i to y_n, respectively. Then $\gamma_n = \pi(\widetilde{\gamma_n})$, and $\rho_n = \pi(\widetilde{\rho_n}) \subset V \looparrowright \mathcal{M}_g$ is the Teichmüller geodesic ray generated by $\mathcal{F}(\omega_n)$. Since $y_n \to y \neq x$, the geodesic $\widetilde{\rho_n}$ stays close to $\widetilde{\gamma_n}$ outside a neighborhood of the cusp x (see Figure 3). Thus equation (5.3) implies that

$$L(\rho_n \cap K) \to \infty. \tag{5.4}$$

Let $A = C(s) = \{A_1, \ldots, A_n\}$ and $B = C(t_1) = \{B_1, \ldots, B_m\}$. Let $B_i(n)$ denote the image of B_i under the natural bijection between B and $C(t_n)$; then $C(t_n) = \{B_1(n), \ldots, B_m(n)\}$. Since the slope s of the cylinders A_i is not zero, their circumferences in the metric $|\omega|$ and their vertical periods are related by

$$c(A_i) = \sqrt{1 + s^{-2}} \int_{A_i} \beta. \tag{5.5}$$

Here the notation means we integrate around an oriented closed geodesic in A_i.

Since any two of the cylinder systems $(B_j(n))$ are related by an affine automorphism of (X, ω), with suitable orientations we also have

$$c(B_j(n)) = \int_{B_j(n)} \alpha_n = \lambda_n \, c(B_j), \tag{5.6}$$

where $\lambda_n > 0$ does not depend on j.

The cohomology class $[B_j(n)]$ is represented by a closed leaf of the foliation $\mathcal{F}(\omega_n)$, which in turn represents a current in $P(\omega_n)$. On the other hand, by equation (5.4) the amount of time the Teichmüller geodesic ray ρ_n generated by ω_n spends in the compact set $K \subset V \looparrowright \mathcal{M}_g$ tends to infinity. Thus by Theorem 5.3, the Hodge norm of the difference between $[B_j(n)]/\|B_j(n)\|$ and β_n tends to zero. Since $\beta_n \to \beta$, it follows that for all j we have

$$\frac{[B_j(n)]}{\|B_j(n)\|} \to [\beta] \in H^1(X, \mathbb{R}).$$

Since $\int \alpha \wedge \beta = \|\beta\|^2 = 1$, and $\alpha_n \to \alpha$, it follows from equation (5.6) that

$$[B_j(n)] \sim c(B_j(n)) \cdot [\beta] = \lambda_n c(B_j) \cdot [\beta].$$

Combined with (5.5), this gives

$$i(A_i, B_j(n)) = \langle A_i, B_j(n) \rangle \sim \lambda_n c(B_j) \int_{A_i} \beta = \lambda_n (1 + s^{-2})^{-1}\, c(A_i)\, c(B_j),$$

and hence

$$\begin{aligned} \lim[I(\gamma_n)] &= \lim[\mathrm{mod}(A_i)\, i(A_i, B_j(n))] = [\mathrm{mod}(A_i)\, c(A_i)\, c(B_j)] \\ &= [h(A_i)\, c(B_j)] = [R(a,b)] \end{aligned}$$

in $\mathbb{P}\,\mathrm{Hom}(\mathbb{R}^B, \mathbb{R}^A)$. This completes the proof that $[I(\sigma_n)] \to [R(a,b)]$ whenever $\sigma_n \to \infty$.

It follows that $\mathcal{I}_{ab}(V) \cup \mathcal{R}_{ab}(V)$ is closed, and hence compact, in the space of matrices up to scale. This union coincides with the closure of $\mathcal{I}_{ab}(V)$ because there exists a sequence $\sigma_n \to \infty$ in $\mathcal{S}_{ab}(V)$. ∎

Proof of Theorem 1.7. Since $\mathcal{S}(V) = \mathcal{S}^0(V) \cup \langle \mathcal{S}^1(V) \rangle$, and $T = I(\mathcal{S}^1(V))$, we have $\langle T \rangle = \mathcal{I}(V)$, so $\overline{T} = \langle T \rangle \cup \mathcal{R}(V)$ by Theorem 5.1. ∎

6 The algebra of a modular symbol

In this section we study the algebraic structure of the family of morphisms $\mathcal{I}(V)$ in more detail.

Let δ be a modular symbol of degree 1 joining a pair of cusps (a, b) of V. We begin by discussing general properties of the intersection matrix $I(\delta) \in \mathrm{Hom}(\mathbb{R}^B, \mathbb{R}^A)$. We then associate to δ an algebra $\mathcal{A}$ of endomorphisms of $\mathbb{R}^A \oplus \mathbb{R}^B$, containing a group G and a distinguished element Q, and show:

Theorem 6.1 *For any $g \in G$, each nonzero block of the product*

$$Qg = \begin{pmatrix} I_{aa} & I_{ba} \\ I_{ab} & I_{bb} \end{pmatrix} \tag{6.1}$$

represents the intersection matrix $[I(\gamma)]$ of a degree 1 modular symbol.

Here $[I_{aa}] \in \mathcal{I}_{aa}(V)$, $[I_{ab}] \in \mathcal{I}_{ab}(V)$, etc. As an application, we will show:

Corollary 6.2 *The space $\mathcal{I}_{aa}(V)$ contains $[P + k^2P^2]$ for all integers k.*

Here $P = I(\delta)I(\delta^*)$, where δ^* is the reverse of δ. This result will be used in the next section to ratify the topological complexity of $\mathcal{I}_{ab}(V)$.

Properties of the intersection matrix. Let $\delta = [x, y] \in \mathcal{S}_{ab}(V)$ be a geodesic joining the pair of cusps a and b of V, and let $(s, t) = (-1/x, -1/y)$ be associated periodic slopes for (X, ω). As usual we let

$$A = C(s) = \{A_1, \ldots, A_n\} \quad \text{and} \quad B = C(t) = \{B_1, \ldots, B_m\}.$$

We also let $J_{ij} = i(A_i, B_j)$, and let Δ_A and Δ_B be the diagonal matrices whose entries are given by

$$m(s) = (a_1, \ldots, a_n) \quad \text{and} \quad m(t) = (b_1, \ldots, b_m)$$

respectively (see equation (1.1)). We then have

$$I(\delta) = \Delta_A J \quad \text{and} \quad I(\delta^*) = \Delta_B J^t,$$

where $\delta^* = [y, x]$. Let

$$P = I(\delta)I(\delta^*) = \Delta_A J \Delta_B J^t \in \operatorname{Hom}(\mathbb{R}^A, \mathbb{R}^A).$$

Here are some basic properties these integral matrices enjoy.

Proposition 6.3 *For any geodesic δ joining a pair of cusps (a, b),*

1. *The matrices P, J and $I(\delta)$ have the same rank.*
2. *If $I(\delta)$ has rank one, then*
$$[I(\delta)] = [h(A_i)c(B_j)] = [R(a, b)] \in \mathbb{P}\operatorname{Hom}(\mathbb{R}^B, \mathbb{R}^A).$$
3. *The matrix P is a diagonalizable over $\mathbb{R}$ and Perron–Frobenius. In particular its largest eigenvalue $\lambda_1 > 0$ is simple.*
4. *The heights of the cylinders (A_i) give a Perron–Frobenius eigenvector for P; that is,*
$$Ph(A_i) = \lambda_1 h(A_i). \tag{6.2}$$
5. *The trace field K of $\mathrm{SL}(X, \omega)$ is given by $\mathbb{Q}(\lambda_1)$. In particular, the rank of $I(\delta)$ is bounded below by the degree of the trace field.*

Proof. Note that P is conjugate over $\mathbb{R}$ to the symmetric matrix LL^t, where $L = \Delta_A^{1/2} J \Delta_B^{1/2}$. It follows that P is diagonalizable over $\mathbb{R}$, and that P and L have the same rank, which agrees with the ranks of J and $I(\delta)$. The fact that $(\bigcup A_i) \cup (\bigcup B_j)$ is connected implies that every entry of P^k is positive for $k \gg 0$, and hence P is a Perron–Frobenius matrix.

Let $\theta \in [0, \pi]$ be the angle between the slopes s and t. Since the cylinders B_j cut A_i into a family of parallelograms, their heights and circumferences are related by

$$c(A_i) \sin \theta = \sum i(A_i, B_j) h(B_j) = J h(B_j).$$

The same identity holds with the roles of A and B reversed. Coupled with the fact that $\Delta_A c(A_i)$ is a positive multiple of $h(A_i)$, it follows that the vector $h(A_i)$ lies in the image of $\Delta_A J$, and $c(B_j)$ lies in the image of its transpose. Thus $[\Delta_A J] = [h_i(A) c(B_j)]$ when J has rank one. Similar reasoning gives equation (6.2).

The final statement regarding trace fields follows from [Mc2, Cor 4.3]. It is a combination of two facts: first, the product $\psi = \tau_A^{-1} \tau_B$ gives an affine pseudo–Anosov automorphism of (X, ω) with $\mathbb{Q}(\operatorname{tr} D\psi) = \mathbb{Q}(\lambda_1)$; and second, $\mathbb{Q}(D\psi)$ coincides with the trace field of $\mathrm{SL}(X, \omega)$ whenever the affine group of (X, ω) contains a pseudo–Anosov element ψ [KS, Theorem 28]. ■

The algebra associated to a modular symbol. Using the matrices above, we can now naturally associate to the modular symbol δ the algebra

$$\mathcal{A} = \mathbb{R}[Q, \pi_A, \pi_B] \subset \mathrm{End}(\mathbb{R}^A \oplus \mathbb{R}^B), \tag{6.3}$$

where π_A and π_B denote projection onto the first and second factors of $\mathbb{R}^A \oplus \mathbb{R}^B$ respectively, and

$$Q = \begin{pmatrix} 0 & I(\delta) \\ I(\delta^*) & 0 \end{pmatrix} = \begin{pmatrix} 0 & \Delta_A J \\ \Delta_B J^t & 0 \end{pmatrix}. \tag{6.4}$$

Let

$$G = \langle T_A, T_B \rangle \subset \mathcal{A}$$

denote the multiplicative group generated by the unipotent matrices

$$T_A = I + \pi_A Q \quad \text{and} \quad T_B = I + \pi_B Q. \tag{6.5}$$

The main result of this section produces, from $\mathcal{A}$, a large supply of intersection matrices.

Proof of Theorem 6.1. Let

$$\Delta = \begin{pmatrix} \Delta_A & 0 \\ 0 & \Delta_B \end{pmatrix}, \quad S = \begin{pmatrix} 0 & J \\ -J^t & 0 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}.$$

Note that conjugation of $\mathcal{A}$ by R sends Q to

$$Q' = RQR = \Delta S = \begin{pmatrix} 0 & \Delta_A J \\ -\Delta_B J^t & 0 \end{pmatrix},$$

leaves G invariant, and keeps the blocks of the matrices appearing in (6.1) the same up to sign. Thus it suffices to prove the Theorem with Q replaced by Q' in the statement of the result and in the definition (6.5) of the generators of G.

The advantage is that, with a suitable choice of parallel orientations, J_{ij} represents the homological intersection number $\langle A_i, B_j \rangle$. Therefore, under the natural map

$$\pi : \mathbb{R}^A \oplus \mathbb{R}^B \to H_1(X, \mathbb{R})$$

these orientations determine, we have

$$\langle \pi(C_1), \pi(C_2) \rangle = C_1^t S C_2$$

for any pair of weighted cylinders systems $C_i \in \mathbb{R}^A \oplus \mathbb{R}^B$. In other words, S is the pullback of the intersection form on $H_1(X, \mathbb{R})$. In addition, the unipotent transformations T_A and T_B are compatible with the fundamental twists $\tau_A, \tau_B \in \mathrm{Aff}^+(X, \omega)$, in the sense that

$$\pi(T_A C) = \tau_A \pi(C) \quad \text{and} \quad \pi(T_B C) = \tau_B \pi(C)$$

for all C (see equation (3.3)). It follows that, for any $g \in G$, there exists a $\phi \in \mathrm{Aff}^+(X, \omega)$ such that $\pi(gC) = \phi(\pi(C))$ for all $C \in \mathbb{R}^A \oplus \mathbb{R}^B$, and hence

$$Q'g = \begin{pmatrix} a_i \langle A_i, \phi A_i \rangle & a_i \langle A_i, \phi B_j \rangle \\ b_j \langle B_j, \phi A_i \rangle & b_j \langle B_j, \phi B_j \rangle \end{pmatrix}.$$

Since (A_i) and (B_j) are given parallel orientations, the orientations of their images under ϕ are also parallel, and hence their homological and geometric intersection numbers agree, up to sign. It follows that each matrix block gives an intersection matrix of the form $I(\gamma)$, up to scale, provided it is nonzero. ∎

Proof of Corollary 6.2. For brevity let $U = I(\delta)$ and $U' = I(\delta^*)$, so $P = UU'$. We then have, for any integer k,

$$\begin{aligned} QT_B^k T_A^k &= \begin{pmatrix} 0 & U \\ U' & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ kU' & I \end{pmatrix} \begin{pmatrix} I & kU \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} kUU' & U + k^2 UU'U \\ U' & kU'U \end{pmatrix}. \end{aligned}$$

Applying the preceding result to the upper right–hand corner, we find that

$$[U + k^2 UU'U] \in \mathcal{I}_{ab}(V).$$

Multiplying by $U' \in \mathcal{I}_{ba}(V)$ on the right gives $[P + k^2 P^2] \in \mathcal{I}_{aa}(V)$. ∎

7 Cylinders and closed geodesics

Let us define the *rank* of a pair of cusps (a, b) by

$$\rho(a, b) = \max\{\operatorname{rank} I(\gamma) \ : \ \gamma \text{ joins } a \text{ to } b\},$$

and the rank of a single cusp by

$$\rho(a) = \max_b \rho(a, b). \tag{7.1}$$

By Proposition 6.3, we have

$$\rho(a, b) \geq \deg(K/\mathbb{Q})$$

where K is the trace field of $\mathrm{SL}(X, \omega)$. In particular, if V has a cusp of rank one, then $K = \mathbb{Q}$.

Let $A = C(s) = \{A_1, \ldots, A_n\}$ be the cylinder system attached to a periodic slope s. In this section we study the space

$$\Pi(A) = \overline{[i(A_i, C)]} \subset \mathbb{PR}^A,$$

where C ranges over all closed geodesics with slope different from s. We will show:

Theorem 7.1 *Assume* $\mathrm{SL}(X, \omega)$ *is lattice. Then either:*

1. $\Pi(A)$ *is homeomorphic to* $\omega^\omega + 1$, $D^\infty \Pi(A) = \{[c(A_i)]\}$, *and the rest of* $\Pi(A)$ *lies in* $\mathbb{PQ}^A$*; or*

2. $\Pi(A)$ *consists of the single point* $[c(A_i)]$, *and the corresponding cusp of* V *has rank one.*

As we will see, Theorem 1.8 follows easily. The more precise statement above will be used in §8 to deduce our main results on limit measures.

Derived sets. We will prove Theorem 7.1 under Condition 1.5; the general case can be handled using Theorem 3.1, since passing to a finite index subgroup of $\mathrm{Aff}^+(X, \omega)$ does not change $\Pi(A)$.

The idea of the proof is to propagate the topology of $\mathcal{S}(V) \cong \omega^\omega$ first to the space of matrices $\mathcal{I}(V)$, and then to the space of matrix columns $\Pi(A)$.

We begin by recalling a basic fact about derived sets: if $f : Y \to Z$ is a continuous surjective map between compact sets, then

$$D^i(Z) \subset f(D^i(Y)) \tag{7.2}$$

for $i = 1, 2, 3 \ldots, \infty$. In other words, topological complexity can only decrease under proper continuous maps. This yields:

Lemma 7.2 *For any pair of cusps* (a, b), *we have*

$$D^\infty(\overline{\mathcal{I}_{ab}(V)}) \subset \mathcal{R}_{ab}(V).$$

Proof. Let Y be the one–point compactification of $\bigcup_{d=1}^\infty \mathcal{S}_{ab}^d(V) \cong \omega^\omega$, with $D^\infty(Y) = \{\infty\}$ the compactifying point, and let $Z = \overline{\mathcal{I}_{ab}(V)}$. Then Y and Z are compact, we have

$$Z = \mathcal{I}_{ab}(V) \cup \mathcal{R}_{ab}(V), \tag{7.3}$$

and by Theorem 5.1, if we extend the definition of the intersection matrix by setting $I(\infty) = R_{ab}$, we obtain a continuous map $I : Y \to Z$. The Lemma then follows from equation (7.2) with $i = \infty$. ∎

The orbit of a geodesic. To study the topology of $\Pi(A)$ it is useful to consider, for a given closed geodesic $C \subset X$, the locus

$$\Pi(A, C) = \overline{\{[a_i \cdot i(A_i, \phi \cdot C)]\}} \subset \mathbb{PR}^A,$$

where ϕ ranges over all elements of $\mathrm{Aff}^+(X, \omega)$ such that these intersection numbers are nonzero, and $m(s) = (a_1, \ldots, a_n)$.

Let t be the slope of C, and let $B = C(t) = (B_1, \ldots, B_m)$ be the corresponding cylinder system. We then have $C \subset B_k$ for a unique index k. Let (a, b) be the cusps of V corresponding to the slopes (s, t), and define

$$\pi_k : \mathbb{P}\,\mathrm{Hom}(\mathbb{R}^B, \mathbb{R}^A) \to \mathbb{PR}^A$$

by $\pi_k(M_{ij}) = [M_{ik}]$. It is then easy to see that

$$\Pi(A, C) = \pi_k(\overline{\mathcal{I}_{ab}(V)}). \tag{7.4}$$

Indeed, $\mathcal{S}^1(V)$ is dense in $\mathcal{S}(V)$, and for each γ connecting a to b we have

$$\pi_k(I(\gamma)) = [a_i \cdot i(A_i, \phi \cdot B_k)]$$

for some affine automorphism ϕ.

Lemma 7.3 *We have $D^\infty\,\Pi(A, C) \subset \{[h(A_i)]\}$, and every other point in $\Pi(A, C)$ lies in $\mathbb{PQ}^A$.*

Proof. Since $\pi_k(R_{ab}) = [h(A_i)]$, the rest of $\Pi(A, C)$ is given by $\pi_k(\mathcal{I}_{ab}(V)) \subset \mathbb{PQ}^A$ by equation (7.3). The calculation of $D^\infty\,\Pi(A, B)$ follows from equation (7.2) and Lemma 7.2. ∎

To produce some topological complexity inside $\Pi(A, C)$, we will use:

Lemma 7.4 *Let $S \subset G$ be a semigroup in a metrizable topological group G. Suppose the derived set DS is nonempty. Then $D^i S \neq \emptyset$ for all $i > 0$.*

Proof. Suppose $y \in DS$. Then S contains a sequence of distinct elements x_k such that $x_k \to y$. Since left multiplication on G is a homeomorphism, and $x_k S \subset S$, it follows that $x_k y \in DS$ for all k and hence $y^2 \in D^2 S$. By similarly reasoning, $y^i \in D^i S$, and hence $D^i S \neq \emptyset$ for all $k > 0$. ∎

Theorem 7.5 *If $\Pi(A, C)$ consists of more than one point, then it is homeomorphic to $\omega^\omega + 1$.*

Proof. Suppose $\Pi(A, C)$ contains a vector $[v] \neq [h(A_i)]$. Let us begin by noting that $\mathcal{I}_{aa}(V)$ is a semigroup, and by (7.4) we have

$$\mathcal{I}_{aa}(V) \cdot [v] \subset \Pi(A, C). \tag{7.5}$$

We will use a suitable semigroup $S \subset \mathcal{I}_{aa}(V)$ to produce a copy of ω^ω in $\Pi(A, C)$.

By equation (7.4), we have $[v] = \pi_k(I(\sigma))$ for some modular symbol σ. Since I is continuous and modular symbols of degree one are dense, we may assume that σ is a single geodesic δ running from a to b, and that $v = I(\delta)_{ik} \geq 0$, $v \neq 0$.

As in §6, let

$$P = I(\delta)I(\delta^*) \in \mathcal{I}_{aa}(V).$$

By Corollary 6.2, the semigroup $S = \mathcal{I}_{aa}(V)$ contains $[P^2]$ and $[P_k] = [P + k^2P^2]$ for all $k \gg 0$.

Since v is in the image of $I(\delta)$, it is also in the image of P. Thus we can write $v = e_1 + \cdots + e_s$ as a sum of distinct eigenvectors for P, satisfying $Pe_i = \lambda_i e_i \neq 0$. Since P is a Perron–Frobenius operator, and v is a non–negative vector, we may assume that λ_1 is the largest eigenvalue for P, and hence $[e_1] = [h(A_i)] \neq [v]$. Thus $s \geq 2$ and the other eigenvalues satisfy $\lambda_i \neq \lambda_1$ by the Perron–Frobenius theorem.

Let $\mathbb{R}^A = V_0 \oplus V_1$, where $V_0 \cong \mathbb{R}^s$ is the span of $e_1, \ldots, e_s$, and V_1 is the unique P–invariant complement to V_0. Note that $v = v_0 \in V_0$ with respect to this splitting.

Let $G_0 \cong (\mathbb{R}^*)^{s-1}$ be the subgroup of $\mathrm{PGL}(V_0)$ consisting of invertible matrices which are diagonal for the basis $(e_1, \ldots, e_s)$. Let $S_0 \subset G_0$ be the semigroup obtained by intersecting $S|V_0$ with G_0. Since $\lambda_1 \neq \lambda_2$, we have a convergent sequence of distinct elements $[P_k|V_0] \to [P^2|V_0]$ in S_0. Thus S_0 contains a copy of ω^ω by Lemma 7.4.

Note that the map $G_0 \to \mathbb{PR}^A$ given by

$$g \mapsto [g \cdot v_0] \in \mathbb{P}V_0 \subset \mathbb{PR}^A$$

is a homeomorphism to its image. It follows that

$$[S_0 \cdot v_0] \subset \mathbb{P}V_0 \subset \mathbb{PR}^A$$

contains a copy of ω^ω. But we have $[S_0 \cdot v_0] = [S \cdot v] \subset \Pi(A, C)$ by equation (7.5), so $\Pi(A, C)$ contains a copy of ω^ω as well.

Since $\Pi(A, C)$ is compact, this implies that $D^\infty \Pi(A, C)$ is nonempty, and hence we have $\Pi(A, C) \cong \omega^\omega + 1$ by Lemma 7.3. ∎

Proof of Theorem 7.1. Let Δ_A be the diagonal matrix with entries $m(s) = (a_i)$. Since the set of maximal cylinders in $(X, |\omega|)$ falls into finitely

many orbits under the action of $\mathrm{Aff}^+(X,\omega)$, we can choose closed geodesics $C_1,\dots,C_s$ such that

$$\Pi(A) = \Delta_A^{-1}\bigcup_1^s \Pi(A, C_i).$$

The desired statements then follow by applying Lemma 7.3 and Theorem 7.5 to each term on the right. Note that if $\Pi(A)$ is a single point, then the columns of every matrix $I(\gamma) \in \mathcal{I}_{ab}(V)$ are all proportional to a single vector, for every b; hence $I(\gamma)$ has rank one, and therefore a has rank one. ∎

Proof of Theorem 1.8. We have $\overline{T}_{ab} = \overline{\mathcal{I}_{ab}(V)}$, and when (a,b) has rank one $\mathcal{I}_{ab}(V) = \mathcal{R}_{ab}(V)$ is a single point. Now suppose (a,b) has rank two or more. By Theorem 7.5 we can find k and C such that

$$\Pi(A, C) = \pi_k(\overline{\mathcal{I}_{ab}(V)}) \cong \omega^\omega + 1.$$

Since $\overline{\mathcal{I}_{ab}(V)}$ is compact, $D^\infty\overline{\mathcal{I}_{ab}(V)}$ is nonempty, and hence equal to the singleton $\mathcal{R}_{ab}(V)$ by Lemma 7.2. In particular, the closure of $\mathcal{I}_{ab}(V)$ is homeomorphic to $\omega^\omega + 1$. ∎

Remark. The related concept of homological dimension – the dimension of the span in $H^1(X,\mathbb{R})$ of the cylinders at a given slope s – is considered in [Fo, Def. 1.4].

8 Limiting measures and currents

In this section we study the space of *currents*

$$Z(\omega) = \{\lim C_n/L(C_n) \,:\, L(C_n) \to \infty \quad \text{and} \quad \theta(C_n) \to 0\}$$

that arise as limits of closed, oriented geodesics, and prove Theorems 1.1, 1.2, 1.3 and 1.4.

Normalize so that $\omega = \alpha + i\beta$ satisfies $\int_X |\omega|^2 = 1$. Under the assumption that $\mathrm{SL}(X,\omega)$ is a lattice, we will show:

Theorem 8.1 *Suppose the horizontal foliation $\mathcal{F}(\omega)$ of X is periodic. Then either:*

1. *We have $Z(\omega) \cong \omega^\omega + 1$ and $D^\infty Z(\omega) = \{\beta\}$; or*

2. *All geodesics with slopes tending to zero equidistribute on X, $Z(\omega) = \{\beta\}$, and the cusp at ∞ for $\mathrm{SL}(X,\omega)$ has rank one.*

(Rank is defined by equation (7.1).) We will also see that the natural map

$$Z(\omega) \to H^1(X, \mathbb{R})$$

is injective, and that each current in $Z(\omega)$ is a linear combination of the currents $\beta|A_i$. The corresponding results on measures follow easily.

Remark. When $\mathcal{F}(\omega)$ is not periodic, it is uniquely ergodic by the Veech dichotomy, and hence $Z(\omega) = \{\beta\}$.

Currents. As in §5, let $P(\omega)$ denote the cone of closed, positive currents carried by $\mathcal{F}(\omega)$, and let

$$P_{\pm}(\omega) = P(\omega) - P(\omega)$$

denote its linear span. Let us say a current ξ is *normalized* if

$$L(\xi) = \left| \int_X \xi \wedge \omega \right| = 1.$$

For example, α and β are normalized currents. The space of normalized currents in $P(\omega)$ is compact.

Let C be an oriented, closed geodesic for $(X, |\omega|)$. Its length and angle with respect to ω are characterized by the relation

$$\int_C \omega = L(C) \exp(i\theta(C));$$

and C determines a normalized closed current of integration

$$C/L(C) \in P(e^{-i\theta(C)}\omega).$$

Limits of closed geodesics. Suppose the horizontal foliation $\mathcal{F}(\omega)$ is periodic. Let $A = C(0) = \{A_1, \ldots, A_n\}$ be the corresponding cylinder decomposition of X.

Each A_i determines a closed, positive current $\beta|A_i \in P(\omega)$, which represents a diffuse linear combination of the closed geodesics foliating A_i. Note that $L(\beta|A_i) = \mathrm{area}(A_i)$ in the metric $|\omega|$.

We now associate, to every closed geodesic C with $\theta(C) \neq 0$, the current in $P(\omega)$ given by:

$$z(C) = \sum_i \frac{i(A_i, C)}{c(A_i)} (\beta|A_i),$$

as well as the normalized current

$$\widehat{z}(C) = z(C)/L(z(C)).$$

These currents account for all limits of periodic cycles. More precisely, we have:

Theorem 8.2 *The currents of the form $\widehat{z}(C)$ are dense in $Z(\omega)$.*

Lemma 8.3 *If $\theta(C_n) \to 0$, then the currents $C_n/L(C_n)$ and $\widehat{z}(C_n)$ have the same limit in $Z(\omega)$.*

Proof. Note that $L(C_n) \to \infty$, since $\theta(C_n) \neq 0$. Let s_n be the slope of C_n. Then for each cylinder A_i, $C_n \cap A_i$ consists of $i(A_i, C_n)$ segments spiraling evenly from one end of the cylinder to the other. Thus the current $C_n|A_i$ is nearly a multiple of $\beta|A_i$. To determine this multiple, note that $L(C_n|A_i) = (1/s_n)h(A_i)i(A_i, C_n) + O(1)$, while $L(\beta|A_i) = \text{area}(A_i) = h(A_i)c(A_i)$. Thus $C_n = \sum(C_n|A_i)$ is well approximated by the current

$$\frac{1}{s_n}\sum_i \frac{i(A_i, C_n)}{c(A_i)}(\beta|A_i) = \frac{z(C_n)}{s_n}.$$

Since $L(C_n) \to \infty$, the difference between $C_n/L(C_n)$ and $\widehat{z}(C_n)$ tends to zero as $n \to \infty$. ∎

Proof of Theorem 8.2. Let C be any closed geodesic with nonzero slope. Let

$$\tau_A \in \text{Aff}^+(X, \omega)$$

be the fundamental twist associated to the cylinder system A_i, and let $C_n = \tau_A^n(C)$. Then $\theta_n(C) \to 0$, but $z(C_n) = z(C)$ for all n since $i(A_i, \tau_A^n(C)) = i(A_i, C)$. It follows that

$$\widehat{z}(C) = \lim C_n/L(C_n) \in Z(\omega).$$

Conversely, if $\xi = \lim C_n/L(C_n) \in Z(\omega)$ with $\theta(C_n) \to 0$, then $z(C_n)/L(z(C_n))$ converges to ξ by Lemma 8.3. Hence currents of the form $z(C)/L(z(C))$ are dense in $Z(\omega)$. ∎

Proof of Theorem 8.1. Define a linear map $w : \mathbb{R}^A \to P_\pm(\omega)$ by

$$w(v) = \sum v_i c(A_i)^{-1}(\beta|A_i). \tag{8.1}$$

Since the forms $\beta|A_i$ are linearly independent, this map is injective. Let $\widehat{w}(v) = w(v)/L(w(v))$ when $L(w(v)) \neq 0$. Then $\widehat{w} : \mathbb{PR}^A \dashrightarrow P(\omega)$ satisfies

$$\widehat{w}(i(A_i, C)) = \widehat{z}(C)$$

for all closed geodesics C with $\theta(C) \neq 0$. Since $\widehat{w}(v)$ is continuous and injective on the compact set $\Pi(A) = \overline{\{i(A_i, C)\}} \subset \mathbb{PR}^A$, Theorem 8.2 implies that

$$\widehat{w} : \Pi(A) \to Z(\omega) \tag{8.2}$$

is a homeomorphism. To complete the proof, observe that $\widehat{w}(c(A_i)) = \beta$, and apply Theorem 7.1 on $\Pi(A)$. ∎

Proposition 8.4 *If* $\mathrm{SL}(X, \omega)$ *is a lattice, then the natural map* $Z(\omega) \to H^1(X, \mathbb{R})$ *is injective.*

Proof. We may assume $\mathcal{F}(\omega)$ is periodic, since otherwise $Z(\omega)$ is a single point. By replacing $i(A_i, C)$ with $\langle A_i, C\rangle$ in the definition of $z(C)$, we obtain a linear map

$$z' : H^1(X, \mathbb{R}) \to P_\pm(\omega)$$

whose image contains $Z(\omega)$. Note that $[\beta|A_i] = h(A_i)[A_i] \in H^1(X, \mathbb{R})$; thus

$$\langle z'(C), C\rangle = \sum \mathrm{mod}(A_i)\langle A_i, C\rangle^2 \geq 0.$$

Now suppose $[z'(C_1)] = [z'(C_2)]$ in $H^1(X, \mathbb{R})$. Setting $C = C_1 - C_2$, we conclude from the equation above that $\langle A_i, C_1\rangle = \langle A_i, C_2\rangle$ for all i, and hence $z'(C_1) = z'(C_2)$ as currents. Thus the image of z' maps injectively into $H^1(X, \mathbb{R})$, so the same is true for $Z(\omega)$. ∎

Remark. It is also known that $P(\omega)$ maps injectively into $H^1(X, \mathbb{R})$ whenever $\mathcal{F}(\omega)$ has a dense leaf [Mc5, Prop. 3.3].

Proof of Theorems 1.2. Our aim is to describe the space of limit measures M_s for a period slope s. We may assume that $s = 0$. There is a natural identification between 2-currents and measures on X, satisfying

$$\alpha \wedge \beta = |\omega|^2 \quad \text{and} \quad \mathrm{Re}(e^{-i\theta(C)\omega}) \wedge C = \rho_C,$$

where ρ_C is arclength measure on an oriented closed geodesic C, itself considered as a current of integration in the wedge product above. It follows easily from these observations that the map $\xi \mapsto \alpha \wedge \xi$ gives a homeomorphism

$$Z(\omega) \cong M_0.$$

Theorem 1.2 is then immediate from Theorem 8.1, using the fact that the trace field of $\mathrm{SL}(X, \omega)$ is $\mathbb{Q}$ whenever V has a cusp of rank one. ∎

Proof of Theorem 1.3. We now wish to describe the measures in M_s. As above we can assume $s = 0$ and $A = C(0) = \{A_1, \ldots, A_n\}$.

First, observe that $i(A_i, C) = |\langle A_i, C \rangle|$ for any closed geodesic C. Thus vectors of the form $v = [iA_i, C]$ with C a closed geodesic are dense in $\Pi(A)$, and every $v \in \Pi(A)$ has this form for some cohomology class C. By Theorem 7.1, we can even assume that C is a rational cohomology class, provided $v \neq [c(A_i)]$.

To complete the proof, we simply transport these observations to M_0, using the fact that $\widehat{\mu}(C) = \alpha \wedge \widehat{w}(|\langle A_i, C \rangle|)$ and $\alpha \wedge \widehat{w}(c(A_i)) = |\omega|^2$. (Here $\widehat{\mu}$ and $\widehat{w}$ are normalized versions of the functions $\mu(C)$ and $w(v)$ defined in equations (1.3) and (8.1).) ∎

Proof of Theorem 1.4. Let K denote the thick part of V. For simplicity suppose $\theta(C_n) \to 0$, C_n is oriented and $C_n / L(C_n) \to \xi \in Z(\omega)$.

The proof that C_n is uniformly distributed if $T_n = L(\gamma_n \cap K) \to \infty$ follows the same lines as the proof that $\sigma_n \to \mathcal{R}(V)$ if $L(\sigma_n \cap K) \to \infty$. Indeed, in this case we have $[\xi] = [\beta]$ in $H^1(X, \mathbb{R})$ by Theorem 5.3, and hence $\xi = \beta$ by Proposition 8.4. Therefore the measures $\rho_{C_n} / L(C_n)$ converges to $\alpha \wedge \beta = |\omega|^2$, so C_n is uniformly distributed in X.

We now prove the converse, under the provision that the trace field of $\mathrm{SL}(X, \omega)$ is not $\mathbb{Q}$. Suppose T_n does not tend to infinity. Pass to a subsequence such that $\sup T_n < \infty$, and let $A = C(0) = \{A_1, \ldots, A_n\}$. We can then choose a sequence of modular symbols σ_n such that $L(\sigma_n \cap K) = T_n + O(1)$ is also bounded, and $v_n = [\mathrm{mod}(A_i) \cdot i(A_i, C_n)]$ is a column of the matrix $I(\sigma_n)$. Since the space of modular symbols with bounded length in the thick part of V is compact, after passing to a subsequence we have $\sigma_n \to \sigma \in \mathcal{S}(V)$, and hence v_n converges to a column v of the rational matrix $I(\sigma)$. If the sequence C_n were uniformly distributed, we also have $v = [h(A_i)] \in \mathbb{PQ}^A$, which is impossible when the trace field is irrational. ∎

Proof of Theorem 1.1. As in Veech's original papers [V1], [V2], one can relate billiard trajectories in the regular polygon P_n to periodic geodesics for the form $\omega = dx/y$ on the curve X defined by $y^2 = x^n - 1$ (or a finite cover of X). The desired results then follow from Theorem 1.2. The condition $n \neq 3, 4$ or 6 insures that the trace field of $\mathrm{SL}(X, \omega)$ is irrational. ■

9 Square–tiled surfaces

In this section we discuss the limit measures on square–tiled surfaces (X, ω). Using the results of §8, Theorem 1.2 can easily be refined to:

Theorem 9.1 *For any periodic slope s associated to a cusp a of V, we have $M_s \cong \omega^\omega + 1$ if a has rank two or more; otherwise, M_s consists of a single point.*

Corollary 9.2 *Provided $\mathrm{SL}(X, \omega)$ is a lattice, the following are equivalent:*

1. *All cusps of $V = \mathbb{H}/\mathrm{SL}(X, \omega)$ have rank one.*
2. *Every sequence of closed geodesics C_n on $(X, |\omega|)$ with lengths tending to infinity is equidistributed.*

We will discuss two well–known examples, showing the variety of behaviors that can occur.

Square–tiled surfaces. Let $(E, dz) = (\mathbb{C}, dz)/\mathbb{Z}[i] \in \Omega\mathcal{M}_1$ denote the square torus. For $g \geq 2$, we say $(X, \omega) \in \Omega\mathcal{M}_g$ is a *square–tiled surface* if the relative periods of ω are contained in $\mathbb{Z}[i]$. In this case, integration of ω gives a canonical map

$$\pi : X \to E,$$

branched only over $z = 0$, such that $\pi^*(dz) = \omega$. The preimages of the vertical and horizontal loops through $z = 0$ on E cut X into $d = \deg(\pi)$ unit squares. Conversely, if (X, ω) can be assembled out of unit squares, it is a square–tiled surface.

It is known [GJ] that when $\mathrm{SL}(X, \omega)$ is a lattice, the following are equivalent: (i) the trace field of $\mathrm{SL}(X, \omega)$ is $\mathbb{Q}$; (ii) $\mathrm{SL}(X, \omega)$ is conjugate to a subgroup of $\mathrm{SL}_2(\mathbb{Z})$; (iii) the orbit $\mathrm{GL}_2^+(\mathbb{R}) \cdot (X, \omega)$ contains a square–tiled surface.

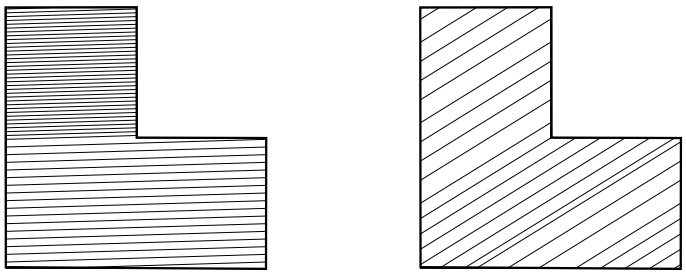

Figure 4. Periodic geodesics near slopes $s = 0$ and $s = 1$.

I. The square *L*. Consider a symmetric L–shaped polygon P made up of three squares. By identifying parallel edges, we obtain a square–tiled surface $(X, \omega) \in \Omega\mathcal{M}_2(2)$ with

$$\mathrm{SL}(X, \omega) = \left\langle \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\rangle \subset \mathrm{SL}_2(\mathbb{Z}).$$

The corresponding Teichmüller curve $V = \mathbb{H}/\mathrm{SL}(X, \omega)$ is the $(2, \infty, \infty)$ orbifold; in the terminology of [Mc1], it is the Weierstrass curve $W_D \subset \mathcal{M}_2$ for discriminant $D = 9$. This is the simplest square–tiled surface of genus $g > 1$.

For this example, Theorem 9.1 implies:

> *$M_{p/q}$ is a single point when p and q are both odd; otherwise, $M_s \cong \omega^\omega + 1$.*

To see this, note that V has two cusps, a and b, corresponding to the slopes 1 and 0 respectively. The first cusp has rank one – indeed, $C(1)$ is a single cylinder; and the corresponding slopes are the ratios of odd integers p/q. The second has rank two; for example, the horizontal and vertical cylinder systems A and B of (X, ω) satisfy

$$i(A_i, B_j) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

By Theorem 9.1, all closed geodesics with slopes $s_n \to 1$ are uniformly distributed, but some with slopes $s_n \to 0$ are not; see Figure 4.

II. The quaternion surface. Our second example is a surface (X, ω) of genus 3 tiled by 8 squares, studied in [HS], [FMZ, Figure 6], [Mo2] and [Mc3,

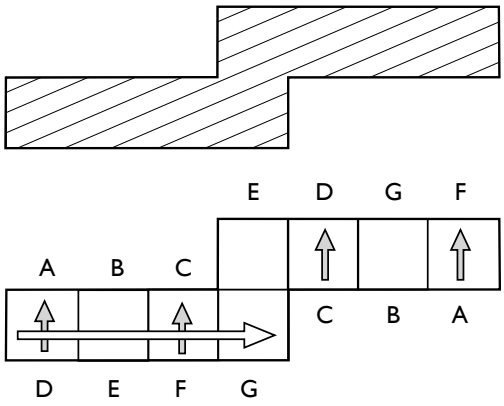

Figure 5. Genus 3 with $\mathrm{Aut}(X,\omega)$ the quaternion group of order 8.

§8]. It arises as a characteristic branched cover of the square torus E, with deck group the quaternion group Q of order 8. All affine automorphisms of E fixing zero lift to X; thus

$$\mathrm{SL}(X,\omega) \cong \mathrm{SL}_2(\mathbb{Z}),$$

and the corresponding Teichmüller curve V is the $(2,3,\infty)$ orbifold, with a unique cusp a.

As can be seen in Figure 5, (X,ω) decomposes into two horizontal cylinders $A = C(0) = \{A_1, A_2\}$; in fact, the core curves of A_1 and A_2 are homologous. Since V has only one cusp, the same is true for all periodic slopes, and hence the cusp a has rank one. For example, the horizontal and vertical cylinder systems satisfy

$$i(A_i, B_j) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.$$

For this square–tiled surface, Theorem 9.1 implies:

> *Every sequence of closed geodesics with $L(C_n) \to \infty$ is uniformly distributed on (X,ω).*

Conditions for rank one. In the first example, it is the fact that there a unique cylinder with slope $s = 1$ that creates a rank one cusp for V. However this mechanism cannot make all cusps rank one.

Proposition 9.3 *There is no square–tiled surface with just a single cylinder in every periodic direction.*

Proof. By the general theory of covering spaces, such a surface (X, ω) is specified by a homomorphism $\rho : \pi_1(E^*) \to S_d$ from the fundamental group of a once–punctured square torus E^* to the symmetric group. We can write $\pi_1(E^*) = \mathbb{Z} * \mathbb{Z} = \langle a, b \rangle$, where a and b represent a pair of simple geodesics on E^* crossing in a single point. Then $a^n b$ is also represented by a simple geodesic, for all $n \in \mathbb{Z}$. The condition on cylinders implies that ρ sends every simple geodesic to a transitive permutation. In particular, if ρ sends (a, b) to (α, β) in S_d, then α, β and $\alpha^n \beta$ must be transitive for all n. But by transitivity, there exists an $n > 0$ such that $\alpha^n(\beta(1)) = 1$, and hence $\alpha^n \beta$ is not transitive. ∎

Question. Can one classify the square–tiled surfaces with all cusps of rank one?

Question. For a general 1–form (X, ω), what can one say about the subspace of $H_1(X, \mathbb{R})$ spanned by the classes $[C]$ of closed geodesics?

10 Pairs of multicurves

In this section we recall Thurston's multicurve construction and prove Theorem 1.9. Thurston's construction was already used implicitly in §6; for more background, see [Th2], [Mc2, §4] and [HL].

Encoding complex geodesics. Let $f : V \to \mathcal{M}_g$ be the complex geodesic generated by a holomorphic 1–form (X, ω) with $\int_X |\omega|^2 = 1$. Assume $\pi_1(V)$ is nonelementary and that V has at least one cusp.

Recall that equation (1.14) gives a natural map

$$\tau : \mathcal{S}^1(V) \to \mathcal{ML}_g(\mathbb{Z}) \times \mathcal{ML}_g(\mathbb{Z}) / \operatorname{Mod}_g,$$

defined by

$$\tau(\gamma) = (\alpha, \beta) = \left(\sum a_i \cdot A_i, \sum b_j \cdot B_j \right),$$

which records the topological configurations of the cylinder systems determined by a pair of periodic slopes.

Thurston's construction allows one to recover the complex geodesic $f : V \to \mathcal{M}_g$, and the modular symbol γ, from the pair of integral laminations (α, β). In particular, any Teichmüller curve can be described by purely topological data; for examples, see e.g. [Lei], [Mc2] and [Ho].

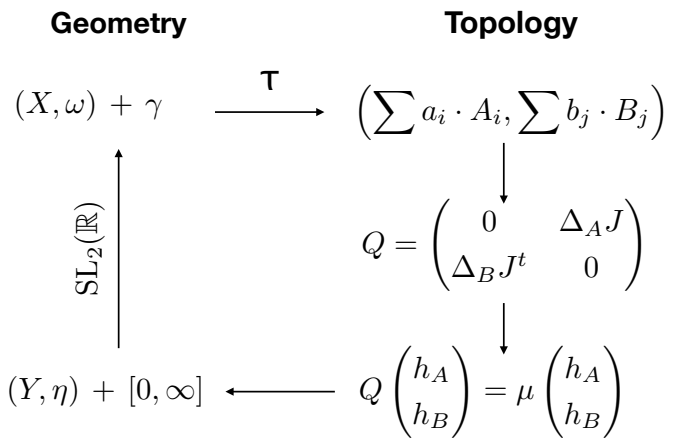

Figure 6. Complex geodesics and Thurston's construction.

From topology to geometry. Here is a sketch of Thurston's construction, summarized in Figure 6.

The pair of integral laminations (α, β) determine an integral matrix Q, whose blocks $\Delta_A J$ and $\Delta_B J^t$ come from the intersection matrix $J = i(A_i, B_j)$, and the diagonal matrices $\Delta_A = (a_i)$ and $\Delta_B = (b_j)$. These blocks are in fact identical to the intersection matrices $I(\gamma)$ and $I(\gamma^*)$; cf. equation (6.4).

The matrix Q has a positive eigenvector $h = (h_A, h_B)$ which is unique up to scale. There is then a unique 1–form (Y, η) such that $C(0) = A$, $C(\infty) = B$, and $(h(A_i), h(B_j)) = (h_A, h_B)$. The form (Y, η) can be constructed explicitly by placing a rectangle $([0, h_i] \times [0, h_j], dz)$ at each crossing of A_i and B_j, and then gluing them together when two crossings are joined by an edge. Suitably scaling h, one can also arrange that $\int_Y |\eta|^2 = 1$.

The difference between (X, ω) and (Y, η) is that the components of $A_i \cap B_j$ are parallelograms in the first case and rectangles in the second. By straightening these parallelograms, one can easily show that

$$\mathrm{SL}_2(\mathbb{R}) \cdot (X, \omega) = \mathrm{SL}_2(\mathbb{R}) \cdot (Y, \eta).$$

In particular, (Y, η) and (X, ω) generate the same Teichmüller curve $f : V \to \mathcal{M}_g$.

Proof of Theorem 1.9. The periodic slopes $(0, \infty)$ for (Y, η) also determine a geodesic joining a pair of cusps of V, allowing one to recover the

original modular symbol γ as well. Consequently τ is injective. Thus modular symbols of degree 1 on the complex geodesic $f : V \to \mathcal{M}_g$ are in bijection with the pairs of multicurves (α, β) which encode it. ∎

The intersection matrix, *reprise*. Since the intersection matrices $I(\gamma)$ and $I(\gamma^*)$ allow one to construct Q, they record important algebraic information about Thurston's construction [Mc2, §4]. For example, one can show that

$$\mathrm{SL}(Y, \eta) \supset \langle D\tau_A, D\tau_B \rangle = \left\langle \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix} \right\rangle,$$

where $\lambda = \mu^2$ is the leading eigenvalue of the Perron–Frobenius matrix $P = I(\gamma)I(\gamma^*)$. In particular, the trace field of $\mathrm{SL}(X, \omega)$ is given by $\mathbb{Q}(\lambda)$. Since P is conjugate to a real symmetric matrix, this field is totally real.

Note that V can be presented by Thurston's construction as soon as $\mathcal{S}^1(V)$ is nonempty. Thus for any $(X, \omega) \in \Omega\mathcal{M}_g$, we have:

> *If* $\mathrm{SL}(X, \omega)$ *is nonelementary and contains a parabolic element, then its trace field is totally real.*

See [HL, Thm 1.1].

Pseudo–Anosov maps and quadratic differentials. The original purpose of Thurston's construction was to provide a rich source of examples of pseudo–Anosov maps, namely the affine maps ψ in $\langle \tau_A, \tau_B \rangle$ such that $|\operatorname{tr} D\psi| > 2$. It is still not known which number fields $\mathbb{Q}(\lambda)$ arise from the expansion factors λ of pseudo–Anosov maps [Th4].

Thurston's construction implicitly uses the inverse of the natural map

$$\iota : \mathcal{QT}_g \to \mathcal{ML}_g \times \mathcal{ML}_g,$$

which records the horizontal and vertical foliations of a marked quadratic differential. However the inverse is not applied to (α, β); rather it is applied to the pair $(\sum h_i A_i, \sum h_j B_j)$ constructed using an eigenvector for Q. The map ι is injective and its image is the space of pairs of measured laminations (α, β) that bind the surface Σ_g; cf. [Le] and [GM, Theorem 3.1]

A Appendix: Modular symbols and the Weil–Petersson metric

For some additional perspective on modular symbols, in this Appendix we give a short proof of:

Theorem A.1 *Let $L \subset \mathbb{R}$ denote the set of lengths of all Weil–Petersson geodesics in $\mathcal{M}_{1,1}$ that begin and end at the cusp. Then $\overline{L}$ is well–ordered, and we have*

$$\overline{L} = \langle L \rangle \cong \omega^\omega.$$

Here $\langle L \rangle$ is the additive semigroup generated by L.

Proof. Let $V = \mathcal{M}_{1,1} \cong \mathbb{H}/\operatorname{SL}_2(\mathbb{Z})$ be the moduli space of hyperbolic Riemann surface of genus one with one cusp, endowed with the Weil–Petersson metric. It is well–known that the corresponding metric on $\mathcal{T}_{1,1} \cong \mathbb{H}$ is negatively curved, convex, and incomplete; and that its completion is given by

$$\mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q}).$$

Moreover $\mathbb{H}^*/\operatorname{SL}_2(\mathbb{Z}) \cong \overline{V} = \mathcal{M}_{1,1} \cup \{p\}$ is a compact metric space, with a single added point p corresponding to $\mathcal{M}_{0,3}$. Near p, $\mathcal{M}_{1,1}$ is well–approximated metrically by the surface of revolution in $\mathbb{R}^3$ obtained by spinning the curve $y^2 = x^3$ about the x–axis. (See e.g. [Wol].)

Let $\ell(x, y)$ denote the length of the unique Weil–Petersson geodesic in $\mathbb{H}$ joining a given pair of distinct points $x, y \in \mathbb{P}^1(\mathbb{Q})$. Since $\ell(gx, gy) = \ell(x, y)$ for all $g \in \operatorname{SL}_2(\mathbb{Z})$, this length gives a map

$$\ell : \mathcal{S}^1(V) \to \mathbb{R}.$$

Extending the definition to all modular symbols by

$$\ell(\gamma_1 * \cdots * \gamma_m) = \sum \ell(\gamma_i),$$

we obtain a functor $\ell : \mathcal{S}(V) \to \mathbb{R}$; this means simply that $\ell(\sigma * \tau) = \ell(\sigma) + \ell(\tau)$. Note that $L = \ell(S^1(V))$.

We now make two geometric observations. Suppose $\gamma_n \to \sigma = \delta_1 * \cdots * \delta_m$ in $\mathcal{S}(V)$. Then:

$$\ell(\sigma) \leq \limsup \ell(\gamma_n),$$

since length can only be lost in the geometric limit. On the other hand, we also have

$$\ell(\sigma) > \ell(\gamma_n) \tag{A.1}$$

for all n sufficiently large. Indeed, for all $n \gg 0$, a representative γ_n' of the homotopy class of γ_n on $\mathcal{M}_{1,1}$ can be obtained by cutting off δ_i at distance ϵ from p, and then connecting δ_i to δ_{i+1} with a curve that spirals finitely many times around the cusp. Due to the shape of the cusp in the Weil–Petersson metric, these spirals each add length on the order of $\epsilon^{3/2} \ll \epsilon$. Choosing ϵ sufficiently small, we obtain $\ell(\gamma_n) \leq \ell(\gamma_n') < \ell(\sigma)$.

Combining these two observations, we find $\ell(\gamma_n) \to \ell(\sigma)$ as $n \to \infty$. It follows easily that the functor $\ell : \mathcal{S}(V) \to \mathbb{R}$ is *continuous*, and comparison to hyperbolic length shows that ℓ is *proper*.

Let $S = \bigcup_{d=1}^{\infty} \mathcal{S}^d(V)$. By basic properties of modular symbols (§2), we have $S = \overline{\mathcal{S}^1(V)} = \langle \mathcal{S}^1(V) \rangle$; and hence, by the properties of ℓ just established, we also have $\overline{L} = \langle L \rangle$. Equation (A.1) implies that $\overline{L}$ is well–ordered.

It remains to show that $\overline{L}$ is homeomorphic to ω^ω; equivalently, that $D^\infty(\overline{L}) = \emptyset$ but $D^n(\overline{L}) \neq \emptyset$ for all finite n. The first point follows from the fact that ℓ is proper and $D^\infty(S) = \emptyset$; while the second follows from equation (A.1), which implies that $D^n(\overline{L})$ contains $\ell(\mathcal{S}^{n+1}(V))$. ∎

Remark. A related result, valid for all $\mathcal{M}_{g,n}$, is announced in [BB, Theorem 1.5].

References

[Bi] B. J. Birch. Elliptic curves over Q: A progress report. In *1969 Number Theory Institute*, volume XX of *Proc. Sympos. Pure Math.*, pages 396–400. Amer. Math. Soc., 1971.

[BM] D. W. Boyd and R. D. Mauldin. The order type of the set of Pisot numbers. *Topology Appl.* **69** (1996), 115–120.

[BB] J. F. Brock and K. W. Bromberg. Inflexibility, Weil-Petersson distance, and volumes of fibered 3–manifolds. *Math. Res. Lett.* **23** (2016), 649–674.

[DL] D. Davis and S. Lelièvre. Periodic paths on the pentagon, double pentagon and golden L. *Preprint, 2018.*

[Fo] G. Forni. A geometric criterion for the nonuniform hyperbolicity of the Kontsevich–Zorich cocycle. *J. Mod. Dyn.* **5** (2011), 355–395.

[FMZ] G. Forni, C. Matheus, and A. Zorich. Square-tiled cyclic covers. *J. Mod. Dyn.* **5** (2011), 285–318.

[GM] F. P. Gardner and H. Masur. Extremal length geometry of Teichmüller space. *Complex Variables Theory Appl.* **16** (1991), 209–237.

[GJ] E. Gutkin and C. Judge. Affine mappings of translation surfaces: geometry and arithmetic. *Duke Math. J.* **103** (2000), 191–213.

[HS] F. Herrlich and G. Schmithüsen. An extraordinary origami curve. *Math. Nachr.* **281** (2008), 219–237.

[Ho] W. P. Hooper. Grid graphs and lattice surfaces. *Int. Math. Res. Not.* 2013 pages 2657–2698.

[HL] P. Hubert and E. Lanneau. Veech groups without parabolic elements. *Duke Math. J.* **133** (2006), 335–346.

[KS] R. Kenyon and J. Smillie. Billiards on rational-angled triangles. *Comment. Math. Helv.* **75** (2000), 65–108.

[La] S. Lang. *Introduction to Modular Forms.* Springer-Verlag, 1995.

[Lei] C. J. Leininger. On groups generated by two positive multi-twists: Teichmüller curves and Lehmer's number. *Geom. Topol.* **8** (2004), 1301–1359.

[Le] G. Levitt. Foliations and laminations on hyperbolic surfaces. *Topology* **22** (1983), 119–135.

[Man] Y. I. Manin. Lectures on modular symbols. In *Arithmetic Geometry*, Clay Math. Proc., pages 137–152. Amer. Math. Soc., 2009.

[Mas] H. Masur. Ergodic theory of translation surfaces. In *Handbook of Dynamical Systems, Vol. 1B*, pages 527–547. Elsevier B. V., 2006.

[MT] K. Matsuzaki and M. Taniguchi. *Hyperbolic Manifolds and Kleinian Groups.* Oxford University Press, 1998.

[Maz] B. Mazur. Courbes elliptiques et symboles modulaires. In *Séminaire Bourbaki, 1971/72*, volume 317, pages 277–294. Springer–Verlag, 1973.

[Mc1] C. McMullen. Billiards and Teichmüller curves on Hilbert modular surfaces. *J. Amer. Math. Soc.* **16** (2003), 857–885.

[Mc2] C. McMullen. Prym varieties and Teichmüller curves. *Duke Math. J.* **133** (2006), 569–590.

[Mc3] C. McMullen. Braid groups and Hodge theory. *Math. Ann.,* **355** (2013), 893–946.

[Mc4] C. McMullen. Cascades in the dynamics of measured foliations. *Ann. scient. Éc. Norm. Sup.* **48** (2015), 1–39.

[Mc5] C. McMullen. Teichmüller dynamics and unique ergodicity via currents and Hodge theory. *J. reine angew. Math.* **768** (2020), 39–54.

[Mc6] C. McMullen. Billiards, heights and the arithmetic of non–arithmetic groups. *Preprint, 2020.*

[Mo1] M. Möller. Affine groups of flat surfaces. In A. Papadopoulos, editor, *Handbook of Teichmüller Theory*, volume II, pages 369–387. Eur. Math. Soc., 2009.

[Mo2] M. Möller. Shimura and Teichmüller curves. *J. Mod. Dyn.* **5** (2011), 1–32.

[MS] S. Mozes and N. Shah. On the space of ergodic invariant measures of unipotent flows. *Ergodic Theory Dynam. Systems* **15** (1995), 149–159.

[Mun] J. R. Munkres. *Elementary Algebraic Topology.* Addison–Wesley, 1984.

[Th1] W. P. Thurston. *Geometry and Topology of Three-Manifolds.* Lecture Notes, Princeton University, 1979.

[Th2] W. P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.* **19** (1988), 417–431.

[Th3] W. P. Thurston. *Three-Dimensional Geometry and Topology*, volume 1. Princeton University Press, 1997.

[Th4] W. P. Thurston. Entropy in dimension one. In *Frontiers in complex dynamics*, volume 51 of *Princeton Math. Ser.*, pages 339–384. Princeton Univ. Press, 2014.

[V1] W. Veech. Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. math.* **97** (1989), 553–583.

[V2] W. Veech. The billiard in a regular polygon. *Geom. Funct. Anal.* **2** (1992), 341–379.

[Wol] S. A. Wolpert. Geometry of the Weil-Petersson completion of Teichmüller space. *Surveys in Differential Geometry* **8** (2003), 357–393.

[Z] A. Zorich. Flat surfaces. In *Frontiers in Number Theory, Physics, and Geometry. I*, pages 437–583. Springer, 2006.

Mathematics Department, Harvard University, Cambridge, MA 02138-2901