COMPOSITIO MATHEMATICA

On Selberg's eigenvalue conjecture for moduli spaces of abelian differentials

Michael Magee

Compositio Math. **155** (2019), 2354–2398.

doi:10.1112/S0010437X1900767X





https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Compositio Math. 155 (2019) 2354–2398
doi:10.1112/S0010437X1900767X

On Selberg's eigenvalue conjecture for moduli spaces of abelian differentials

Michael Magee

Abstract

J.-C. Yoccoz proposed a natural extension of Selberg's eigenvalue conjecture to moduli spaces of abelian differentials. We prove an approximation to this conjecture. This gives a qualitative generalization of Selberg's $\frac{3}{16}$ theorem to moduli spaces of abelian differentials on surfaces of genus $\geqslant 2$.

1. Introduction

Let $\Lambda := \mathrm{SL}_2(\mathbf{Z})$ be the modular group. Then Λ acts on the hyperbolic upper half plane $\mathbb{H}$ by Möbius transformations and the quotient $X := \Lambda\backslash\mathbb{H}$ is an orbifold Riemann surface. We denote by $\Lambda(q)$ the *principal congruence subgroup* of Λ given by the kernel of the reduction modulo q map $\Lambda \to \mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$. Then $\Lambda(q)$ is a normal subgroup of Λ and for $q \geqslant 2$

$$X(q) := \Lambda(q)\backslash\mathbb{H}$$

is a Riemann surface.

If we parameterize points in $\mathbb{H}$ by $x + iy$ with $x, y \in \mathbf{R}$ and $y > 0$, then the Laplacian on $\mathbb{H}$ is given by

$$\Delta = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right).$$

This operator is invariant under Möbius transformations and hence descends to an operator on smooth functions on $X(q)$. The surface $X(q)$ also has a measure μ_q induced from the Λ-invariant volume form $(dx \wedge dy)/y^2$ on $\mathbb{H}$. The Laplacian extends to an unbounded operator $\Delta_{X(q)}$ on $L^2(X, \mu_q)$. For all $q \geqslant 2$, $\Delta_{X(q)}$ has a simple eigenvalue at 0 and the spectrum of $\Delta_{X(q)}$ below $\frac{1}{4}$ is discrete. Therefore we may write $\lambda_1(X(q))$ for the smallest non-zero eigenvalue of $\Delta_{X(q)}$. In a celebrated 1965 paper [Sel65], Selberg proved the following.

Theorem 1.1 (Selberg's $\frac{3}{16}$ theorem). *For all $q \geqslant 2$, $\lambda_1(X(q)) \geqslant \frac{3}{16}$.*

At the same time, Selberg made the following conjecture.

Conjecture 1.2 (Selberg's eigenvalue conjecture). For all $q \geqslant 2$, $\lambda_1(X(q)) \geqslant \frac{1}{4}$.

This conjecture cannot be pushed any further since there are examples [Maa49] of q such that $\Delta_{X(q)}$ has an eigenvalue at $\frac{1}{4}$. Progress on Conjecture 1.2 has been made by several authors over the interim decades, including works of Gelbart and Jacquet ($\lambda_1 > \frac{3}{16}$) [GJ78],

Received 3 May 2019, accepted in final form 17 July 2019, published online 30 October 2019.
2010 Mathematics Subject Classification 11F72, 22E40, 32G15, 30F60, 37D20, 37D25, 37D35 (primary).
Keywords: abelian differentials, spectral gap, congruence covers.
This journal is © Foundation Compositio Mathematica 2019.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Luo, Rudnick and Sarnak ($\lambda_1 \geqslant \frac{171}{784}$) [LRS95], and Iwaniec ($\lambda_1 \geqslant \frac{10}{49}$) [Iwa96]. The current best result is due to Kim and Sarnak [Kim03] who proved for all $q \geqslant 2$,

$$\lambda_1(X(q)) \geqslant \tfrac{975}{4096} \approx 0.238.$$

Selberg's conjecture remains one of the fundamental open questions of automorphic forms; see the expository articles of Sarnak [Sar95, Sar05].

Selberg's conjecture can also be stated in terms of representation theory. This is fitting with Selberg's original motivation[1] of Conjecture 1.2 as an archimedean analog of the Ramanujan–Petersson conjectures. The equivalence classes of irreducible unitary representations (*unitary dual*) of $\mathrm{SL}_2(\mathbf{R})$ were classified by Bargmann [Bar47]: one has the trivial representation, the principal series, complementary series, discrete series and limits of discrete series. Of particular interest to us are the complementary series Comp^u that are indexed by a parameter $u \in (0,1)$; see [Kna01, p. 36] for a precise description of these representations.

For each $q \geqslant 2$, we obtain a unitary representation of $\mathrm{SL}_2(\mathbf{R})$ on $L^2(\Lambda(q)\backslash \mathrm{SL}_2(\mathbf{R}))$ by right translation. This representation can be decomposed as a direct integral over a projection-valued measure on the unitary dual of $\mathrm{SL}_2(\mathbf{R})$. Conjecture 1.2 is equivalent to the following.

Conjecture 1.2*. For all $q \geqslant 2$, the measure on the unitary dual of $\mathrm{SL}_2(\mathbf{R})$ that decomposes $L^2(\Lambda(q)\backslash \mathrm{SL}_2(\mathbf{R}))$ is supported away from complementary series representations.

The point of view taken in this work is that $\Lambda\backslash \mathrm{SL}_2(\mathbf{R})$ is the moduli space of unit area translation surfaces of genus 1 and hence Selberg's eigenvalue conjecture is a conjecture about moduli spaces and their covering spaces. A *translation surface* is a closed topological surface S with a finite subset Σ, together with a set of complex charts on $S - \Sigma$ such that all transition functions are translations, and the charts extend to conical singularities at Σ. Let $\Sigma = \{A_1, \ldots, A_s\}$. The conical singularity at A_i is required to have cone angle $2\pi(\kappa_i + 1)$ with $\kappa_i \in \mathbf{Z}_+$ and the Gauss–Bonnet theorem forces the relation

$$\sum_{i=1}^{s} \kappa_i = 2g - 2.$$

Translation surfaces can be equivalently be thought of as abelian differentials with respect to a complex structure on S. The zeros of the differential correspond to the conical singularities of the translation surface.

The moduli space of translation surfaces of genus $g \geqslant 2$ is stratified according to the partitions $\kappa = (\kappa_1, \ldots, \kappa_s)$. A stratum $\mathcal{H}(\kappa)$ need not be connected, but there are finitely many connected components that are understood by work of Kontsevich and Zorich [KZ03]. We let $\mathcal{H}^{(1)}(\kappa) \subset \mathcal{H}(\kappa)$ denote the unit area translation surfaces in $\mathcal{H}(\kappa)$. In this paper, $\mathcal{M}$ will be a connected component of $\mathcal{H}^{(1)}(\kappa)$. Since $\mathcal{H}^{(1)}(\kappa)$ can be obtained as a quotient of a Teichmüller space by the mapping class group $\Gamma = \Gamma(S, \Sigma)$ of (S, Σ) (see § 2.1), we may define *congruence covers* via the natural family of maps

$$\Pi_q : \Gamma \mapsto \mathrm{Aut}(H_1(S, \mathbf{Z}/q\mathbf{Z})). \tag{1}$$

The *principal congruence subgroup* $\Gamma(q)$ is defined to be the kernel of Π_q. By considering moduli only up to $\Gamma(q)$, and not Γ, for each connected component $\mathcal{M}$ of $\mathcal{H}^{(1)}(\kappa)$ we obtain a congruence cover $\mathcal{M}(q)$ generalizing $\Lambda(q)\backslash \mathrm{SL}_2(\mathbf{R})$. The details of this construction are given in § 2.1.

Each component $\mathcal{M}$ has the following associated objects generalizing those attached to $\Lambda\backslash \mathrm{SL}_2(\mathbf{R})$.

[1] See [HS08].

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

- There is an action of $\mathrm{SL}_2(\mathbf{R})$ on $\mathcal{M}$. The restriction of the $\mathrm{SL}_2(\mathbf{R})$ action to the one parameter diagonal subgroup gives a flow on $\mathcal{M}$ called the *Teichmüller flow* that generalizes the geodesic flow on the unit tangent bundle of X.
- There is a unique probability measure $\nu_{\mathcal{M}}$ on $\mathcal{M}$ that is $\mathrm{SL}_2(\mathbf{R})$-invariant, ergodic for the Teichmüller flow, and in the Lebesgue class with respect to a natural affine orbifold structure on $\mathcal{M}$. This is due to works of Masur [Mas82] and Veech [Vee82].
- The space $\mathrm{SO}(2)\backslash\mathcal{M}$ is locally foliated by $\mathbb{H}$ and hence it is possible to define a *foliated Laplacian* $\Delta_{\mathcal{M}}$ on $\mathrm{SO}(2)\backslash\mathcal{M}$ generalizing Δ_X. This operator has a simple eigenvalue at zero and by a result of Avila and Gouëzel [AG13] its spectrum below $\frac{1}{4}$ has no accumulation points other than possibly at $\frac{1}{4}$.

Each of these objects lifts to $\mathcal{M}(q)$, so there is an $\mathrm{SL}_2(\mathbf{R})$ action on $\mathcal{M}(q)$ preserving a finite measure $\nu_{\mathcal{M}(q)}$, and a foliated Laplacian $\Delta_{\mathcal{M}(q)}$ whose spectrum below $\frac{1}{4}$ does not accumulate[2] away from $\frac{1}{4}$. Hence we can write $\lambda_1(\mathcal{M}(q))$ for the infimum of the non-zero spectrum[3] of $\Delta_{\mathcal{M}(q)}$. The following extension of Selberg's conjecture to genus $g \geqslant 2$ was proposed by Yoccoz.[4]

Conjecture 1.3 (Yoccoz). For all $q \geqslant 2$, and any connected component $\mathcal{M}$ of a stratum, we have the following:

(A) $\lambda_1(\mathcal{M}(q)) \geqslant \frac{1}{4}$;
(B) the measure on the unitary dual of $\mathrm{SL}_2(\mathbf{R})$ that decomposes $L^2(\mathcal{M}(q), \nu_{\mathcal{M}(q)})$ is supported away from complementary series representations.

The main theorem of this paper gives an approximation to Conjecture 1.3.

Theorem 1.4. *For any connected component $\mathcal{M}$ of a stratum, there exists $\epsilon, \eta > 0$ and Q_0[5] $\in \mathbf{Z}_+$ such that for all q coprime to Q_0 the following hold:*

(A) $\lambda_1(\mathcal{M}(q)) \geqslant \epsilon$;
(B) *the measure on the unitary dual of* $\mathrm{SL}_2(\mathbf{R})$ *that decomposes $L^2(\mathcal{M}(q), \nu_{\mathcal{M}(q)})$ is supported away from complementary series representations* Comp^u *with $u \in (1-\eta, 1)$;*
(C) *the Teichmüller flow on $\mathcal{M}(q)$ has exponential decay of correlations on compactly supported C^1 observables with a rate of decay that is independent of q.*

The corresponding theorem for $\mathcal{M}$, i.e. without any congruence aspect, was obtained by Avila, Gouëzel and Yoccoz in [AGY06]. In an earlier version of this manuscript, for certain types of components $\mathcal{M}$, Theorem 1.4 was conditional on a conjecture of Zorich [Zor99] that has since been proved by Gutiérrez-Romo [Gut19].

It is known that parts (A), (B), and (C) of Theorem 1.4 are equivalent. That part (B) implies part (C), namely, that one can use representation theory to deduce rates of mixing of the diagonal flow, is due to Ratner [Rat87]. The argument that part (C) implies part (B) is given

[2] By [AG13, Remark 2.4] this result also applies to $\mathcal{M}(q)$.

[3] In contrast to the situation with X, where it is known [Sel56] that there are infinitely many eigenvalues of Δ_X, we do not know whether $\Delta_{\mathcal{M}}$ or $\Delta_{\mathcal{M}(q)}$ have any non-zero eigenvalues.

[4] The formulation of the conjecture appears in print in [AG13], although Avila and Gouëzel stopped short of making the conjecture because of lack of evidence. We learned from Matheus that Yoccoz had made this conjecture in private.

[5] See the proof of Proposition 3.12 for the origin of Q_0.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

by the 'reverse Ratner estimates' in [AGY06, Appendix B]. The equivalence between parts (A) and (B) is due to the interpretation of the foliated Laplacian as a Casimir operator. This is discussed in detail in [AG13, § 3.4].

So it is sufficient to prove the dynamical statement of part (C). This is made into a precise statement in Theorem 3.5.

We mention that in recent work [MR19], joint with Rühr and Gutiérrez-Romo, we extend Theorem 1.4 to congruence covers coming from relative homology of (S, Σ), and apply both Theorem 1.4 and the extended result to the problem of counting saddle connections in a homology class modulo q.

The current paper deals with Masur–Veech measures on components of strata of abelian differentials. As pointed out to us by Eskin, one can ask whether the results of the current paper extend to arbitrary $\mathrm{SL}_2(\mathbf{R})$-invariant and ergodic probability measures[6] on strata. The natural starting point for such an investigation is the paper of Avila and Gouëzel [AG13] where a spectral gap result is obtained for such measures. It is not clear how to combine the methods of the current paper with [AG13], since the methods here make crucial use of a symbolic coding of the Teichmüller flow. On the other hand, the methods of [AG13] are based on anisotropic Sobolev spaces and do not involve a symbolic model.

1.1 The ideas of the proof

The reader is invited to read this section before the rest of the paper for the main ideas of the proof.

While we will prove Theorem 1.4 in dynamical terms, the philosophy of the proof goes back to works of Brooks [Bro86] and Burger [Bur86, Bur88] that were originally stated in terms of the first non-zero eigenvalue λ_1. Both Brooks and Burger realized that if one has a Galois covering $Y \to X$ of Riemann surfaces, with deck transformation group G, then one can transfer bounds on the spectral gap of the Cayley graph of G with respect to certain generators, to bounds on the first non-zero eigenvalue $\lambda_1(Y)$ of the Laplacian on Y. In particular, if X is fixed, and Y ranges over a family of Galois covers, if the associated Cayley graphs have a uniform spectral gap, then $\lambda_1(Y)$ is uniformly bounded below away from zero.

The classic construction of an infinite family of graphs of bounded degree with a uniform spectral gap, known as an *expander family*, is take a fixed generating set U in an arithmetic lattice $G(\mathbf{Z})$ that has Kazhdan's property (T), and then form the Cayley graphs for $G(\mathbf{Z}/q\mathbf{Z})$ with respect to the projection of U modulo q. This construction is due to Margulis [Mar73].

Since the covering spaces $\mathcal{M}(q)$ of this paper have deck transformation groups contained in $\mathrm{Sp}(H_1(S, \mathbf{Z}/q\mathbf{Z}), \cap) \cong \mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$, and $\mathrm{Sp}_{2g}(\mathbf{Z})$ has property (T) for $g \geqslant 2$, one might expect the Brooks–Burger philosophy to apply directly here, as long as one can prove that the deck transformation group is all of $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$, or in other words, $\mathcal{M}(q)$ is connected. However even if the issue of $\mathcal{M}(q)$ being connected is resolved,[7] the Brooks–Burger philosophy does not obviously apply. The core issue is that the foliated Laplacian is not elliptic and only measures fluctuations of functions in the direction of $\mathrm{SL}_2(\mathbf{R})$-leaves.

[6] Such measures were classified by Eskin and Mirzakhani in [EM18].

[7] And this issue can be resolved as follows; however, these arguments are not used in the paper. The image $\Gamma_{\mathcal{M}}$ of the natural representation of the fundamental group of $\mathcal{M}$ in $\mathrm{Aut}(H_1(S, \mathbf{Z}))$ is known to be a Zariski-dense subgroup of $\mathrm{Sp}_{2g}(\mathbf{Z})$ by a result of Filip from [Fil17, Corollary 1.3]. Then one has the strong approximation result of Matthews, Vaserstein and Weisfeiler [MVW84] that says if $\Gamma_{\mathcal{M}}$ is Zariski-dense in $\mathrm{Sp}_{2g}(\mathbf{Z})$ then $\Gamma_{\mathcal{M}}$ maps onto $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$ for all q coprime to some fixed modulus q_0 and hence that $\mathcal{M}(q)$ is connected for the same q.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Instead we take a dynamical viewpoint. We think of functions on $\mathcal{M}(q)$ as sections of a $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$ principal bundle over $\mathcal{M}$, and make this precise in Lemma 2.1. We know the dynamics on $\mathcal{M}$ is exponentially mixing by the work of Avila, Gouëzel and Yoccoz [AGY06]. The key point for obtaining uniform exponential mixing as in Theorem 1.4(C) is to exploit the following fact: when one travels along the Teichmüller flow and returns close to the initial point, we move in the fibre by a monodromy element of $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$. This monodromy is globally defined in the sense that for a given approximate loop, the monodromy at different levels q are obtained by reduction mod q of some element of $\mathrm{Sp}_{2g}(\mathbf{Z})$. Moreover, if one can argue that the dynamics on the base $\mathcal{M}$ is sufficiently combinatorially complicated, then we can obtain many monodromy elements in this way. Then we hope to use property (T) to prove this dynamics in the fibre spreads out exponentially fast. So one has exponential mixing in the base, and some form of exponential mixing in the fibres, and hopes to combine these two. The problem is that the two processes are not independent. So we will use hyperbolicity of the dynamics on the base $\mathcal{M}$ to 'decouple' these aspects of the dynamics. However, the base dynamics is not uniformly hyperbolic, so one needs to perform 'time acceleration' as in [AGY06] to induce uniform hyperbolicity and then incorporate this into the method.

The previous paragraph was a high level overview of the approach. Now we give details of how this is implemented.

Our framework for understanding the dynamics of $\mathcal{M}$ is that same as Avila, Gouëzel and Yoccoz in [AGY06]. Namely, instead of working with $\mathcal{M}$, we pass to a finite cover called the moduli space of zippered rectangles $\mathrm{Rect}_{\mathcal{M}}$ for $\mathcal{M}$. This finite cover carries a lift of the Teichmüller flow that has some very nice properties that were worked out in [AGY06]. A key insight of [AGY06] is that by carefully choosing a cross section, one obtains a model of the flow on $\mathrm{Rect}_{\mathcal{M}}$ as a suspension flow over a hyperbolic skew product $\hat{Z} : \hat{\Xi} \to \hat{\Xi}$ with a base transformation $Z : \Xi \to \Xi$ that is a uniformly expanding Markoff map (Lemmas 3.2 and Proposition 3.1). Moreover the roof function for this suspension model has desirable properties, it is 'good' in the sense of [AGY06] (Lemma 3.3), and it has exponential tails (Theorem 3.4). The latter statement is quite hard and relies on exponential recurrence estimates for the Teichmüller flow that were first obtained by Athreya [Ath06].

This suspension model has another key property that is not explicitly used in [AGY06]: the symbolic coding is very well adapted to keeping track of what happens to the homology of the surface when we follow the flow. Indeed, there is a linear group G attached to $\mathcal{M}$ called the *Rauzy–Veech group* that is defined purely in terms of the symbolic dynamics of $\mathrm{Rect}_{\mathcal{M}}$ and the return maps on the base of the suspension model. This group G performs the desired function of keeping track of monodromy in homology around approximate loops and is defined precisely in § 2.7. It was a conjecture of Zorich [Zor99] that G is Zariski-dense in its ambient symplectic group. Recently, it has been proven in works of Avila, Matheus and Yoccoz [AMY18], and Gutiérrez-Romo [Gut19], that the Rauzy–Veech group is finite index in $\mathrm{Sp}_{2g}(\mathbf{Z})$. Therefore, in particular, it has property (T). The precise statement about the Rauzy–Veech group that we use is given in Theorem 2.3.

Other than discussing the Rauzy–Veech group, the main purpose of § 2 is to go through the setup of [AGY06] and explain how to keep track of what happens to homology along the flow, as well as stating the results we need from [AGY06].

In § 3, we follow the strategy of Avila, Gouëzel and Yoccoz of reducing Theorem 3.5, the precise formulation of our main theorem, to exponential mixing of the flow on $\mathrm{Rect}_{\mathcal{M}}$ (Theorem 3.6), and then to exponential mixing of a suspension flow over the base Ξ of the hyperbolic skew product (Theorem 3.9). These statements must now be uniform in q.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

A well-known technique for proving exponential mixing of suspension flows is to take a Laplace transform of the correlation function, and express this transform in terms of iterates of *transfer operators*. To deal with the q aspect, one uses *skew transfer operators*, one operator for each q. The transfer operators act on vector-valued C^1 functions on Ξ and one needs spectral estimates for the transfer operators that are uniform in q. This strategy of proving uniform exponential mixing via q-uniform bounds on transfer operators originates in work of Oh and Winter [OW16]. One needs estimates for the transfer operators in two regimes: high frequency (given by Proposition 3.11) and low frequency (given by Proposition 3.12).

The technique for carrying out the necessary high frequency estimates are due to Dolgopyat [Dol98] and extended to the current setting, with no q-aspect, by Avila, Gouëzel and Yoccoz [AGY06]. The use of the Dolgopyat argument to establish q-uniform versions of the high frequency estimates was first done by Oh and Winter [OW16], and then in a different setting by Magee, Oh, and Winter [MOW19]. In § 4 we explain how to extend the arguments of Avila, Gouëzel and Yoccoz in this regime to skew transfer operators.

The technique for proving q-uniform low frequency estimates for skew transfer operators goes back to the work of Bourgain, Gamburd, and Sarnak[8] [BGS11]. The philosophy here, mirroring the Brooks–Burger philosophy, is that an iterate of the transfer operator looks somewhat like an iterate of the adjacency operator of a Cayley graph of $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$. In work of Bourgain, Kontorovich, and Magee [MOW19, Appendix], an improvement was made to this method that allows one to use uniform expansion of Cayley graphs (in the current setting, furnished by property (T)) as a 'black box'[9] to prove q-uniform estimates for transfer operators.

We give the details of how this method can be extended to the current setting in § 5. It requires not only the uniform expansion of certain Cayley graphs as an input, but also an extra input that the dimensions of representations of $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$ that do not arise from representations of $\mathrm{Sp}_{2g}(\mathbf{Z}/q'\mathbf{Z})$ with $q'|q$ have a lower bound that is polynomial in q. This is a version of *quasirandomness*[10] for $\mathrm{Sp}(\mathbf{Z}/q\mathbf{Z})$. The reason for needing this kind of bound is that it allows us to obtain information on the spectral radius of a complex-valued measure μ on $\Gamma_q = \mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$ acting by convolution on a certain subspace of $\ell^2(\Gamma_q)$ if we have information on the spectral radius of a real-valued measure μ' that majorizes $|\mu|$. This is a key idea in § 5. We state the precise quasirandomness estimate we need in Proposition 6.1 and then prove it following an argument of Kelmer and Silberman [KS13].

2. Background

2.1 Abelian differentials and translation surfaces

Let $g \geqslant 1$ and let $S = S_g$ be a fixed closed topological surface of genus g. Let $\Sigma = \{A_1, \ldots, A_s\}$ be a finite subset of S. An *abelian differential* on (S, Σ) is a pair $(\mathcal{J}, \omega)$ where $\mathcal{J}$ is a complex structure on S and ω is a holomorphic one form with respect to $\mathcal{J}$, and with zeros contained in Σ.

[8] Bourgain, Gamburd, and Sarnak were interested in spectral bounds for transfer operators for reasons that are related to exponential mixing but in [BGS11] phrased in terms of counting problems.

[9] The original argument of Bourgain, Gamburd, and Sarnak in [BGS11] involved unraveling the proof that the associated Cayley graphs are uniform expanders.

[10] Gowers [Gow08, Theorem 4.5] made the definition that a finite group G should be regarded as quasirandom relative to an ambient parameter C if the dimension of any non-trivial irreducible representation of G has dimension $\geqslant C$. Prior to this formal notion, the concept had been used in the construction of Ramanujan graphs by Lubotzky, Phillips and Sarnak [LPS88], the work of Sarnak and Xue on multiplicities of automorphic representations [SX91], and the construction of uniformly expanding Cayley graphs of $\mathrm{SL}_2(\mathbb{F}_p)$ by Bourgain and Gamburd [BG08].

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

As is well known, an abelian differential ω on (S, Σ) gives S the structure of a translation surface with conical singularities in Σ; the complex structure comes from integrating the differential. Hence we may speak about the area of an abelian differential as the area of the corresponding translation surface.

One may further specify that the abelian differential has a zero of order $\kappa_i \in \mathbf{Z}_+$ at A_i. This is possible whenever $\sum \kappa_i = 2g - 2$. For such $\kappa = (\kappa_1, \ldots, \kappa_s)$ we let $\mathcal{X}(\kappa)$ denote the collection of abelian differentials on (S, Σ) with zeros of orders $\kappa_1, \ldots, \kappa_s$ at $A_1, \ldots, A_s$, up to isotopies of S preserving Σ. This Teichmüller space has a natural affine manifold structure arising through period coordinates as described in [AGY06, § 2.2.1]. Let $\mathcal{X}^{(1)}(\kappa) \subset \mathcal{X}(\kappa)$ be the abelian differentials whose corresponding translation surface has unit area, up to isotopy. Then $\mathcal{X}^{(1)}(\kappa)$ is a submanifold of $\mathcal{X}(\kappa)$.

The modular group $\Gamma = \Gamma(S, \Sigma)$ is defined to be the homeomorphisms of S that fix Σ pointwise, modulo homeomorphisms that are isotopic to the identity relative to Σ. Thus Γ acts on $\mathcal{X}(\kappa)$, preserving $\mathcal{X}^{(1)}(\kappa)$, and we define $\mathcal{H}(\kappa)$ to be $\Gamma\backslash\mathcal{X}(\kappa)$ and $\mathcal{H}^{(1)}(\kappa) = \Gamma\backslash\mathcal{X}^{(1)}(\kappa)$. This $\mathcal{H}^{(1)}(\kappa)$ is often referred to as a *stratum* of the moduli space of unit area abelian differentials. The connected components of these strata have been classified by Kontsevich and Zorich [KZ03]. Throughout the paper we write $\mathcal{M}$ for a connected component of $\mathcal{H}^{(1)}(\kappa)$.

Any connected component $\mathcal{M}$ of $\mathcal{H}^{(1)}(\kappa)$ inherits, from the manifold structure of $\mathcal{X}^{(1)}(\kappa)$, the structure of an orbifold. We define $\mathcal{H}^{(1)}(\kappa; q) = \Gamma_g(q)\backslash\mathcal{X}^{(1)}(\kappa)$ where $\Gamma_g(q)$ is the kernel of Π_q defined in (1). We thus have a covering map $\mathcal{H}^{(1)}(\kappa; q) \to \mathcal{H}^{(1)}(\kappa)$. We define $\mathcal{M}(q)$ to be the preimage of $\mathcal{M}$ under this map. For each q the lift of $\mathcal{M}(q)$ to $\mathcal{X}(\kappa)$ is a submanifold.

Recall that a Finsler manifold is a smooth manifold together with a continuous assignment of norm on each tangent fibre. The norm is called a Finsler metric. As described in [AGY06, § 2.2.2] there is a Γ-invariant Finsler metric on $\mathcal{X}(\kappa)$ arising from period coordinates making $\mathcal{X}(\kappa)$ into a Finsler manifold. This induces a Finsler manifold structure on $\mathcal{X}^{(1)}(\kappa)$.

2.2 The Hodge bundle

The Hodge bundle is defined to be the fibred product

$$H_1(\mathcal{H}^{(1)}(\kappa)) := \Gamma\backslash(\mathcal{X}^{(1)}(\kappa) \times H_1(S, \mathbf{R})) \to \Gamma\backslash\mathcal{X}^{(1)}(\kappa) = \mathcal{H}^{(1)}(\kappa),$$

where the mapping class group Γ acts diagonally. Let $\mathcal{H}^{(1)}(\kappa)^0$ be the complement of the orbifold points in $\mathcal{H}^{(1)}(\kappa)$. The Hodge bundle restricts to a vector bundle $H_1(\mathcal{H}^{(1)}(\kappa)^0)$ over $\mathcal{H}^{(1)}(\kappa)^0$. At any orbifold point $[(\mathcal{J}, \omega)]$ of $\mathcal{H}^{(1)}(\kappa)$ the fibre degenerates to $\mathrm{Aut}(\mathcal{J}, \omega)\backslash H_1(S, \mathbf{R})$. Note that by Hurwitz's automorphisms theorem, $\mathrm{Aut}(\mathcal{J}, \omega)$ is a finite group.

The total space of the Hodge bundle contains as a discrete subset the lattice bundle

$$\Gamma\backslash(\mathcal{X}^{(1)}(\kappa) \times H_1(S, \mathbf{Z})).$$

Then one may specify the *Gauss–Manin* connection on the Hodge bundle by the requirement that lattice-valued continuous sections be parallel. This gives a flat vector bundle connection on $H_1(\mathcal{H}^{(1)}(\kappa)^0)$ that extends to a flat connection on $H_1(\mathcal{H}^{(1)}(\kappa))$ in the following sense. A section of $H_1(\mathcal{H}^{(1)}(\kappa))$ can be viewed as a function $\sigma : \mathcal{X}^{(1)}(\kappa) \to H_1(S, \mathbf{R})$ that transforms according to

$$\sigma(\gamma.x) = \gamma_* \sigma(x), \quad \gamma \in \Gamma.$$

Then a local section is parallel by definition if it takes values in $H_1(S, \mathbf{Z})$ and this specifies the connection on general sections.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

The action of Γ on $H_1(S,\mathbf{Z})$ lies in the integral symplectic group $\mathrm{Sp}(H_1(S,\mathbf{Z}),\cap)$ where $\cap$ is the (symplectic) intersection form on integral homology. Therefore for any unitary representation $(\rho,\mathcal{W})$ of $\mathrm{Sp}(H_1(S,\mathbf{Z}),\cap)$ we obtain an *associated orbifold vector bundle*[11] $H_1(\mathcal{H}^{(1)}(\kappa);\rho)$. The total space of this bundle is

$$H_1(\mathcal{H}^{(1)}(\kappa);\rho)=\Gamma\backslash(\mathcal{X}^{(1)}(\kappa)\times\mathcal{W}) \tag{2}$$

where the action of Γ on $\mathcal{X}^{(1)}(\kappa)\times\mathcal{W}$ is given by $\gamma.(\omega,v)=(\gamma.\omega,\rho(\gamma_*).v)$, where $\gamma_*\in\mathrm{Sp}(H_1(S,\mathbf{Z}),\cap)$ is the map induced by γ on homology. This bundle also has a flat connection, in the same sense as before, coming from the fibred product structure in (2).

Of course, for any connected component $\mathcal{M}$ of the stratum $\mathcal{H}^{(1)}(\kappa)$ we may restrict $H_1(\mathcal{H}^{(1)}(\kappa))$ or $H_1(\mathcal{H}^{(1)}(\kappa);\rho)$ to $\mathcal{M}$. We denote by $H_1(\mathcal{M};\rho)$ the obtained orbifold vector bundle.

For a lot of the rest of the paper we deal with abstract unitary ρ but in reality we are interested in the following specific examples. Recall the map Π_q from (1). Because the symplectic intersection product $\cap$ on $H_1(S,\mathbf{Z}/q\mathbf{Z})$ is preserved by the mapping class group, we have

$$\Pi_q:\Gamma\to\mathrm{Sp}(H_1(S;\mathbf{Z}/q\mathbf{Z}),\cap).$$

We let $\Gamma_q=\mathrm{Sp}(H_1(S;\mathbf{Z}/q\mathbf{Z}),\cap)$. Let $\ell^2_0(\Gamma_q)$ be the subspace of functions in $\ell^2(\Gamma_q)$ that are orthogonal to constant functions with respect to the ℓ^2 inner product. This gives a subrepresentation $(\rho_q,\ell^2_0(\Gamma_q))$ of the action of Γ on $\ell^2(\Gamma_q)$ by reduction mod q and then left translation.[12]

We will also consider the subspace of $\ell^2_0(\Gamma_q)$ consisting of functions that are orthogonal to all functions lifted from $\Gamma_{q'}$ with $q'|q$ via the natural mapping of reduction modulo q'

$$\Gamma_q\to\Gamma_{q'}.$$

We denote by $\ell^2_{\mathrm{new}}(\Gamma_q)$ this *new subspace* of functions. This gives a subrepresentation $(\rho_q^{\mathrm{new}},\ell^2_{\mathrm{new}}(\Gamma_q))$ of $(\rho_q,\ell^2_0(\Gamma_q))$.

2.3 The Teichmüller flow on moduli space

There is a postcomposition action of $\mathrm{SL}_2(\mathbf{R})$ on the space of abelian differentials on S as follows. For $h\in\mathrm{SL}_2(\mathbf{R})$ we define

$$h.(\mathcal{J},\omega)=(\mathcal{J}_h,\omega_h),$$

where

$$\omega_h=h\begin{pmatrix}\Re(w)\\ \Im(w)\end{pmatrix}$$

and $\mathcal{J}_h$ is the unique complex structure on S that makes ω_h holomorphic. This action preserves the area of abelian differentials. As this action also commutes with any homeomorphism of S, it descends to both the Teichmüller spaces $\mathcal{X}(\kappa)$, $\mathcal{X}^{(1)}(\kappa)$ and $\mathcal{H}(\kappa)$, $\mathcal{H}^{(1)}(\kappa)$ and $\mathcal{M}$. The *Teichmüller geodesic flow* on any of these objects is the restriction of the $\mathrm{SL}_2(\mathbf{R})$ action to the diagonal subgroup:

$$\mathcal{T}_t(\mathcal{J},\omega):=\begin{pmatrix}e^t & 0\\ 0 & e^{-t}\end{pmatrix}.(\mathcal{J},\omega).$$

The Teichmüller flow also preserves each connected component $\mathcal{M}$. By results of Masur [Mas82] and Veech [Vee82] there is a unique probability measure $\nu_{\mathcal{M}}$ that is invariant and ergodic for the Teichmüller flow on $\mathcal{M}$. This measure is in the Lebesgue class with respect to period coordinates

[11] By *orbifold vector bundle* we mean that the fibres are vector spaces of constant rank away from the orbifold points of the base space, where the fibres degenerate only to a quotient of a vector space by a finite group.

[12] In other words, it is the inflation of the left regular representation of Γ_q to Γ.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

on $\mathcal{M}$. We pull back the measure $\nu_{\mathcal{M}}$ on $\mathcal{M}$, using the counting measure on the fibres of the covering map, to obtain a measure $\nu_{\mathcal{M}(q)}$ on $\mathcal{M}(q)$. Note that $\nu_{\mathcal{M}(q)}$ is not a probability measure.

Since in § 2.2 we specified a connection on each of $H_1(\mathcal{H}^{(1)}(\kappa))$, $H_1(\mathcal{H}^{(1)}(\kappa);\rho)$ the Teichmüller flow acts on sections of each of these bundles by pullback along parallel transport. For example, viewing a section of $H_1(\mathcal{H}^{(1)}(\kappa);\rho)$ as a $\mathcal{W}$-valued function σ on $\mathcal{X}^{(1)}(\kappa)$ satisfying $\sigma(\gamma.x) = \rho(\gamma)\sigma(x)$ for each $\gamma \in \Gamma$, we have the following defining equation for $\mathcal{T}_t^*$:

$$[\mathcal{T}_t^*\sigma](\mathcal{J},\omega) := \sigma(\mathcal{T}_t(\mathcal{J},\omega)). \tag{3}$$

This action also restricts to an action on sections of $H_1(\mathcal{M};\rho)$.

We now explain the relationship between sections of $H_1(\mathcal{M};\rho_q)$ and functions on $\mathcal{M}(q)$. Let $L^2_\star(\mathcal{M}(q))$ be the subspace of functions in $L^2(\mathcal{M}(q))$ orthogonal to lifts from $L^2(\mathcal{M})$, with respect to the measure $\nu_{\mathcal{M}(q)}$. Let $L^2(H_1(\mathcal{M};\rho_q))$ denote the L^2 sections of $H_1(\mathcal{M};\rho_q)$ with respect to the natural Hermitian fibre metric and measure $\nu_{\mathcal{M}}$. We say that a function f on $\mathcal{M}(q)$ or a section σ of $H_1(\mathcal{M};\rho)$ is C^1 if its lift to $\tilde{f} : \mathcal{X}^{(1)}(\kappa) \to \mathbf{C}$ (respectively $\tilde{\sigma} : \mathcal{X}^{(1)}(\kappa) \to \mathcal{W}$) is C^1 (bounded with bounded derivative[13]) with respect to the Finsler manifold structure on $\mathcal{X}^{(1)}(\kappa)$. Define $\|f\|_{C^1} = \|\tilde{f}\|_\infty + \|D\tilde{f}\|_\infty$ and similarly $\|\sigma\|_{C^1}$. Write $C^1(\mathcal{M}(q))$ for the C^1 complex-valued functions on $\mathcal{M}(q)$ and $C^1(H_1(\mathcal{M};\rho))$ for the C^1 sections of $H_1(\mathcal{M};\rho)$. These are Banach spaces with respect to the respective C^1 norms.

Lemma 2.1. *We have the following correspondences.*

(i) *For each q there is a natural linear isometry*

$$\Phi_q : L^2_\star(\mathcal{M}(q)) \to L^2(H_1(\mathcal{M};\rho_q)).$$

If $f \in L^2_\star(\mathcal{M}(q))$, then $\Phi_q[f]$ is the $\ell^2_0(\Gamma_q)$-valued function on $\mathcal{X}^{(1)}(\kappa)$ given by

$$\Phi_q[f](x) = \sum_{\Gamma(q)\gamma\in\Gamma_q} f(\gamma.x)\delta_\gamma,$$

where in the right-hand side we view f as a $\Gamma(q)$-invariant function on $\mathcal{X}^{(1)}(\kappa)$.

(ii) *The map Φ_q intertwines the maps $\mathcal{T}_t^*$ defined by pullback on $L^2_\star(\mathcal{M}(q))$ and by (3) on $L^2(H_1(\mathcal{M};\rho_q))$.*

(iii) *The restriction*

$$\Phi_q : C^1(\mathcal{M}(q)) \cap L^2_\star(\mathcal{M}(q)) \to C^1(H_1(\mathcal{M};\rho_q))$$

preserves C^1 norms.

2.4 Combinatorial data and Rauzy classes

Now we begin an account of the dynamics of the Teichmüller flow, viewed through the lens of Veech's zippered rectangles construction. We draw in the following sections from the sources [AGY06, Via08] that both build on work of Marmi, Moussa and Yoccoz [MMY05].

The relevant combinatorial objects are as follows. Let $\mathcal{A}$ denote a finite alphabet with $|\mathcal{A}| = d$. Eventually, $\mathcal{A}$ will be chosen depending on g, κ and the component $\mathcal{M}$. We let $\mathfrak{S}(\mathcal{A})$ denote the set of pairs

$$(\pi_t, \pi_b),$$

[13] In case of $\mathcal{W}$-valued F on a Finsler manifold X with $\mathcal{W}$ a Hilbert space, to define the norm of the derivative we view the derivative at $x \in X$ as a map $DF_x : T_xX \to T_{F(x)}\mathcal{W} \cong \mathcal{W}$ then use the operator norm with respect to the Finsler metric at x and the Hilbert space norm on $\mathcal{W}$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

where each $\pi_\epsilon : \mathcal{A} \to \{1, \dots, d\}$ is a bijection. Henceforth, ϵ will index one of the symbols t, b ('top' or 'bottom'). As in [AGY06] it is convenient to visualize (π_t, π_b) as a pair of rows each of which contains the elements of $\mathcal{A}$ in some order, where the top corresponds to π_t and the bottom to π_b. We say (π_t, π_b) is *irreducible* if there is no $d' < d$ such that the set of the first d' elements of the top row is the same as the first d' elements of the bottom. Let $\mathfrak{S}^0(\mathcal{A}) \subset \mathfrak{S}(\mathcal{A})$ denote the irreducible combinatorial data.

We now define 'top' and 'bottom' operations on $\mathfrak{S}^0(\mathcal{A})$. For the next paragraph, let α and β denote the last elements of the top and bottom rows of $\pi \in \mathfrak{S}^0(\mathcal{A})$ respectively. The top operation on π modifies the bottom row by moving the occurrence of β to the immediate right of the occurrence of α. The bottom operation modifies the top row by moving α to the right of β. As in [AGY06] we say that the last element of the unchanged row is the *winner* and the last element of the row of π that is to be changed the *loser*.

By adding directed 'top' and 'bottom' labeled edges according to these operations we obtain an edge-labeled directed graph on the vertex set of irreducible combinatorial data $\mathfrak{S}^0(\mathcal{A})$. Each vertex has exactly one incoming top (respectively bottom) and one outgoing top (respectively bottom) edge. A *Rauzy diagram* is a connected component of this graph and a *Rauzy class* is the vertex set of a Rauzy diagram.

2.5 Suspension data and zippered rectangles

Let $\mathfrak{R}$ be a Rauzy class. For each $\pi \in \mathfrak{R}$ we form a cell

$$X_\pi = \{\pi\} \times \mathbf{R}_+^{\mathcal{A}} \times \mathcal{K}_\pi,$$

where

$$\mathcal{K}_\pi = \left\{ \tau \in \mathbf{R}^{\mathcal{A}} : \sum_{\pi_t(\xi) \leqslant k} \tau_\xi > 0, \sum_{\pi_b(\xi) \leqslant k} \tau_\xi < 0 \text{ for all } 1 \leqslant k \leqslant d-1 \right\}.$$

The set $\mathcal{K}_\pi$ is an open convex cone. Let $X_{\mathfrak{R}} = \cup_{\pi \in \mathfrak{R}} X_\pi$. We may drop the dependence on $\mathfrak{R}$ since we usually view it as fixed. We associate to each $\pi \in \mathfrak{R}$ a linear map $\Omega_\pi : \mathbf{R}^{\mathcal{A}} \to \mathbf{R}^{\mathcal{A}}$ given by

$$[\Omega_\pi]_{\alpha,\beta} = \begin{cases} +1 & \text{if } \pi_t(\alpha) > \pi_t(\beta), \pi_b(\alpha) < \pi_b(\beta), \\ -1 & \text{if } \pi_t(\alpha) < \pi_t(\beta), \pi_b(\alpha) > \pi_b(\beta), \\ 0 & \text{else.} \end{cases}$$

There is a construction due to Veech [Vee82] that builds a point in the moduli space of translation surfaces from suspension data (π, λ, τ). This mapping is called the *zippered rectangles* construction that we denote by

$$\mathrm{zip} : X_{\mathfrak{R}} \to \mathcal{H}(\kappa), \quad \kappa = \kappa(\mathfrak{R}).$$

The explicit details of this construction are clearly described in lecture notes of Viana [Via08, ch. 2]. In the current paper it will be better to simply work with the properties of the map zip that we give below. Henceforth a superscript $^{(1)}$ on any set of suspension data refers to the subset whose associated zippered rectangles have unit area: for example $X_{\mathfrak{R}}^{(1)}, X_\pi^{(1)}$ etc.

Theorem 2.2 (Veech [Vee82]). *For any connected component $\mathcal{M}$ of the stratum $\mathcal{H}^{(1)}(\kappa)$ there is a Rauzy class $\mathfrak{R} = \mathfrak{R}(\mathcal{M})$ such that $\mathrm{zip}(X_{\mathfrak{R}}^{(1)}) \subset \mathcal{M}$ and $\mathrm{zip}(X_{\mathfrak{R}}^{(1)})$ has full measure with respect to $\nu_{\mathcal{M}}$.*

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

There is a natural identification

$$\mathbf{R}^{\mathcal{A}}/\ker \Omega_\pi \cong H_1(\mathrm{zip}(\pi,\lambda,\tau),\mathbf{R}) \tag{4}$$

for each $(\pi,\lambda,\tau)\in X_\pi$. For a detailed discussion of this map see Viana [Via08, § 2.9]. The bilinear form

$$(v_1,v_2)\mapsto \langle v_1,-\Omega_\pi v_2\rangle$$

descends to a non-degenerate symplectic form ω_π on $\mathbf{R}^{\mathcal{A}}/\ker\Omega_\pi$. Under the identification (4), the form ω_π is precisely the intersection form on homology. Furthermore, the map (4) arises from an isomorphism of integral symplectic lattices

$$(\mathbf{Z}^{\mathcal{A}}/\ker(\Omega_\pi|_{\mathbf{Z}^{\mathcal{A}}}),\omega_\pi)\cong (H_1(S,\mathbf{Z}),\cap). \tag{5}$$

Therefore the pullback of the Hodge bundle to X_π via zip is naturally trivialized:

$$[\mathrm{zip}^*\, H_1(\mathcal{H}(\kappa))]|_{X_\pi}\cong X_\pi\times \mathbf{R}^{\mathcal{A}}/\ker\Omega_\pi. \tag{6}$$

We also note here that the area of $\mathrm{zip}(\pi,\lambda,\tau)$ is given by

$$\mathrm{area}(\mathrm{zip}(\pi,\lambda,\tau))=\langle\lambda,-\Omega_\pi\tau\rangle. \tag{7}$$

2.6 The Rauzy induction map

Given π, let α be the last element of the top row of π and β the last element of the bottom row. Say that a pair (π,λ) has type *top* if $\lambda_\alpha>\lambda_\beta$. Say it has type *bottom* if $\lambda_\beta<\lambda_\alpha$. This splits each cell into two pieces of the form

$$X_{\pi,\epsilon}=\{(\pi,\lambda,\tau)\in\{\pi\}\times\mathbf{R}^{\mathcal{A}}_+\times\mathcal{K}_\pi : (\pi,\lambda)\text{ of type }\epsilon\},\quad \epsilon\in\{t,b\}$$

together with a hyperplane. We also introduce $Y_{\pi,\epsilon}=\{(\pi,\lambda)\in\{\pi\}\times\mathbf{R}^{\mathcal{A}}_+$ of type $\epsilon\}$, so that

$$X_{\pi,\epsilon}=Y_{\pi,\epsilon}\times\mathcal{K}_\pi.$$

We now give an assignment of a linear map $\Theta_{\pi,\epsilon}:\mathbf{R}^{\mathcal{A}}\to\mathbf{R}^{\mathcal{A}}$ to each pair (π,ϵ). This is given by [Via08, (1.9), (1.10)]

$$[\Theta_{\pi,\epsilon}]_{\alpha,\beta}:=\begin{cases}1 & \text{if } \alpha=\beta,\\ 1 & \text{if } \alpha \text{ loses and } \beta \text{ wins in type } \epsilon \text{ move at } \pi,\\ 0 & \text{else.}\end{cases} \tag{8}$$

If π' is obtained from π by a type ϵ move, then the map[14] $\Theta^*_{\pi,\epsilon}$ maps $Y_{\pi'}:=\{\pi'\}\times\mathbf{R}^{\mathcal{A}}_+$ homeomorphically to $Y_{\pi,\epsilon}$. Furthermore $(\Theta^*_{\pi,\epsilon})^{-1}$ maps $\mathcal{K}_\pi$ injectively into $\mathcal{K}_{\pi'}$ [Via08, Lemma 2.13]. We also have the intertwining relation

$$\Theta_{\pi,\epsilon}\Omega_\pi\Theta^*_{\pi,\epsilon}=\Omega_{\pi'}. \tag{9}$$

The *Rauzy induction map* on suspension data is given by

$$\hat{Q}(\pi,\lambda,\tau):=(\pi',(\Theta^*_{\pi,\epsilon})^{-1}\lambda,(\Theta^*_{\pi,\epsilon})^{-1}\tau)$$

[14] Here and henceforth a $*$ denotes a transpose with respect to the standard basis of $\mathbf{R}^{\mathcal{A}}$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

when $(\pi, \lambda, \tau) \in X_{\pi,\epsilon}$; here again π' is obtained from π by an operation of type ϵ. Using the same notation, notice that $\hat{Q}$ is a skew extension of the map[15]

$$Q(\pi, \lambda) = (\pi', (\Theta^*_{\pi,\epsilon})^{-1}\lambda).$$

Equation (9) together with the area formula (7) shows that $\hat{Q}$ preserves the area of the associated zippered rectangles. Hence $\hat{Q}$ preserves $X^{(1)}$.

The zippered rectangles associated to (π, λ, τ) define the same point in $\mathcal{H}(\kappa)$ as the zippered rectangles associated to $\hat{Q}(\pi, \lambda, \tau)$, that is,

$$\text{zip} \circ \hat{Q} = \text{zip}.$$

See Viana [Via08, § 2.8] for a clear explanation of this fact.

We now define cylinders for the Rauzy induction map. Let γ be a path in the Rauzy diagram associated to the class $\mathfrak{R}$. Throughout the rest of the paper, we consider oriented paths that follow the given direction of the edges.[16] Suppose that γ traverses vertices $\pi(0), \pi(1), \ldots, \pi(N)$ in that order. Then define

$$X_\gamma := X_{\pi(0)} \cap \hat{Q}^{-1}(X_{\pi(1)}) \cap \hat{Q}^{-2}(X_{\pi(2)}) \cap \cdots \cap \hat{Q}^{-N}(X_{\pi(N)}).$$

Notice that $X_{\pi,\epsilon}$ is the same as X_γ where γ is the outgoing type ϵ arrow from π. We then define Θ_γ in terms of the $\Theta_{\pi,\epsilon}$ by stating that for $(\pi, \lambda, \tau) \in X_\gamma$ we have

$$\hat{Q}^N(\pi(0), \lambda, \tau) = (\pi(N), (\Theta^*_\gamma)^{-1}\lambda, (\Theta^*_\gamma)^{-1}\tau).$$

We define $Y_\gamma = Y_{\pi(0)} \cap \cdots \cap Q^{-N}(Y_{\pi(N)})$ the analogous cylinder for Q. If γ begins at π, then we define the subcone of $\mathcal{K}_\pi$

$$\mathcal{K}_\gamma := (\Theta^*_\gamma)^{-1}\mathcal{K}_\pi.$$

2.7 The Rauzy–Veech group

Observe that $\Theta^*_{\pi,\epsilon}$ induces a map $\mathbf{Z}^{\mathcal{A}}/\ker \Omega_{\pi'} \to \mathbf{Z}^{\mathcal{A}}/\ker \Omega_\pi$ in light of (9) and the fact that $\Theta_{\pi,\epsilon}$ is integral from (8). These facts are discussed by Viana in [Via08, § 2.8]. As a consequence, (9) implies that if γ begins and ends at π, Θ^*_γ induces a symplectic endomorphism of $(\mathbf{Z}^{\mathcal{A}}/\ker(\Omega_\pi|_{\mathbf{Z}^{\mathcal{A}}}), \omega_\pi) \overset{(5)}{\cong} (H_1(S, \mathbf{Z}), \cap)$. In fact it is easy to check from (8) that Θ^*_γ is an automorphism. We therefore view each

$$\Theta^*_\gamma \in \text{Sp}(\mathbf{Z}^{2g}, \omega_\pi).$$

For each $\pi \in \mathfrak{R}$ let G_π be the subgroup of $\text{Sp}(\mathbf{Z}^{2g}, \omega_\pi)$ generated by the Θ^*_γ obtained as γ ranges over loops in $\mathfrak{R}$ beginning and ending at π. This group G_π is called the *Rauzy–Veech group* at π.

The key property of G_π that we rely on is the following recent theorem of Gutiérrez-Romo [Gut19, Theorem 1.1] that was previously known for certain *hyperelliptic* components by work of Avila, Matheus and Yoccoz [AMY18, Theorem 1.1].

Theorem 2.3 (Gutiérrez-Romo, Avila–Matheus–Yoccoz). *For any Rauzy class $\mathfrak{R}$ there exists $\pi \in \mathfrak{R}$ such that G_π contains the principal congruence subgroup of level 2 of $\text{Sp}(\mathbf{Z}^{2g}, \omega_\pi)$. Recall the principal congruence subgroup of level 2 is the kernel of reduction modulo 2.*

Theorem 2.3 resolved, in a strong form, the conjecture of Zorich [Zor99, Appendix A.3 Conjecture 5] that G_π should be Zariski-dense.

[15] As a comment for the initiated, the map Q is the Rauzy induction map on Interval Exchange Transformations. See [Rau79] for Rauzy's original analysis of this map.

[16] While it is not immediately obvious, the equivalence classes induced by identifying end points of oriented paths coincide with the Rauzy classes [Via08, Lemma 1.23].

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

2.8 Relationship to the Hodge bundle

Let $\mathcal{M}$ be a connected component of $\mathcal{H}^{(1)}(\kappa)$ and let σ be a section of the Hodge bundle $H_1(\mathcal{M})$. The pullback of σ to any $X_\pi^{(1)}$ under the zippered rectangles map can be naturally viewed as a

$$\mathbf{R}^{\mathcal{A}}/\ker \Omega_\pi$$

valued function $\tilde{\sigma}$ via the identifications (4) and (6). Since Rauzy induction does not change the modulus of zippered rectangles, the fibre of $\mathrm{zip}^* H_1(\mathcal{M})$ at (π, λ, τ) should be identified with the fibre at $\hat{Q}(\pi, \lambda, \tau)$. In fact, the identification involves the previously defined map Θ_γ and requires for $(\pi, \lambda, \gamma) \in X_{\pi,\epsilon}^{(1)}$ that if π' is the result of applying a type ϵ move to π then

$$\tilde{\sigma}(\pi, \lambda, \tau) = \Theta_{\pi,\epsilon}^* \tilde{\sigma}(\hat{Q}(\pi, \lambda, \tau)). \tag{10}$$

The iterated form of the compatibility equation (10) that we will use is the following. If γ is a path of N edges in a Rauzy diagram that begins and ends at π, then for $(\pi, \lambda, \tau) \in X_\gamma^{(1)}$ we have

$$\tilde{\sigma}(\pi, \lambda, \tau) = \Theta_\gamma^* \tilde{\sigma}(\hat{Q}^N(\pi, \lambda, \tau)).$$

This is an important point of this paper as it describes the equivariance properties of sections of the Hodge bundle in the suspension model. We now extend this formula to the setting of associated orbifold vector bundles $H_1(\mathcal{M}; \rho)$. After fixing π, using the isomorphism (5) we identify

$$\Gamma_q \cong \mathrm{Sp}((\mathbf{Z}/q\mathbf{Z})^{2g}, \omega_\pi)$$

so we may view ρ_q and ρ_q^{new} as representations of $\mathrm{Sp}(\mathbf{Z}^{2g}, \omega_\pi)$ that are submodules of $\ell^2(\mathrm{Sp}((\mathbf{Z}/q\mathbf{Z})^{2g}, \omega_\pi))$. More generally, using (5) we may pull back any unitary representation $(\rho, \mathcal{W})$ of $\mathrm{Sp}(H_1(S; \mathbf{Z}), \cap)$ to a representation of $\mathrm{Sp}(\mathbf{Z}^{2g}, \omega_\pi)$ that we also call ρ.

We may now argue by analogy with the Hodge bundle that if σ is any section of the associated bundle $H_1(\mathcal{M}; \rho)$, then the pullback $\tilde{\sigma}$ of this section to a $\mathcal{W}$-valued function on $X_\pi^{(1)}$ satisfies

$$\tilde{\sigma}(\pi, \lambda, \tau) = \rho(\Theta_\gamma^*) \tilde{\sigma}(\hat{Q}^N(\pi, \lambda, \tau)), \quad (\pi, \lambda, \tau) \in X_\gamma^{(1)} \tag{11}$$

for each path γ in $\mathfrak{R}$ of length N beginning and ending at π.

2.9 A fundamental domain

There is a nice fundamental domain for Rauzy induction on X described in [AGY06, p. 159]. We let $\mathcal{F} = \mathcal{F}_{\mathfrak{R}}$ denote the set of (π, λ, τ) such that either:

(i) $\hat{Q}(\pi, \lambda, \tau) = (\pi', \lambda', \tau')$ is defined and $\|\lambda'\| < 1 \leqslant \|\lambda\|$; or
(ii) $\hat{Q}(\pi, \lambda, \tau)$ is not defined and $1 \leqslant \|\lambda\|$; or
(iii) $\hat{Q}^{-1}(\pi, \lambda, \tau)$ is not defined and $\|\lambda\| < 1$.

The norm we use is $\|\lambda\| := \sum_{\alpha \in \mathcal{A}} |\lambda_\alpha|$. The fibres of the zippered rectangles map

$$\mathrm{zip} : \mathcal{F}^{(1)} \to \mathcal{M}, \quad \mathcal{M} = \mathcal{M}(\mathfrak{R})$$

are almost everywhere finite with constant cardinality depending on $\mathcal{M}$.

2.10 The Teichmüller flow on suspension data

Recall that $\mathcal{M}$ is a connected component of $\mathcal{H}^{(1)}(\kappa)$ and $\mathfrak{R}$ the associated Rauzy class. The *Teichmüller flow* is a one parameter flow on $X_{\mathfrak{R}}$ that commutes with $\hat{Q}$ and is given by $T_t(\pi, \lambda, \tau) = (\pi, e^t\lambda, e^{-t}\tau)$. Note that this preserves each $X_\pi^{(1)}$ and $X^{(1)}$. The flow T_t lifts the Teichmüller flow on $\mathcal{M}$, that is,

$$\mathcal{T}_t \circ \mathrm{zip} = \mathrm{zip} \circ T_t.$$

Evidently, T_t preserves Lebesgue measure on X. The flow T_t also preserves Lebesgue measure on $X^{(1)}$, the pushforward of which under zip is a multiple of $\nu_{\mathcal{M}}$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

2.11 Time acceleration and renormalization

The approach of Avila, Gouëzel and Yoccoz [AGY06] to the Teichmüller flow is to consider the first return time to an appropriately chosen cross section. This cross section involves the choice of $\pi \in \mathfrak{R}$ and a path γ_0 that begins and ends at π. We give details on the choice of γ_0 in §§ 2.13 and 5.1. For now, assume we have chosen π and γ_0.

We will use the notation $\mathcal{F}_\pi = \mathcal{F} \cap X_\pi$ and $\mathcal{F}_\gamma = \mathcal{F} \cap X_\gamma$. We consider the regions

$$\hat{\Xi} := \{(\pi, \lambda, \tau) \in \mathcal{F}^{(1)}_{\gamma_0} : \|\lambda\| = 1\} \cap (Y_{\gamma_0} \times \mathcal{K}_{\gamma_0})$$

and the closely related

$$\Xi := \{(\pi, \lambda) \in Y_{\gamma_0} : \|\lambda\| = 1\}.$$

Let $\hat{m}$ (respectively m) denote the normalized natural Lebesgue measure on $\hat{\Xi}$ (respectively Ξ). It is known that almost all orbits of the Teichmüller flow pass through $\hat{Q}^{\mathbf{Z}}(\Xi)$, this is stated in [AGY06, 4.1.3] as a consequence of the ergodicity of the Veech flow.[17] For each $x \in \hat{\Xi}$ we denote by $r(x)$ the first return time of x to $\hat{Q}^{\mathbf{Z}}(\Xi)$ under the Teichmüller flow. That is, $r(x)$ is the smallest positive value such that

$$T_{r(x)}(x) \in \hat{Q}^{-n}(\hat{\Xi})$$

for some positive[18] integer n. This means there is some value $\hat{Z}(x) \in \hat{\Xi}$ such that

$$T_{r(x)}\hat{Q}^n(x) = \hat{Q}^n T_{r(x)}(x) = \hat{Z}(x). \tag{12}$$

Suppose that $x = (\pi, \lambda, \tau) \in X_\gamma$ with $\hat{Q}^n(x) = (\pi, (\Theta^*_\gamma)^{-1}\lambda, (\Theta^*_\gamma)^{-1}\tau) \in \hat{\Xi}$. Then

$$r(x) = -\log\|(\Theta^*_\gamma)^{-1}\lambda\|.$$

Note here that $r(\pi, \lambda, \tau)$ depends only on the coordinates (π, λ) and we can view r also as a function on Ξ.

We will write $\gamma_1.\gamma_2$ or just $\gamma_1\gamma_2$ for the concatenation of two oriented paths γ_1 and γ_2 in $\mathfrak{R}$ with compatible endpoints. In $\gamma_1.\gamma_2$, γ_1 is the first path traversed. Consider γ with the property that the γ_0 subpaths of $\gamma.\gamma_0$ are precisely the beginning and the end segment. We say that such a γ is γ_0-*adapted*. For such a γ, if $x \in X_{\gamma.\gamma_0} \cap \hat{\Xi}$, then

$$\hat{Z}(x) = \left(\pi, \frac{(\Theta^*_\gamma)^{-1}\lambda}{\|(\Theta^*_\gamma)^{-1}\lambda\|}, \|(\Theta^*_\gamma)^{-1}\lambda\|(\Theta^*_\gamma)^{-1}\tau\right).$$

The domain of $\hat{Z}$ is therefore $\cup_{\gamma_0}$-adapted $\gamma\hat{\Xi}_{\gamma\gamma_0}$ where

$$\hat{\Xi}_{\gamma\gamma_0} := \hat{\Xi} \cap (Y_{\gamma\gamma_0} \times \mathcal{K}_{\gamma_0}).$$

We extend this definition to $\hat{\Xi}_{\gamma_1\cdots\gamma_N\gamma_0} := \hat{\Xi} \cap (Y_{\gamma_1\cdots\gamma_N\gamma_0} \times \mathcal{K}_{\gamma_0})$ where $\gamma_1, \ldots, \gamma_N$ are a sequence of γ_0-adapted paths with both endpoints equal to π.

Notice that the mapping $\hat{Z}$ has the following properties.

(i) The map $\hat{Z}$ is a skew extension of the mapping $Z : \Xi \to \Xi$ defined Lebesgue almost everywhere by

$$Z(\pi, \lambda) = \left(\pi, \frac{(\Theta^*_\gamma)^{-1}\lambda}{\|(\Theta^*_\gamma)^{-1}\lambda\|}\right), \quad (\pi, \lambda) \in Y_{\gamma.\gamma_0}.$$

The connected components of the domain of Z are the sets

$$\Xi_{\gamma\gamma_0} := \Xi \cap Y_{\gamma\gamma_0}.$$

[17] The Veech flow is not discussed in the current paper.

[18] Notice that from (8) that Θ does not decrease norms, so if $(\pi', \lambda', \tau') = \hat{Q}(\pi, \lambda, \tau)$ then $\|\lambda'\| \leqslant \|\lambda\|$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

(ii) The maps $\hat{Z}$ and Z preserve $\|\lambda\| = 1$. This is usually referred to as *renormalization*.

(iii) The maps $\hat{Z}$ and Z involve many iterations of Rauzy induction and this is usually referred to as *time acceleration*. This is first due to Zorich [Zor96]; see also [Zor06, § 5.3] for further discussion.

(iv) The map $\hat{Z}$ (respectively Z) preserves the Lebesgue measure $\hat{m}$ (respectively m).

Following [AGY06, § 4.2.1], in order to enforce hyperbolicity of the map $\hat{Z}$ (cf. Proposition 3.1 and Lemma 3.2) one puts adapted metrics on Ξ and $\hat{\Xi}$. On Ξ we put the Hilbert metric d_Ξ coming from the inclusion $\Xi \to Y_\pi$ and on $\hat{\Xi}$ we consider the product metric

$$d_{\hat{\Xi}}((\pi,\lambda,\tau),(\pi,\lambda',\tau')) := d_\Xi((\pi,\lambda),(\pi,\lambda')) + d_{\mathcal{K}_\pi}(\tau,\tau'),$$

where $d_{\mathcal{K}_\pi}$ is the Euclidean distance in $\mathcal{K}_\pi$. These metrics induce Finsler metric structures on Ξ and $\hat{\Xi}$ that make them into Finsler manifolds.

2.12 Flow on sections of associated bundles in the suspension model

We may now map

$$\hat{\Xi}_r := \{(x,s) : x \in \hat{\Xi}, s \in [0, r(x))\}$$

homeomorphically to a part of $X^{(1)}_\pi$ by the map

$$P : (x,s) \mapsto T_s(x). \tag{13}$$

The image $X'^{(1)}_\pi$ of P is up to a Lebesgue-null set, a fundamental domain for the action of $\hat{Q}$ on $X^{(1)}$. Given a section of $H_1(\mathcal{M};\rho)$, its pullback to $X^{(1)}$ is therefore determined (up to a zero measure set) by its values on $X'^{(1)}_\pi \subset X^{(1)}_\pi$.

The pushforward of Lebesgue measure under the mapping in (13) is Lebesgue measure. We write $\hat{m}_r = \hat{m} \otimes \mathrm{Leb}$ for the Lebesgue measure on $\hat{\Xi}_r$.

As explained in § 2.3, $\mathcal{T}_t$ acts by $\mathcal{T}_t^*$ on sections of $H_1(\mathcal{M};\rho)$. If (after pullback) we view a section $\tilde{\sigma}$ as a $\mathcal{W}$-valued function satisfying (11) and then view $\tilde{\sigma}$ as a $\mathcal{W}$-valued function $\hat{\sigma}$ on $\hat{\Xi}_r$ by the mapping in (13) then the action of T_t^* on $\hat{\sigma}$ will be denoted by $\hat{T}_t^*$ and defined as follows. Let γ be γ_0-adapted with $l(\gamma) = n$. If $x \in X^{(1)}_{\gamma.\gamma_0} \cap \hat{\Xi}$ and $t + s \in [r(x), r(x) + r(\hat{Z}(x)))$, then

$$\begin{aligned}
[\hat{T}_t^*\hat{\sigma}](x,s) &= \hat{\sigma}(x, t+s) \\
&= \tilde{\sigma}(T_{t+s}x) \overset{(11)}{=} \rho(\Theta_\gamma^*)\tilde{\sigma}(\hat{Q}^n T_{t+s}x) \\
&\overset{(12)}{=} \rho(\Theta_\gamma^*)\tilde{\sigma}(T_{t+s-r(x)}\hat{Z}(x)) = \rho(\Theta_\gamma^*)\hat{\sigma}(\hat{Z}(x), t+s-r(x)).
\end{aligned}$$

Let

$$r^{(N)}(x) := r(x) + r(\hat{Z}(x)) + \cdots + r(\hat{Z}^{N-1}(x)).$$

For $\gamma_1, \gamma_2, \ldots, \gamma_N$ each γ_0-adapted, $x \in X^{(1)}_{\gamma_1.\gamma_2\cdots\gamma_N.\gamma_0} \cap \hat{\Xi}$, and $t + s \in [r^{(N)}(x), r^{(N+1)}(x))$, we have then

$$[\hat{T}_t^*\hat{\sigma}](x,s) = \rho(\Theta^*_{\gamma_1})\rho(\Theta^*_{\gamma_2})\cdots\rho(\Theta^*_{\gamma_N}).\hat{\sigma}(\hat{Z}^N(x), t+s-r^{(N)}(x)). \tag{14}$$

This is the master equation for the Teichmüller flow on sections of $H_1(\mathcal{M};\rho)$ in our suspension model. Notice that the argument of $\hat{\sigma}$ in the right-hand side of (14) defines a mapping we call

$$\hat{T}_t : \hat{\Xi}_r \to \hat{\Xi}_r, \quad \hat{T}_t(x,s) := (\hat{Z}^N(x), t+s-r^{(N)}(x))$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

for $x \in X^{(1)}_{\gamma_1.\gamma_2\cdots\gamma_N.\gamma_0} \cap \hat{\Xi}$ and $t+s \in [r^{(N)}(x), r^{(N+1)}(x))$. Then $\hat{T}_t$ is the suspension flow over $\hat{Z}$ with roof function r. The flow $\hat{T}_t$ lifts the Teichmüller flow under the mapping in (13) and as a consequence, Lebesgue measure $\hat{m}_r$ on $\hat{\Xi}_r$ is invariant under $\hat{T}_t$.

Since the roof function r depends only on a coordinate in Ξ we may also define

$$\Xi_r = \{(y,s) : y \in \Xi, s \in [0, r(y))\}.$$

We write m_r for the Lebesgue measure on Ξ_r. We also define for $r \in Z^{-(N-1)}(\Xi)$

$$r^{(N)}(y) := r(y) + r(Z(y)) + \cdots + r(Z^{N-1}(y)).$$

We may define a similar operator to $\hat{T}_t^*$ that we will call T_t^* that will act on $\mathcal{W}$-valued functions on Ξ_r. For $\sigma : \Xi_r \to \mathcal{W}$, $\gamma_1, \gamma_2, \ldots, \gamma_N$ each γ_0-adapted, $t+s \in [r^{(N)}(y), r^{(N+1)}(y))$ and $y \in \Xi_{\gamma_1.\gamma_2\cdots\gamma_N.\gamma_0}$ we define

$$[T_t^*\sigma](y,s) = \rho(\Theta^*_{\gamma_1})\rho(\Theta^*_{\gamma_2})\cdots\rho(\Theta^*_{\gamma_N}).\sigma(Z^N(y), t+s-r^{(N)}(y)). \tag{15}$$

We give Ξ_r and $\hat{\Xi}_r$ Finsler metrics that are the product of the Finsler metric on Ξ (respectively $\hat{\Xi}$) with the usual metric in the s direction.

2.13 Preliminary choice of γ_0

Recall γ_0 is a path in $\mathfrak{R}$ beginning and ending in π. We now explain the choice of γ_0 that is made in [AGY06]. Avila, Gouëzel and Yoccoz require that the following hold.

(*Strongly Positive*) γ_0 is a *strongly positive* path, meaning that all the entries of Θ^*_γ are positive and moreover $(\Theta^*_{\gamma_0})^{-1}$ maps $\overline{\mathcal{K}}_\pi - \{0\}$ into $\mathcal{K}_\pi$.

(*Neat*) γ_0 is *neat*, meaning that $\gamma_0 = \gamma'\gamma_e = \gamma_s\gamma'$ implies γ' is trivial or $\gamma' = \gamma_0$. This means in any path, occurrences of γ_0 are (edge) disjoint. Therefore γ_0-adapted γ are precisely those of the form

$$\gamma = \gamma_0\gamma',$$

where γ' does not contain γ_0 as a subpath.

According to [AGY06, § 4.13], such a choice of γ_0 is possible. However, in the present paper, we must choose γ_0 more carefully, while still making sure γ_0 is strongly positive and neat. This is done in § 5.1. For now, assume that γ_0 is strongly positive and neat.

3. Decay of correlations

In this section we state in more precise terms and then prove Theorem 1.4(C) on uniform exponential decay of correlations.

3.1 Dynamical setup

The following definitions and results are from [AGY06]. Recall the maps $\hat{Z}$ and Z introduced in § 2.11. Throughout we use the Finsler metric on the tangent bundle to Ξ defined in § 2.11. We write D for the total derivative of a function. We write $C^0(\Xi)$ for the uniform norm. For a $\mathcal{W}$-valued function F, $\|DF\|$ refers to operator norm with respect to the Finsler metric on the fibres and the Hilbert space metric on $\mathcal{W}$. When we write $\bigcup^*_\gamma$ or $\sum^*_\gamma$ it means that we restrict the indexing to γ_0-adapted γ. We assume here that γ_0 is strongly positive and neat as in § 2.13, since these conditions are required for the results of Avila, Gouëzel and Yoccoz [AGY06].

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Proposition 3.1 [AGY06, Proof of Proposition 4.3]. *The map Z is a* uniformly expanding Markov map *with respect to Lebesgue measure m and the Finsler metric structure defined in § 2.11. That is to say, we have the following.*

(i) *The union*

$$\bigcup_{\gamma}^{*} \Xi_{\gamma\gamma_0}$$

is a countable union of open sets that are m-conull in Ξ.

(ii) *If γ is γ_0-adapted, then Z maps $\Xi_{\gamma\gamma_0}$ diffeomorphically to Ξ and there are constants $\Lambda > 1$ and $c_1(\gamma) > 0$ such that for all $x \in \Xi_{\gamma\gamma_0}$ and v in the tangent fibre to x*

$$\Lambda\|v\| \leqslant \|[DZ]_x.v\| \leqslant c_1(\gamma)\|v\|.$$

(iii) *Let J denote the inverse of the Jacobian of Z with respect to m. The function $\log J$ is C^1 on each $\Xi_{\gamma\gamma_0}$ and there is some $C > 0$ such that for any inverse branch α of Z,*

$$\sup_{y\in\Xi} \|D(\log J \circ \alpha)(y)\| \leqslant C.$$

Lemma 3.2 [AGY06, Lemma 4.3]. *The pair $(\hat{Z}, \hat{m})$ is a* hyperbolic skew product *over (Z, m). This means, with all norms and distances coming from the Finsler metric on $\hat{\Xi}$ defined in § 2.11, we have the following.*

(i) *The projection $\mathrm{pr} : \hat{\Xi} \to \Xi$ defined by*

$$\mathrm{pr}(\pi, \lambda, \tau) = (\pi, \lambda)$$

satisfies $Z \circ \mathrm{pr} = \mathrm{pr} \circ \hat{Z}$ whenever both sides of the equality are defined.

(ii) *The measure $\hat{m}$ gives full mass to the domain of definition of $\hat{Z}$.*

(iii) *There is a family of probability measures $\{\hat{m}_y\}_{y\in\Xi}$ on $\hat{\Xi}$ which is a disintegration of $\hat{m}$ over m in the following sense: $y \mapsto \hat{m}_y$ is measurable, $\hat{m}_y$ is supported on $\mathrm{pr}^{-1}(y)$ and for any measurable $U \subset \hat{\Xi}$, $\hat{m}(U) = \int_{y\in\Xi} \hat{m}_y(U)\, dm(y)$. Moreover, there is a constant $C > 0$ such that for any open $V \subset Z^{-1}(\Xi)$, for any $u \in C^1(\mathrm{pr}^{-1}(V))$ the function $\bar{u}(x) = \int u(x)\, d\hat{m}_y(x)$ is in $C^1(V)$ with*

$$\sup_{y\in V} \|D\bar{u}(x)\| \leqslant C \sup_{x\in\mathrm{pr}^{-1}(V)} \|Du(y)\|.$$

(iv) *There is a constant $K > 1$ such that for all $x_1, x_2 \in \hat{\Xi}$ with $\mathrm{pr}(x) = \mathrm{pr}(y)$ we have*

$$d_{\hat{\Xi}}(\hat{Z}(x_1), \hat{Z}(x_2)) \leqslant K^{-1} d_{\hat{\Xi}}(x_1, x_2).$$

Lemma 3.3 [AGY06, Lemma 4.5]. *The roof function r is* good. *This means the following.*

(i) *There is $\epsilon_1 > 0$ such that $r \geqslant \epsilon_1$.*

(ii) *There is $C > 0$ such that for any inverse branch α of Z one has*

$$\sup_{y\in\Xi} \|D(r \circ \alpha)(y)\| \leqslant C.$$

(iii) *There is no C^1 function ϕ on $\bigcup_{\gamma}^{*} \Xi_{\gamma\gamma_0}$ such that*

$$r - \phi \circ T + \phi$$

is constant on each $\Xi_{\gamma\gamma_0}$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Theorem 3.4 [AGY06, Theorem 4.6]. *The roof function r has* exponential tails. *This means there is $\sigma_0 > 0$ such that*

$$\int_\Xi \exp(\sigma_0 r)\, dm < \infty.$$

3.2 The main technical results

The following will be the precise version of Theorem 1.4(C). Recall the definition of $\mathcal{M}(q)$ from § 2.1.

Theorem 3.5. *There exists $\delta, \eta > 0$ and $Q_0 \in \mathbf{Z}_+$ such that for all q coprime to Q_0, for all u, $v \in C^1(\mathcal{M}(q))$ whose supports project into a compact set $K \subset \mathcal{M}$, there exists $C = C(K) > 0$ such that for all $t \geqslant 0$*

$$\left| \int u.v \circ \mathcal{T}_t \, d\nu_{\mathcal{M}(q)} - |\Gamma_q|^{-1} \left(\int u \, d\nu_{\mathcal{M}(q)} \right) \left(\int v \, d\nu_{\mathcal{M}(q)} \right) \right| \leqslant C(K) \|u\|_{C^1} \|v\|_{C^1} q^\eta e^{-\delta t}.$$

The key feature of this estimate is that δ does not depend on q.

For any Finsler manifold X and Hilbert space $\mathcal{W}$ we may define the Banach space of C^1 $\mathcal{W}$-valued functions on X as in § 2.3. Recall from §§ 2.11 and 2.12 that there are Finsler metric structures on Ξ, $\hat{\Xi}$, Ξ_r, $\hat{\Xi}_r$. If $(\rho, \mathcal{W})$ is a unitary representation we write e.g. $C^1(\Xi; \rho)$ for the C^1 $\mathcal{W}$-valued functions on Ξ, with respect to the Finsler metric. We make a reduction of Theorem 3.5 to the following that is analogous to [AGY06, Theorem 2.7].

Theorem 3.6. *There exists $C, \delta, \eta > 0$ and $Q_0 \in \mathbf{Z}_+$ such that for all q coprime to Q_0, for all $U, V \in C^1(\hat{\Xi}_r; \rho_q)$ and all $t \geqslant 0$*

$$\left| \int \langle U, T_t^* V \rangle \, d\hat{m}_r \right| \leqslant C \|U\|_{C^1} \|V\|_{C^1} q^\eta e^{-\delta t}.$$

We now explain how Theorem 3.5 reduces to Theorem 3.6.

Passage from Theorem 3.6 to Theorem 3.5. Note that in the context of Theorem 3.5, we can write $u = \tilde{u}_0 + u'$, $v = \tilde{v}_0 + v'$ with $u', v' \in L^2_\star(\mathcal{M}(q))$ and $\tilde{u}_0, \tilde{v}_0$ given by lifts of functions from $\mathcal{M}$. In other words, if $\mathrm{cover}_q : \mathcal{M}(q) \to \mathcal{M}$ is the covering map, there are functions u_0 and v_0 such that $\tilde{u}_0 = u_0 \circ \mathrm{cover}_q$ and $\tilde{v}_0 = v_0 \circ \mathrm{cover}_q$. Since $\tilde{u}_0$ (respectively $\tilde{v}_0$) is obtained from u (respectively v) by averaging over Γ_q, and the Finsler metric on $\mathcal{X}(\kappa)$ is Γ-invariant, we have estimates

$$\|\tilde{u}_0\|_{C^1} = \|u_0\|_{C^1} \leqslant \|u\|_{C^1}, \quad \|\tilde{v}_0\|_{C^1} = \|v_0\|_{C^1} \leqslant \|v\|_{C^1}$$

and hence also by the triangle inequality

$$\|u'\|_{C^1} \leqslant 2\|u\|_{C^1}, \quad \|v'\|_{C^1} \leqslant 2\|v\|_{C^1}.$$

Also note that $\int u' \, d\nu_{\mathcal{M}(q)} = \int v' \, d\nu_{\mathcal{M}(q)} = 0$. Moreover since the supports of u, v project to K in $\mathcal{M}$, the same holds for $\tilde{u}_0, \tilde{v}_0, u', v', u_0, v_0$. Since $\mathcal{T}_t$ preserves $L^2_\star(\mathcal{M}(q))$ and its orthogonal complement, we have

$$\int u.v \circ \mathcal{T}_t \, d\nu_{\mathcal{M}(q)} = \int \tilde{u}_0.\tilde{v}_0 \circ \mathcal{T}_t \, d\nu_{\mathcal{M}(q)} + \int u'.v' \circ \mathcal{T}_t \, d\nu_{\mathcal{M}(q)}.$$

We can replace the first term by

$$|\Gamma_q| \int u_0.v_0 \circ \mathcal{T}_t \, d\nu_{\mathcal{M}}$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

which by exponential mixing on $\mathcal{M}$ [AGY06, Theorem 2.14] is for some $\delta' > 0$

$$|\Gamma_q|\left(\int u_0\, d\nu_{\mathcal{M}}\right)\left(\int v_0\, d\nu_{\mathcal{M}}\right) + O_K(|\Gamma_q|\|u\|_{C^1}\|v\|_{C^1}e^{-\delta' t}).$$

Notice that since δ' depends only on $\mathcal{M}$ and, for some $\eta > 0$, $|\Gamma_q| \ll q^\eta$ for all q, the error term here is of the form as in Theorem 3.5. This also explains why the error term of Theorem 3.5 must contain a q^η factor.

Since $\int u_0\, d\nu_{\mathcal{M}} = |\Gamma_q|^{-1}\int \tilde{u}_0\, d\nu_{\mathcal{M}(q)}$ and similarly for $v_0, \tilde{v}_0$, we have by putting the previous arguments together

$$\int u.v \circ \mathcal{T}_t\, d\nu_{\mathcal{M}(q)} = |\Gamma_q|^{-1}\left(\int u\, d\nu_{\mathcal{M}(q)}\right)\left(\int v\, d\nu_{\mathcal{M}(q)}\right) + O_K(q^\eta\|u\|_{C^1}\|v\|_{C^1}e^{-\delta' t}) + \int u'.v' \circ \mathcal{T}_t\, d\nu_{\mathcal{M}(q)}.$$

This reduces Theorem 3.5 to the case of $u = u'$, $v = v' \in L^2_\star(\mathcal{M}(q))$. Now assume this is the case.

We apply Lemma 2.1 to obtain sections $u^*, v^* \in L^2(H_1(\mathcal{M}; \rho_q))$ that have the same C^1 norms as u and v. To apply Theorem 3.6 to the correlation function of u^* and v^* and conclude the proof, one needs to use the correspondence from § 2.12 to lift u^* and v^* to continuously differentiable $\mathcal{W}$-valued functions u^{**} and v^{**} on $\hat{\Xi}_r$. However, u^{**} and v^{**} may not have bounded C^1 norms, because of distortion between the Finsler metric structures on $\hat{\Xi}_r$ and $\mathcal{M}$. So one needs to perform some 'chopping' and 'smoothing' to conclude the result and it is at this stage that the condition on the support of u^* and v^* must be used. One may obtain estimates for L^p norms of u^* and v^* in terms of their C^1 norms and the compact set K. Once this is done, the rest of the argument is as in [AGY06, pp. 166–169]. It applies in the same way to vector-valued functions as to scalar-valued functions. □

3.3 Entrance of the transfer operator

We now recall the definition of the spaces $\mathcal{B}_0$ and $\mathcal{B}_1$ from [AGY06].

Definition 3.7. A function $U : \Xi_r \to \mathcal{W}$ is in $\mathcal{B}_0(\Xi_r; \rho)$ if it is bounded, continuously differentiable on each set

$$(\Xi_r)_{\gamma\gamma_0} := \{(y,t) : y \in \Xi_{\gamma\gamma_0}, t \in (0, r(y))\} \quad \gamma \text{ is } \gamma_0\text{-adapted}$$

and also $\sup_{(y,t)\in\bigcup^*(\Xi_r)_{\gamma\gamma_0}} \|DU(y,t)\| < \infty$. Define the norm

$$\|U\|_{\mathcal{B}_0(\Xi_r;\rho)} := \sup_{(y,t)\in\bigcup^*(\Xi_r)_{\gamma\gamma_0}} \|U(y,t)\| + \sup_{(y,t)\in\bigcup^*(\Xi_r)_{\gamma\gamma_0}} \|DU(y,t)\|.$$

Definition 3.8. A function $U : \Xi_r \to \mathcal{W}$ is in $\mathcal{B}_1(\Xi_r; \rho)$ if it is bounded and there exists a constant $C > 0$ such that for all fixed $y \in \cup^*\Xi_{\gamma\gamma_0}$, the function $t \mapsto U(y,t)$ is of bounded variation[19] on the interval $(0, r(y))$ and its variation $\mathrm{Var}_{(0,r(y))}(t \mapsto U(y,t))$ is bounded by $Cr(y)$. Let

$$\|U\|_{\mathcal{B}_1} = \sup_{(y,t)\in\bigcup^*(\Xi_r)_{\gamma\gamma_0}} \|U(y,t)\| + \sup_{y\in\bigcup^*(\Xi)_{\gamma\gamma_0}} \frac{\mathrm{Var}_{(0,r(y))}(t \mapsto U(y,t))}{r(y)}.$$

[19] We make the obvious extension of bounded variation to $\mathcal{W}$-valued functions using the norm induced by the inner product on the Hilbert space $\mathcal{W}$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

As in [AGY06] we reduce to decay of correlations for the ρ-skew extension of Ξ_r rather than $\hat{\Xi}_r$.

Theorem 3.9 (Decay of correlations). *There exists $C, \delta, \eta > 0$ and $Q_0 \in \mathbf{Z}_+$ such that for all q coprime to Q_0, for all $U \in \mathcal{B}_0(\Xi_r; \rho_q)$ and $V \in \mathcal{B}_1(\Xi; \rho_q)$, for all $t \geqslant 0$,*

$$\left| \int \langle U, T_t^* V \rangle \, dm_r \right| \leqslant C q^\eta \|U\|_{\mathcal{B}_0} \|V\|_{\mathcal{B}_1} e^{-\delta t}.$$

This is proved for scalar-valued functions in [AGY06, Theorem 7.3]. The key point of Theorem 3.9 is the uniformity in q. The passage from Theorem 3.9 to Theorem 3.6 is handled as in [AGY06, § 8]. In fact, the arguments of [AGY06, § 8] are followed closely and extended to the skew setting by Oh and Winter in [OW16, Proof of Theorem 1.5]. So we have presently explained the reduction of Theorem 3.5 to Theorem 3.9 whose proof we now take up.

From now on, all integrals are taken with respect to the relevant Lebesgue measure. Following [AGY06] let

$$A_t = \{(y, a) \in \Xi_r : a + t \geqslant r(y)\}$$

and $B_t = \Xi_r \backslash A_t$. We bound

$$\int_{B_t} \langle U, T_t^* V \rangle \leqslant \|U\|_{\mathcal{B}_0} \|V\|_{\mathcal{B}_1} \int_{y \in \Xi} \max(r(y) - t, 0) \leqslant \|U\|_{\mathcal{B}_0} \|V\|_{\mathcal{B}_1} \int_{y : r(y) \geqslant t} r(y).$$

By the Cauchy–Schwarz inequality and since r has exponential tails (Theorem 3.4) the above contributes $\leqslant C' \|U\|_{\mathcal{B}_0} \|V\|_{\mathcal{B}_1} \exp(-\delta' t)$ for some $\delta' > 0$ and $C' > 0$ that do not depend on U, V or ρ. Therefore the proof of Theorem 3.9 reduces to estimating the quantity

$$I(t) := \int_{A_t} \langle U, T_t^* V \rangle$$

on the order of

$$I(t) \leqslant C q^\eta \|U\|_{\mathcal{B}_0} \|V\|_{\mathcal{B}_1} \exp(-\delta t) \tag{16}$$

for some absolute constants $C, \delta, \eta > 0$.

We now begin the proof of (16). We will estimate the Laplace transform

$$\hat{I}(s) := \int_0^\infty \exp(-st) I(t) \, dt. \tag{17}$$

This is convergent for $\Re(s) > 0$ since I is bounded using the finiteness of m_r. The estimation of $\hat{I}_s(t)$ is closely related to certain skew transfer operators as follows. Using notation of [AGY06], if $F : \Xi_r \to \mathcal{W}$ and $s \in \mathbf{C}$, let

$$\hat{F}_s(y) := \int_0^{r(y)} F(y, \tau) \exp(-s\tau) \, d\tau.$$

Then following the proof of [AGY06, Lemma 7.17] and adapting to our ρ-skew setting we have

$$\begin{aligned} \hat{I}(s) &= \int_{y \in \Xi} \int_{\tau=0}^{r(y)} \int_{t+\tau \geqslant r(y)} e^{-st} \langle U(y, \tau), [T_t^* V](y, \tau) \rangle \, dt \, d\tau \, dy \\ &= \sum_{k=1}^\infty \int_{y \in \Xi} \int_{\tau=0}^{r(y)} \int_{\tau'=0}^{r(Z^k y)} e^{-s(r^{(k)}(y) + \tau' - \tau)} \langle U(y, \tau), [T^*_{r^{(k)}(y) + \tau' - \tau} V](y, \tau) \rangle \, d\tau' \, d\tau \, dy. \end{aligned} \tag{18}$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

The manipulation above follows from writing for each y, $t+\tau = r^{(k)}(y)+\tau'$ with $\tau' \in [0, r(Z^k x))$. For each y and t there is a unique k and τ' for which this is possible. Supposing more specifically that $y \in \Xi_{\gamma_1 \cdots \gamma_k \gamma_0}$ with each γ_i γ_0-adapted, we get from (15) that

$$[T^*_{r^{(k)}(y)+\tau'-\tau} V](y,\tau) = \rho(\Theta^*_{\gamma_1})\rho(\Theta^*_{\gamma_2}) \cdots \rho(\Theta^*_{\gamma_k}).V(Z^k(y),\tau'). \tag{19}$$

Inserting this into (18) gives that (throwing out a measure zero set)

$$\begin{aligned} \hat{I}(s) &= \sum_{k=1}^{\infty} \sum_{\gamma_1,\ldots,\gamma_k}^{*} \int_{y \in \Xi_{\gamma_1 \cdots \gamma_k \gamma_0}} \int_{\tau=0}^{r(y)} \int_{\tau'=0}^{r(Z^k y)} e^{-s(r^{(k)}(y)+\tau'-\tau)} \langle U(y,\tau), (19) \rangle \, d\tau' \, d\tau \, dy \\ &= \sum_{k=1}^{\infty} \sum_{\gamma_1,\ldots,\gamma_k}^{*} \int_{y \in \Xi_{\gamma_1 \cdots \gamma_k \gamma_0}} e^{-s r^{(k)}(y)} \langle \hat{U}_{-s}(y), \rho(\Theta^*_{\gamma_1})\rho(\Theta^*_{\gamma_2}) \cdots \rho(\Theta^*_{\gamma_k}) \hat{V}_s(Z^k(y)) \rangle \, dy. \end{aligned} \tag{20}$$

Here, we write a $\sum^*$ to indicate that the γ_i being summed over are all γ_0-adapted. The expression (20) is best understood by the *skew transfer operator* that we now introduce. Recall that $y \in \Xi$ can be written $y = (\pi, \lambda)$. The inverse branches of Z are indexed by γ_0-adapted γ and are given explicitly by

$$\alpha_\gamma : (\pi, \lambda) \mapsto \left(\pi, \frac{\Theta^*_\gamma \lambda}{\|\Theta^*_\gamma \lambda\|} \right), \quad \Xi \to \Xi_{\gamma\gamma_0}. \tag{21}$$

The skew transfer operator $\mathcal{L}_{s,\rho}$ is defined for arbitrary unitary $(\rho, \mathcal{W})$ and $f : \Xi \to \mathcal{W}$ by

$$\mathcal{L}_{s,\rho}[f](y) := \sum_{\gamma}^{*} e^{-sr \circ \alpha_\gamma(y)} J \circ \alpha_\gamma(y) \rho(\Theta^*_\gamma)^{-1}.f \circ \alpha_\gamma(y).$$

Recall that J is the inverse of the Jacobian of Z with respect to Lebesgue measure. By results of [AGY06] the summation involved in $\mathcal{L}_{s,\rho}$ is convergent (cf. Theorem 3.10 and the discussion afterwards). With the operator $\mathcal{L}_{s,\rho}$ in hand, by making a change of variables of the form $y \mapsto Z^k(y)$ one obtains from (20)

$$\hat{I}(s) = \sum_{k=1}^{\infty} \int_{y \in \Xi} \langle \mathcal{L}^k_{s,\rho}[\hat{U}_{-s}](y), \hat{V}_s(y) \rangle \, dy, \quad \Re(s) > 0. \tag{22}$$

It is clear from inspection of the above that spectral bounds for the operator $\mathcal{L}_{s,\rho}$ will be helpful in estimating $\hat{I}$. More precisely, we will aim to analytically continue $\hat{I}(s)$ to a strip $\Re(s) > -\sigma'$ with $\sigma' > 0$.

3.4 Spectral bounds for transfer operators

It will be useful at times to compare $\mathcal{L}_{s,\rho}$ to the operator on scalar functions on Ξ given by

$$\mathcal{L}_s[f](y) := \sum_{\gamma}^{*} e^{-sr \circ \alpha_\gamma(y)} J \circ \alpha_\gamma(y) f(\alpha_\gamma(y))$$

that features in [AGY06, formula (7.13)]. Recall that σ_0 is such that $\int \exp(\sigma_0 r) \, dm < \infty$ given by Theorem 3.4. The following is given in [AGY06, p. 188].

Theorem 3.10. *There is some $0 < \sigma_1 < \sigma_0$ such that for s with $|\Re(s)| < \sigma_1$, $\mathcal{L}_\sigma$ is a bounded operator on $C^1(\Xi)$. Moreover we have the following properties after suitable choice of σ_1.*

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

(i) *The transfer operator $\mathcal{L}_0$ has a simple eigenvalue at 1 and the rest of the spectrum of $\mathcal{L}_0$ is contained in a ball around 0 of radius <1.*

(ii) *For real σ with $|\sigma| < \sigma_1$ the largest eigenvalue λ_σ of $\mathcal{L}_\sigma$ is simple and varies real analytically in σ. In particular for all $\eta > 0$ there is $\sigma_2(\eta) > 0$ such that for real σ with $|\sigma| \leqslant \sigma_2$ we have $e^{-\eta} < \lambda_\sigma \leqslant e^{\eta}$.*

(iii) *The corresponding eigenfunctions h_σ (normalized so $\int h_\sigma = 1$) are positive and also vary real analytically as $C^1(\Xi)$-valued functions on $(-\sigma_1, \sigma_1)$. The functions h_σ are uniformly bounded below when $|\sigma| \leqslant \sigma_1$.*

As a corollary to Theorem 3.10 we may note that for real σ with $|\sigma| < \sigma_1$, the infinite sum

$$\sum_{\gamma}^{*} e^{-\sigma r \circ \alpha_\gamma(y)} J \circ \alpha_\gamma(y) = \mathcal{L}_\sigma[1](y)$$

converges to a C^1 function of $y \in \Xi$. Moreover (see [AGY06, paragraph following Proposition 7.8]) since for $\sigma < \sigma_1$, $\mathcal{L}_\sigma$ is a continuous perturbation of $\mathcal{L}_0$, by possibly decreasing σ_1, we can ensure the sum above is uniformly bounded for all $y \in \Sigma$ and all $\sigma \in (-\sigma_1, \sigma_1)$. This will be useful later.

We now give spectral estimates for $\mathcal{L}_{s,\rho}$ in two regimes: for large imaginary part of s (corresponding to high frequency aspects of the dynamics) and small (bounded) imaginary part of s (corresponding to low frequencies).

(a) $|\Im(s)| \gg 1$. Here we give spectral bounds for transfer operators $\mathcal{L}_{s,\rho}$, where ρ is an arbitrary unitary representation, that come from the method of Dolgopyat [Dol98]. In the case of scalar-valued functions on Ξ these bounds were obtained by Avila, Gouëzel and Yoccoz in [AGY06] by adapting Dolgopyat's argument to the Teichmüller setting.

To state the next result we introduce the warped norm on $C^1(\Xi; \rho)$ by

$$\|u\|_{1,t} = \sup_{y \in \Xi} \|u(y)\| + \frac{1}{\max(1, |t|)} \sup_{y \in \Xi} \|Du(y)\|.$$

Proposition 3.11. *There is $0 < \sigma_0' \leqslant \sigma_0$, $T_0 > 0$, $C > 0$ and $\beta < 1$ such that for all $s = \sigma + it$ with $|\sigma| \leqslant \sigma_0'$ and $|t| \geqslant T_0$, for any unitary $(\rho, \mathcal{W})$, $u \in C^1(\Xi; \rho)$ and for all $k \in \mathbf{N}$*

$$\|\mathcal{L}_{s,\rho}^k u\|_{L^2(\Xi)} \leqslant C\beta^k \|u\|_{1,t}.$$

The version of Proposition 3.11 with no twist by ρ can be found in [AGY06, Proposition 7.7]. We prove Proposition 3.11 in § 4.

(b) $|\Im(s)| \ll 1$. Here we give spectral bounds for $\mathcal{L}_{s,\rho_q}$ that are good when $|\Im(s)|$ is below a fixed constant.

Proposition 3.12. *Let $s = \sigma + it$. For all $t_0 > 0$ there are constants $C, \eta, \epsilon > 0$, $Q_0 \in \mathbf{Z}_+$, and $0 < \sigma_1' < \sigma_0$ such that when $|\sigma| < \sigma_1'$ and $|t| < t_0$ then for all $u \in C^1(\Xi; \rho_q)$, all $k \in \mathbf{N}$, all q coprime to Q_0, for $s = \sigma + it$,*

$$\|\mathcal{L}_{s,\rho_q}^k u\|_{C^1} \leqslant C(1-\epsilon)^k q^\eta \|u\|_{C^1}.$$

Proposition 3.12 is proved in § 5. Propositions 3.11 and 3.12 together with the expression (22) imply Theorem 3.9 by the arguments that we give now.

Proof of Theorem 3.9. In the first part of the argument we follow [AGY06, Lemma 7.18], and extend the argument to vector-valued functions. Recall we aim to prove (16). We write $s = \sigma + it$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Suppose $|\sigma| \leqslant \sigma_1/4$ for $\sigma_1 > 0$ as in Theorem 3.10. By integration by parts in the flow direction, we have for some $c > 0$ and all $x \in \Xi$

$$\|\hat{U}_{-s}(x)\|_{\mathcal{W}} \leqslant \frac{ce^{\sigma_1 r(x)/2}}{\max(1,|t|)}\|U\|_{\mathcal{B}_0}, \quad \|\hat{V}_s(x)\| \leqslant \frac{ce^{\sigma_1 r(x)/2}}{\max(1,|t|)}\|V\|_{\mathcal{B}_1}. \tag{23}$$

We can estimate

$$\|\mathcal{L}_{s,\rho}\hat{U}_{-s}(x)\|_{\mathcal{W}} \leqslant \sum_{\gamma}^{*} |e^{-sr\circ\alpha_\gamma(x)} J\circ\alpha_\gamma(x)| \|\hat{U}_{-s}(\alpha_\gamma x)\|_{\mathcal{W}}.$$

We were able to remove the factors here coming from ρ since the representation is unitary. By the estimate for $\|\hat{U}_{-s}(x)\|_{\mathcal{W}}$ in (23), this is

$$\leqslant \frac{c\|U\|_{\mathcal{B}_0}}{\max(1,|t|)} \sum_{\gamma}^{*} e^{(\sigma_1/4)r\circ\alpha_\gamma(x)} J\circ\alpha_\gamma(x) e^{(\sigma_1/2)r\circ\alpha_\gamma(x)}.$$

The sum is bounded by a constant $c' > 0$ given $|\sigma| \leqslant \sigma_1/4$ and Theorem 3.10. Hence by increasing c if necessary

$$\|\mathcal{L}_{s,\rho}\hat{U}_{-s}(x)\|_{\mathcal{W}} \leqslant \frac{c\|U\|_{\mathcal{B}_0}}{\max(1,|t|)}$$

for all $x \in \Xi$. We have (recalling footnote 3)

$$\|D(\mathcal{L}_{s,\rho}\hat{U}_{-s})(x)\| = \left\| \sum_{\gamma}^{*} D[e^{-sr\circ\alpha_\gamma} J\circ\alpha_\gamma \rho(\Theta_\gamma^*)^{-1}.\hat{U}_{-s}\circ\alpha_\gamma](x) \right\|.$$

We have to differentiate $e^{-sr\circ\alpha_\gamma}$, $J\circ\alpha_\gamma$, the limits of the integral defining $\hat{U}_{-s}$, or $\rho(\Theta_\gamma^*)^{-1}.\hat{U}_{-s}\circ\alpha_\gamma$. The latter is the only deviation from [AGY06, Lemma 7.18]. Since ρ is locally constant, we have

$$\|D[\rho(\Theta_\gamma^*)^{-1}.\hat{U}_{-s}\circ\alpha_\gamma](x)\| = \|\rho(\Theta_\gamma^*)^{-1} D[\hat{U}_{-s}\circ\alpha_\gamma](x)\| = \|D[\hat{U}_{-s}\circ\alpha_\gamma](x)\|.$$

Therefore, since this is the same estimate [AGY06] obtain for the analogous term, the same arguments as in [AGY06, Lemma 7.18] imply $\|D(\mathcal{L}_{s,\rho}\hat{U}_{-s})(x)\| \leqslant c$.

Hence putting the previous estimates together, we have $\mathcal{L}_{s,\rho}\hat{U}_{-s} \in C^1(\Xi;\rho)$ with

$$\|\mathcal{L}_{s,\rho}\hat{U}_{-s}\|_{C^1(\Xi;\rho)} \leqslant c\|U\|_{\mathcal{B}_0}, \quad \|\mathcal{L}_{s,\rho}\hat{U}_{-s}\|_{1,t} \leqslant \frac{c\|U\|_{\mathcal{B}_0}}{\max(1,|t|)}. \tag{24}$$

As a clarifying remark, we would have liked to obtain these bounds for $\hat{U}_{-s}$, but it was not possible, so we used $\mathcal{L}_{s,\rho}\hat{U}_{-s}$ instead. We also have from the bound for $\|\hat{V}_s\|_\infty$ from (23),

$$\|\hat{V}_s\|^2_{L^2(\Xi)} \leqslant \int_{x\in\Xi} \|\hat{V}_s(x)\|^2\, dm(x) \leqslant \frac{c^2\|V\|^2_{\mathcal{B}_1}}{\max(1,|t|)^2} \int_{x\in\Xi} e^{\sigma_1 r(x)}\, dm(x) \leqslant \frac{c'\|V\|^2_{\mathcal{B}_1}}{\max(1,|t|)^2}. \tag{25}$$

These are all the functional norm bounds we need for the proof. We now proceed to use the spectral bounds for the transfer operator.

Let $\sigma'' = \min(\sigma_1/4, \sigma_0', \sigma_1')$ where σ_0' and σ_1' are the constants from Propositions 3.11 and 3.12. We now specialize ρ to $\rho = \rho_q$. Writing $\mathcal{L}^k_{s,\rho} = \mathcal{L}^{k-1}_{s,\rho}\mathcal{L}_{s,\rho}$ we obtain from Proposition 3.11 and (24) that for $|\sigma| \leqslant \sigma''$, if $|t| \geqslant T_0$,

$$\|\mathcal{L}^k_{s,\rho}\hat{U}_{-s}\|_{L^2(\Xi)} \leqslant C\beta^{k-1}\|\mathcal{L}_{s,\rho}\hat{U}_{-s}\|_{1,t} \leqslant \frac{C'\beta^k}{\max(1,|t|)}.$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Using (25) and the Cauchy–Schwarz inequality, we can bound the terms defining $\hat{I}$ in (22) by

$$|\langle \mathcal{L}^k_{s,\rho}[\hat{U}_{-s}](y), \hat{V}_s(y)\rangle| \leqslant \frac{c\beta^k \|U\|_{\mathcal{B}_0}\|V\|_{\mathcal{B}_1}}{t^2}$$

for $|t| \geqslant T_0$ and some $c > 0$. Hence for $|t| \geqslant T_0$ we have

$$\hat{I}(s) \leqslant \sum_{k=0}^{\infty} \frac{c\beta^k \|U\|_{\mathcal{B}_0}\|V\|_{\mathcal{B}_1}}{t^2} \leqslant \frac{c'\|U\|_{\mathcal{B}_0}\|V\|_{\mathcal{B}_1}}{t^2}.$$

For $|t| < T_0$ we apply Proposition 3.12 with $t_0 = T_0$ and $u = \mathcal{L}_{s,\rho_q}\hat{U}_{-s}$ to obtain

$$\|\mathcal{L}^k_{s,\rho_q}\hat{U}_{-s}\|_{C^1} \leqslant C(1-\epsilon)^{k-1}q^\eta \|\mathcal{L}_{s,\rho_q}\hat{U}_{-s}u\|_{C^1} \leqslant C'(1-\epsilon)^k q^\eta \|U\|_{\mathcal{B}_0},$$

where the last inequality used (24). Hence for $|t| < T_0$, using the Cauchy–Schwarz inequality again to bound $|\langle \mathcal{L}^k_{s,\rho}[\hat{U}_{-s}](y), \hat{V}_s(y)\rangle|$,

$$\hat{I}(s) \leqslant \sum_{k=0}^{\infty} C'(1-\epsilon)^k q^\eta \|U\|_{\mathcal{B}_0}\|V\|_{\mathcal{B}_1} \leqslant c' q^\eta \|U\|_{\mathcal{B}_0}\|V\|_{\mathcal{B}_1}.$$

These estimates prove that the expression defining $\hat{I}(s)$ is absolutely uniformly convergent on compact sets in $|\sigma| \leqslant \sigma''$. Since each of the terms are analytic, this establishes analytic continuation of $\hat{I}(s)$ to $\Re(s) > -\sigma''$. Since we have established the estimate

$$\hat{I}(s) \leqslant \frac{c' q^\eta}{1+|t|^2}\|U\|_{\mathcal{B}_0}\|V\|_{\mathcal{B}_1}, \quad |\sigma| \leqslant \sigma'',$$

by inverting the Laplace transform, using a contour integral over the vertical line $\Re(s) = -\sigma''/2$ as in [OW16, Proposition 5.5], we obtain for some $\delta > 0$

$$I(t) \leqslant c'' q^\eta e^{-\delta t}\|U\|_{\mathcal{B}_0}\|V\|_{\mathcal{B}_1}.$$

This completes the proof of Theorem 3.9. □

Now, the only outstanding proofs required for Theorem 3.5 are those of Propositions 3.11 and 3.12. These are given in §§ 4 and 5 respectively.

4. The Dolgopyat argument for twisted transfer operators

In this section we explain the necessary modifications to [AGY06, § 7] in order to prove Proposition 3.11.

The key idea of the proof, due to Dolgopyat, is to systematically exploit oscillations of the roof function r. As illustrated in the work of Oh and Winter [OW16] and Magee, Oh and Winter [MOW19], Dolgopyat's argument works for skew transfer operators, provided the twisting unitary cocycle is constant on cylinders of length 1. The reason is that because the cocycle is locally constant, it should not interfere with the oscillations of r during the argument, which is what is being exploited. In the current setting, the values of the cocycle $\rho(\Theta_\gamma)$ only depend on the cylinder $\Xi_{\gamma.\gamma_0}$, so the same arguments should in principle apply. We establish this rigorously below.

It will be useful to make the following normalization of the transfer operator as in [AGY06]. Recall from Theorem 3.10 that for σ real with $|\sigma| \leqslant \sigma_1$, λ_σ is the leading eigenvalue of $\mathcal{L}_\sigma$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

and h_σ the corresponding positive eigenfunction. The h_σ are uniformly bounded below. We write $s = \sigma + it$ throughout this section, assume $|\sigma| \leqslant \sigma_1$, and define

$$\mathbf{L}_{s,\rho}[f] := \lambda_\sigma^{-1} h_\sigma^{-1} \mathcal{L}_{s,\rho}[h_\sigma f]$$

for $f \in C^1(\Xi;\rho)$ and similarly $\mathbf{L}_s[f] := \lambda_\sigma^{-1} h_\sigma^{-1} \mathcal{L}_s[h_\sigma f]$ for $f \in C^1(\Xi;\mathbf{C})$. The purpose of this normalization is that for real σ, $\mathbf{L}_\sigma[1] = [1]$, i.e., $\mathbf{L}_\sigma$ is a Markoff operator. The operator $\mathbf{L}_{s,\rho}$ acts on $C^1(\Xi;\rho)$ by

$$\mathbf{L}_{s,\rho}[f](y) := \sum_\gamma{}^{*} e^{R_s(\alpha_\gamma y)} \rho(\Theta_\gamma^*)^{-1} f(\alpha_\gamma(y)),$$

where

$$R_s := -sr + \log J - \log(h_\sigma \circ Z) + \log h_\sigma - \log \lambda_\sigma. \tag{26}$$

Many extensions of the results in [AGY06, § 7] rely on two simple observations. Let $(\rho, \mathcal{W})$ be a unitary representation. When we consider $\mathbf{L}_{s,\rho}^n[f](x)$ we obtain sums of terms containing factors $\rho(\Theta_{\gamma_n}^*)^{-1} \cdots \rho(\Theta_{\gamma_1}^*)^{-1} f(\alpha_{\gamma_1 \cdots \gamma_n} x)$ (cf. (41) below). The two observations that we will use several times are the following.

(i) Since $(\rho, \mathcal{W})$ is assumed to be unitary, any time we apply the triangle inequality to $\mathbf{L}_{s,\rho}^n f$, or expressions derived from $\mathbf{L}_{s,\rho}^n f$ (for example, by taking a derivative), we can use $\|\rho(\Theta_{\gamma_n}^*)^{-1} \cdots \rho(\Theta_{\gamma_1}^*)^{-1} f(y)\| \leqslant \|f(y)\|$.

(ii) If we take a derivative of $\mathbf{L}_{s,\rho}^n f$, since the factors $\rho(\Theta_{\gamma_n}^*)^{-1} \cdots \rho(\Theta_{\gamma_1}^*)^{-1}$ are constant for each $\gamma_1, \ldots, \gamma_n$, we may commute the operator D with $\rho(\Theta_{\gamma_n}^*)^{-1} \cdots \rho(\Theta_{\gamma_1}^*)^{-1}$. Then the previous point may be used.

The following is the extension of [AGY06, Lemma 7.8] to vector-valued functions. Let σ_1 be as in Theorem 3.10. Recall $\Lambda > 1$ is the constant from Proposition 3.1.

Lemma 4.1. *There is $\mathcal{K} > 0$ such that for all $n \geqslant 1$, for all $s = \sigma + it$ with $\sigma \in [-\sigma_1, \sigma_1]$, $t \in \mathbf{R}$, for all unitary representations $(\rho, \mathcal{W})$, for all $u \in C^1(\Xi;\rho)$, and for all $x \in \Xi$,*

$$\|D[\mathbf{L}_{s,\rho}^n u](x)\| \leqslant \mathcal{K}(1+|t|)\mathbf{L}_\sigma^n[\|u\|](x) + \Lambda^{-n}\mathbf{L}_\sigma^n[\|Du\|](x).$$

Proof. The proof is the same as that of [AGY06, Lemma 7.8] with the addition of points (i) and (ii) above. □

We now fix $\mathcal{K} > 5$ satisfying Lemma 4.1. The next lemma is the extension of [AGY06, Lemma 7.9] that shows the iterates of $\mathbf{L}_{s,\rho}$ are bounded in the operator norm of $\|\bullet\|_{1,t}$.

Lemma 4.2. *There is $C > 1$ such that for all $s = \sigma + it$ with $\sigma \in [-\sigma_1, \sigma_1]$ and $|t| \geqslant 10$, for all $k \in \mathbf{N}$, for all unitary $(\rho, \mathcal{W})$, and for all $u \in C^1(\Xi;\rho)$*

$$\|\mathbf{L}_{s,\rho}^k u\|_{1,t} \leqslant C\|u\|_{C^0} + \frac{\Lambda^{-k}}{|t|}\|Du\|_{C^0}.$$

Therefore $\|\mathbf{L}_{s,\rho}^k u\|_{1,t} \leqslant C\|u\|_{1,t}$.

Proof. Again, the proof is a straightforward extension of [AGY06, Lemma 7.9] incorporating point (i) and Lemma 4.1 in place of [AGY06, Lemma 7.8]. □

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Next we note the extension of [AGY06, Lemma 7.10] to vector-valued functions.

Lemma 4.3. *There is $N_0 \in \mathbf{N}$ such that for any $n \geqslant N_0$ the following hold. Let $s = \sigma + it$ with $\sigma \in [-\sigma_1, \sigma_1]$ and $|t| \geqslant 10$. Let $(\rho, \mathcal{W})$ be a unitary representation. Let $v \in C^1(\Xi; \rho)$ satisfy*

$$\sup \|Dv\| \geqslant 2\mathcal{K}|t| \sup \|v\|.$$

Then

$$\|\mathbf{L}^n_{s,\rho} v\|_{1,t} \leqslant \tfrac{9}{10}\|v\|_{1,t}.$$

Proof. Follow the proof of [AGY06, Lemma 7.10] and use point (i) and the replacement of [AGY06, Lemma 7.8] by Lemma 4.1. □

This tells us that to establish contraction of $\mathbf{L}^n_{s,\rho}$ it remains to deal with functions with $\sup \|Dv\| \leqslant 2\mathcal{K}|t| \sup \|v\|$. Now we make the following natural modification to [AGY06, Definition 7.11].

Definition 4.4. For $t \in \mathbf{R}$ and $(\rho, \mathcal{W})$ a unitary representation, we say a pair of functions (u, v) on Ξ is in $\mathcal{E}^{\mathcal{W}}_t$ if $u : \Xi \to \mathbf{R}_+$ is C^1, $v : \Xi \to \mathcal{W}$ is C^1, $0 \leqslant \|v\| \leqslant u$, and for all $x \in \Xi$,

$$\max(\|Du(x)\|, \|Dv(x)\|) \leqslant 2\mathcal{K}|t|u(x).$$

The next lemma is the current analog of [AGY06, Lemma 7.12].

Lemma 4.5. *There exists $N_1 \in \mathbf{N}$ such that for any $n \geqslant N_1$ the following hold. Let $s = \sigma + it$ with $\sigma \in [-\sigma_1, \sigma_1]$ and $|t| \geqslant 10$. Let $(\rho, \mathcal{W})$ be unitary and $(u, v) \in \mathcal{E}^{\mathcal{W}}_t$. Let $\chi \in C^1(\Xi)$ with $\|D\chi\| \leqslant |t|$ and $\frac{3}{4} \leqslant \chi \leqslant 1$. Assume for all $x \in \Xi$, $\|\mathbf{L}^n_{s,\rho} v(x)\| \leqslant \mathbf{L}^n_\sigma(\chi u)(x)$. Then*

$$(\mathbf{L}^n_\sigma(\chi u), \mathbf{L}^n_{s,\rho} v) \in \mathcal{E}^{\mathcal{W}}_t.$$

Proof. The proof is the same as that of [AGY06, Lemma 7.12] after replacing all uses of [AGY06, Lemma 7.8] by Lemma 4.1. □

By the arguments of [AGY06, p. 191], there exists $n \geqslant \max(N_0, N_1)$, α_1, α_2 inverse branches of Z^n, and a smooth vector field y on Ξ with $1 \leqslant \|y\| \leqslant 2$, such that for all $x \in \Xi$,

$$\|D[r^{(n)} \circ \alpha_1](x)y(x) - D[r^{(n)} \circ \alpha_2](x)y(x)\| \geqslant 100\mathcal{K} \max(\|D\alpha_1(x)y(x)\|, \|D\alpha_2(x)y(x)\|). \quad (27)$$

Note we replaced the constant 9 from [AGY06] by 100, for technical reasons. This is permissible by the same arguments leading to the constant 9 in [AGY06] (9 was arbitrary). The estimate (27) lies at the very core of the Dolgopyat argument.

We now fix this n, α_1, α_2, and y throughout the rest of this section.

For $\gamma_1, \ldots, \gamma_k$ γ_0-adapted let $\alpha_{\gamma_1 \cdots \gamma_k}$ denote the inverse branch of Z^k that maps Ξ to $\Xi_{\gamma_1 \cdots \gamma_k \gamma_0}$. Then recalling the previously defined α_γ from (21) one has the composition law

$$\alpha_{\gamma_1 \cdots \gamma_k} = \alpha_{\gamma_1} \circ \alpha_{\gamma_2} \circ \cdots \circ \alpha_{\gamma_k}.$$

If we write for each of α_1, α_2, $\alpha_j = \alpha_{\gamma_1^{(j)} \cdots \gamma_n^{(j)}}$, then let us define

$$\Theta_1 := (\Theta^*_{\gamma_n^{(1)}})^{-1} \cdots (\Theta^*_{\gamma_1^{(1)}})^{-1}, \quad \Theta_2 := (\Theta^*_{\gamma_n^{(2)}})^{-1} \cdots (\Theta^*_{\gamma_1^{(2)}})^{-1}.$$

Then (and this is the reason for the definition), $\rho(\Theta_i)$ is the unitary matrix appearing in the summand of $\mathbf{L}^n_{s,\rho}$ corresponding to the inverse branch α_i.

The following lemma, analogous to [AGY06, Lemma 7.13], is the main point where a new idea is needed to extend the methods of [AGY06, § 7] to vector-valued functions. Recall the notation $r^{(n)}$ from § 2.12 and the functions h_σ from Theorem 3.10.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Lemma 4.6. *There are constants $\delta > 0$ and $\zeta > 0$ such that the following hold. Let $s = \sigma + it$ with $\sigma \in [-\sigma_1, \sigma_1]$ and $|t| \geqslant 10$. For any unitary $(\rho, \mathcal{W})$ let $(u, v) \in \mathcal{E}_t^{\mathcal{W}}$. For all $x_0 \in \Xi$ such that $B(x_0, (\zeta + \delta)/|t|)$ is compactly included in Ξ, there exists a point x_1 with $d(x_0, x_1) \leqslant \zeta/|t|$ such that one of the following holds:*

– *either, for all $x \in B(x_1, \delta/|t|)$*

$$\begin{aligned}&\|e^{-sr^{(n)}\circ\alpha_1(x)}J(\alpha_1 x)\rho(\Theta_1)[v.h_\sigma](\alpha_1 x) + e^{-sr^{(n)}\circ\alpha_2(x)}J(\alpha_2 x)\rho(\Theta_2)[v.h_\sigma](\alpha_2 x)\|\\ &\quad\leqslant \tfrac{3}{4}e^{-\sigma r^{(n)}\circ\alpha_1(x)}J(\alpha_1 x)[u.h_\sigma](\alpha_1 x) + e^{-\sigma r^{(n)}\circ\alpha_2(x)}J(\alpha_2 x)[u.h_\sigma](\alpha_2 x);\end{aligned}$$

– *or, for all $x \in B(x_1, \delta/|t|)$*

$$\begin{aligned}&\|e^{-sr^{(n)}\circ\alpha_1(x)}J(\alpha_1 x)\rho(\Theta_1)[v.h_\sigma](\alpha_1 x) + e^{-sr^{(n)}\circ\alpha_2(x)}J(\alpha_2 x)\rho(\Theta_2)[v.h_\sigma](\alpha_2 x)\|\\ &\quad\leqslant e^{-\sigma r^{(n)}\circ\alpha_1(x)}J(\alpha_1 x)[u.h_\sigma](\alpha_1 x) + \tfrac{3}{4}e^{-\sigma r^{(n)}\circ\alpha_2(x)}J(\alpha_2 x)[u.h_\sigma](\alpha_2 x).\end{aligned}$$

Proof. We follow [AGY06, Proof of Lemma 7.13] and split into two cases.

Case 1. Assume there is $x_1 \in B(x_0, \zeta/|t|)$ such that either $\|v \circ \alpha_1(x_1)\| \leqslant u \circ \alpha_1(x_1)/2$ or $\|v \circ \alpha_2(x_1)\| \leqslant u \circ \alpha_2(x_1)/2$. The same arguments as in [AGY06], incorporating point (i) above, prove that the lemma holds in this case by choosing δ sufficiently small.

The harder case is the alternative one, wherein we must extend the arguments of Magee, Oh, and Winter [MOW19, Proof of Lemma 29] to higher dimensions.

Case 2. Assume for all $x \in B(x_0, \zeta/|t|)$, $\|v \circ \alpha_1(x_1)\| > u \circ \alpha_1(x_1)/2$ and $\|v \circ \alpha_2(x_1)\| > u \circ \alpha_2(x_1)/2$. This implies on $\alpha_1(B(x_0, \zeta/|t|)) \cup \alpha_2(B(x_0, \zeta/|t|))$ the function v is non-vanishing.

As in [AGY06] let $\phi : [0, \zeta/(2|t|)) \to \Xi$ be the solution of the differential equation $\phi'(\tau) = y(\phi(\tau))$ with $\phi(0) = x_0$. Let $x^\tau := \phi(\tau)$. Define for $x \in \Xi$

$$\begin{aligned}F_1(x) &:= e^{-sr^{(n)}\circ\alpha_1(x)}J(\alpha_1 x)\rho(\Theta_1)[v.h_\sigma](\alpha_1 x),\\ F_2(x) &:= e^{-sr^{(n)}\circ\alpha_2(x)}J(\alpha_2 x)\rho(\Theta_2)[v.h_\sigma](\alpha_2 x).\end{aligned}$$

Our goal is to find cancelation between these two functions. We must follow a slightly different approach to [AGY06] because we do not have exactly the same concept of 'phase'. Instead we consider the complex-valued function

$$\Phi(x) := \frac{\langle F_1(x), F_2(x)\rangle}{\|F_1(x)\|\|F_2(x)\|}.$$

Our strategy of proof will be to establish the following claim.

Claim. *There is a choice of $\zeta > 0$ such that for any x_0 as above, there is $\tau \in [0, \zeta/(8|t|))$ such that $\Re\Phi(x^\tau) \leqslant \frac{1}{8}$.*

Before proving the claim, let us see how it implies Lemma 4.6. Indeed, given τ as in the claim, we let $x_1 = x^\tau$. We have $x^\tau \in B(x_0, \zeta/(4|t|))$. We need to argue as we perturb from x^τ in any direction, Φ does not change too much.

First we control the sizes of F_1 and F_2 and their rates of change. Following [AGY06, p. 193], one obtains a constant $C > 0$, independent of δ, such that for all $x \in B(x_0, \zeta/|t|)$ we have

$$\|DF_i(x)\| \leqslant C|t|\|F_i(x)\| \tag{28}$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

from which it follows from Gronwall's inequality that for all x, $x' \in B(x_0, \zeta/(3|t|))$, we have

$$\|F_i(x')\| \leqslant e^{C|t|\, d(x,x')}\|F_i(x)\|. \tag{29}$$

Note from (29) it follows that

$$|D\|F_i\|| \leqslant C|t|\|F_i\| \tag{30}$$

on the domain $B(x_0, \zeta/(3|t|))$.

Next we have that $D\langle F_1, F_2\rangle = \langle DF_1, F_2\rangle + \langle F_1, DF_2\rangle$ so by the Schwarz inequality and (28)

$$|D\langle F_1, F_2\rangle| \leqslant \|DF_1\|\|F_2\| + \|F_1\|\|DF_2\| \leqslant 2C|t|\|F_1\|\|F_2\|. \tag{31}$$

Now,

$$D\Phi = \frac{D\langle F_1, F_2\rangle}{\|F_1\|\|F_2\|} - \frac{\langle F_1, F_2\rangle D\|F_1\|}{\|F_1\|^2\|F_2\|} - \frac{\langle F_1, F_2\rangle D\|F_2\|}{\|F_2\|\|F_2\|}.$$

By using (30), (31), the Schwarz inequality and the triangle inequality we obtain

$$|D\Phi| \leqslant 2C|t| + C|t| + C|t| = 4C|t|$$

on $B(x_0, \zeta/(3|t|))$. Therefore, if δ is small enough with $\delta < \zeta/12$, for all $x \in B(x_1, \delta/|t|)$ we have $\Re\Phi(x) \leqslant \frac{1}{4}$.

Assume that $\|F_1(x_1)\| \geqslant \|F_2(x_1)\|$. This is without loss of generality since the other case is symmetrical. Then by choosing δ small enough, and using (29), we may further assume that for all $x \in B(x_1, \delta/|t|)$, $\|F_1(x)\| \geqslant \|F_2(x)\|/2$. This implies for all $x \in B(x_1, \delta/|t|)$

$$\begin{aligned}\|F_1(x) + F_2(x)\|^2 &= \|F_1(x)\|^2 + \|F_2(x)\|^2 + 2\Re\langle F_1(x), F_2(x)\rangle \\ &= \|F_1(x)\|^2 + \|F_2(x)\|^2 + 2\Re\Phi(x)\|F_1\|\|F_2\| \\ &\leqslant \|F_1(x)\|^2 + \|F_2(x)\|^2 + \tfrac{1}{2}\|F_1\|\|F_2\| \\ &= (\|F_1(x)\| + \tfrac{3}{4}\|F_2(x)\|)^2 + \|F_2(x)\|(\tfrac{7}{16}\|F_2(x)\| - \|F_1(x)\|)\end{aligned}$$

which is $\leqslant(\|F_1(x)\|+\frac{3}{4}\|F_2(x)\|)^2$ for all $x \in B(x_1, \delta/|t|)$ since $\|F_1\| \geqslant \|F_2\|/2$ there. *This concludes the proof of Lemma 4.6 modulo the proof of our claim.*

Now we prove the claim. Note first that we can assume we have $|\Phi(x^\tau)| > 1/8$ for all $\tau \leqslant \zeta/(8|t|)$, otherwise the claim is established.

For $i = 1, 2$, let $s_i(x) := v \circ \alpha_i(x)$. Since v is non-vanishing on $B(x_0, \zeta/|t|)$ we can write $s_i = \|s_i\|s_i^*$ where $\|s_i\|$ and s_i^* are continuously differentiable. Then we have the expression

$$\Phi(x^\tau) = e^{-itr^{(n)}\circ\alpha_1(x^\tau)}\overline{e^{-itr^{(n)}\circ\alpha_2(x^\tau)}}\langle\rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\rangle.$$

Note that $|\Phi(x^\tau)| > 1/8$ for all $\tau \leqslant \zeta/(8|t|)$ implies $|\langle\rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\rangle| > 1/8$ for τ in the same range.

We have for $i = 1, 2$ and any vector field $\mathbf{w}$

$$D_{\mathbf{w}}s_i = (D_{\mathbf{w}}\|s_i\|)s_i^* + \|s_i\|(D_{\mathbf{w}}s_i^*)$$

and from $\langle s_i^*, s_i^*\rangle_{\mathcal{W}} = 1$ we obtain $\langle D_{\mathbf{w}}s_i^*, s_i^*\rangle_{\mathcal{W}} = 0$, so we have

$$\|D_{\mathbf{w}}s_i\|^2 = (D_{\mathbf{w}}\|s_i\|)^2 + \|s_i\|^2\|D_{\mathbf{w}}s_i^*\|^2.$$

This yields

$$\|D_{\mathbf{w}}s_i^*\| \leqslant \frac{\|D_{\mathbf{w}}s_i\|}{\|s_i\|} \tag{32}$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

which is a version of 'Kato's inequality'. The numerator on the right-hand side is estimated using $\|Dv\| \leqslant 2\mathcal{K}|t|u$ giving

$$\|D_{\mathbf{w}} s_i\|(x) \leqslant 2\mathcal{K}|t|u(x)\|D\alpha_i(x).\mathbf{w}(x)\|. \tag{33}$$

The denominator is estimated using the assumptions of the current case giving $\|s_i\| \geqslant (u \circ \alpha_i)/2$. Together this gives

$$\|D_{\mathbf{w}} s_i^*(x)\| \leqslant 4\mathcal{K}|t|\|D\alpha_i(x).\mathbf{w}(x)\|. \tag{34}$$

Consider now the function

$$\Upsilon(\tau) := \frac{\langle \rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\rangle}{|\langle \rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\rangle|}.$$

Since h_σ is always non-zero, we can locally write $\Phi(x^\tau) = e^{i\theta(\tau)}\|\Phi(\tau)\|$. Similarly, for some function $\arg\Upsilon$ we can write $\Upsilon(\tau) = \exp(i \arg\Upsilon(\tau))$. Then

$$\theta(\tau) = -t(r^{(n)} \circ \alpha_1(x^\tau) - r^{(n)} \circ \alpha_2(x^\tau)) + \arg\Upsilon(\tau). \tag{35}$$

By the same arguments as led to (32) we have

$$|\Upsilon'(\tau)| \leqslant \frac{|(d/d\tau)\langle \rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\rangle|}{|\langle \rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\rangle|}. \tag{36}$$

Using (34) and the triangle and Schwarz inequalities gives

$$\begin{aligned}
&\left|\frac{d}{d\tau}\langle \rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\rangle\right| \\
&\quad = \left|\left\langle \rho(\Theta_1)\frac{d}{d\tau}s_1^*(x^\tau), \rho(\Theta_2)s_2^*(x^\tau)\right\rangle + \left\langle \rho(\Theta_1)s_1^*(x^\tau), \rho(\Theta_2)\frac{d}{d\tau}s_2^*(x^\tau)\right\rangle\right| \\
&\quad \leqslant 4\mathcal{K}|t|(\|D\alpha_1(x).y(x)\| + \|D\alpha_2(x).y(x)\|).
\end{aligned}$$

This bounds the numerator of (36). The denominator is $>1/8$ by our current assumptions. Hence

$$|\Upsilon'(\tau)| \leqslant 64\mathcal{K}|t| \max_{i=1,2}(\|D\alpha_i(x).y(x)\|). \tag{37}$$

The inequality (37) implies, since the values of Υ have absolute value one,

$$|(\arg\Upsilon)'(\tau)| \leqslant 64\mathcal{K}|t| \max_{i=1,2}(\|D\alpha_i(x).y(x)\|). \tag{38}$$

Therefore using (27) and (38) together with (35) we obtain

$$\begin{aligned}
|\theta'(\tau)| &\geqslant 100\mathcal{K}|t| \max_{i=1,2}(\|D\alpha_i(x).y(x)\|) - 64\mathcal{K}|t| \max_{i=1,2}(\|D\alpha_i(x).y(x)\|) \\
&= 36\mathcal{K}|t| \max_{i=1,2}(\|D\alpha_i(x).y(x)\|).
\end{aligned}$$

Following [AGY06] there is a constant $\gamma_0 > 0$ such that for all $x \in \Xi$ we have $\max_{i=1,2}(\|D\alpha_i(x).y(x)\|) \geqslant \gamma_0$. Hence $|\theta'(\tau)| \geqslant c|t|$ for some $c > 0$. We now choose $\zeta = 16\pi/c$ so that there will be $\tau \in [0, \zeta/(8|t|)]$ with $\theta = -\pi \bmod 2\pi$. Note this choice of ζ only depends on constants defined before this proof so could have been made *a priori*. This gives $\Phi(x^\tau) \in \mathbf{R}$ and $\Phi(x^\tau) \leqslant -1/8$. *This proves the claim.* □

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

We now fix the constants ζ and δ given by Lemma 4.6. In [AGY06, p. 194] it is explained that there are constants C_0 and ϵ_0 such that for all $\epsilon < \epsilon_0$, for all $x \in \Xi$, there exists $x' \in \Xi$ such that $d(x, x') \leqslant C_0\epsilon$ and $B(x', \epsilon)$ is compactly included in Ξ. We now choose $T_0 \geqslant 10$ such that $2(\zeta + \delta)/T_0 < \epsilon_0$.

The following lemma is the replacement of [AGY06, Lemma 7.15].

Lemma 4.7. *There exist $\beta_0 < 1$ and $0 < \sigma_2 < \sigma_1$ such that the following hold. Let $s = \sigma + it$ with $\sigma \in [-\sigma_2, \sigma_2]$ and $|t| \geqslant T_0$. Let $(\rho, \mathcal{W})$ be unitary and $(u, v) \in \mathcal{E}_t^{\mathcal{W}}$. Then there exists $\tilde{u} : \Xi \to \mathbf{R}$ such that $(\tilde{u}, \mathbf{L}_{s,\rho}^n v) \in \mathcal{E}_t^{\mathcal{W}}$ and $\int \tilde{u}^2\, dm \leqslant \beta_0 \int u^2\, dm$.*

Proof. Following the same arguments of [AGY06, Lemma 7.15] with Lemma 4.6 in place of [AGY06, Lemma 7.13] we construct a function χ on Ξ with $3/4 \leqslant \chi \leqslant 1$, $\|D\chi\| \leqslant |t|$, and

$$\|\mathbf{L}_{s,\rho}^n v\| \leqslant \mathbf{L}_\sigma^n(\chi u). \tag{39}$$

We let $\tilde{u} = \mathbf{L}_\sigma^n(\chi u)$. By (39) combined with Lemma 4.5 we obtain $(\tilde{u}, \mathbf{L}_{s,\rho}^n v) \in \mathcal{E}_t^{\mathcal{W}}$. It remains to show for some $\beta_0 < 1$, $\int \tilde{u}^2\, dm \leqslant \beta_0 \int u^2\, dm$ when σ is sufficiently small. This can be done using the same arguments as in [AGY06], since it has nothing to do with $\mathcal{W}$, only the construction of χ which is basically the same as in [AGY06]. □

Finally, to conclude this section, we note that Proposition 3.11 follows from Lemmas 4.2, 4.3, and 4.7 exactly in the same way that [AGY06, Proposition 7.7] is proved from the analogous [AGY06, Lemmas 7.9, 7.10, and 7.15].

5. Expansion and the twisted transfer operator

This section contains a proof of Proposition 3.12.

5.1 Refining the choice of γ_0

We now assume that π is the member of $\mathfrak{R}$ specified by Theorem 2.3. Let S denote a fixed finite set of generators of G_π, this is possible since we know G_π is finite index in $\mathrm{Sp}(\mathbf{Z}^{2g}, \omega_\pi)$. Choose a finite set Υ_0 of γ that are paths in $\mathfrak{R}$ beginning and ending in π and such that

$$\{\Theta_\gamma^* : \gamma \in \Upsilon_0\}$$

together with their inverses generate S. Now let

$$\Upsilon = \Upsilon_0 \cup \{\gamma.\gamma : \gamma \in \Upsilon_0\}.$$

We note for later on that this definition guarantees the following.

Lemma 5.1. *The elements*

$$\Theta_\gamma^*.(\Theta_{\gamma'}^*)^{-1} \in \mathrm{Sp}(\mathbf{Z}^{2g}, \omega_\pi) \quad \gamma, \gamma' \in \Upsilon$$

generate G_π.

Proof. For a given $\tilde{\gamma} \in \Upsilon_0$, we have $\Theta_{\tilde{\gamma}.\tilde{\gamma}}^*(\Theta_{\tilde{\gamma}}^*)^{-1} = \Theta_{\tilde{\gamma}}^*$ and $\Theta_{\tilde{\gamma}}^*(\Theta_{\tilde{\gamma}.\tilde{\gamma}}^*)^{-1} = (\Theta_{\tilde{\gamma}}^*)^{-1}$. On the other hand, the $\Theta_{\tilde{\gamma}}^*$ with $\tilde{\gamma} \in \Upsilon_0$ together with their inverses generate S and hence G_π'. □

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

We will now choose γ_0 such that no $\gamma \in \Upsilon$ contains γ_0 as a substring and moreover γ_0 is strongly positive and neat (recall these properties from § 2.13). This can be done simply by ensuring that γ_0 is strongly positive and neat and longer than all $\gamma \in \Upsilon$. We now give the details of this construction.

Before stating the next lemma we introduce some language. A path in $\mathfrak{R}$ is *complete* if every $\alpha \in \mathcal{A}$ is the winner of some arrow in γ. It follows from a result of [MMY05, § 1.2.3] (see also [AGY06, Lemma 3.2]) that there exists a complete path γ_* beginning and ending at π. A path in $\mathfrak{R}$ is said to be k-complete if it is the concatenation of k complete paths. Write γ_*^k for the k-fold concatenation of γ_* with itself. Then for example, if γ_* is complete, then γ_*^k is k-complete.

Lemma 5.2 [AGY06, Lemma 4.2]. *A k-complete path with $k \geqslant 3|\mathcal{A}| - 4$ is strongly positive.*

As noted in [AGY06, p. 162, footnote], a path is neat if it ends with a type ϵ arrow and begins with a string of opposite type arrows at least half the length of the path. Suppose that γ_* ends with a bottom arrow. Choose then k such that

$$l(\gamma_*).k \geqslant \max_{\gamma \in \Upsilon} l(\gamma), \quad k \geqslant 3|\mathcal{A}| - 4. \tag{40}$$

Next choose γ' beginning and ending at π with $l(\gamma_*).k + |\mathfrak{R}|$ top arrows at its beginning and $\leqslant |\mathfrak{R}|$ arrows afterwards (this is always possible since whatever the endpoint of the first top arrows, one can quickly return to π). Then

$$\gamma_0 := (\gamma' \gamma_*) \gamma_*^{k-1}$$

begins with

$$l(\gamma_*).k + |\mathfrak{R}| = \tfrac{1}{2}(l(\gamma_*).k + |\mathfrak{R}| + |\mathfrak{R}| + l(\gamma_*).k) \geqslant \tfrac{1}{2} l(\gamma_0)$$

top arrows so is therefore neat. Also, clearly $\gamma' \gamma_*$ is complete so γ_0 is k-complete. Therefore γ_0 is strongly positive by Lemma 5.2. Finally, by choice of k in (40) γ_0 is longer than any element of Υ. We have shown the following.

Lemma 5.3. *It is possible to choose γ_0 so that no element $\gamma \in \Upsilon$ contains γ_0 as a substring and moreover γ_0 is strongly positive and neat.*

We fix such a γ_0 for the remainder of the paper (and retroactively for the previous sections). From the discussion in § 2.13 this has the consequence that the elements of the set

$$\gamma_0.\Upsilon := \{\gamma_0 \gamma : \gamma \in \Upsilon\}$$

are all γ_0-adapted. We will use this later.

5.2 Decoupling I: releasing the convolution

We now perform the decoupling argument of [MOW19] with the first part of the argument based on [BGS11] and the latter part of the argument coming from [MOW19, Appendix]. One key difference here is the fact that the symbolic dynamics takes place on an infinite alphabet.

We understand during this section that all γ_i are γ_0-adapted in all sums and so forth. It is possible to show by adapting the proof of [AGY06, Lemma 7.8] that $\mathcal{L}_{s,\rho}$ and $\mathbf{L}_{s,\rho}$ act on $C^1(\Xi; \rho)$ for $|\sigma| < \sigma_1$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Recall the definition of the function R_s from (26). It will be convenient to introduce the function $R_s^{(n)}(y) := \sum_{i=0}^{n-1} R_s(Z^i y)$ for y in the domain of Z^{n-1}. Also recall the notation $\alpha_{\gamma_1 \cdots \gamma_k}$ from § 4. With these notations, for $F \in C^1(\Xi; \rho)$,

$$\mathbf{L}^N_{s,\rho}[F](y) = \sum_{\gamma_1,\ldots,\gamma_N} e^{R_s^{(N)}(\alpha_{\gamma_1\cdots\gamma_N} y)} \rho(\Theta^*_{\gamma_N})^{-1} \cdots \rho(\Theta^*_{\gamma_1})^{-1} F(\alpha_{\gamma_1\cdots\gamma_N} y). \tag{41}$$

We prepare a preliminary lemma.

Lemma 5.4. *There is a constant $C > 0$ such that for all $M \geqslant 1$, all σ with $|\sigma| \leqslant \sigma_1$ and $s = \sigma + it$, and all $y \in \Xi$,*

$$\|D[R_s^{(M)} \circ \alpha_{\gamma_1\cdots\gamma_M}](y)\| \leqslant C(1 + |t|). \tag{42}$$

Proof. Using Proposition 3.1, Lemma 3.3, and Theorem 3.10 we obtain

$$\|D(R_s \circ \alpha_\gamma)(y)\| \leqslant c(1 + |t|)$$

for all γ and y. Furthermore, for $k \geqslant 1$, by the chain rule and Proposition 3.1 part (ii) we have

$$\|D(R_s \circ \alpha_{\gamma_1\cdots\gamma_k})(y)\| \leqslant c(1 + |t|)\Lambda^{-k+1}$$

for $\Lambda > 1$. Now,

$$\begin{aligned}\|D(R_s^{(M)} \circ \alpha_{\gamma_1\cdots\gamma_M})(y)\| &= \left\| \sum_{i=0}^{M-1} D(R_s \circ \alpha_{\gamma_{1+i}\cdots\gamma_M})(y) \right\| \\ &\leqslant \sum_{i=0}^{M-1} c(1 + |t|)\Lambda^{-k+1} \leqslant C(1 + |t|)\end{aligned}$$

as required, by summing the geometric series. □

We now perform the same initial decoupling arguments as in Bourgain, Gamburd and Sarnak [BGS11]. Let

$$N = M + \tilde{M}$$

and let o be an arbitrary point in Ξ. Write d for the distance on Ξ coming from the Hilbert metric, induced by the Finsler metric on Ξ. Note that Ξ has bounded diameter with respect to d (since Ξ is a John domain in the sense of [AGY06] by [AGY06, Lemma 4.4]).

Lemma 5.5. *We have*

$$\mathbf{L}^N_{s,\rho}[F](y) = \sum_{\gamma_1,\ldots,\gamma_M} \mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho).\rho(\Theta^*_{\gamma_M})^{-1} \cdots \rho(\Theta^*_{\gamma_1})^{-1} F(\alpha_{\gamma_1\cdots\gamma_M} o) + O(\|F\|_{C^1}\Lambda^{-M})$$

where

$$\mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho) := \sum_{\gamma_{M+1},\ldots,\gamma_N} e^{R_s^{(N)}(\alpha_{\gamma_1\cdots\gamma_N} y)} \rho(\Theta^*_{\gamma_N})^{-1} \cdots \rho(\Theta^*_{\gamma_{M+1}})^{-1}$$

is a member of the algebra generated by the $\rho(\Theta^*_{\gamma_i})^{-1}$ acting on $\mathcal{W}$. The error term is in the norm of $\mathcal{W}$. We also have

$$\begin{aligned}D(\mathbf{L}^N_{s,\rho}[F])(y) = &\sum_{\gamma_1,\ldots,\gamma_M} \mathrm{Op}^{\tilde{M},\partial}_{\gamma_1\cdots\gamma_M;y}(\rho).\rho(\Theta^*_{\gamma_M})^{-1} \cdots \rho(\Theta^*_{\gamma_1})^{-1} F(\alpha_{\gamma_1\cdots\gamma_M} o) \\ &+ O((1 + |t|)\|F\|_{C^1}\Lambda^{-M}),\end{aligned} \tag{43}$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

where

$$\mathrm{Op}^{\tilde{M},\partial}_{\gamma_1\cdots\gamma_M;y}(\rho) := \sum_{\gamma_{M+1},\ldots,\gamma_N} D[e^{R_s^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}}](y)\otimes\rho(\Theta^*_{\gamma_N})^{-1}\cdots\rho(\Theta^*_{\gamma_{M+1}})^{-1}$$

is a member of $\mathrm{Hom}(T_y\Xi,\mathbf{R})\otimes\mathrm{End}(\mathcal{W})\cong\mathrm{Hom}(T_y\Xi,\mathrm{End}(\mathcal{W}))$ *and the big O term is interpreted with respect to the operator norm between the Finsler metric norm on* $T_y\Xi$ *and* $\mathrm{End}(\mathcal{W})$ *with its own operator norm. Write* $\|\bullet\|_{T_y\Xi,\mathrm{End}(\mathcal{W})}$ *for this norm and* $\|\bullet\|_{\mathrm{End}(\mathcal{W})}$ *for the operator norm on* $\mathrm{End}(\mathcal{W})$.

Proof. We begin by inspecting (41) and noting that $F(\alpha_{\gamma_1\cdots\gamma_N}y)$ and $F(\alpha_{\gamma_1\cdots\gamma_M}o)$ are distance $\ll\Lambda^{-M}$ apart, where $\Lambda>1$ is the constant from Proposition 3.1 part (ii). Hence we have

$$\|F(\alpha_{\gamma_1\cdots\gamma_N}y)-F(\alpha_{\gamma_1\cdots\gamma_M}o)\|_{\mathcal{W}}\ll\|F\|_{C^1}\Lambda^{-M}. \tag{44}$$

This gives

$$\begin{aligned}\mathbf{L}^N_{s,\rho}[F](y) = &\sum_{\gamma_1,\ldots,\gamma_N} e^{R_s^{(N)}(\alpha_{\gamma_1\cdots\gamma_N}y)}\rho(\Theta^*_{\gamma_N})^{-1}\cdots\rho(\Theta^*_{\gamma_1})^{-1}F(\alpha_{\gamma_1\cdots\gamma_M}o)\\ &+\sum_{\gamma_1,\ldots,\gamma_N} e^{R_s^{(N)}(\alpha_{\gamma_1\cdots\gamma_N}y)}\rho(\Theta^*_{\gamma_N})^{-1}\cdots\rho(\Theta^*_{\gamma_1})^{-1}(F(\alpha_{\gamma_1\cdots\gamma_N}y)-F(\alpha_{\gamma_1\cdots\gamma_M}o)).\end{aligned}$$

Using that ρ is unitary, the second line above can be bounded in $\|.\|_{\mathcal{W}}$ using (44) by

$$\|F\|_{C^1}\Lambda^{-M}\sum_{\gamma_1,\ldots,\gamma_N} e^{R_\sigma^{(N)}(\alpha_{\gamma_1\cdots\gamma_N}y)} = \|F\|_{C^1}\Lambda^{-M}\mathbf{L}^N_\sigma[1](y) = \|F\|_{C^1}\Lambda^{-M}.$$

This proves the first part of the lemma. For the second part, note

$$\begin{aligned}D(\mathbf{L}^N_{s,\rho}[F])(y) = &\sum_{\gamma_1,\ldots,\gamma_N} D[e^{R_s^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}}](y)\rho(\Theta^*_{\gamma_N})^{-1}\cdots\rho(\Theta^*_{\gamma_1})^{-1}F(\alpha_{\gamma_1\cdots\gamma_N}y)\\ &+\sum_{\gamma_1,\ldots,\gamma_N} e^{R_s^{(N)}(\alpha_{\gamma_1\cdots\gamma_N}y)}\rho(\Theta^*_{\gamma_N})^{-1}\cdots\rho(\Theta^*_{\gamma_1})^{-1}D(F\circ\alpha_{\gamma_1\cdots\gamma_N})(y).\end{aligned} \tag{45}$$

Since $\|D(F\circ\alpha_{\gamma_1\cdots\gamma_N})(y)\|\leqslant\Lambda^{-N}\|F\|_{C^1}$, the second term can be bounded in the norm of $\mathcal{W}$ by $\|F\|_{C^1}\Lambda^{-N}$ by the arguments from the first part of the lemma. To deal with the first line of (45), we argue as before, replacing $F(\alpha_{\gamma_1\cdots\gamma_N}y)$ with $F(\alpha_{\gamma_1\cdots\gamma_M}o)$. By the same arguments, we incur an error that can be bounded by

$$\|F\|_{C^1}\Lambda^{-M}\sum_{\gamma_1,\ldots,\gamma_N}\|D[e^{R_s^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}}](y)\|. \tag{46}$$

We must estimate the sum here. We calculate

$$D(e^{R_s^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}})(y) = D(R_s^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N})(y)e^{R_s^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}(y)}$$

so by Lemma 5.4 we have

$$\|D(e^{R_s^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}})(y)\|\leqslant C(1+|t|)e^{R_\sigma^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}(y)}.$$

Therefore the sum in (46) is $\leqslant C(1+|t|)\sum_{\gamma_1,\ldots,\gamma_N}e^{R_\sigma^{(N)}\circ\alpha_{\gamma_1\cdots\gamma_N}(y)} = C(1+|t|)$. This concludes the proof. □

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

We will now aim to give operator norm bounds for $\mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$ and $\mathrm{Op}^{\tilde{M},\partial}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$ that involve power decay in q.

Proposition 5.6. *Let $s=\sigma+it$. There is $D>0$ such that for all $t_0>0$, there are σ_1, c, C, L, $q_0>0$ such that for $|\sigma|<\sigma_1$, $|t|\leqslant t_0$, q odd with $q>q_0$ and $\tilde{M}=L\lfloor c\log q\rfloor$, we have*

$$\|\mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})\|_{\mathrm{End}(\mathcal{W})}\leqslant Ce^{R_\sigma^{(M)}(\alpha_{\gamma_1\cdots\gamma_M}o)}q^{-D},$$
$$\|\mathrm{Op}^{\tilde{M},\partial}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})\|_{T_y\Xi,\mathrm{End}(\mathcal{W})}\leqslant Ce^{R_\sigma^{(M)}(\alpha_{\gamma_1\cdots\gamma_M}o)}q^{-D}.$$

The bound for $\mathrm{Op}^{\tilde{M},\partial}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$ is similar to that for $\mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$ with no added difficulties,[20] so we treat only $\mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$. The reader can consult [MOW19, § 5.3] for more details. The proof of Proposition 5.6 will take up the remaining §§ 5.3, 5.4 of the present section.

Proof of Proposition 3.12 from Proposition 5.6. Import all the constants from Proposition 5.6. Recall we are given t_0 such that we assume $s=\sigma+it$ with $|t|\leqslant t_0$. We also assume $|\sigma|<\sigma_1$. We choose $M=\lfloor c'\log q\rfloor$ where $c'>0$ is chosen such that $\Lambda^{-M}\ll q^{-D}$. Then[21] $N\approx c_0\log q$. Note that

$$\|\rho_q^{\mathrm{new}}(\Theta^*_{\gamma_M})^{-1}\cdots\rho(\Theta^*_{\gamma_1})^{-1}F(\alpha_{\gamma_1\cdots\gamma_M}o)\|_{\mathcal{W}}\leqslant\|F\|_{C^1}$$

as ρ_q^{new} is unitary.

Lemma 5.5 and using the triangle inequality gives a constant $C>0$ such that for any $F\in C^1(\Xi;\rho_q^{\mathrm{new}})$

$$\|\mathbf{L}^N_{s,\rho_q^{\mathrm{new}}}F\|_{C^1}\leqslant Cq^{-D}\|F\|_{C^1}\sum_{\gamma_1,\ldots,\gamma_M}e^{R_\sigma^{(M)}(\alpha_{\gamma_1\cdots\gamma_M}o)}=Cq^{-D}\|F\|_{C^1}$$

since the sum is $\mathbf{L}^M_\sigma[1]=1$. Now by increasing q_0 if necessary, we ensure that when $q>q_0$, $Cq^{-D}\leqslant q^{-D/2}$.

Now given an arbitrary N', we can write $N'=aN+b$ with $0\leqslant b<N\approx c_0\log q$. Since the operator norm of $\mathbf{L}_{s,\rho_q^{\mathrm{new}}}$ is bounded (by comparison to $\mathbf{L}_\sigma$) by a constant K depending on t_0, we obtain for any $F\in C^1(\Xi;\rho_q^{\mathrm{new}})$

$$\|\mathbf{L}^{N'}_{s,\rho_q^{\mathrm{new}}}F\|_{C^1}\leqslant K^bq^{-aD/2}\|F\|_{C^1}\leqslant q^\eta(1-\epsilon)^{N'}\|F\|_{C^1}$$

for some $\epsilon,\eta>0$.

To deal with ρ_q in place of ρ_q^{new}, we consider the groups $\Gamma_q(q')$ that are defined to be the kernels of reduction modulo q' on Γ_q.

We decompose ρ_q as $\bigoplus_{1\neq q'|q}\rho^q_{q'}$ where $\rho^q_{q'}$ is the subrepresentation of $\ell^2_0(\Gamma_q)$ corresponding to functions invariant under $\Gamma_q(q')$ but not invariant by any $\Gamma_q(q'')$ with $q''|q'$, $q''\neq q'$. This gives a splitting

$$C^1(\Xi;\rho_q)=\bigoplus_{1\neq q'|q}C^1(\Xi;\rho^q_{q'}).\tag{47}$$

[20] After applying $\mathrm{Op}^{\tilde{M},\partial}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$ to a test tangent vector v in $T_y\Xi$ one obtains an element of $\mathrm{End}(\mathcal{W})$ with the task of bounding its operator norm, which can be done in exactly the same way as we will treat $\mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$. On the other hand, it is worth pointing out that the bound for $\mathrm{Op}^{\tilde{M},\partial}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$ relies crucially on the fact that $|t|\leqslant t_0$ whereas this is not a factor in bounding $\mathrm{Op}^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}(\rho_q^{\mathrm{new}})$.

[21] Here and henceforth we say $f\approx g$ if there is a positive constant k_1 such that $|f-g|<k_1$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

The action of the transfer operator $\mathbf{L}_{s,\rho_q}$ on $C^1(\Xi;\rho_{q'}^q)$ is intertwined with the action of $\mathbf{L}_{s,\rho_{q'}^{\mathrm{new}}}$ on $C^1(\Xi;\rho_{q'}^{\mathrm{new}})$. Thus if $f \in C^1(\Xi;\rho_q)$ we can decompose $f = \sum_{1\neq q'|q} f_{q'}$ according to (47).

This is the point in the paper where the modulus Q_0 of Theorem 1.4 comes into play. We now assume Q_0 is the product of primes less than or equal to $\leqslant q_0$, where q_0 is the constant fixed during this proof. In particular, if q is coprime to Q_0, then any $q'|q$ has no proper divisors $\leqslant q_0$. Under this assumption, $f_{q'} = 0$ if $q' \leqslant q_0$.

Now we have

$$\|\mathbf{L}_{s,\rho_q}^{N'} f\|_{C^1} \leqslant \sum_{q_0<q'|q} \|\mathbf{L}_{s,\rho_{q'}}^{N'} f_{q'}\| \leqslant \sum_{q_0<q'|q} q^{\eta}(1-\epsilon)^{N'}\|f_{q'}\|_{C^1},$$

where we used the bound we previously obtained for the operator norm of $\mathbf{L}_{s,\rho_{q'}^{\mathrm{new}}}^{N'}$ to bound $\|\mathbf{L}_{s,\rho_{q'}}^{N'} f_{q'}\|_{C^1}$. Since $\|f_{q'}\|_{C^1} \leqslant \|f\|_{C^1}$ for each q', and q has fewer than q^{ζ} divisors for some $\zeta > 0$ and all q, by increasing η if necessary the above can be bounded by

$$q^{\eta}(1-\epsilon)^{N'}\|f\|_{C^1}.$$

This proves Proposition 3.12 with $\mathbf{L}_{s,\rho_q^{\mathrm{new}}}$ in place of $\mathcal{L}_{s,\rho_q^{\mathrm{new}}}$. To convert between estimates for the unnormalized and normalized transfer operators, note that $\mathcal{L}_{s,\rho_q}^N = \lambda_\sigma^N h_\sigma \mathbf{L}_{s,\rho_q}^N h_\sigma^{-1}$. Multiplication and division by h_σ is a uniformly bounded operator in $|\sigma| < \sigma_1$ by Theorem 3.10. Moreover, by Theorem 3.10 we can choose $\sigma_1' < \sigma_1$ such that $\lambda_\sigma < (1-\epsilon)^{1/2}$ for all $|\sigma| < \sigma_1'$. Therefore under these assumptions on σ we have for some $C > 0$, for all $N' \geqslant 1$,

$$\|\mathcal{L}_{s,\rho_q}^{N'} f\|_{C^1} \leqslant C q^{\eta}(1-\epsilon)^{N'/2}$$

which concludes the proof. □

5.3 Bounding the operator norm of convolution operators

Let $\pi_q : \mathrm{Sp}(\mathbf{Z}^{2g},\omega_\pi) \to \Gamma_q$ be the reduction mod q map. To improve the readability of the following argument we will write for γ_0-adapted γ

$$h_\gamma := \pi_q(\Theta_\gamma^*)^{-1} \in \Gamma_q.$$

Recall $N = M + \tilde{M}$, $s = \sigma + it$, and o is an arbitrary but fixed point in Ξ. We are tasked with estimating the operator norm of the group algebra element

$$\mu_{\gamma_1\cdots\gamma_M;y}^{\tilde{M}} := \sum_{\gamma_{M+1},\ldots,\gamma_N} e^{R_s^{(N)}(\alpha_{\gamma_1\cdots\gamma_N} y)} h_{\gamma_N} h_{\gamma_{N-1}} \cdots h_{\gamma_{M+1}} \in \mathbf{C}[\Gamma_q]$$

as it acts by convolution on $\ell^2_{\mathrm{new}}(\Gamma_q)$. Indeed, this is precisely the operator $\mathrm{Op}_{\gamma_1\cdots\gamma_M;y}^{\tilde{M}}(\rho_q^{\mathrm{new}})$ when restricted to $\ell^2_{\mathrm{new}}(\Gamma_q)$. We view elements of $\mathbf{C}[\Gamma_q]$ interchangeably as complex-valued measures on Γ_q. We write $*$ for the convolution of measures, this corresponds to multiplication in $\mathbf{C}[\Gamma_q]$. Given $\mu \in \mathbf{C}[\Gamma_q]$ we write $|\mu|$ for the non-negative real measure obtained by taking absolute values of coefficients. We let $\tilde{\mu}$ be the measure defined by $\tilde{\mu}(g) := \overline{\mu(g^{-1})}$. If $\mu_1, \mu_2 \in \mathbf{R}[\Gamma_q]$ we write $\mu_1 \leqslant \mu_2$ if $\mu_1(g) \leqslant \mu_2(g)$ for all $g \in \Gamma_q$.

LEMMA 5.7. *We have*

$$|\mu_{\gamma_1\cdots\gamma_M;y}^{\tilde{M}}| \leqslant C e^{R_\sigma^{(M)}(\alpha_{\gamma_1\cdots\gamma_M} o)} \mu_1 \tag{48}$$

where $C > 0$ is a constant and

$$\mu_1 = \sum_{\gamma_{M+1},\ldots,\gamma_N} e^{R_\sigma^{(\tilde{M})}(\alpha_{\gamma_{M+1}\cdots\gamma_N} o)} h_{\gamma_N} h_{\gamma_{N-1}} \cdots h_{\gamma_{M+1}}. \tag{49}$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Proof. The proof is the same as [MOW19, Lemma A.2]. □

Note that although μ_1 depends on $M, \tilde{M}$, we suppress this dependence from now on. We now organize the ingredients for the proof of Proposition 5.6.

Proposition 5.8 (Majorization of μ_1). *There is a constant $\epsilon > 0$ such that for any $B > 1$, there exists an integer $L > 0$ such that for all $K > 0$, if $\tilde{M} = LK$, there is a measure μ_2 such that*

$$\mu_1 \leqslant \mu_2, \tag{50}$$

for all $\phi \in \ell^2_0(\Gamma_q)$,

$$\|\mu_2 * \phi\|_{\ell^2} \leqslant (1-\epsilon)^K \|\mu_2\|_1 \|\phi\|_{\ell^2}, \tag{51}$$

and

$$\|\mu_2\|_1 \leqslant B^K \|\mu_1\|_1. \tag{52}$$

The proof of this proposition is deferred to the next section. Note that we would like to have (51) for μ_1, or even better, the analogous result for $\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}$. However, we only know $|\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}| \ll e^{R^{(M)}_\sigma(\alpha_{\gamma_1\cdots\gamma_M} o)}\mu_2$ from which it is not obvious how to convert (51) into Proposition 5.8.

The solution is to first use Proposition 5.8 to deduce that the ℓ^2 norm of $\widetilde{\mu_2} * \mu_2$ is small, and hence the ℓ^2 norm of $\widetilde{\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}} * \mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}$ is small. This will be done using the following lemma.

Lemma 5.9 [MOW19, Proposition A.8]. *For any measure ν on Γ_q, we have*

$$\|\tilde{\nu} * \nu\|_2 \leqslant \frac{\|\nu\|_1^2}{|\Gamma_q|^{1/2}} + \|\nu\|_1 \|\nu\|_{\ell^2_0(\Gamma_q)}.$$

Here $\|\nu\|_{\ell^2_0(\Gamma_q)}$ is the operator norm of ν acting by convolution on $\ell^2_0(\Gamma_q)$.

Proof. This is proved in [MOW19, proof of Proposition A.8]. □

In the previous lemma, we will take $\nu = \mu_2$. When we succeed in proving $\|\widetilde{\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}} * \mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|_2$ is small, we will then need a way to convert this into information on the operator norm of $\mu_{\gamma_1\cdots\gamma_M;y}$. It is here that the quasirandomness property of $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$ is crucially used.

Lemma 5.10. *For some absolute $C, D > 0$*

$$\|\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|_{\ell^2_{\mathrm{new}}(\Gamma_q)} \leqslant C\left(\frac{|\Gamma_q| \|\widetilde{\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}} * \mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|_2^2}{q^D}\right)^{1/4}.$$

Here $\|.\|_2^2$ denotes the ℓ^2 norm of the measure on Γ_q and $\|\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|_{\ell^2_{\mathrm{new}}(\Gamma_q)}$ is the operator norm of $\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}$ acting on the new subspace of $\ell^2(\Gamma_q)$.

Proof. We need to use the lower bound for the degree of new irreducible representations of $\mathrm{Sp}((\mathbf{Z}/q\mathbf{Z})^{2g}, \omega_\pi)$ that is given in Proposition 6.1. Supposing that the smallest new irreducible representation has $\gg q^D$, then by the trace formula argument of [MOW19, Lemma A.7] the largest eigenvalue of A^*A where $A := \mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y} *$ acting on $\ell^2_{\mathrm{new}}(\Gamma_q)$ satisfies

$$\lambda^2 q^D \leqslant C'|\Gamma_q| \|\widetilde{\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}} * \mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|_2^2.$$

The crucial point is that the eigenvalue appears with high multiplicity in the trace formula, an idea that goes back to Sarnak and Xue [SX91]. Since $\|A\| = \lambda^{1/2}$ the lemma follows. □

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Now we can prove Proposition 5.6, modulo the deferred proof of Proposition 5.8.

Proof of Proposition 5.6. We now make precise the argument we outlined before. Let $\epsilon > 0$ be the constant from Proposition 5.8. Choose $B > 1$ such that for all $q \geqslant 2$,

$$\log B \leqslant -\log(1-\epsilon)\frac{D}{2}\frac{\log q}{\log|\Gamma_q|},$$

where D is the constant from Lemma 5.10. The reason for this choice will be pointed out shortly. Apply Proposition 5.8 for this B to obtain a constant L and measure μ_2 such that $\mu_1 \leqslant \mu_2$. We let $M = LK$ as in Proposition 5.8. Combining Proposition 5.8 and Lemma 5.9 we obtain

$$\|\widetilde{\mu_2} * \mu_2\|_2 \leqslant \|\mu_2\|_1^2\left(\frac{1}{|\Gamma_q|^{1/2}} + (1-\epsilon)^K\right) \leqslant B^{2K}\|\mu_1\|_1^2\left(\frac{1}{|\Gamma_q|^{1/2}} + (1-\epsilon)^K\right).$$

We can evaluate $\|\mu_1\|_1$ by

$$\|\mu_1\|_1 = \sum_{\gamma_{M+1},\dots,\gamma_N} e^{R_\sigma^{(\tilde{M})}(\alpha_{\gamma_{M+1}\cdots\gamma_N} o)} = \mathbf{L}_\sigma^{\tilde{M}}[1](o) = 1.$$

Hence

$$\|\widetilde{\mu_2} * \mu_2\|_2 \leqslant B^{2K}\left(\frac{1}{|\Gamma_q|^{1/2}} + (1-\epsilon)^K\right).$$

From (48) and $\mu_1 \leqslant \mu_2$ we obtain

$$\|\widetilde{\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}} * \mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|_2 \leqslant C^2 e^{2R_\sigma^{(M)}(\alpha_{\gamma_1\cdots\gamma_M} o)} B^{2K}\left(\frac{1}{|\Gamma_q|^{1/2}} + (1-\epsilon)^K\right).$$

Using this as input to Lemma 5.10 gives

$$\begin{aligned}\|\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|^2_{\ell^2_{\mathrm{new}}(\Gamma_q)} &\ll \frac{|\Gamma_q|^{1/2}}{q^{D/2}}\|\widetilde{\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}} * \mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|_2 \\ &\ll e^{2R_\sigma^{(M)}(\alpha_{\gamma_1\cdots\gamma_M} o)}\frac{B^{2K}}{q^{D/2}}(1 + |\Gamma_q|^{1/2}(1-\epsilon)^K).\end{aligned}$$

We now choose our constant $c > 0$ such that for $K = \lfloor c\log q\rfloor$, $1 \ll |\Gamma_q|^{1/2}(1-\epsilon)^K \ll 1$. The choice of B ensures that for this K, $B^{2K}/q^{D/2} \ll q^{-D/4}$. Hence

$$\|\mu^{\tilde{M}}_{\gamma_1\cdots\gamma_M;y}\|^2_{\ell^2_{\mathrm{new}}(\Gamma_q)} \ll q^{-D/4} e^{2R_\sigma^{(M)}(\alpha_{\gamma_1\cdots\gamma_M} o)}.$$

This proves the first inequality of Proposition 5.6 (replacing $D/8$ by D). As remarked before, the second inequality uses essentially the same argument. □

5.4 Decoupling II: majorizing μ_1

In this section we prove Proposition 5.8 by adapting arguments from [MOW19, Appendix] to the infinite alphabet setting, using also a different spectral gap input from property (T) that relies on our preparation of the set Υ and its relation to the Rauzy–Veech group G_π. The key idea in the proof is that while μ_1 is not a convolution, it can be majorized by a carefully chosen sum of convolutions.

We further decompose

$$\tilde{M} = LK, \tag{53}$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

where L is going to be chosen to be a large constant, and decompose $\{M+1,\dots,N\}$ into blocks of size either 1, $L-1$ or L. Let

$$I_{i,j} = [\gamma_i, \gamma_{i+1}, \dots, \gamma_j]$$

denote the block of all $\gamma_{i'}$ with $i \leqslant i' \leqslant j$. Rewrite the summation in (49) as

$$\mu_1 = \sum_{I_{M+1,M+L-1}, I_{M+L+1,M+2L-1}, \dots, I_{N-L+1,N-1}} \sum_{\gamma_{M+L}, \gamma_{M+2L}, \dots, \gamma_N} e^{R_\sigma^{(\tilde{M})}(\alpha_{\gamma_{M+1}\cdots\gamma_N} o)} h_{\gamma_N} h_{\gamma_{N-1}} \cdots h_{\gamma_{M+1}}. \tag{54}$$

This reordering of summation is permitted since the sums are suitably absolutely convergent by Theorem 3.10 and the following discussion. Following [MOW19, (A.15)], using contraction properties of α_{γ_i} and the bound (42) for the derivative of $R_\sigma^{(\tilde{M})}$, one has the bounds

$$\exp(-c\Lambda^{-L})^{K-1} \beta_1 \beta_2 \cdots \beta_K \leqslant e^{R_\sigma^{(\tilde{M})}(\alpha_{\gamma_{M+1}\cdots\gamma_N} o)} \leqslant \exp(c\Lambda^{-L})^{K-1} \beta_1 \beta_2 \cdots \beta_K, \tag{55}$$

where

$$\beta_K = e^{R_\sigma^{(L)}(\alpha_{\gamma_{N-L+1}\cdots\gamma_N} o)}, \quad \beta_j = e^{R_\sigma^{(L)}(\alpha_{\gamma_{M+(j-1)L+1}\cdots\gamma_{M+(j+1)L-1}} o)}, \quad 1 \leqslant j \leqslant K-1$$

and $c > 0$ is a constant. Notice the important feature that each β_j depends on only one of γ_{M+jL}. Inserting the second inequality of (55) into (54) gives

$$\mu_1 \leqslant \mu_2 := \exp(c\Lambda^{-L})^{K-1} \sum_{I_{M+1,M+L-1}, I_{M+L+1,M+2L-1}, \dots, I_{N-L+1,N-1}} \eta_K * \eta_{K-1} * \cdots * \eta_1, \tag{56}$$

where the $\eta_j = \eta_j(I_{M+1,M+L-1}, I_{M+L+1,M+2L-1}, \dots, I_{N-L+1,N-1})$ are given by

$$\eta_K := \sum_{\gamma_N} \beta_K(\gamma_{N-L}, \dots, \gamma_N) h_{\gamma_N} \cdots h_{\gamma_{N-L+1}},$$
$$\eta_j := \sum_{\gamma_{M+jL}} \beta_j(\gamma_{M+(j-1)L+1}, \dots, \gamma_{M+(j+1)L-1}) h_{\gamma_{M+jL}} \cdots h_{\gamma_{M+(j-1)L+1}}, \quad 1 \leqslant j \leqslant K-1.$$

We point out for the readers' convenience that *we have now defined* μ_2. This proves (50).

To prove (51) we now aim for bounds on the operator norms of the measures η_j acting by convolution on $\ell_0^2(\Gamma_q)$. We write $\|\eta_j\|_{\mathrm{op}}$ for this operator norm. Consider, taking for example $1 \leqslant j \leqslant K-1$,

$$\eta_j * \tilde{\eta}_j = \sum_{\gamma_{M+jL}, \gamma'_{M+jL}} \beta_j(\dots, \gamma_{M+jL}, \dots) \beta_j(\dots, \gamma'_{M+jL}, \dots) h_{\gamma_{M+jL}} (h_{\gamma'_{M+jL}})^{-1}. \tag{57}$$

Since

$$\|\eta_j\|_{\mathrm{op}} = \|\tilde{\eta}_j\|_{\mathrm{op}} = \sup_{\phi \in \ell_0^2(\Gamma_q) : \|\phi\| = 1} \langle \eta_j * \tilde{\eta}_j \phi, \phi \rangle^{1/2} \tag{58}$$

we turn to estimating the operator norm of $\eta_j * \tilde{\eta}_j$ on $\ell_0^2(\Gamma_q)$. We need to both:

(i) estimate the values of β_j; and
(ii) discuss the group elements $h_{\gamma_{M+jL}} (h_{\gamma_{M+jL}})^{-1}$.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

These are both points of departure from [MOW19, Appendix], so we give more details.

(1) Continuing with $1 \leqslant j \leqslant K-1$ (the edge case $j = K$ is similar) we have

$$\begin{aligned}\beta_j(\dots,\gamma_{M+jL},\dots) &= e^{R_\sigma^{(L)}(\alpha_{\gamma_{M+(j-1)L+1}\cdots\gamma_{M+(j+1)L-1}}o)}\\ &= \exp\left(\sum_{i=0}^{L-2} R_\sigma(\alpha_{\gamma_{M+(j-1)L+1+i}\cdots\gamma_{M+(j+1)L-1}}o)\right)\exp(R_\sigma(\alpha_{\gamma_{M+jL}\cdots\gamma_{M+(j+1)L-1}}o))\\ &= \exp\left(\sum_{i=0}^{L-2} R_\sigma(\alpha_{\gamma_{M+(j-1)L+1+i}\cdots\gamma_{M+jL-1}}o) + O(\Lambda^{-i})\right)\\ &\quad\times\exp(R_\sigma(\alpha_{\gamma_{M+jL}\cdots\gamma_{M+(j+1)L-1}}o))\\ &\asymp \kappa(\gamma_{M+(j-1)L+1+i},\dots,\gamma_{M+jL-1})\exp(R_\sigma(\alpha_{\gamma_{M+jL}\cdots\gamma_{M+(j+1)L-1}}o)),\end{aligned}\tag{59}$$

where $\asymp$ means bounded above and below by a constant independent of all γ_i and L, and

$$\kappa(\gamma_{M+(j-1)L+1+i},\dots,\gamma_{M+jL-1}) := \exp(R_\sigma^{(L-1)}(\alpha_{\gamma_{M+(j-1)L+1+i}\cdots\gamma_{M+jL-1}}o)).$$

Note the arguments of κ are fixed given η_j. Also note that for fixed η_j the values

$$\alpha_{\gamma_{M+jL+1}\cdots\gamma_{M+(j+1)L-1}}o$$

lie in the same cylinder of depth one. On the other hand, the derivative of $R_\sigma\circ\alpha_{\gamma_{M+jL}}$ is uniformly bounded by (42), so the values in the exponent of (59) fluctuate by at most $O(1)$ while $\alpha_{\gamma_{M+jL}}$ is fixed. Therefore

$$\beta_j(\dots,\gamma_{M+jL},\dots) \asymp \kappa(\gamma_{M+(j-1)L+1+i},\dots,\gamma_{M+jL-1})\exp(R_\sigma(\alpha_{\gamma_{M+jL}}o)).\tag{60}$$

In light of this estimate and the discussion after Theorem 3.10 concerning convergence of infinite sums, we see that η_j and $\eta_j * \tilde{\eta}_j$ have finite ℓ_1 norms. This supports our earlier justification of reordering of summations.

(2) Recall Υ from Lemma 5.1. We can write

$$\eta_j * \tilde{\eta}_j = \nu + \tilde{\nu},$$

where ν is the contribution to (57) from $\gamma_{M+jL}, \gamma'_{M+jL} \in \gamma_0.\Upsilon$ and $\tilde{\nu}$ are the remaining contributions. Then the support of ν is the reduction mod q of the set

$$\Sigma = \{\Theta^*_\gamma.(\Theta^*_{\gamma'})^{-1} : \gamma,\gamma' \in \gamma_0.\Upsilon\}.$$

By Lemma 5.1, the set Σ generates the conjugate of G_π by $\Theta^*_{\gamma_0}$. Call this conjugate group G'_π.

We now bring these arguments (1) and (2) together. Let $\nu = \eta_j * \tilde{\eta}_j$. Note that the operator formed from convolution by ν on $\ell^2_0(\Gamma_q)$ is self-adjoint and positive. Therefore the operator norm of $\|\nu\|$ acting by convolution on $\ell^2_0(\Gamma_q)$ is

$$\|\nu\|_{\mathrm{op}} = \sup_{\phi\in\ell^2_0(\Gamma_q),\|\phi\|=1} \langle\nu * \phi, \phi\rangle.\tag{61}$$

We need to use the following property of the action of G'_π on Γ_q.

Lemma 5.11 (No almost invariant vectors). *There is some $\epsilon > 0$ such that for all odd q, for all $\phi \in \ell^2_0(\Gamma_q)$ with $\|\phi\|_{\ell^2} = 1$ there is some $g \in \Sigma$ such that if $g_q := g \mod q$, then*

$$\|g_q * \phi - \phi\|_{\ell^2} > \epsilon.$$

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Proof. By Theorem 2.3, when q is odd, G'_π maps onto Γ_q. Hence $\ell^2_0(\Gamma_q)$ has no invariant vectors. The statement of the lemma is then a consequence of Kazhdan's property (T) for finite index subgroups of $\mathrm{Sp}(\mathbf{Z}^{2g}, \omega_\pi)$ [Kaž67] applied to G'_π. □

Write

$$\nu = \sum_{g_q \in \Gamma_q} \nu_{g_q} g_q.$$

Let ϵ, g_q^0 be the constant (respectively group element) provided by Lemma 5.11 on inputting ϕ with $\|\phi\| = 1$. Then it is straightforward to check that $|\Re(\langle g_q^0 * \phi, \phi\rangle)| < (1 - \epsilon')$ where $\epsilon' = \epsilon^2/2$. Returning to (61), using $\nu_{g_q} = \nu_{(g_q)^{-1}}$ from (57) we get

$$\begin{aligned}\langle \nu * \phi, \phi\rangle &= \sum_{g_q \in \Gamma_q} \nu_{g_q} \langle g_q * \phi, \phi\rangle = \sum_{g_q \in \Gamma_q} \nu_{g_q} \Re\langle g_q * \phi, \phi\rangle \\ &= \nu_{g_q^0} \Re\langle g_q^0 * \phi, \phi\rangle + \sum_{g_q \neq g_q^0} \nu_{g_q} \Re\langle g_q * \phi, \phi\rangle \\ &\leqslant (1 - \epsilon')\nu_{g_q^0} + \sum_{g_q \neq g_q^0} \nu_{g_q} = \|\nu\|_1 - \epsilon' \nu_{g_q^0}.\end{aligned}$$

Also, from (60), and that $\sum_{\gamma_{M+jL}} \exp(R_\sigma(\alpha_{\gamma_{M+jL}} o)) = \mathbf{L}_\sigma[1](o) = 1$, we get

$$\nu_{g_q^0} \geqslant C\|\nu\|_1$$

with constant C independent of $\nu_{g_q^0}$ and η_j. So combining this with the preceding estimate and (61) we get

$$\|\nu\|_{\mathrm{op}} \leqslant \|\nu\|_1(1 - \epsilon'')$$

for some $\epsilon'' > 0$. Inserting this into (58) gives

$$\|\eta_j\|_{\mathrm{op}} \leqslant \|\eta_j\|_{\ell^1}(1 - \epsilon'')^{1/2}. \tag{62}$$

Using (62) in (56) gives for any $\phi \in \ell^2_0(\Gamma_q)$

$$\|\mu_2 * \phi\|_{\ell^2} \leqslant (1 - \epsilon'')^{K/2}\|\mu_2\|_1\|\phi\|_{\ell^2}.$$

This proves (51). It remains to choose L so that (52) holds. To do so, we estimate $\|\mu_2\|_1$. We have

$$\|\mu_2\|_1 \leqslant \exp(c\Lambda^{-L})^{K-1} \sum_{I_{M+1,M+L-1}, I_{M+L+1,M+2L-1}, \ldots, I_{N-L+1,N-1}} \sum_{\gamma_{M+L}, \gamma_{M+2L}, \ldots, \gamma_N} \beta_1 \cdots \beta_K.$$

We now use the first inequality of (55) to get

$$\|\mu_2\|_1 \leqslant \exp(2c\Lambda^{-L})^{K-1} \sum_{I_{M+1,M+L-1}, I_{M+L+1,M+2L-1}, \ldots, I_{N-L+1,N-1}} \sum_{\gamma_{M+L}, \gamma_{M+2L}, \ldots, \gamma_N} e^{R_\sigma^{(\tilde{M})}(\alpha_{\gamma_{M+1} \cdots \gamma_N} o)}.$$

But from inspection of (54), the above is $\exp(2c\Lambda^{-L})^{K-1}\|\mu_1\|$. Recall B is the quantifier from Proposition 5.8. We now choose L large enough so that

$$\exp(2c\Lambda^{-L}) \leqslant B.$$

This gives

$$\|\mu_2\|_1 \leqslant \exp(2c\Lambda^{-L})^{K-1}\|\mu_1\| \leqslant B^K\|\mu_1\|$$

and completes the proof of Proposition 5.8.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

6. Quasirandomness

In this section we show that 'new' representations of $\mathrm{Sp}((\mathbf{Z}/q\mathbf{Z})^{2g}, \omega_{\mathcal{M}})$ have large dimension. This is a version of the *quasirandomness* property of a group that takes into account the level structure of the family of groups $\mathrm{Sp}((\mathbf{Z}/q\mathbf{Z})^{2g}, \omega_{\mathcal{M}})$.

Proposition 6.1 (Quasirandomness estimates). *There is $C > 0$ and $D > 0$ such that any irreducible representation of $\mathrm{Sp}((\mathbf{Z}/q\mathbf{Z})^{2g}, \omega_{\mathcal{M}})$ that does not factor through*

$$\mathrm{Sp}((\mathbf{Z}/q\mathbf{Z})^{2g}, \omega_{\mathcal{M}}) \to \mathrm{Sp}((\mathbf{Z}/q_1\mathbf{Z})^{2g}, \omega_{\mathcal{M}})$$

for some $q_1|q$ has dimension $\geqslant Cq^D$.

We follow the type of argument given by Kelmer and Silberman in [KS13, § 4] for rank-one groups (see also [Mag15] for a small improvement to that argument). We may treat the group $\mathrm{Sp}_{2g}(\mathbf{Z})$ without loss of generality, that is, we assume the symplectic form is the standard one. Let $g \geqslant 2$. Let $q \in \mathbf{N}$ and let $(\rho, \mathcal{W})$ be an irreducible unitary representation of $\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$ that is not obtained by a composition

$$\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z}) \to \mathrm{Sp}_{2g}(\mathbf{Z}/q_1\mathbf{Z}) \xrightarrow{\rho'} U(\mathcal{W})$$

with $q_1|q$. We refer to this property as ρ being *new*.

6.1 The case when q is prime

For p an odd prime, let $\mathbb{F}_p$ denote the finite field with p elements. The table of Seitz and Zalesskii in [SZ93, Table 1] implies that $P\,\mathrm{Sp}_{2g}(\mathbb{F}_p)$ has no projective complex irreducible representation of dimension $< \frac{1}{2}(p^g - 1)$ and hence this is also a lower bound for the dimension of an irreducible representation of $\mathrm{Sp}_{2g}(\mathbb{F}_p)$.

6.2 The case $q = p^r$

In this section we prove the following.

Proposition 6.2. *There is some $C > 0$ depending only on g such that for all $r \geqslant 2$, letting $R := \lfloor r/2 \rfloor$ any new representation $(\rho, \mathcal{W})$ of $\mathrm{Sp}_{2g}(\mathbf{Z}/p^r\mathbf{Z})$ has dimension at least*

$$\dim \rho \geqslant Cp^R.$$

Let $q = p^r$. Write $H_q := \mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z})$ and for $q'|q$ let $H_q(q')$ be the kernel of the reduction modulo q' map

$$H_q \to H_{q'}.$$

Let $\mathfrak{g}(\mathbf{Z}/q\mathbf{Z})$ denote the Lie algebra of Sp_{2g} over $\mathbf{Z}/q\mathbf{Z}$. We view this as an abelian group. Let $R = \lfloor r/2 \rfloor$. The congruence subgroup $H_{p^r}(p^{r-R})$ is an abelian normal subgroup of H_{p^r} that is naturally isomorphic to $\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$. The action of H_{p^r} on $H_{p^r}(p^{r-R})$ by conjugation descends to an action of H_{p^R}. After using the isomorphism $H_{p^r}(p^{r-R}) \cong \mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$ this conjugation action is identified with the Adjoint action of H_{p^R} on $\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$, i.e.

$$\mathrm{Ad}(g)v = gvg^{-1}, \quad g \in H_{p^R}, \quad v \in \mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z}).$$

Let $(\rho, \mathcal{W})$ be a unitary representation of H_q. Suppose $R \geqslant 1$. If ρ is trivial when restricted to $H_q(p^R)$, then ρ is not a new representation. More generally, if ρ is new, then the restriction

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

of ρ to $H_q(p^{r-R})$ must not be trivial on any $H_q(p^{r-R+\eta})$ with $\eta \in \mathbf{Z}_+$ since these are also normal subgroups with $H_q/H_q(p^{r-R+\eta}) \cong H_{p^{r-R+\eta}}$. Notice $H_q(p^{r-R+\eta}) \leqslant H_q(p^{r-R})$ corresponds to the inclusion $p^\eta \mathfrak{g}(\mathbf{Z}/p^{R-\eta}\mathbf{Z}) \leqslant \mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$.

The strategy is to consider the H_{p^R} invariant set of characters of $\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$ that appear when restricting ρ to $H_{p^r}(p^{r-R}) \cong \mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$, since the size of this set gives a lower bound for the dimension of ρ.

The Killing form on $\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$ is non-degenerate which allows us to identify the unitary dual $\widehat{\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})}$ with $\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$. Under this identification, the co-Adjoint action on characters becomes an Adjoint action on $\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$. Moreover any character that is non-trivial on each $H_q(p^{r-R+\eta})$, $\eta \in \mathbf{Z}_+$, becomes an element of $\mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$ which is not congruent to 0 modulo p.

We have therefore reduced Proposition 6.2 to the following lemma.

Lemma 6.3. *There is some $C > 0$ depending only on g such that for all $R \geqslant 1$ the H_{p^R}-Adjoint orbit of any $X \in \mathfrak{g}(\mathbf{Z}/p^R\mathbf{Z})$ with $X \not\equiv 0 \bmod p$ has size*

$$|\mathrm{Ad}(H_{p^R}).X| \geqslant Cp^R.$$

Proof. By the orbit-stabilizer theorem the orbit has size at least

$$\frac{|H_{p^R}|}{|C_{H_{p^R}}(X)|} \tag{63}$$

where we write C to stand for centralizer, and therefore $C_{H_{p^r}}(X) = \{h \in H_{p^R} : hXh^{-1} = X\}$. Since H_{p^R} is an $(R-1)$-fold extension of H_p by groups isomorphic to $\mathfrak{g}(\mathbb{F}_p)$ we know $|H_{p^R}| = |H_p||\mathfrak{g}(\mathbb{F}_p)|^{R-1} \gg p^{R.\dim(\mathrm{Sp}_{2g})} = p^{g(2g+1)R}$. This gives the bound we will use for the numerator of (63).

Considering next the denominator of (63), by an elementary induction argument appearing in [KS13, Proof of Proposition 4.3]

$$|C_{H_{p^r}}(X)| \leqslant |C_{H_p}(X \bmod p)|\,|C_{\mathfrak{g}(\mathbb{F}_p)}(X \bmod p)|^{R-1}, \tag{64}$$

where the latter centralizer is $C_{\mathfrak{g}(\mathbb{F}_p)}(X \bmod p) = \{y \in \mathfrak{g}(\mathbb{F}_p) : [y, X \bmod p] \equiv 0\}$. According to Springer and Steinberg [SS70, II. 4.1, 4.2, IV. 2.26], the algebraic group $C_{\mathrm{Sp}_{2g}}(X \bmod p)$ defined over $\mathbb{F}_p$ has a number of components bounded by a constant depending only on g. By a bound of Nori [Nor87, Lemma 3.5] each component can have at most $(p+1)^{\dim C_{\mathrm{Sp}_{2g}}(X \bmod p)}$ points over $\mathbb{F}_p$. But $\dim C_{\mathrm{Sp}_{2g}}(X \bmod p)$ is also the dimension of the centralizer of $X \bmod p$ in $\mathfrak{g}(\mathbb{F}_p)$, so we have now reduced the estimation of the right-hand side of (64) to a bound for

$$\dim C_{\mathfrak{g}(\mathbb{F}_p)}(X'),$$

where $X' = X \bmod p$ is a non-zero element of $\mathfrak{g}(\mathbb{F}_p)$.

Assume the bound $\dim C_{\mathfrak{g}(\mathbb{F}_p)}(X') \leqslant \dim(\mathrm{Sp}_{2g}) - e = g(2g+1) - e$. Then putting our previous estimates together we obtain

$$|\mathrm{Ad}(H_{p^R}).X| \gg \frac{p^{g(2g+1)R}}{(p+1)^{g(2g+1)-e}(p^{g(2g+1)-e})^{R-1}} \gg p^{eR}.$$

Since it is not particularly important here to optimize e, we give the easy argument that one may take $e = 1$ since $\mathfrak{g}(\mathbb{F}_p)$ has no non-trivial center.[22] This gives the result stated in the lemma. □

[22] One may definitely do better here, and it would be good to work out the best possible bound, but it is not the purpose of the current paper.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

6.3 The case of general moduli

If p_i are primes and

$$q = \prod_{i=1}^{M} p_i^{m_i}$$

is the prime factorization of q, then we have by the Chinese remainder theorem

$$\mathrm{Sp}_{2g}(\mathbf{Z}/q\mathbf{Z}) \cong \prod_{i=1}^{M} \mathrm{Sp}_{2g}(\mathbf{Z}/p_i^{m_i}\mathbf{Z}).$$

Then ρ splits as a tensor product

$$\rho = \bigotimes_{i=1}^{M} \rho_i,$$

where ρ_i are irreducible representations of $\mathrm{Sp}_{2g}(\mathbf{Z}/p_i^{m_i}\mathbf{Z})$. Since ρ is new, all of the ρ_i are new. Now using Proposition 6.2 and the bounds for the case of prime modulus from § 6.1 gives

$$\dim \rho \geqslant \prod_{i:m_i=1} \frac{1}{2}(p_i^g - 1) \prod_{i:m_i>1} C p_i^{\lfloor m_i/2 \rfloor} \geqslant q^{1/2}(C')^{-\omega(q)}$$

given $g \geqslant 2$ for some $C' > 1$ and $\omega(q)$ standing for the number of distinct prime factors of q. But $(C')^{\omega(q)} \ll_\epsilon q^\epsilon$ for any $\epsilon > 0$. This concludes the proof of Proposition 6.1; in fact, our proof shows that one may take D as close as one likes to $1/2$ provided one chooses C appropriately.

Acknowledgements

We thank Alex Eskin, Ilya Gekhtman, Sébastien Gouëzel, Rodolfo Gutiérrez-Romo, Carlos Matheus, Babak Modami, Hee Oh and Giulio Tiozzo for helpful discussions related to this work.

References

Ath06 J. S. Athreya, *Quantitative recurrence and large deviations for Teichmüller geodesic flow*, Geom. Dedicata **119** (2006), 121–140.

AG13 A. Avila and S. Gouëzel, *Small eigenvalues of the Laplacian for algebraic measures in moduli space, and mixing properties of the Teichmüller flow*, Ann. of Math. (2) **178** (2013), 385–442.

AGY06 A. Avila, S. Gouëzel and J.-C. Yoccoz, *Exponential mixing for the Teichmüller flow*, Publ. Math. Inst. Hautes Études Sci. **104** (2006), 143–211.

AMY18 A. Avila, C. Matheus and J.-C. Yoccoz, *Zorich conjecture for hyperelliptic Rauzy–Veech groups*, Math. Ann. **370** (2018), 785–809.

Bar47 V. Bargmann, *Irreducible unitary representations of the Lorentz group*, Ann. of Math. (2) **48** (1947), 568–640.

BG08 J. Bourgain and A. Gamburd, *Uniform expansion bounds for Cayley graphs of* $\mathrm{SL}_2(\mathbb{F}_p)$, Ann. of Math. (2) **167** (2008), 625–642.

BGS11 J. Bourgain, A. Gamburd and P. Sarnak, *Generalization of Selberg's* $\frac{3}{16}$ *theorem and affine sieve*, Acta Math. **207** (2011), 255–290.

Bro86 R. Brooks, *The spectral geometry of a tower of coverings*, J. Differential Geom. **23** (1986), 97–107.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

Bur86 M. Burger, *Estimation de petites valeurs propres du laplacien d'un revêtement de variétés riemanniennes compactes*, C. R. Math. Acad. Sci. Paris Sér. I **302** (1986), 191–194.

Bur88 M. Burger, *Spectre du laplacien, graphes et topologie de Fell*, Comment. Math. Helv. **63** (1988), 226–252.

Dol98 D. Dolgopyat, *On decay of correlations in Anosov flows*, Ann. of Math. (2) **147** (1998), 357–390.

EM18 A. Eskin and M. Mirzakhani, *Invariant and stationary measures for the* $\mathrm{SL}(2,\mathbf{R})$ *action on Moduli space*, Publ. Math. Inst. Hautes Études Sci. **127** (2018), 95–324.

Fil17 S. Filip, *Zero Lyapunov exponents and monodromy of the Kontsevich–Zorich cocycle*, Duke Math. J. **166** (2017), 657–706.

GJ78 S. Gelbart and H. Jacquet, *A relation between automorphic representations of* GL(2) *and* GL(3), Ann. Sci. Éc. Norm. Supér. (4) **11** (1978), 471–542.

Gow08 W. T. Gowers, *Quasirandom groups*, Combin. Probab. Comput. **17** (2008), 363–387.

Gut19 R. Gutiérrez-Romo, *Classification of Rauzy–Veech groups: proof of the Zorich conjecture*, Invent. Math. **215** (2019), 741–778.

HS08 D. Hejhal and P. Sarnak, *Some commentary on Atle Selberg's mathematics*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), 485–487.

Iwa96 H. Iwaniec, *The lowest eigenvalue for congruence groups*, in *Topics in geometry*, Progress in Nonlinear Differential Equations and their Applications, vol. 20 (Birkhäuser Boston, Boston, MA, 1996), 203–212.

Kaž67 D. A. Každan, *On the connection of the dual space of a group with the structure of its closed subgroups*, Funktsional. Anal. i Prilozhen **1** (1967), 71–74.

KS13 D. Kelmer and L. Silberman, *A uniform spectral gap for congruence covers of a hyperbolic manifold*, Amer. J. Math. **135** (2013), 1067–1085.

Kim03 H. H. Kim, *Functoriality for the exterior square of* GL_4 *and the symmetric fourth of* GL_2, J. Amer. Math. Soc. **16** (2003), 139–183.

Kna01 A. W. Knapp, *An overview based on examples*, in *Representation theory of semisimple groups*, Princeton Landmarks in Mathematics, reprint of the 1986 original (Princeton University Press, Princeton, NJ, 2001).

KZ03 M. Kontsevich and A. Zorich, *Connected components of the moduli spaces of Abelian differentials with prescribed singularities*, Invent. Math. **153** (2003), 631–678.

LPS88 A. Lubotzky, R. Phillips and P. Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), 261–277.

LRS95 W. Luo, Z. Rudnick and P. Sarnak, *On Selberg's eigenvalue conjecture*, Geom. Funct. Anal. **5** (1995), 387–401.

Maa49 H. Maass, *Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **121** (1949), 141–183.

Mag15 M. Magee, *Quantitative spectral gap for thin groups of hyperbolic isometries*, J. Eur. Math. Soc. (JEMS) **17** (2015), 151–187.

MOW19 M. Magee, H. Oh and D. Winter, *Uniform congruence counting for Schottky semigroups in* $\mathrm{SL}_2(\mathbf{Z})$, J. Reine Angew. Math. **753** (2019), 89–135.

MR19 M. Magee and R. Rühr, *Counting saddle connections in a homology class modulo q. With an Appendix by R. Gutiérrez-Romo*, J. Mod. Dyn. **15** (2019), 237–262.

Mar73 G. A. Margulis, *Explicit constructions of expanders*, Problemy Peredachi Informatsii **9** (1973), 71–80.

MMY05 S. Marmi, P. Moussa and J.-C. Yoccoz, *The cohomological equation for Roth-type interval exchange maps*, J. Amer. Math. Soc. **18** (2005), 823–872 (electronic).

Mas82 H. Masur, *Interval exchange transformations and measured foliations*, Ann. of Math. (2) **115** (1982), 169–200.

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press

MVW84 C. R. Matthews, L. N. Vaserstein and B. Weisfeiler, *Congruence properties of Zariski-dense subgroups. I*, Proc. Lond. Math. Soc. (3) **48** (1984), 514–532.

Nor87 M. V. Nori, *On subgroups of* $\mathrm{GL}_n(\mathbf{F}_p)$, Invent. Math. **88** (1987), 257–275.

OW16 H. Oh and D. Winter, *Uniform exponential mixing and resonance free regions for convex cocompact congruence subgroups of* $\mathrm{SL}_2(\mathbb{Z})$, J. Amer. Math. Soc. **29** (2016), 1069–1115.

Rat87 M. Ratner, *The rate of mixing for geodesic and horocycle flows*, Ergodic Theory Dynam. Systems **7** (1987), 267–288.

Rau79 G. Rauzy, *Échanges d'intervalles et transformations induites*, Acta Arith. **34** (1979), 315–328.

Sar95 P. Sarnak, *Selberg's eigenvalue conjecture*, Notices Amer. Math. Soc. **42** (1995), 1272–1277.

Sar05 P. Sarnak, *Notes on the generalized Ramanujan conjectures*, in *Harmonic analysis, the trace formula, and Shimura varieties*, Clay Mathematics Proceedings, vol. 4 (American Mathematical Society, Providence, RI, 2005), 659–685.

SX91 P. Sarnak and X. X. Xue, *Bounds for multiplicities of automorphic representations*, Duke Math. J. **64** (1991), 207–227.

SZ93 G. M. Seitz and A. E. Zalesskii, *On the minimal degrees of projective representations of the finite Chevalley groups. II*, J. Algebra **158** (1993), 233–243.

Sel56 A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) **20** (1956), 47–87.

Sel65 A. Selberg, *On the estimation of Fourier coefficients of modular forms*, Proceedings of Symposia in Pure Mathematics, vol. 8 (American Mathematical Society, Providence, RI, 1965), 1–15.

SS70 T. A. Springer and R. Steinberg, *Conjugacy classes*, in *Seminar on Algebraic Groups and Related Finite Groups (The Institute for Advanced Study, Princeton, NJ, 1968/69)*, Lecture Notes in Mathematics, vol. 131 (Springer, Berlin, 1970), 167–266.

Vee82 W. A. Veech, *Gauss measures for transformations on the space of interval exchange maps*, Ann. of Math. (2) **115** (1982), 201–242.

Via08 M. Viana, *Dynamics of interval exchange transformations and Teichmüller flows*, Lecture Notes, IMPA (2008), http://w3.impa.br/∼viana/out/ietf.pdf.

Zor96 A. Zorich, *Finite Gauss measure on the space of interval exchange transformations. Lyapunov exponents*, Ann. Inst. Fourier (Grenoble) **46** (1996), 325–370.

Zor99 A. Zorich, *How do the leaves of a closed 1-form wind around a surface?* in *Pseudoperiodic topology*, American Mathematical Society Translations Series 2, vol. 197 (American Mathematical Society, Providence, RI, 1999), 135–178.

Zor06 A. Zorich, *Flat surfaces*, in *Frontiers in number theory, physics, and geometry I* (Springer, Berlin, 2006), 437–583.

Michael Magee michael.r.magee@durham.ac.uk
Department of Mathematical Sciences, Durham University, Lower Mountjoy,
Stockton Rd, Durham DH1 3LE, UK

https://doi.org/10.1112/S0010437X1900767X Published online by Cambridge University Press