EFFECTIVE COUNTING ON TRANSLATION SURFACES

AMOS NEVO, RENE RÜHR, AND BARAK WEISS

Abstract. We prove an effective version of a celebrated result of Eskin and Masur: for any $\mathrm{SL}_2(\mathbb{R})$-invariant locus $\mathcal{L}$ of translation surfaces, there exists $\kappa > 0$, such that for almost every translation surface in $\mathcal{L}$, the number of saddle connections with holonomy vector of length at most T, grows like $cT^2 + O(T^{2-\kappa})$. We also provide effective versions of counting in sectors and in ellipses.

1. Introduction

The main goal of this paper is the effectivization of a celebrated result of Eskin and Masur [EM01] which we recall. A translation surface $\mathbf{x}$ is a compact oriented surface equipped with an atlas of planar charts, whose transition maps are translations, where the charts are defined at every point of the surface except finitely many singular points at which the planar structure completes to form a cone point of angle an integer multiple of 2π. Such structures arise in many contexts in geometry, complex analysis and dynamics, and have various equivalent definitions, see the surveys [MT, Z] for more details. The collection of all translation surfaces of a fixed genus, fixed number of singular points, and fixed cone angle at each singular point is called a *stratum*, and has a natural structure of a linear orbifold. Furthermore each connected component of the subset of area one surfaces in a stratum is the support of a natural smooth probability measure which we will call *flat measure*.

A *saddle connection* on a translation surface $\mathbf{x}$ is a segment connecting two singular points which is linear in each planar chart and contains no singular points in its interior. The *holonomy vector* of a saddle connection is the vector in the plane obtained by integrating the pullback of the planar form (dx, dy), along the saddle connection. We denote the collection of all holonomy vectors for $\mathbf{x}$ by $V(\mathbf{x})$. The large scale geometry of $V(\mathbf{x})$ has been intensively studied, and one of the main results of [EM01] is that there is $c > 0$ such that for a.e. $\mathbf{x}$ (with respect to the flat measure), the number $N(T, \mathbf{x}) = |V(\mathbf{x}) \cap B(0, T)|$ satisfies

$$N(T, \mathbf{x}) = cT^2 + o\left(T^2\right). \tag{1.1}$$

When this holds we will say that $\mathbf{x}$ *satisfies quadratic growth*.

The main purpose of this paper is to estimate the error term in the above result, that is to establish that

$$N(T, \mathbf{x}) = cT^2 + O\left(T^{2(1-\kappa)}\right)$$

for some $\kappa > 0$. In order to state our result in its full generality we need to introduce more precise terminology.

Let $\mathcal{H}$ be a stratum of translation surfaces, let $G = \mathrm{SL}_2(\mathbb{R})$ and let $\mathcal{L} \subset \mathcal{H}$ be the closure of a G-orbit in $\mathcal{H}$. By recent breakthrough results of Eskin, Mirzakhani and Mohammadi [EMi15, EMiMo15], $\mathcal{L}$ is the intersection of $\mathcal{H}$ with a linear suborbifold, and is the support of a smooth ergodic probability measure μ, which we will call the *flat measure of* $\mathcal{L}$. We will refer to $(\mathcal{L}, \mu)$ as a *locus* (the terminology 'affine invariant manifold' is also in common use).

A *cylinder* on a translation surface is an isometrically embedded image of the annulus $[a_1, a_2] \times \mathbb{R}/c\mathbb{Z}$, for some $a_1 < a_2$ and $c > 0$. The image of a curve $\{b\} \times \mathbb{R}/c\mathbb{Z}$ for $a_1 < b < a_2$ is called a *waist curve* of the cylinder and the integral along a waist curve of the pullback of (dx, dy) is called the *holonomy vector* of the cylinder. One can also study the asymptotic growth of $V^{\mathrm{cyl}}(\mathbf{x}) \cap B(0, T)$, where $V^{\mathrm{cyl}}(\mathbf{x})$ is the collection of holonomy vectors of cylinders on $\mathbf{x}$. Furthermore, in [EMZ, §3], Eskin, Masur and Zorich defined *configurations* which are a common generalization of saddle connections and cylinders. We will not need to repeat the definition of a configuration in this paper; in order to give the idea, we note three other examples of configurations: (i) $\mathcal{C}$ consists of a saddle connection joining some fixed singularity to itself, (ii) $\mathcal{C}$ is a saddle connection joining distinct fixed singularities, (iii) $\mathcal{C}$ consists of two homologous saddle connections joining two distinct fixed singularities, and forming a slit which disconnects the surface into components with a fixed topology. For each configuration $\mathcal{C}$ one can then define a collection of holonomy vectors $V^{\mathcal{C}}(\mathbf{x})$ of the saddle connections or cylinders comprising the configuration, and study the asymptotic growth of $N^{\mathcal{C}}(T, \mathbf{x}) = |V^{\mathcal{C}}(\mathbf{x}) \cap B(0, T)|$. A remarkable feature of [EM01, EMZ] is the authors' foresight: they proved their results in an abstract framework which later (in [EMi15]) was proved to be sufficient to cover all G-invariant ergodic measures and all configurations. Namely, they proved that for any locus $(\mathcal{L}, \mu)$ and any configuration $\mathcal{C}$ there is $c = c(\mathcal{L}, \mathcal{C})$ such that for μ-a.e. $\mathbf{x} \in \mathcal{L}$ one has $N^{\mathcal{C}}(T, \mathbf{x}) = cT^2 + o(T^2)$. Furthermore, [EMZ] also discussed counting with multiplicities (that is, vectors in $\mathbb{R}^2$ are counted according to the number of saddle connections which have them as holonomy vectors). In the notation of this paper $N^{\mathcal{C}}(T, \mathbf{x})$ may refer to counting either with or without multiplicity, i.e. the count in question is assumed to be a part of the data associated with $\mathcal{C}$. Finally, for the case $\mathcal{L} = \mathcal{H}$, an algorithm for computing the constants c in the above asymptotic was described, in terms of so-called Siegel-Veech constants introduced by Veech in [Vee98].

An additional improvement, due to Vorobets [Vo05, Thm. 1.9], concerns counting in sectors. Let $\varphi_1 < \varphi_2$ with $\varphi_2 - \varphi_1 \leq 2\pi$ and let $N(T, \mathbf{x}, \varphi_1, \varphi_2)$ denote the cardinality of the intersection of $V(\mathbf{x})$ with the sector

$$S_{T,\varphi_1,\varphi_2} = \{r(\cos\varphi, \sin\varphi) \,:\, 0 \leq r \leq T,\ \varphi_1 \leq \varphi \leq \varphi_2\} \subset \mathbb{R}^2.$$

Vorobets showed that there is $c > 0$ such that for a.e. $\mathbf{x} \in \mathcal{H}$ (with respect to the flat measure on $\mathcal{H}$), $N(T, \mathbf{x}, \varphi_1, \varphi_2) = c(\varphi_2 - \varphi_1)T^2 + o(T^2)$. Our main result is an effective version of the above-mentioned results. Setting $N^{\mathcal{C}}(T, \mathbf{x}, \varphi_1, \varphi_2)$ for the number of holonomy vectors corresponding to the configuration $\mathcal{C}$ on $\mathbf{x}$ with holonomy vector in $S_{T,\varphi_1,\varphi_2}$, we have:

Theorem 1.1. *For any locus $(\mathcal{L}, \mu)$ there is a constant $\kappa > 0$ such that for any configuration $\mathcal{C}$ there is a constant $c > 0$ such that for any $\varphi_1 < \varphi_2$ with $\varphi_2 - \varphi_1 \leq 2\pi$, for μ-a.e. $\mathbf{x}$ we have*

$$N^{\mathcal{C}}(T, \mathbf{x}, \varphi_1, \varphi_2) = \frac{c}{2}(\varphi_2 - \varphi_1)T^2 + O_{\mathbf{x},\varphi_2-\varphi_1}\left(T^{2(1-\kappa)}\right). \tag{1.2}$$

Here, in the basic case that $\mathcal{L} = \mathcal{H}$ is a stratum and $\mathcal{C}$ is one saddle connection (i.e. $V^{\mathcal{C}}(\mathbf{x}) = V(\mathbf{x})$), the constant c is the Siegel-Veech constant of [Vee98] (this is the reason for the denominator 2 appearing in (1.2)). As we shall see below, κ can be estimated explicitly in terms of the size of the spectral gap in the unitary representation of G in $L^2(\mathcal{L})$; in fact our proof shows that it can be taken to be $\frac{\lambda}{5.5+5.5\lambda}$ for any λ smaller than the size of the gap. We have chosen to normalize our power savings exponent κ so that the error is written in the form $2(1 - \kappa)$ rather than $2 - \kappa$, that is to estimate the error as a power of the area growth, in order to permit easier comparisons with other bounds appearing in the literature on related problems. Note that in (1.2), the dependence of the implicit constant in the O-notation on $\mathbf{x}$ is unavoidable given the existence of surfaces with different quadratic growth coefficients.

The proof of Theorem 1.1 does not give any insight into the set of full measure of $\mathbf{x}$ which satisfy (1.2). In fact it is expected that *every* translation surface $\mathbf{x}$ satisfies quadratic growth (see [EMiMo15] for a remarkable result in this direction). Thus it is of interest to exhibit explicit surfaces which satisfy quadratic growth with an effective error estimate (in particular, where κ is known). It is also of interest to count points in the intersection of $V(\mathbf{x})$ with more general subsets of $\mathbb{R}^2$. These questions will be discussed in the forthcoming work [BNRW].

The expectation that any translation surface satisfies quadratic growth, and that the constant c appearing in (1.1) depends only on the orbit closure $\overline{G\mathbf{x}}$, leads to the expectation that the set of surfaces satisfying (1.1) is G-invariant. Since the assignment $\mathbf{x} \mapsto V(\mathbf{x})$

satisfies $V(g\mathbf{x}) = gV(\mathbf{x})$, this can be equivalently stated as a problem on counting in ellipses: in the definition of $N(T, \mathbf{x})$, one should be able to replace Euclidean balls of radius T, with dilates of any fixed ellipse centered at the origin, and the same should be true for $N(T, \mathbf{x}, \varphi_1, \varphi_2)$. The issue of existence of a full measure G-invariant set of surfaces with quadratic growth was not discussed in [EM01], but could probably be derived from the arguments in [EM01, Vo05]. Moreover, it can be derived from a recent result of Athreya, Cheung and Masur [ACM], in combination with an argument of Veech [Vee98, Thm. 14.11]. Using our technique we obtain the following effective strengthening.

Theorem 1.2. *For any locus $(\mathcal{L}, \mu)$, there is $\kappa > 0$ such that for every configuration $\mathcal{C}$ there is $c > 0$ such that for μ-a.e. $\mathbf{x}$, for every $\varphi_1 < \varphi_2$ with $\varphi_2 - \varphi_1 \leq 2\pi$, and for every $g \in G$,*

$$N^{\mathcal{C}}(T, g\mathbf{x}, \varphi_1, \varphi_2) = \frac{c}{2}(\varphi_2 - \varphi_1)T^2 + O_{\mathbf{x}, \varphi_2 - \varphi_1, g}\left(T^{2(1-\kappa)}\right).$$

Note that Theorem 1.2 implies Theorem 1.1, but we present the proof of Theorem 1.1 separately. This is because the proof of Theorem 1.2 presents additional technicalities which may obscure the main ideas, and also because our proof of Theorem 1.2 gives slightly weaker estimates on κ.

1.1. **Ingredients of the proofs.** Our proof of Theorem 1.1 follows the strategy of [EM01] (which in turn was inspired by [EMaMo98, Vee98]) of reducing the counting problem to an ergodic theoretic problem regarding the convergence of the translated circle averages $\pi_{\mathcal{L}}(\Sigma_t)f(\mathbf{x}) = \int_K f(a_t k \mathbf{x}) dm_K$ (the notation is introduced in §2.2), as $t \to \infty$. In the treatment of [EM01], f is the Siegel-Veech transform of an indicator of a rectangle in $\mathbb{R}^2$, and the required convergence of $\pi_{\mathcal{L}}(\Sigma_t)f(\mathbf{x})$ was proved by replacing f with a smoothed version of f, developing various estimates to bound the amount of time the translated circle average spends outside large compact subsets of $\mathcal{H}$, and appealing to a pointwise ergodic theorem of the first-named author (see [N]).

Our proof of Theorem 1.1 uses all of the above ingredients and more. The essential new ingredient is the fact that any $(\mathcal{L}, \mu)$ possesses a spectral gap (see §3 for the definition). This was proved by Avila, Gouëzel and Yoccoz [AGY06] for the case of strata, and by Avila and Gouëzel [AG13] for general loci (again, in an abstract framework, as [AG13] also preceded [EMi15]). Using the spectral gap it is possible to obtain an effective estimate of the difference $|\pi_{\mathcal{L}}(\Sigma_t)f(\mathbf{x}) - \int_{\mathcal{L}} f d\mu|$, in case f is a K-smooth function and t is large enough (depending on $\mathbf{x}$ and f). See §3 for the definition of K-smooth functions. The estimate is valid for $\mathbf{x}$ in a set of large measure depending on f and t. Using a Borel-Cantelli argument (see Theorem 3.4) we upgrade this to a set of full measure and a countable collection of K-smooth functions, which we then use in order to estimate effectively the integrals appearing in the counting problem, and thus the numbers $N^{\mathcal{C}}(T_n, \mathbf{x}, \varphi_1, \varphi_2)$ for a countable collections of radii (T_n). In order to pass from a countable collection of functions to the results, it is advantageous to replace the rectangle used in [EM01], or the trapezoid used in [E06], with a triangle with an apex at the origin.

Theorem 1.2 improves Theorem 1.1 in two ways: uniform counting with an error term in all sectors and in all ellipses. These improvements require two additional ingredients. First we note that the same Borel-Cantelli argument, and further approximation arguments, make it possible to use countably many functions in order to approximate all sectors and all ellipses simultaneously. That is, instead of working only with a countable set of radii, we work with a countable set of radii, a countable set of ellipses, and a countable collections of sectors. Furthermore, for uniform counting in ellipses, we replace the circle averages with ellipse averages $\pi_{\mathcal{L}}\left(\Sigma_t^{(g)}\right) f(\mathbf{x}) = \int_K f(a_t k g \mathbf{x}) dm_K$, and obtain an estimate on the rate at which $\pi_{\mathcal{L}}\left(\Sigma_t^{(g)}\right)(\mathbf{x}) \to \int_{\mathcal{L}} f d\mu$, which is uniform as g ranges over compact subsets of G.

1.2. **Acknowledgements.** The authors were supported by ERC starter grant DLGAPS 279893 and ISF grant 2095/15.

2. Preliminaries

In this section we will collect results which we will need concerning the moduli space of translation surfaces.

2.1. The Siegel-Veech formula and the function ℓ.

We recall the Siegel-Veech summation formula:

Theorem 2.1. [Vee98, Thm. 0.5] *For any locus $\mathcal{L}$ with flat measure μ, and any configuration $\mathcal{C}$, there exists $c = c(\mathcal{L}, \mathcal{C}) > 0$ (called a Siegel-Veech constant) such that for any $\psi \geq 0$ Borel measurable on $\mathbb{R}^2$, if we let $\widehat{\psi}(\mathbf{x}) = \sum_{v \in V^{\mathcal{C}}(\mathbf{x})} \psi(v)$, then*

$$\int_{\mathcal{L}} \widehat{\psi}(\mathbf{x}) d\mu(\mathbf{x}) = c \int_{\mathbb{R}^2} \psi(x) dx.$$

We stress that the definition of $\widehat{\psi}$ depends on a choice of configuration $\mathcal{C}$, but this choice will not play an important role in what follows, and will be suppressed from the notation.

Let $\ell(\mathbf{x})$ be the Euclidean length of a shortest saddle connection in $\mathbf{x}$. Building on earlier work in [M] and [EMaMo98], a fundamental bound on the number of saddle connections in a compact set was established by Eskin and Masur as follows.

Theorem 2.2. [EM01, Theorem 5.1] *For any stratum $\mathcal{H}$, any configuration $\mathcal{C}$, any compact set $B \subset \mathbb{R}^2$, any $\mathbf{x} \in \mathcal{H}$, and any $\alpha_1 > 1$,*

$$\left| V^{\mathcal{C}}(\mathbf{x}) \cap B \right| \ll_{\mathcal{H}, B, \alpha_1} \ell(\mathbf{x})^{-\alpha_1}.$$

Note that in [EM01], the bound was only stated for the set $V(\mathbf{x})$ of all saddle connection holonomies, that is the case in which the configuration $\mathcal{C}$ consists of any saddle connection; however since any cylinder contain saddle connections along its boundary, the bound for $V(\mathbf{x})$ implies the same bound for $V^{\mathcal{C}}(\mathbf{x})$ for any configuration $\mathcal{C}$.

2.2. Translated circle averages.

Consider the elements

$$a_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \quad r_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{2.1}$$

and let $K = \{r_\theta : \theta \in \mathbb{R}\} \subset G$. When G acts ergodically by measure preserving transformations on a standard Borel probability space (X, μ), we will say that (X, μ) is an ergodic p.m.p. G-space. We let π_X denote the unitary representation of G in $L^2(X)$, given by $\pi_X(g)f(x) = f(g^{-1}x)$. We extend π_X to a representation of the convolution algebra $M(G)$ of bounded complex Borel measures on G. Each $\sigma \in M(G)$ acts as an operator on $L^2(X)$ via the formula

$$\pi_X(\sigma)f(x) = \int_G f(g^{-1}x) d\sigma(g), \quad \text{for } f \in L^2(X).$$

For any two measures $\sigma_1, \sigma_2 \in M(G)$, we have $\pi_X(\sigma_1 * \sigma_2) = \pi_X(\sigma_1) \circ \pi_X(\sigma_2)$.

Let m_K denote the probability Haar measure on the circle K given in coordinates by $\frac{1}{2\pi} d\theta$, and denote the probability measure $m_K * \delta_{a_{-t}}$ by Σ_t. Thus for $f : \mathcal{L} \to \mathbb{R}$

$$\pi_X(\Sigma_t) f(\mathbf{x}) = \int_K f(a_t k \mathbf{x}) dm_K(k).$$

An important property of integrability of the function ℓ, and a bound on its translated circle averages, were established by Eskin and Masur:

Theorem 2.3 (See [EM01] Thm. 5.2, Lem. 5.5 and [Vee98], Cor 2.8)**.** *For any $\mathbf{x} \in \mathcal{L}$, and for any $1 \leq \alpha_2 < 2$,*

$$\sup_{t > 0} \pi_{\mathcal{L}}(\Sigma_t) \left(\ell(\mathbf{x})^{-\alpha_2} \right) < \infty. \tag{2.2}$$

The bound can be taken to be uniform as $\mathbf{x}$ ranges over compact sets in $\mathcal{L}$. Furthermore, for any locus $(\mathcal{L}, \mu)$, $\ell(\cdot)^{-\alpha_2} \in L^1(\mathcal{L}, \mu)$.

To account for sectors, we will use the family of measures on the circle

$$\pi_X(\Sigma_{\nu,t})f(\mathbf{x}) = \int_K f(a_t k\mathbf{x})\nu(k)dm_K(k)$$

where ν is a bounded density on K. In fact, in this paper ν will be a characteristic function of the angular sector $[\varphi_1, \varphi_2]$ that is either slightly contracted or slightly thickened. It is clear that (2.2) also holds for such $\Sigma_{\nu,t}$, uniformly for all $\nu \leq 1$.

3. Spectral gap and pointwise ergodic theorem

Let (X, μ) be a p.m.p. G-space, and denote by $L^2_0(X, \mu)$ the zero mean functions in $L^2(X, \mu)$. The action is said to have a *spectral gap* if the associated unitary representation of G is isolated from the trivial representation. That is, there does not exist a sequence of unit vectors $(u_n)_{n\in\mathbb{N}}$ in $L^2_0(X, \mu)$ which is asymptotically invariant under the representation, namely such that $\lim_{n\to 0} \|\pi_X(g)u_n - u_n\| = 0$ for every g in G.

In the special case $G = \mathrm{SL}_2(\mathbb{R})$ the spectral gap condition implies the following quantitative estimate (see [C, R, CHH88] and references therein). It is formulated in terms of K-eigenvectors, namely functions $f \in L^2(X)$ satisfying that there exists a character χ of K such that $\pi_X(k)f = \chi(k)f$ for all $k \in K$.

Theorem 3.1. *Let $G = \mathrm{SL}_2(\mathbb{R})$ and let (X, μ) be a p.m.p. G-space with a spectral gap. Then there are positive C, λ such that for any $f_1, f_2 \in L^2_0(X)$ which are eigenvectors for K, and for any $g \in G$, written in Cartan polar coordinates as $g = k_1 a_t k_2$, we have*

$$|\langle \pi_X(g)f_1, f_2\rangle| \leq Ce^{-|t|\lambda}\|f_1\|_2\|f_2\|_2\,. \tag{3.1}$$

The supremum of $\lambda > 0$ for which one can find C such that (3.1) is satisfied for K-eigenvectors f_1, f_2, will be denoted by λ_X and will be called the *size of the spectral gap*.

We briefly review some basic results and explain the relation of our notation to that used in other papers. The functions $g \mapsto \langle \pi_X(g)f_1, f_2\rangle$, for $f_i \in L^2(X, \mu)$, are known as *matrix coefficients* for the action on (X, μ). If f is a finite linear combination of K-eigenvectors, it is called *K-finite*. It is known (see [C, Thm. 2.5.3]) that if the action of G on (X, μ) has a spectral gap, then there is $q < \infty$ such that all its matrix coeffients are in $L^q(G)$. Suppose (X, μ) has a spectral gap, and let p be the infimum of $q < \infty$ such that for every $\varepsilon > 0$ and every pair of K-finite functions f_1, f_2, the matrix coefficient $g \mapsto \langle \pi_X(g)f_1, f_2\rangle$ is in $L^q(G)$; this parameter is often called the *integrability exponent* of π_X. By [C, Cor. 2.2.4], the integrability exponent p is also the smallest number such that $\langle \pi_X(g)f_1, f_2\rangle = O(\|g\|^{-2/(p+\varepsilon)})$ for all $\varepsilon > 0$, with implicit constant depending on ε, f_1, f_2, and on the norm chosen on $\mathrm{Mat}_2(\mathbb{R})$. It follows from more detailed estimates obtained in [CHH88, Corollary (b)], that in terms of the integrability exponent, the size of the spectral gap is $\lambda_X = 2/p$. The case in which the unitary representation is tempered, i.e. absence of the complementary series in π_X (see the discussion in [AGY06, Prop. B.2]), one has $p = 2$ which in our notation corresponds to $\lambda_X = 1$.

Our results are based on the following important result:

Theorem 3.2. [AGY06, AG13] *The representation of G on $L^2_0(\mathcal{L}, \mu)$ possesses a spectral gap.*

Note that the results of [AGY06, AG13] do not give explicit bounds on the size of the spectral gap.

A standard norm-comparison argument implies that for any choice of norm on $\mathrm{Mat}_2(\mathbb{R})$, for any $f_1, f_2 \in L^2_0(X)$ which are eigenvectors for K, and for $0 < \lambda < \lambda_X$,

$$|\langle \pi_X(g)f_1, f_2\rangle| \ll \|g\|^{-\lambda}\, \|f_1\|_2\|f_2\|_2. \tag{3.2}$$

Fix $\omega = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ as a generator of the Lie algebra of K. A function $f \in L^2(X)$ is called *K-smooth of degree one* if

$$\pi_X(\omega)f = \lim_{t\to 0} \frac{1}{t}\left(\pi_X(\exp(t\omega)f - f\right) \tag{3.3}$$

exists, where the convergence is with respect to the $L^2(X)$-norm (one may also consider the obvious extension to smoothness of degree d for ω, but we will not need this). Define the (degree one) Sobolev norm by

$$\mathcal{S}_K(f)^2 = \|f\|_2^2 + \|\pi_X(\omega)f\|_2^2.$$

We denote the space of K-Sobolev functions with finite $\mathcal{S}_K(f)$-norm by $\mathcal{S}_K(X)$, and set

$$\mathcal{S}_{K,0}(X) = \mathcal{S}_K(X,\mu) \cap L_0^2(X,\mu).$$

The estimates (3.1) and (3.2), which are stated in terms of K-eigenvectors, can also be formulated in terms of K-Sobolev functions $f_1, f_2 \in \mathcal{S}_{K,0}(X)$, as follows. For $g = k_1 a_t k_2$

$$|\langle \pi_X(g) f_1, f_2\rangle| \leq C'(\lambda) e^{-|t|\lambda} \mathcal{S}_K(f_1)\mathcal{S}_K(f_2),$$

and for any matrix norm on $\mathrm{Mat}_2(\mathbb{R})$ and any $0 < \lambda < \lambda_X$,

$$\langle \pi_X(g) f_1, f_2\rangle \ll \|g\|^{-\lambda} \mathcal{S}_K(f_1)\mathcal{S}_K(f_2) \tag{3.4}$$

which follows from a standard proof of the Sobolev inequality via Fourier series (see e.g. [Ei10]).

As shown by Eskin, Margulis and Mozes, from estimates such as those in Theorem 3.1, one can derive an estimate for the norm of the operator $\pi_X(\Sigma_{\nu,t})$, $\nu \leq 1$, viewed as operator from the K-Sobolev space $\mathcal{S}_{K,0}(X)$ to $L^2(X)$.

Theorem 3.3 (See [EMaMo98](3.32) and [Vee98], §14)**.** *Let $G = \mathrm{SL}_2(\mathbb{R})$ and let (X,μ) be a p.m.p. G-space with a spectral gap of size λ_X. Then for any $\lambda < \lambda_X$, there exists $C > 0$ such that for any interval $I \subset \mathbb{S}^1 \simeq K$ of length $|I| \neq 0$, any $f \in \mathcal{S}_{K,0}(X)$, and all $t > \frac{1}{2}\log\frac{1}{|I|}$,*

$$\|\pi_X(\Sigma_{\nu,t})f\|_2^2 \leq C e^{-2\lambda\eta t} \mathcal{S}_K(f)^2 \, |I|^{2-\lambda\eta}\,, \tag{3.5}$$

where $\eta = \frac{1}{\lambda+1}$ and ν is the indicator function of the interval I.

We note that if one normalizes ν to be the density of a probability measure then the quality of the rate in (3.5) diminishes as the length of the interval decreases. We will not normalize ν in this way because it will turn out to be less natural for some geometric considerations involved in the counting problem.

For completeness, and in order to have precise control of constants, we repeat the argument found in [EMaMo98, Vee98].

Proof. Writing $\nu^* * \nu(k) = \int_K \nu(k')\nu(k'k)dm_K(k') = D_\nu(k)$, by G-invariance of the measure μ one has

$$\|\pi_X(\Sigma_{\nu,t})f\|_2^2 = \langle \pi_X(\delta_{a_t} * \nu^* * \nu * \delta_{a_{-t}})f, f\rangle = \int_X \int_K f(a_t k a_{-t} x)\bar{f}(x) D_\nu(k) dm_K(k) d\mu(x). \tag{3.6}$$

Replacing I with a slightly smaller interval will only require a small increase in C, so with no loss of generality we can assume that $|I| \leq \pi/4$. Then using a rotation we can assume that $I = [0, \phi_I] \subset [0, \pi/4]$. We identify K with $\mathbb{S}^1$ using (2.1) so that $\nu(K) = \frac{|I|}{2\pi} \leq |I|$. For a parameter $0 < \lambda' < 1$ to be fixed below, we set $J = \{\phi \in I : |\sin\phi| e^{2t} < e^{2\lambda' t}\}$. To put ourselves in the case that J is a proper subset of I, we assume that $\sin\phi_I > e^{2t(\lambda'-1)}$, which implies that $\phi_I = |I| > e^{2t(\lambda'-1)}$, and thus $\lambda' < \frac{\log|I|}{2t} + 1$.

Write $k = r_\phi$ so that, using the supremum norm in (3.4), we have $\|a_t k a_{-t}\| \geq e^{2\lambda' t}$ for $k \in I \setminus J$. Since $I \subset [0, \frac{\pi}{4}]$, $\nu^* * \nu$ can be computed using convolution on $\mathbb{R}$. Since $\phi \leq 2\sin\phi$ in the interval I, and since $\|D_\nu\|_\infty \leq |I|$, we conclude that $\int_J D_\nu(k) dm_K < 2\,|I|\, e^{2t(\lambda'-1)}$. Furthermore, clearly $\int_{I\setminus J} D_\nu(k) dm_K \leq |I|^2$. By Fubini, the matrix coefficient (3.6) is equal to

$$\int_J \langle \pi_X(a_t k a_{-t})f, f\rangle D_\nu(k) dm_K(k) + \int_{I\setminus J} \langle \pi_X(a_t k a_{-t})f, f\rangle D_\nu(k) dm_K(k).$$

We apply the previous estimate to the integral over J, and apply (3.4) to the integral over $I \smallsetminus J$, to arrive at

$$\|\pi_X(\Sigma_{\nu,t})f\|_2^2 \ll |I|\, e^{2t(\lambda'-1)}\|f\|_2^2 + |I|^2\, e^{-2\lambda\lambda' t}\mathcal{S}_K(f)^2 \le \left(|I|\, e^{2t(\lambda'-1)} + |I|^2\, e^{-2\lambda\lambda' t}\right)\mathcal{S}_K(f)^2.$$

To obtain (3.5), note that both terms on the right hand side are exponentially decreasing in t, but one term increases with λ' and the other decreases. Thus the best choice is to take λ' for which both terms are equal, and this yields (3.5). More precisely, we set:

$$\lambda' = \frac{1}{\lambda+1}\left(\frac{1}{2t}\log|I| + 1\right). \tag{3.7}$$

With this choice, using $t > \frac{1}{2}\log\frac{1}{|I|}$ and $|I| < 1$ we find $0 < \lambda' < 1$. □

The next result follows from the bound (3.5) combined with the Borel-Cantelli Lemma and the Markov inequality.

Theorem 3.4. *Let (X,μ) be a p.m.p. G-space with a spectral gap of size λ_X. Let $\lambda < \lambda_X$, let $t_n \in \mathbb{R}_+$, let $\eta = \frac{1}{1+\lambda}$ and let η_1 be such that*

$$\sum_{n\in\mathbb{N}} e^{-\lambda\eta_1 t_n} < \infty. \tag{3.8}$$

Let $0 \le \nu_n \le 1$ be a sequence of functions on K as in Theorem 3.3, satisfying $\nu_n(K) = \int \nu_n dm_K > e^{-2t_n}$. Let $(f_n)_{n\in\mathbb{N}}$ be a collection of functions in $\mathcal{S}_{K,0}(X)$. Then for almost all $x \in X$ there exists $n_0 = n_0(x)$ such that if $n \ge n_0$ then

$$|\pi_X(\Sigma_{\nu_n,t_n})f_n(x)| \le e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}\mathcal{S}_K(f_n)\nu_n(K)^{1-\frac{\lambda\eta}{2}}. \tag{3.9}$$

Here η is as in (3.5).

Note that we will only be interested in the nontrivial case where the right hand side of (3.9) decays with t, i.e. when η_1 satisfies $0 < \frac{\eta_1}{2} < \eta$.

Proof. Using (3.5), there is $C > 0$ such that for $f_n \in \mathcal{S}_{K,0}(X)$ we have

$$\|\pi_X(\Sigma_{\nu_n,t_n})f_n\|_2^2 \le Ce^{-2\lambda\eta t_n}C_n, \text{ where } C_n = \mathcal{S}_K(f_n)^2\nu_n(K)^{2-\lambda\eta}. \tag{3.10}$$

Consider for each n the set of 'bad points'

$$U_n = \left\{x \,:\, e^{-\lambda\eta_1 t_n/2}\,|\pi_X(\Sigma_{\nu_n,t_n})f_n(x)| \ge e^{-\lambda\eta t_n}C_n^{1/2}\right\}.$$

By Markov's inequality and (3.10),

$$\mu(U_n) \le e^{-\lambda\eta_1 t_n}\frac{\|\pi_X(\Sigma_{\nu_n,t_n})f_n\|_2^2}{e^{-2\lambda\eta t_n}C_n} \le Ce^{-\lambda\eta_1 t_n}.$$

By (3.8) $\sum_{n\in\mathbb{N}}\mu(U_n) < \infty$, so by the Borel-Cantelli lemma, almost every $x \in X$ belongs to at most finitely many of the sets U_n. We conclude that for almost every $x \in X$, there exists n_0 such that for all $n \ge n_0$ we have $x \notin U_n$. □

Now let

$$\pi_X\left(\Sigma_{\nu,t}^{(g)}\right)f(x) = \pi_X(\Sigma_{\nu,t})f(gx) = \int_K f(a_t kgx)\nu(k)dm_K(k)$$

denote the 'dilated ellipse average' associated with $g \in G$. For the proof of Theorem 1.2 we will need the following uniform versions of Theorems 3.3 and 3.4:

Theorem 3.5. *With the notations of Theorems 3.3 and 3.4, for every $\lambda < \lambda_X$ there exists $C > 0$ such that for all $t > 1$, any interval $I \subset \mathbb{S}^1$ with $|I| > e^{-2t}$, any $f \in \mathcal{S}_{K,0}(X)$, and any $g \in G$, we have*

$$\left\|\pi_X\left(\Sigma_{\nu,t}^{(g)}\right)f\right\|_2^2 \le Ce^{-2\lambda\eta t}\mathcal{S}_K(f)^2|I|^{2-\lambda\eta}. \tag{3.11}$$

Furthermore, if $(t_n) \subset \mathbb{R}_+$, $\eta_1 > 0$ satisfy (3.8)*, $0 \le \nu_n \le 1$ is a sequence of characteristic functions on K satisfying $\nu_n(K) > e^{-2t_n}$, (f_n) is a sequence of functions in $\mathcal{S}_{K,0}$, and (g_n)*

is a countable subset of G, then for almost all $x \in X$ there is n_0 such that for all $n \geq n_0$ we have

$$\left|\pi_X\left(\Sigma^{(g_n)}_{t_n,\nu_n}\right) f(x)\right| \leq e^{-(\eta-\frac{\eta_1}{2})\lambda t_n} \mathcal{S}_K(f_n)\nu_n(K)^{1-\frac{\lambda\eta}{2}}.$$

Proof. A change of variables $y = gx$ shows that $\|\pi_X(\Sigma^{(g)})_{t,\nu} f\|_2 = \|\pi_X(\Sigma_{t,\nu})f\|_2$ and thus (3.11) follows from the same argument as for (3.5).

The proof of the second assertion for any fixed choice of sequence (g_n) is similar to the proof of Theorem 3.4, using (3.11) instead of (3.5). □

4. Control over the cusp

The results of the previous section apply to every action with a spectral gap, and we will want to apply them to the action on the moduli space of flat surfaces, taking the functions f_n to be Siegel-Veech transforms of compactly supported functions on $\mathbb{R}^2$. However in this setting, the Sobolev norms $\mathcal{S}_K(f_n)$ might not be bounded, owing to a large contribution coming from surfaces in the thin part, i.e. surfaces $\mathbf{x}$ with $\ell(\mathbf{x})$ small. When dealing with this issue it is helpful to note that the Sobolev norm we have used above involves only differentiation in the K-direction, and as we shall now see, this fact will allow us to use a simple argument for "cutting off the cusp". We let $M_\varepsilon = \ell^{-1}([\varepsilon,\infty))$. By a well-known compactness criterion (see [MT]) the sets M_ε are an exhaustion of $\mathcal{H}$ by compact sets. Theorems 2.2 and 2.3 give bounds on the measure of the complement $M_\varepsilon^c = \mathcal{H} \smallsetminus M_\varepsilon$, and on the time a translated circle spends in M_ε^c. We will use these to cut off any function at the cusp without affecting its asymptotic behavior.

Before proceeding with this argument, note that since we used the Euclidean metric in the definition of the function ℓ, the set M_ε^c is K-invariant, and hence its characteristic function is K-smooth. Below we let ∂_θ denote the partial derivative in the spherical direction in polar coordinates. In terms of the action of K on the plane, it is defined as $\pi_{\mathbb{R}^2}(\omega)$ in the notation (3.3). Equivalently, at a point $\mathbf{y} \in \mathbb{R}^2$,

$$\partial_\theta\varphi(\mathbf{y}) = \left.\frac{d}{dt}\right|_{t=0} \varphi(\exp(t\omega)\mathbf{y}).$$

Lemma 4.1. *Suppose $R > 0$ and $\psi : \mathbb{R}^2 \to \mathbb{R}$ is a non-negative bounded function which is supported in the ball $B(0,R)$, such that $\partial_\theta\psi$ is also bounded, and denote by $f = \widehat{\psi}$ its Siegel-Veech transform as in Theorem 2.1, with respect to some configuration $\mathcal{C}$. Let χ_ε denote the characteristic function of the cusp M_ε^c. Then the decomposition*

$$f = f_{\mathrm{main}} + f_\varepsilon, \quad \text{where } f_{\mathrm{main}} = f(1-\chi_\varepsilon) \quad \text{and } f_\varepsilon = f\chi_\varepsilon$$

satisfies for any $1 < \alpha_1 < \alpha_2 < 2$,

$$\mathcal{S}_K(f_{\mathrm{main}}) \ll_{R,\alpha_1} \max_{r\leq R}\left(\int_K \left(\psi^2 + |\partial\psi|^2\right)(kr\mathbf{e}_1)dm_K(k)\right)^{1/2} \varepsilon^{-\alpha_1}, \tag{4.1}$$

$$\int_{\mathcal{L}} f_\varepsilon \, d\mu \ll_{R,\alpha_1,\alpha_2} \|\psi\|_\infty \varepsilon^{\alpha_2-\alpha_1}, \tag{4.2}$$

and

$$\pi_{\mathcal{L}}(\Sigma_t)f_\varepsilon(\mathbf{x}) \ll_{\mathbf{x},R,\alpha_1,\alpha_2} \|\psi\|_\infty \varepsilon^{\alpha_2-\alpha_1}. \tag{4.3}$$

Moreover the implicit constant in (4.3) can be taken to be uniform as $\mathbf{x}$ ranges over compact subsets of $\mathcal{L}$.

Proof. We first bound the L^2-norm of f_{main}. Since the measure μ and the set M_ε are K-invariant, and $y \mapsto V(y)$ is K-equivariant, we have

$$\begin{aligned}
\|f_{\mathrm{main}}\|_2^2 &= \int_{\mathcal{L}} |f(1-\chi_\varepsilon)|^2 d\mu = \int_{\ell(\mathbf{y})\geq\varepsilon} \left| \sum_{v\in V(\mathbf{y})} \psi(v) \right|^2 d\mu(\mathbf{y}) \\
&= \int_{\ell(\mathbf{y})\geq\varepsilon} \int_K \left| \sum_{v\in V(k\mathbf{y})} \psi(v) \right|^2 dm_K(k)\, d\mu(\mathbf{y}) \\
&\leq \int_{\ell(\mathbf{y})\geq\varepsilon} \left(\max_{r\leq R} \int_K |\psi(rk\mathbf{e}_1)|^2\, dm_K(k) \right) |V(\mathbf{y})\cap B(0,R)|^2\, d\mu(\mathbf{y}).
\end{aligned}$$

In the last inequality above we have used Cauchy-Schwarz for the estimate $\left|\sum_{v\in V(k\mathbf{y})}\psi(v)\right|^2 \leq \left(\sum_{v\in V(k\mathbf{y})} |\psi(v)|^2\right) |V(\mathbf{y})\cap B(0,R)|$, and then exchanged summation and integration.

Using Theorem 2.2, we conclude that

$$\|f_{\mathrm{main}}\|_2^2 \ll_{R,\alpha_1} \varepsilon^{-2\alpha_1} \max_{r\leq R} \int_K |\psi(kr\mathbf{e}_1)|^2 dm_K(k) \tag{4.4}$$

We repeat this calculation for the angular derivative of ψ. Here we also use the fact that since the set of saddle connections satisfies $V(g\mathbf{y}) = gV(\mathbf{y})$ for any $\mathbf{y}\in\mathcal{L}$ and $g\in G$, which implies that taking derivatives in the K direction commutes with the Siegel-Veech transform. Namely, for any compactly supported $\psi : \mathbb{R}^2 \to \mathbb{R}$ for which $\partial_\theta\psi$ exists everywhere,

$$\pi_{\mathcal{L}}(\omega)\widehat{\psi} = \lim_{t\to 0}\frac{1}{t}\left(\pi_{\mathcal{L}}(\exp(t\omega)\widehat{\psi} - \widehat{\psi}\right) = \lim_{t\to 0}\frac{1}{t}\left(\sum_{u\in V(\exp(t\omega)\mathbf{y})} \psi(u) - \sum_{v\in V(\mathbf{y})} \psi(v) \right)$$

$$= \lim_{t\to 0}\frac{1}{t}\left(\sum_{v\in V(\mathbf{y})} (\psi(\exp(t\omega)v) - \psi(v)) \right) = \sum_{v\in V(\mathbf{y})} \partial_\theta\psi(v)$$

(where we have used the fact that ψ is compactly supported to ensure that the sum is finite and hence we can switch the order of summation and differentiation). Thus

$$\pi_{\mathcal{L}}(\omega)\hat{\psi} = \widehat{\partial_\theta\psi}.$$

By K-invariance of $(1-\chi_\varepsilon)$,

$$\pi_{\mathcal{L}}(\omega) f_{\mathrm{main}} = (1-\chi_\varepsilon)\pi_{\mathcal{L}}(\omega) f + f\pi_{\mathcal{L}}(\omega)(1-\chi_\varepsilon) = (1-\chi_\varepsilon)\widehat{\partial_\theta\psi} + 0$$

and consequently, applying the argument used to prove inequality (4.4) to $\pi_{\mathcal{L}}(\omega) f_{\mathrm{main}}$, we obtain (4.1).

Now we set $\beta = \alpha_2 - \alpha_1$, and proceed to bound $\int f_\varepsilon\, d\sigma$ in the two cases $\sigma = \mu$, $\sigma = \pi_{\mathcal{L}}(\Sigma_t) = \int_K \delta_{a_t k\mathbf{x}} dm_K(k)$. By Theorem 2.2,

$$\int f_\varepsilon\, d\sigma = \int_{\ell(\mathbf{y})<\varepsilon} \widehat{\psi} d\sigma \leq \int_{\ell(\mathbf{y})<\varepsilon} \|\psi\|_\infty |V(\mathbf{y})\cap B(0,R)| d\sigma$$

$$\ll_{R,\alpha_1} \|\psi\|_\infty \int_{\ell(\mathbf{y})<\varepsilon} \ell(\mathbf{y})^{-\alpha_1} d\sigma < \|\psi\|_\infty \int_{\ell(\mathbf{y})<\varepsilon} \frac{\varepsilon^\beta}{\ell(\mathbf{y})^\beta} \ell(\mathbf{y})^{-\alpha_1} d\sigma.$$

The term $\int_{\ell(\mathbf{y})<\varepsilon} \ell(\mathbf{y})^{-\alpha_1-\beta} d\sigma$ is bounded by $\|\ell(\cdot)^{-\alpha_2}\|_{L^1(\sigma)}$ for $\alpha_2 = \alpha_1+\beta$, which in turn is bounded by Theorem 2.3 for any $\alpha_2 < 2$ (where for $\sigma = \pi_{\mathcal{L}}(\Sigma_t)$ the bound depends on $\mathbf{x}$ uniformly on compact subsets of $\mathcal{L}$, and is independent of t). $\square$

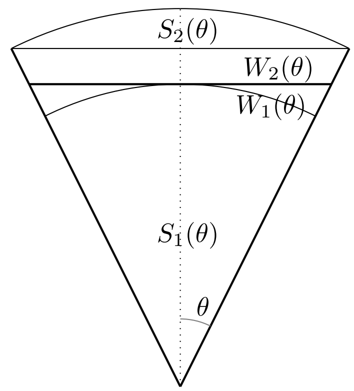

FIGURE 1. The four planar domains $S_1(\theta) \subset W_1(\theta) \subset W_2(\theta) \subset S_2(\theta)$.

5. Effective counting of saddle connections

We now give the proof of Theorem 1.1, dividing the argument into three steps. In the first we use a geometric counting method introduced in [EMaMo98, Lem. 3.6] and [EM01, Lem. 3.4] to estimate the quantity $N(e^t, \mathbf{x}, \varphi_1, \varphi_2)$ by orbit integrals $\pi_{\mathcal{L}}(\Sigma_t)f(\mathbf{x})$, where f is a Siegel-Veech transform of the indicator of a triangle. We will follow the simplified approach outlined in the survey [E06], but replacing a trapezoid used in [E06] with a triangle. In the second step we will replace f with certain smooth approximations and use Theorem 3.4 and Lemma 4.1 to estimate the resulting orbit integrals. Since it relies on the Borel-Cantelli lemma, Theorem 3.4 only gives information about $N(T, \mathbf{x}, \varphi_1, \varphi_2)$ for a countable number of values of T. In the third and final step we use an interpolation argument to pass from countably many values, which get denser and denser on a logarithmic scale, to all T.

Step 1. Triangles, and reduction of counting to orbit integrals. We fix a configuration $\mathcal{C}$ and use it to define a Siegel-Veech transform as in Theorem 2.1. For $\theta \in (0,1)$ we define two triangles $W_1 = W_1(\theta)$ and $W_2 = W_2(\theta)$ in the plane as follows. Let $e_2 = (0,1)$ and let W_1 have vertices $(0,0), r_\theta e_2, r_{-\theta} e_2$ and W_2 have vertices $(0,0), \frac{1}{\cos\theta} r_\theta e_2, \frac{1}{\cos\theta} r_{-\theta} e_2$. That is, W_1 and W_2 are similar isosceles triangles with apex at the origin, apex angle 2θ, symmetric around the positive y-axis, and with height $\cos\theta$ and 1 respectively. In particular $W_1 \subset W_2$. See Figure 1.

Now let $t > 1$ be a parameter. Applying the diagonal flow a_{-t} transforms W_1, W_2 into triangles with a narrow apex angle and large height, specifically the apex angle $2\theta_t$ of both $a_{-t}W_1$ and $a_{-t}W_2$ satisfies

$$\tan\theta_t = e^{-2t}\tan\theta. \tag{5.1}$$

We will obtain lower and upper bounds for $N^{\mathcal{C}}(e^t, \mathbf{x}, \varphi_1, \varphi_2)$ using radial averages over shrinking versions of these triangles.

Let $\varphi_1 < \varphi_2$ be as in Theorem 1.1. By a rotation, assume with no loss of generality that $\varphi_2 = \varphi > 0$ and $\varphi_1 = -\varphi$ so that $I = [-\varphi, \varphi]$ is symmetric around 0 and $\varphi_2 - \varphi_1 = 2\varphi$. Recall the notation r_s for an element of K (see (2.1)). We will identify angles in $\mathbb{R}$ with their image modulo $2\pi\mathbb{Z}$ and functions on K with functions on $\mathbb{R}/2\pi\mathbb{Z}$ without further mention.

Define

$$I_t^- = [-(\varphi - \theta_t), \varphi - \theta_t], \quad I_t^+ = [-(\varphi + \theta_t), \varphi + \theta_t], \tag{5.2}$$

so that $I_t^- \subset I \subset I_t^+$, and let ν_t^-, ν, ν_t^+ denote respectively the indicator functions of I_t^-, I, I_t^+ (note that the dependence of these indicators on $\varphi_2 - \varphi_1$ is suppressed from the notation). Also let $\mathbb{1}_{W_1}, \mathbb{1}_{W_2}$ denote the indicators of W_1 and W_2.

We claim that for any $\mathbf{x}$,

$$\pi_{\mathcal{L}}(\Sigma_{\nu_t^-,t})\widehat{\mathbb{1}_{W_1(\theta)}}(\mathbf{x}) \le \frac{\theta_t}{\pi} N^{\mathcal{C}}(e^t, \mathbf{x}, \varphi_1, \varphi_2) \le \pi_{\mathcal{L}}(\Sigma_{\nu_t^+,t})\widehat{\mathbb{1}_{W_2(\theta)}}(\mathbf{x}). \tag{5.3}$$

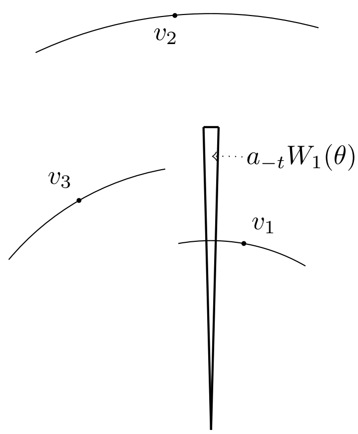

FIGURE 2. The arc $K(I_t^-)v_1$ cuts through $a_{-t}W_1$ but the arcs $K(I_t^-)v_2$ and $K(I_t^-)v_3$ miss $a_{-t}W_1(\theta)$.

To see the left hand inequality, recall that by the definition of the Siegel-Veech transform and the operator $\Sigma_{\nu,t}$ we have

$$\pi_{\mathcal{L}}(\Sigma_{\nu_t^-,t})\widehat{\mathbb{1}_{W_1(\theta)}}(\mathbf{x}) = \sum_{v\in V^{\mathcal{C}}(\mathbf{x})} \int_K \mathbb{1}_{W_1(\theta)}(a_t k v)\nu_t^-(k)dm_K(k). \tag{5.4}$$

We will estimate the contribution of each individual $v \in V^{\mathcal{C}}(\mathbf{x})$ to the sum (5.4). For any $v \in \mathbb{R}^2$,

$$\int_K \mathbb{1}_{W_1}(a_t k v)\nu_t^-(k)dm_K(k) = \frac{1}{2\pi}\int_{-\varphi+\theta_t}^{\varphi-\theta_t} \mathbb{1}_{a_{-t}W_1}(r_s v)ds \tag{5.5}$$

is at most $\frac{\theta_t}{\pi}$, since the apex angle of $a_{-t}W_1$ is $2\theta_t$. The quantity (5.5) vanishes if $\|v\| \geq e^t$ or $\measuredangle(v, e_2) \notin I$, since in these cases the arc $K(I_t^-)v = \{r_\beta v : \beta \in I_t^-\}$ never enters the triangle $a_{-t}W_1$ (see Figure 2). Furthermore, if $\|v\| \leq e^t$ and $\measuredangle(v, e_2) \in I$ then the arc $K(I_t^+)v$ intersects $a_{-t}W_2$ along its entire apex angle, and so $\int_K \mathbb{1}_{W_2}(a_t k v)\nu_t^+(k)dm_K = \frac{\theta_t}{\pi}$, and this implies the right hand inequality.

Step 2. Smooth approximations, ergodic theorem, and cutting off the cusp. Our goal will be to estimate the left and right hand sides of (5.3). To this end we will replace $\widehat{\mathbb{1}_{W_1}}, \widehat{\mathbb{1}_{W_2}}$ with smooth approximations and apply Theorem 3.4 and Lemma 4.1 to the approximating functions, where the n-th function will give a bound on (5.3) for a certain time $T_n = e^{t_n}$. Our approximation depends on functions and parameters which we now describe (omitting their dependence on n). The first parameter θ controls the apex angle of the triangles W_i, as above. We will bound $\mathbb{1}_{W_i}$ from above and below by K-smooth functions $\psi_{(-,\delta)}, \psi_{(+,\delta)}$ on the plane, supported respectively in a slightly contracted (resp. expanded) copy of W_1 (resp. W_2), where the dilation is controlled by a smoothing parameter δ. The corresponding Siegel-Veech transforms $\widehat{\psi_{(\pm,\delta)}}$ will be denoted by $f_{(\pm,\delta)}$. They will be truncated using Lemma 4.1 along with a cutoff parameter ε, for an appropriate choice of parameters α_1, α_2. Finally Theorem 3.4 will be applied to the main term $(f_{(\pm,\delta)})_{\text{main}}$, and all resulting errors will be collected and bounded.

We now make this discussion more precise and record some estimates for the errors incurred at the various stages. After collecting these bounds we will choose our parameters and optimize the error terms in the next step. Our optimization gives $\theta = \delta^{1/2}$, and so in order to reduce the number of parameters we will use this dependence of θ and δ throughout.

Let $\theta < 1$ and $R = \cos(1)^{-1}$. We approximate both triangles $W_1(\theta), W_2(\theta)$ by sectors around the positive vertical axis, that is by sets of the form

$$S_{r_0,\varphi_0} = \{r(\cos\beta, \sin\beta) : 0 \leq r \leq r_0, |\beta - \pi/2| \leq \varphi_0\}.$$

The sector $S_1(\theta) = S_{\cos\theta,\theta}$ is contained in $W_1(\theta)$ and the sector $S_2(\theta) = S_{(\cos\theta)^{-1},\theta}$ contains $W_2(\theta)$ (see Figure 1). Let $\delta = \theta^2$ and let $H : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function which vanishes outside $[-\theta, \theta]$, is equal to 1 on $[-\theta+\delta, \theta-\delta]$, and such that $\|H'\|_\infty \leq 2\delta^{-1}$, and define

$$\psi_{(-,\delta)}(r\cos\beta, r\sin\beta) = \begin{cases} H(\beta - \pi/2) & r \leq \cos\theta \\ 0 & r > \cos\theta \end{cases} \tag{5.6}$$

Then $\psi_{(-\delta)}$ is supported in $B(0,R)$ and we have pointwise inequalities

$$\psi_{(-,\delta)} \leq \mathbb{1}_{S_1(\theta)} \leq \mathbb{1}_{W_1(\theta)},$$

and an estimate

$$\|\partial_\theta \psi_{(-,\delta)}\|_\infty \ll \delta^{-1}.$$

Similarly, we define functions $\psi_{(+,\delta)}$ which satisfy a pointwise inequality $\mathbb{1}_{W_2(\theta)} \leq \psi_{(+,\delta)}$, are supported in $B(0,R)$, and also satisfy $\|\partial_\theta \psi_{(+,\delta)}\|_\infty \ll \delta^{-1}$. Since $\partial_\theta \psi_{(\pm,\delta)}(kr\mathbf{e}_1)$ is only supported on an arc of angular width 2δ, this implies that

$$\max_{r \leq R} \left(\int_K |\partial_\theta \psi_{(\pm,\delta)}(kr\mathbf{e}_1)|^2 dm_K(k) \right)^{1/2} \ll \delta^{-1/2}. \tag{5.7}$$

Since we also have pointwise bounds $\mathbb{1}_{S_1(\theta-\delta)} \leq \psi_{(-,\delta)} \leq \psi_{(+,\delta)} \leq \mathbb{1}_{S_2(\theta+\delta)}$, we obtain a bound

$$\int_{\mathbb{R}^2} (\psi_{(+,\delta)} - \psi_{(-,\delta)}) dx \leq \mathrm{Area}\left(S_2(\theta+\delta) \smallsetminus S_1(\theta-\delta)\right) \ll \delta + \theta^3 \ll \delta. \tag{5.8}$$

Similarly, we obtain the bounds

$$\int_{\mathbb{R}^2} (\mathbb{1}_{W_1(\theta)} - \psi_{(-,\delta)}) dx \ll \delta \quad \text{and} \quad \int_{\mathbb{R}^2} (\psi_{(+,\delta)} - \mathbb{1}_{W_2(\theta)}) dx \ll \delta. \tag{5.9}$$

Since

$$\mathrm{Area}(W_1(\theta)) = \cos\theta \sin\theta \quad \text{and} \quad \mathrm{Area}(W_2(\theta)) = \tan\theta,$$

(5.9) also implies that

$$\left| \int_{\mathbb{R}^2} \psi_{(-,\delta)} dx - \cos\theta\sin\theta \right| \ll \delta, \quad \left| \int_{\mathbb{R}^2} \psi_{(+,\delta)} dx - \tan\theta \right| \ll \delta.$$

Using (5.1) and expanding the Taylor series for sin, cos, arctan we find

$$\frac{\sin\theta\cos\theta}{e^{2t}\theta_t} = \frac{\theta + O(\theta^3)}{\theta + O(\theta^3)} = 1 + O(\theta^2), \quad \text{and also} \quad \frac{\tan\theta}{e^{2t}\theta_t} = 1 + O(\theta^2),$$

so that (using $e^{2t}\theta_t = \theta + O(\theta^3)$)

$$\frac{1}{e^{2t}\theta_t} \int_{\mathbb{R}^2} \psi_{(\pm,\delta)} dx = 1 + \frac{1}{e^{2t}\theta_t} O(\theta^2 + \delta) = 1 + O(\delta^{1/2}).$$

Note that the appearance of $\theta^2 + \delta$ in this bound explains our choice $\theta^2 = \delta$.

We have (provided $0 < \theta \leq \frac{\pi}{4}$)

$$\left| \int_K \nu_t^{\pm} dm_K - \frac{\varphi_2 - \varphi_1}{2\pi} \right| \leq \frac{\theta_t}{\pi} \leq \frac{\tan\theta_t}{\pi} \leq e^{-2t}\tan\theta \leq e^{-2t},$$

and we will make our choices so that

$$e^{-2t} = o(\delta^{1/2}), \tag{5.10}$$

so that we have

$$\frac{1}{e^{2t}\theta_t} \left(\int_{\mathbb{R}^2} \psi_{(\pm,\delta)} dx \right) \left(\int_K \nu_t^{\pm} dm_K \right) = \frac{\varphi_2 - \varphi_1}{2\pi} + O(\delta^{1/2}). \tag{5.11}$$

Let $f_{(\pm,\delta)} = \widehat{\psi_{(\pm,\delta)}}$ be the Siegel-Veech transform of the functions defined in (5.6). The transform preserves pointwise inequalities of functions, and so (5.3) implies

$$\frac{\pi}{\theta_t} \pi_{\mathcal{L}}(\Sigma_{\nu_t^-,t}) f_{(-,\delta)}(\mathbf{x}) \leq N^{\mathcal{C}}(e^t, \mathbf{x}, \varphi_1, \varphi_2) \leq \frac{\pi}{\theta_t} \pi_{\mathcal{L}}(\Sigma_{\nu_t^+,t}) f_{(+,\delta)}(\mathbf{x}). \tag{5.12}$$

We apply Lemma 4.1 with parameters $\alpha_1, \alpha_2, \varepsilon$. Decompose $f_{(\pm,\delta)}$ as the sum $(f_{(\pm,\delta)})_{\mathrm{main}} + (f_{(\pm,\delta)})_\varepsilon$, where $(f_{(\pm,\delta)})_{\mathrm{main}} = f_{(\pm,\delta)}(1-\chi_\varepsilon)$ and $(f_{(\pm,\delta)})_\varepsilon = f_{(\pm,\delta)}\chi_\varepsilon$, and let

$$f_{(\pm,\delta,\varepsilon)} = f_{(\pm,\delta)}(1-\chi_\varepsilon) - I^{\pm}_{(\delta,\varepsilon)}, \text{ where } I^{\pm}_{(\delta,\varepsilon)} = \int f_{(\pm,\delta)}(1-\chi_\varepsilon)d\mu,$$

i.e. $f_{(\pm,\delta,\varepsilon)}$ is the projection $\left(f_{(\pm,\delta)}\right)_{\mathrm{main}}$ to the space of zero integral functions. Inequality (4.4) in the proof of (4.1) shows that

$$\|f_{(\pm,\delta,\varepsilon)}\|_2^2 = \|(f_{(\pm,\delta)})_{\mathrm{main}}\|_2^2 - (I^{\pm}_{(\delta,\varepsilon)})^2 \le \|(f_{(\pm,\delta)})_{\mathrm{main}}\|_2^2 \ll \varepsilon^{-2\alpha_1}\theta^2 = \varepsilon^{-2\alpha_1}\delta. \tag{5.13}$$

Similarly, using (4.4) and (5.7),

$$\mathcal{S}_K(f_{(\pm,\delta,\varepsilon)})^2 \ll \varepsilon^{-2\alpha_1}\delta^{-1}. \tag{5.14}$$

Note that the implicit constants in (5.13) and (5.14) depend only on the support of $\psi_{(\pm,\delta)}$, which is uniformly bounded.

Having recorded these bounds, we turn to the application of Theorem 3.4. We will choose a sequence $t_n \to \infty$, choose parameters $\lambda \in (0, \lambda_{\mathcal{L}})$, set $\eta = \frac{1}{\lambda+1}$, and for each n, define parameters δ_n, ε_n, thus giving functions

$$f_n^{\pm} = f_{(\pm,\delta_n,\varepsilon_n)}.$$

The theorem will be applied twice, to each of the two sequences f_n^+, f_n^-. We will choose $0 < \eta_1 < 2\eta$ so that (3.8) is satisfied. Then, since the $\nu_{t_n}(K)$ are bounded, plugging (5.13) and (5.14) into (3.9) we will obtain the bound

$$\left|\pi_{\mathcal{L}}(\Sigma_{\nu^{\pm}_{t_n},t_n})f_n^{\pm}(\mathbf{x})\right| \ll e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}D_n, \tag{5.15}$$

where

$$D_n = \varepsilon_n^{-\alpha_1}\delta_n^{-\frac{1}{2}}. \tag{5.16}$$

In what follows we continue with the set of full measure of $\mathbf{x}$ for which (5.15) holds, and thus the implicit constants in the $\ll$ and $O(\cdot)$ notations may depend on $\mathbf{x}$. Let $I_n^{\pm} = I^{\pm}_{(\delta_n,\varepsilon_n)}$. Since $f_{(\pm,\delta_n)}(1-\chi_{\varepsilon_n}) = f_n^{\pm} + I_n^{\pm}$, (4.2) and (5.15) imply

$$\begin{aligned}\pi_{\mathcal{L}}(\Sigma_{\nu^{\pm}_{t_n},t_n})(f_{(\pm,\delta_n)}(1-\chi_{\varepsilon_n}))(\mathbf{x}) &= \int_K I_n^{\pm}\nu^{\pm}_{t_n}dm_K + O(e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}D_n)\\ &= \left(\int f_{(\pm,\delta_n)}d\mu\right)\left(\int_K \nu^{\pm}_{t_n}dm_K\right) + O(\varepsilon_n^{\beta} + e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}D_n),\end{aligned} \tag{5.17}$$

where

$$\beta = \alpha_2 - \alpha_1.$$

Moreover (4.3) implies

$$\pi_{\mathcal{L}}(\Sigma_{\nu^{\pm}_{t_n},t_n})(f_{(\pm,\delta_n)}\chi_{\varepsilon_n})(\mathbf{x}) \le \pi_{\mathcal{L}}(\Sigma_{t_n})(f_{(\pm,\delta_n)}\chi_{\varepsilon_n})(\mathbf{x}) \ll_{\mathbf{x},\alpha_1,\alpha_2} \varepsilon_n^{\beta}. \tag{5.18}$$

By Theorem 2.1 we have $\int f_{(\pm,\delta_n)}d\mu = c\int_{\mathbb{R}^2}\psi_{(\pm,\delta_n)}dx$, where $c = c(\mathcal{L},\mathcal{C})$ is the Siegel-Veech constant. Combining this with (5.17) and (5.18) we obtain

$$\left|\pi_{\mathcal{L}}(\Sigma_{\nu^{\pm}_{t_n},t_n})f_{(\pm,\delta_n)}(\mathbf{x}) - c\left(\int_{\mathbb{R}^2}\psi_{(\pm,\delta_n)}dx\right)\left(\int_K \nu^{\pm}_{t_n}dm_K\right)\right| \ll_{\mathbf{x}} \varepsilon_n^{\beta} + e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}D_n. \tag{5.19}$$

Combining (5.11) and (5.19), and using again $e^{2t}\theta_t = \theta + O(\theta^3) = \delta^{1/2} + O(\delta^{3/2})$, we get

$$\left|\frac{1}{e^{2t_n}\theta_{t_n}}\pi_{\mathcal{L}}(\Sigma_{\nu^{\pm}_{t_n},t_n})f_{(\pm,\delta_n)}(\mathbf{x}) - \frac{c(\varphi_2-\varphi_1)}{2\pi}\right| \ll \delta_n^{1/2} + \delta_n^{-1/2}(\varepsilon_n^{\beta} + e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}D_n).$$

Plugging this estimate into (5.12) and using (5.16), we find that for any n,

$$\left|\frac{N^{\mathcal{C}}(e^{t_n},\mathbf{x},\varphi_1,\varphi_2)}{e^{2t_n}} - \frac{c}{2}(\varphi_2-\varphi_1)\right| \ll \delta_n^{1/2} + \delta_n^{-1/2}\varepsilon_n^{\beta} + e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}\varepsilon_n^{-\alpha_1}\delta_n^{-1}. \tag{5.20}$$

Step 3. Choosing parameters and deriving bounds for any T. Let $\lambda_{\mathcal{L}}$ be the size of the spectral gap for $\mathcal{L}, \mu$ as in §3, and let $\lambda < \lambda_{\mathcal{L}}$. We will show that if $\sigma = 5.5 + 3.5\lambda$ then $\kappa = \frac{\lambda}{\sigma}$ satisfies the conclusion of the theorem (see the discussion following the statement of Theorem 1.1). Let t_n be the sequence defined by the equation

$$e^{t_n} = n^{\frac{\sigma}{\lambda}}. \tag{5.21}$$

Fix $1 < \alpha_1 < \alpha_2 < 2$ and $\beta = \alpha_2 - \alpha_1$ as in Lemma 4.1, and set

$$\eta_1 = \frac{\alpha_1}{\sigma} \tag{5.22}$$

(note that with this choice, $0 < \eta_1 < 2\eta$). This choice is necessary if we want to satisfy (3.8), since $\alpha_1 > 1$ implies

$$\sum_{n\in\mathbb{N}} e^{-\lambda\eta_1 t_n} = \sum_{n\in\mathbb{N}} n^{-\sigma\eta_1} = \sum_{n\in\mathbb{N}} n^{-\alpha_1} < \infty.$$

The right hand side of (5.20) is optimized when the three terms appearing in it are equal to each other. After some algebraic manipulations, this leads to

$$(\delta_n)^{\frac{3}{2}+\frac{\alpha_1}{\beta}} = e^{-(\eta-\frac{\eta_1}{2})\lambda t_n}$$

and therefore

$$\delta_n = e^{\frac{-(\eta-\frac{\eta_1}{2})\lambda t_n}{\frac{3}{2}+\frac{\alpha_1}{\beta}}}$$

and the right hand side of (5.20) becomes (up to constants)

$$\left(e^{2t_n}\right)^{-\kappa}, \quad \text{where } \kappa = \lambda\frac{\eta - \eta_1/2}{6+4\alpha_1/\beta}. \tag{5.23}$$

This expression is decreasing in α_1 and increasing in β, and we are allowed to take arbitrary $\alpha_1 > 1$ and $\beta < 2 - \alpha_1$. Thus we may let $\alpha_1 \to 1$ and $\beta \to 1$, plugging in (5.22), to obtain κ arbitrarily close to

$$\kappa(\sigma) = \frac{\lambda}{10}\left(\eta - \frac{1}{2\sigma}\right),$$

and thus for times $T_n = e^{t_n}$, (5.20) can be written as

$$\left|N^{\mathcal{C}}(T_n, \mathbf{x}, \varphi_1, \varphi_2) - \frac{c}{2}(\varphi_2-\varphi_1)T_n^2\right| \ll T_n^{2(1-\kappa(\sigma))}.$$

Now for arbitrary T, let n satisfy $T_n < T \le T_{n+1}$. By monotonocity,

$$\frac{c(\varphi_2-\varphi_1)}{2}T_n^2\left(1 - O\left(T_n^{-2\kappa(\sigma)}\right)\right) \le N^{\mathcal{C}}(T, \mathbf{x}, \varphi_1, \varphi_2) \le \frac{c(\varphi_2-\varphi_1)}{2}T_{n+1}^2\left(1 + O\left(T_{n+1}^{-2\kappa(\sigma)}\right)\right). \tag{5.24}$$

Since $e^{t_n} = n^{\frac{\sigma}{\lambda}}$ we have

$$\max\left(\frac{T^2}{T_n^2}, \frac{T_{n+1}^2}{T^2}\right) \ll \frac{T_{n+1}^2}{T_n^2} = \left(1+\frac{1}{n}\right)^{\frac{2\sigma}{\lambda}} = 1+O\left(\frac{1}{n}\right) = 1+O\left(e^{-\frac{t_n\lambda}{\sigma}}\right) = 1+O\left(T^{2(-\frac{\lambda}{2\sigma})}\right).$$

So both sides of (5.24) are $\frac{c(\varphi_2-\varphi_1)}{2}T^2\left(1+O\left(T^{-\frac{\lambda}{2\sigma}}\right)^2 + O\left(T^{-\kappa(\sigma)}\right)^2\right)$. Thus we can do no better than to set $\kappa(\sigma) = \frac{\lambda}{2\sigma}$. Solving this equation leads to $\sigma = 5.5\eta = 5.5 + 5.5\lambda$, as claimed. We leave it to the reader to verify that with this choice of σ, (5.10) is satisfied. □

6. Effective counting in all sectors and all dilates of an ellipse

In order to change the order of quantifiers and obtain an estimate simultaneously true for all φ_1, φ_2 and all $\{g\mathbf{x} : g \in G\}$, we will use two distinct techniques. Firstly we will use Theorem 3.5 instead of 3.4, as this will allow us to control countably many ellipses. Secondly we will give an additional approximation argument which shows how to use countably many functions, approximating a countable dense set of sectors, along with a countable dense set of ellipses, of countably many radii, to simultaneously control all ellipses and all sectors. We proceed to the details.

Proof of Theorem 1.2. We will use the notations and estimates as in the proof of Theorem 1.1, and follow the same steps. We fix a configuration $\mathcal{C}$ and use it throughout, and let c be the corresponding Siegel-Veech constant. In analogy with (5.2), for fixed φ_1, φ_2 we set

$$I = [\varphi_1, \varphi_2], \quad I^-_{t,\varphi_1,\varphi_2} = [\varphi_1 + \theta_t, \varphi_2 - \theta_t], \quad I^+_{t,\varphi_1,\varphi_2} = [\varphi_1 - \theta_t, \varphi_2 + \theta_t],$$

and let ν, $\nu^-_{t,\varphi_1,\varphi_2}$, $\nu^+_{t,\varphi_1,\varphi_2}$ denote respectively the indicator functions of these intervals (where we have selected a different notation to reflect the dependence on φ_1, φ_2). Using (5.3), and using that $\pi_{\mathcal{L}}\left(\Sigma_{\nu_t^{\pm},\varphi_1,\varphi_2}\right)(g\mathbf{x}) = \pi_{\mathcal{L}}\left(\Sigma^{(g)}_{\nu_t^{\pm},\varphi_1,\varphi_2}\right)(\mathbf{x})$, we find that for every $\mathbf{x} \in \mathcal{H}$ and every $g \in G$,

$$\pi_{\mathcal{L}}\left(\Sigma^{(g)}_{\nu^-_{t,\varphi_1,\varphi_2},t}\right)\widehat{\mathbb{1}_{W_1(\theta)}}(\mathbf{x}) \leq \frac{\theta_t}{\pi} N^{\mathcal{C}}(e^t, g\mathbf{x}, \varphi_1, \varphi_2) \leq \pi_{\mathcal{L}}\left(\Sigma^{(g)}_{\nu^+_{t,\varphi_1,\varphi_2},t}\right)\widehat{\mathbb{1}_{W_2(\theta)}}(\mathbf{x}).$$

This estimate generalizes (5.3) and constitutes the first step of the proof.

In the second step of the proof we again need to record certain bounds, but this time we will record their dependence on three additional parameters. Namely, as before, we will have parameters $1 < \alpha_1 < \alpha_2 < 2$, $\beta = \alpha_2 - \alpha_1$, $\eta_1 > 0$, as well as sequences of times $t_n \nearrow \infty$, smoothing parameters δ_n and cutoff parameters ε_n. In addition we will have sequences of 'ellipse parameters' $(g_n) \subset G$ and 'angular sector parameters' $\varphi_1^{(n)} < \varphi_2^{(n)}$ with $\varphi_2^{(n)} - \varphi_1^{(n)} \leq 2\pi$.

Using a smoothing parameter $\delta = \delta_n$, defining the functions $f_{(\pm,\delta)}$ (Siegel-Veech transforms of smooth approximations of $\mathbb{1}_{W_1}, \mathbb{1}_{W_2}$) as before, and in analogy with (5.12), we obtain

$$\frac{\pi}{\theta_t}\pi_{\mathcal{L}}\left(\Sigma^{(g)}_{\nu^-_{t,\varphi_1,\varphi_2}}\right) f_{(-,\delta)}(\mathbf{x}) \leq N^{\mathcal{C}}(e^t, g\mathbf{x}, \varphi_1, \varphi_2) \leq \frac{\pi}{\theta_t}\pi_{\mathcal{L}}\left(\Sigma^{(g)}_{\nu^+_{t,\varphi_1,\varphi_2}}\right) f_{(+,\delta)}(\mathbf{x}).$$

Note that these upper and lower bounds are valid for any $g \in G$ and any φ_1, φ_2.

In the proof of (5.20), there are two sources for the dependence of the estimate depending on $\mathbf{x}$. The first arises in deriving (5.18) by way of (4.3), and gives rise to an estimate which is uniform as $\mathbf{x}$ ranges over a compact subset of $\mathcal{H}$, and the second arises from Theorem 3.4, and gives rise to a condition $n \geq n_0(\mathbf{x})$. Thus the same argument (with Theorem 3.5 instead of Theorem 3.4) gives

$$\begin{aligned}&\left|\frac{N^{\mathcal{C}}\left(e^{t_n}, g_n\mathbf{x}, \varphi_1^{(n)}, \varphi_2^{(n)}\right)}{e^{2t_n}} - \frac{c}{2}\left(\varphi_2^{(n)} - \varphi_1^{(n)}\right)\right| \\ \ll&\delta_n^{1/2} + \delta_n^{-1/2}\varepsilon_n^{\beta} + \left(\varphi_2^{(n)} - \varphi_1^{(n)}\right)^{1-\lambda\eta/2} e^{-(\eta - \frac{\eta_1}{2})\lambda t_n}\varepsilon_n^{-\alpha_1}\delta_n^{-1},\end{aligned} \tag{6.1}$$

as long as $n \geq n_0(\mathbf{x})$ and where the implicit constant depends on g_n and $\mathbf{x}$ and can be taken to be uniform in compact subsets of G and $\mathcal{L}$. This completes the second step of the proof.

We now choose $\lambda, \alpha_1, \alpha_2$ satisfying $\lambda < \lambda_{\mathcal{L}}$, $1 < \alpha_1 < \alpha_2 < 2$. For each $n \in \mathbb{N}$, we define an auxiliary variable

$$m = m_n = \lfloor n^{1/7} \rfloor,$$

which we will refer to as the *scale* of n. For fixed m, let

$$\mathcal{N}_m = \{n : m_n = m\}$$

denote the indices of scale m. Note that as $n \to \infty$, the scales m_n also tend to infinity at a slower rate, and the cardinality of $\mathcal{N}_m$ is approximately m^6. Now choose $\varphi_1^{(n)}, \varphi_2^{(n)}, g_n$ so that for all large enough m, the collection of triples

$$\left\{\left(\varphi_1^{(n)}, \varphi_2^{(n)}, g_n\right) : n \in \mathcal{N}_m\right\}$$

is $\frac{1}{m(\log m)}$-dense in

$$\left\{(\varphi_1, \varphi_2, g) : \varphi_1 \in [0, 2\pi],\ \varphi_2 - \varphi_1 \in [0, 2\pi],\ g \in G,\ \max(\|g\|, \|g^{-1}\|) < \log m\right\} \tag{6.2}$$

(with respect to the sup-norm in the first two coordinates and the operator norm in the third coordinate). This is possible since (6.2) defines a 5-dimensional manifold of diameter $O(\log m)$.

Now following (5.21) we choose t_n so that $e^{t_n} = m_n^{\frac{\sigma}{\lambda}}$, where σ is a parameter we will optimize. It will turn out that the optimal value will be

$$\sigma = 8.5 + 8.5\lambda. \tag{6.3}$$

Let

$$\eta_1 = \frac{7\alpha_1}{\sigma} > \frac{7}{\sigma}, \tag{6.4}$$

so that

$$\sum_{n\in\mathbb{N}} e^{-\lambda\eta_1 t_n} = \sum_{m\in\mathbb{N}} \sum_{n\in\mathcal{N}_m} m^{-\sigma\eta_1} \ll \sum_{m\in\mathbb{N}} m^{6-\sigma\eta_1} < \infty,$$

so that (3.8) holds. Also note that the lengths of intervals at scale m_n is bounded below by $e^{-\frac{\lambda t_n}{\sigma}}/t_n$ and in particular, since $\frac{\lambda}{\sigma} < 2$, satisfies the lower bound $|I| > e^{-2t_n}$ for all large enough n. Thus we can apply Theorem 3.5, and deduce (6.1).

As before, we optimize the right hand side of (6.1) by setting all three summands equal to each other, and we obtain that it is bounded by a constant (depending on $\|g_n\|$) multiplied by the expression (5.23). Letting $\alpha, \beta \to 1$ in (5.23) and using (6.4) instead of (5.22), we find that the right hand side of (6.1) is on the order of $\left(e^{2t_n}\right)^{-\kappa}$ where

$$\kappa = \kappa(\sigma) = \frac{\lambda}{10}\left(\eta - \frac{7}{2\sigma}\right),$$

Denote

$$S(r_0, \varphi_1, \varphi_2) = \{r(\cos\beta, \sin\beta) : \beta \in [\varphi_1, \varphi_2], 0 \le r \le r_0\}.$$

Let $g \in G$ and $\varphi_1 \in [0, 2\pi]$ and $\varphi_2 \in \mathbb{R}$ with $\varphi_2 - \varphi_1 \le 2\pi$. When $\|g_j - g\| < \frac{1}{m(\log m)}$, and g, g_j are as in (6.2), then $\max(\|\mathrm{Id} - g_j g^{-1}\|, \|\mathrm{Id} - g g_j^{-1}\|) < \frac{1}{m}$. Thus there is a constant c_1 such that for all large enough m there are $k, \ell \in \mathcal{N}_m$ such that

$$\varphi_1^{(\ell)} < \varphi_1 < \varphi_1^{(k)} < \varphi_2^{(k)} < \varphi_2 < \varphi_2^{(\ell)},$$
$$\left|\varphi_i^{(\ell)} - \varphi_i^{(k)}\right| < \frac{c_1}{m},\ i = 1, 2,$$

and for any r_0 we have the inclusions

$$g_k^{-1} S\left(r_0\left(1 - \frac{c_1}{m}\right), \varphi_1^{(k)}, \varphi_2^{(k)}\right) \subset g^{-1} S\left(r_0, \varphi_1, \varphi_2\right) \subset g_\ell^{-1} S\left(r_0\left(1 + \frac{c_1}{m}\right), \varphi_1^{(\ell)}, \varphi_2^{(\ell)}\right).$$

Hence for all T,

$$N^{\mathcal{C}}\left(T\left(1 - \frac{c_1}{m}\right), g_k\mathbf{x}, \varphi_1^{(k)}, \varphi_2^{(k)}\right) \le N^{\mathcal{C}}\left(T, g\mathbf{x}, \varphi_1, \varphi_2\right) \le N^{\mathcal{C}}\left(T\left(1 + \frac{c_1}{m}\right), g_\ell\mathbf{x}, \varphi_1^{(\ell)}, \varphi_2^{(\ell)}\right).$$

Choosing n so that $e^{t_n} \le T < e^{t_{n+1}}$, and assuming T and hence m are large enough so that the preceding estimates are all satisfied, arguing as in the preceding proof, we obtain the following analogue of (5.24):

$$\begin{aligned} &\left(\varphi_2^{(\ell)} - \varphi_1^{(\ell)}\right)\frac{c}{2}e^{2t_n}\left(1 - O((e^{2t_n})^{-\kappa})\right)\left(1 - O\left(\frac{1}{m}\right)\right) \\ \le &N^{\mathcal{C}}(T, g\mathbf{x}, \varphi_1, \varphi_2) \\ \le &\left(\varphi_2^{(k)} - \varphi_1^{(k)}\right)\frac{c}{2}e^{2t_{n+1}}\left(1 + O((e^{2t_{n+1}})^{-\kappa})\right)\left(1 + O\left(\frac{1}{m}\right)\right) \end{aligned} \tag{6.5}$$

(with implicit constants depending on $\|g\|$). As before

$$\frac{e^{2t_{n+1}}}{e^{2t_n}} = 1 + O\left(T^{2(-\frac{\lambda}{2\sigma})}\right),$$

and since

$$\max\left[\frac{\varphi_2 - \varphi_1}{\varphi_2^{(k)} - \varphi_1^{(k)}}, \frac{\varphi_2^{(\ell)} - \varphi_1^{(\ell)}}{\varphi_2 - \varphi_1}\right] = 1 + O\left(\frac{1}{m}\right) = 1 + O\left((e^{2t_n})^{\frac{-\lambda}{2\sigma}}\right),$$

both sides of (6.5) are $(\varphi_2 - \varphi_1)\frac{c}{2}T^2\left(1 + O\left(T^{-\kappa'}\right)\right)$, where

$$\kappa' = \min\left\{\kappa(\sigma), \frac{\lambda}{2\sigma}\right\}.$$

Setting both of these terms equal to each other and computing σ gives (6.3) and completes the proof. □

References

[ACM] J. Athreya, Y. Cheung and H. Masur, *Siegel-Veech transforms are in L^2*, CIRM lecture, Feb. 16, 2017.

[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. Math. **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):143–211, 2006.

[BNRW] C. Burrin, A. Nevo, R. Rühr and B. Weiss, *in preparation.*

[C] M. Cowling, *Sur les coefficients des représentations unitaires des groups de Lie simples*, Lecture Notes in Math. **739**, Springer (1979) 132–178.

[CHH88] M. Cowling, U. Haagerup, and R. Howe, *Almost L^2 matrix coefficients,* J. Reine Angew. Math., 387:97–110, 1988.

[Ei10] M. Einsiedler, *Effective equidistribution and spectral gap*, European Congress of Mathematics, Eur. Math. Soc., Zürich (2010) 31-–51.

[E06] A. Eskin, *Counting Problens in Moduli Space,* in **Handbook of Dynamical Systems**, Vol 1B, 2006.

[EMaMo98] A. Eskin, G. Margulis, and S. Mozes. *Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture,* Ann. of Math. (2), 147(1):93–141, 1998.

[EM01] A. Eskin and H. Masur, *Asymptotic formulas on flat surfaces,* Ergodic Theory Dynam. Systems, 21(2) 443–478, 2001.

[EMZ] A. Eskin, H. Masur and A. Zorich, *Moduli spaces of abelian differentials: the principal boundary, counting problems, and the Siegel-Veech constants*, Publ. I.H.E.S. 97 (2003) 631–678.

[EMi15] A. Eskin and M. Mirzakhani, *Invariant and stationary measures for the* $\mathrm{SL}(2,\mathbb{R})$ *action on moduli space*, Ann. Math., to appear.

[EMiMo15] A. Eskin, M. Mirzakhani and A. Mohammadi, *Isolation, Equidistribution, and Orbit Closures for the* $\mathrm{SL}(2,\mathbb{R})$ *action on moduli space,* Ann. Math. **182** (2015), no. 2, 673–721.

[M] H. Masur, *The growth rate of trajectories of a quadratic differential,* Erg. Th. Dyn. Sys. **10** (1990) 151–176.

[MT] H. Masur and S. Tabachnikov, *Rational billiards and flat structures*, in **Handbook of dynamical systems,** Vol. 1A, 1015–1089. North-Holland, Amsterdam, 2002

[N] A. Nevo, *Equidistribution in measure-preserving actions of semisimple groups: case of* $\mathrm{SL}_2(\mathbb{R})$, (2017) preprint `arXiv:1708.03886`

[R] M. Ratner, *The rate of mixing for geodesic and horocycle flows,* Erg. Theory Dynam. Sys. **7** (1987) 267-–288.

[Vee98] W. A. Veech, *Siegel measures,* Ann. of Math.(2), 148(3):895–944, 1998.

[Vo05] Y. Vorobets, *Periodic geodesics on generic translation surfaces*, in **Algebraic and topological dynamics,** Cont. Math. Vol. 385, 205– 258, 2005.

[Z] A. Zorich, *Flat surfaces,* In **Frontiers in number theory, physics, and geometry.** Vol. I, pages 437–583. Springer, Berlin, 2006.

Department of Mathematics, Technion
E-mail address: `put email here`

Department of Mathematics, Tel Aviv University, Tel Aviv, Israel
E-mail address: `put email here`

Department of Mathematics, Tel Aviv University, Tel Aviv, Israel
E-mail address: `barakw@post.tau.ac.il`