J. Eur. Math. Soc. 28, 5427–5468 (2026) DOI 10.4171/JEMS/1563

Claire Burrin · Samantha Fairchild
(with an appendix by Jon Chaika)

Pairs in discrete lattice orbits with applications to Veech surfaces

Received November 26, 2022; revised July 5, 2023

Abstract. Let Λ_1, Λ_2 be two discrete orbits under the linear action of a lattice $\Gamma < \mathrm{SL}_2(\mathbf{R})$ on the Euclidean plane. We prove a Siegel–Veech-type integral formula for the averages

$$\sum_{\mathbf{x}\in\Lambda_1}\sum_{\mathbf{y}\in\Lambda_2} f(\mathbf{x},\mathbf{y}),$$

from which we derive new results for the set S_M of holonomy vectors of saddle connections of a Veech surface M. This includes an effective count for generic Borel sets with respect to linear transformations, and upper bounds on the number of pairs in S_M with bounded determinant and on the number of pairs in S_M with bounded distance. This last estimate is used in the appendix to prove that for almost every $(\theta,\psi) \in S^1 \times S^1$ the translation flows F_θ^t and F_ψ^t on any Veech surface M are disjoint.

Keywords: translation surfaces, saddle connections, Siegel–Veech transform, non-uniform lattice.

1. Introduction

Let $\Gamma \subset G = \mathrm{SL}_2(\mathbf{R})$ be a lattice acting linearly on the Euclidean plane $\mathbf{R}^2$. Any orbit of Γ is either a dense or a discrete subset of $\mathbf{R}^2$. When the orbit is dense, a limiting distribution was computed by Ledrappier [31] and extended to more general lattices in locally compact groups by Gorodnik and Weiss [19]. When the orbit is discrete, the lattice must be non-uniform and the problem of understanding the distribution is a venerable one, going back to Gauss.

Claire Burrin: Institute of Mathematics, University of Zürich, 08057 Zürich, Switzerland; claire.burrin@math.uzh.ch

Samantha Fairchild: Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; s.k.fairchild@tue.nl

Jon Chaika: Department of Mathematics, University of Utah, Salt Lake City, UT 84112-0090, USA; jonchaika@gmail.com

Mathematics Subject Classification 2020: 11F72 (primary); 22E40, 37E35 (secondary).

A more recent incentive to understand the distribution of discrete lattice orbits is motivated by the study of rational polygonal billiards and of the linear flow F_θ^t (with direction $\theta \in S^1$) on translation surfaces. A complicating aspect in the study of the translation flow is the presence of saddle points – for the billiard flow, the future trajectory of a ball hitting a corner of the polygonal table is ill-defined. Corner-to-corner trajectories correspond to finite geodesics on the translation surface called saddle connections. The set S_M of holonomy vectors of saddle connections on a translation surface M is a discrete planar set in $\mathbf{R}^2$ that records the length and direction of each saddle connection. For a typical translation surface, the number $|S_M \cap B_R|$ of saddle connections of length $\|\mathbf{x}\| < R$ grows quadratically, with a growth asymptotic of the form

$$|S_M \cap B_R| = c_M R^2 + O(R^{2-\delta}) \tag{1.1}$$

[13, 35], where $\delta > 0$ is a non-trivial but non-explicit power-saving.

In the case of Veech surfaces[1] the set S_M of holonomy vectors of saddle connections is a finite disjoint union of discrete lattice orbits [40], which may then be accessed using a combination of tools and ideas from dynamics, ergodic theory, spectral theory, or metric geometry. Hence one can hope to say more, and in particular to specify results to individual surfaces. This is of interest considering for instance that the set of surfaces constructed from rational billiard polygons have null measure in the stratum. As such, results for typical surfaces do not yield new information on billiards in rational polygons.

Our main theorem is an explicit mean value formula for the Siegel–Veech transform of pairs of discrete lattice orbits in the plane. We postpone the statement of the main theorem to Section 1.2 to first describe some applications to the study of holonomies for Veech surfaces. The main novelty of this paper is to obtain estimates towards counting pairs of vectors in S_M. For this we build on tools from the geometry of numbers, the spectral theory of automorphic forms, the weak mixing of the translation flow for non-arithmetic Veech surfaces established by Avila and Delecroix [4], and previous works of the authors [6, 11, 14].

1.1. Applications

Let M be a Veech surface, Γ_M its Veech group, and S_M the set of holonomy vectors of its saddle connections. The results we present here for S_M are proven in the main text in the more general setting of discrete lattice orbits in the plane. References to those more general statements are indicated in Section 1.3.

1.1.1. Counting in Borel sets. In the case of Veech surfaces, we have

$$|S_M \cap B_R| = c_M R^2 + O(R^{2-\delta}), \tag{1.2}$$

[1] We recall that a translation surface is a *Veech surface* if the image in $\mathrm{PSL}_2(\mathbf{R})$ of its stabilizer group under the action of $\mathrm{SL}_2(\mathbf{R})$ on the moduli space of all translation surfaces is a lattice. Veech surfaces are sometimes called *lattice surfaces*.

[6, 40] where this time the power-saving $\delta = \delta(\Gamma_M)$ is explicit,[2] in contrast to (1.1). In fact, the same asymptotic holds when replacing S_M by gS_M for any $g \in G$ – this amounts to counting points in S_M that lie in an ellipsoid centered at the origin. For more general shapes it is typical of analytic approaches that a lack of regularity of the shape's boundary leads to weaker power-savings; see [6, Theorems 2.6, 2.7] or Section 10 for examples. The following theorem recovers a (nearly) optimal count for typical shapes with respect to linear transformations.

Theorem 1.1. *Let $\Omega \subset \mathbf{R}^2$ be a bounded Borel set that contains the origin, and consider its dilates $\Omega_R = R \cdot \Omega$. Then for almost every*[3] *linear transformation A we have*

$$|A(S_M) \cap \Omega_R|\,|\det(A)| = c_M|\Omega|R^2 + O(R^{2-\delta}\log^{3/2}(R)),$$

where $\delta = \delta(\Gamma_M)$ is as in (1.2)*.*

Remark 1.2. The restriction to dilates is cosmetic; see Theorem 10.3 for a counting asymptotic that holds for every element of a linearly ordered family of Borel sets in the plane.

1.1.2. Weak uniform discreteness. Recall that a discrete planar set is *η-uniformly discrete* if $|\Lambda \cap B_\eta(\mathbf{x})| \leq 1$ for all $\mathbf{x} \in \mathbf{R}^2$. Answering a question of Barak Weiss, Wu showed that S_M is not uniformly discrete when M is a non-arithmetic Veech surface [43]; in other words, there exists a pair of vectors that are arbitrarily close. On the other hand, uniform discreteness is easy to prove when M is an arithmetic Veech surface; see Proposition 11.3. The next theorem quantifies the failure of uniform discreteness when M is non-arithmetic.

Theorem 1.3. *Let M be a Veech surface. For each $\varepsilon > 0$ there is an $\eta > 0$ such that*

$$\limsup_{R\to\infty} \frac{|\{\mathbf{x} \in S_M \cap B_R : |S_M \cap B_\eta(\mathbf{x})| \geq 2\}|}{|S_M \cap B_R|} < \varepsilon.$$

The theorem shows that upon discarding an ε-dense set of saddle connections, S_M is η-uniformly discrete for some $\eta > 0$.

1.1.3. Disjointness of flows. The translation flow F_θ^t is uniquely ergodic [30] for Lebesgue almost every direction θ. On a *typical* translation surface (with respect to the Lebesgue measure class on a given stratum) of genus $g \geq 2$, the flow F_θ^t is weakly mixing for almost every θ [5] (but never strongly mixing [26]). For a typical surface the flows F_θ^t and F_ψ^t are disjoint (for almost every pair (θ, ψ) of directions) [11]. An immediate consequence of disjointness is that F_θ^t and F_ψ^t are not isomorphic.

[2]We refer the reader to Section 4 for an explicit description of δ, which is determined by the bottom of the residual spectrum of Γ_M. It is worth noting that there are known constructions of lattices Γ for which δ can be arbitrarily close to 0 [38].

[3]With respect to the Euclidean metric induced by the matrix representation of A.

Regarding individual surfaces, the work of [11] can be applied to establish disjointness for branched covers of the torus in typical directions. So far, covers of tori were the only examples of individual surfaces for which disjointness is shown. In Appendix A, the third author presents a direct proof of the following theorem based on Theorem 1.3.

Theorem 1.4. *Let M be any Veech surface. Then for Lebesgue almost every pair $(\theta, \psi) \in S^1 \times S^1$, the translation flows F_θ^t and F_ψ^t are disjoint.*

This provides the first family of surfaces other than branched covers of tori for which the flows in almost every pair of directions are not isomorphic. If Theorem 1.4 could be extended to all surfaces, we would even be able to recover the result of Chaika and Forni that there is a weakly mixing billiard in a polygon [9].

1.1.4. Counting pairs with bounded determinant. Theorem 1.3 is a consequence of the following stronger density theorem that also accounts for multiplicities. Here we denote $B_\eta^*(\mathbf{x}) = B_\eta(\mathbf{x}) \setminus \{\mathbf{x}\}$.

Theorem 1.5. *Let M be a Veech surface. There exists a constant C such that*

$$\frac{|\{(\mathbf{x}, \mathbf{y}) \in S_M \times S_M : \mathbf{x} \in B_R,\ \mathbf{y} \in B_\eta^*(\mathbf{x})\}|}{|S_M \cap B_R|} < C\eta^2.$$

The same arguments also recover an upper bound on pairs of saddle connections with bounded determinant. For a vector $\mathbf{x} \in \mathbf{R}^2$, set

$$\mathcal{D}_{D,1}(\mathbf{x}) = \{\mathbf{y} \in \mathbf{R}^2 : |\mathbf{y}| \le |\mathbf{x}| \text{ and } |\mathbf{x} \wedge \mathbf{y}| \le D\}.$$

For a typical surface M, the second author with Athreya and Masur [2] showed that for $D > 0$ there is a *non-explicit* constant $C_D > 0$ such that

$$\lim_{R\to\infty} \frac{|\{(\mathbf{x}, \mathbf{y}) \in S_M \times S_M : \mathbf{x} \in B_R,\ \mathbf{y} \in \mathcal{D}_{D,1}(\mathbf{x})|}{R^2} = C_D.$$

Theorem 1.6. *Let M be a Veech surface. For any $D > 0$, there are constants C_M and c depending only on M such that*

$$\limsup_{R\to\infty} \frac{|\{(\mathbf{x}, \mathbf{y}) \in S_M \times S_M : \mathbf{x} \in B_R,\ \mathbf{y} \in \mathcal{D}_{D,1}(\mathbf{x})\}|}{R^2} \le C_M(D + c).$$

Note we have two terms in the upper bound. This comes from the fact that when D is small, there are essentially only parallel pairs, and thus a constant multiple of the Siegel–Veech constant c_Γ will be dominating (cf. [2, Theorem 1.2]). However for D large, we have an upper bound which is asymptotically linear in D.

1.1.5. Pair correlations. In view of these results it is natural to consider the two-dimensional pair correlation function

$$R_2(B_s, S_M, R) := \frac{|\{(\mathbf{x}, \mathbf{y}) \in S_M \times S_M : \mathbf{x} \in B_R,\ \mathbf{y} \in B^*_{s/\sqrt{c_M}}(\mathbf{x})\}|}{|S_M \cap B_R|} \qquad (s > 0)$$

with $s > 0$ and a rescaling determined by the mean spacing $\frac{|S_M \cap B_R|}{|B_R|} \sim c_M$. The pair correlation is said to be *Poissonian* if $R_2(B_s, S_M, R) \to |B_s|$ as $R \to \infty$. Poissonian expresses that we see the same asymptotic behavior for the pair correlation function as if we were looking at points in the plane generated by a two-dimensional Poisson process.

Our last application shows the pair correlation function to be Poissonian on average. To make this precise we need some notation. Let μ denote the probability Haar measure on G/Γ_M and let C be the cone

$$C = \{A = \nu^{1/2} g : g \in G/\Gamma_M,\ \det(A) = \nu \in (0,1]\}$$

equipped with the probability measure given by

$$\int_C f(A)\, dm(A) = \int_0^1 \int_{G/\Gamma_M} f(\nu^{1/2} g)\, d\mu(g)\, d\nu.$$

Theorem 1.7. *For each $s > 0$ we have*

$$\lim_{R\to\infty} \int_C R_2(B_s, A(S_M), R) \det(A)\, dm(A) = |B_s|.$$

We will study the pair correlation function in follow-up work.

1.2. A mean value formula for pairs

Let Γ be a (non-uniform) lattice in G that contains $-I$ with discrete orbit Λ, let $B_c(\mathbf{R}^2)$ denote the space of Borel measurable bounded compactly supported functions on the Euclidean plane.

In analogy to Siegel's mean value formula in the geometry of numbers, Veech [41] introduced the Siegel–Veech transform

$$B_c(\mathbf{R}^2) \ni f \mapsto \Theta_{\Lambda;f}(g) := \sum_{\mathbf{x}\in\Lambda} f(g\mathbf{x})$$

and proved that $\Theta_{\Lambda;f} \in L^\infty(G/\Gamma)$ with first moment

$$\int_{G/\Gamma} \Theta_{\Lambda;f}(g)\, d\mu(g) = c_\Lambda \int_{\mathbf{R}^2} f(\mathbf{x})\, d\mathbf{x}. \tag{1.3}$$

The constant c_Λ is explicit. For each discrete lattice orbit Λ there exists a preferred scaling of Λ in its homothety class $\{\lambda\Lambda : \lambda \neq 0\}$ for which $c_\Lambda = c_\Gamma := \frac{2}{\pi V}$, where V is the hyperbolic covolume of Γ. We say that Λ is a *scaled discrete orbit* if its constant has this form. (The precise scaling λ can be explicitly computed; see Proposition 4.1.)

Similarly for two (not necessarily distinct) discrete Γ-orbits Λ_1, Λ_2, and any $f \in B_c(\mathbf{R}^2 \times \mathbf{R}^2)$, we consider the Siegel–Veech theta-transform

$$\Theta_{\Lambda_1,\Lambda_2;f} : G/\Gamma \to \mathbf{R}, \quad \Theta_{\Lambda_1,\Lambda_2;f}(g) = \sum_{\substack{\mathbf{x}\in\Lambda_1\\ \mathbf{y}\in\Lambda_2}} f(g\mathbf{x}, g\mathbf{y}).$$

To simplify our formulas, we assume that Λ_1, Λ_2 are scaled so that $c_{\Lambda_i} = c_\Gamma$. The general statement can be recovered from the observation that if $\Lambda_i' = \lambda_i \Lambda_i$, then $\Theta_{\Lambda_1',\Lambda_2';f} = \Theta_{\Lambda_1,\Lambda_2;f\circ\lambda}$, with $f \circ \lambda(\mathbf{x},\mathbf{y}) := f(\lambda_1\mathbf{x}, \lambda_2\mathbf{y})$.

Let $\mathcal{N} = \{\det(\mathbf{x}\,|\,\mathbf{y}) : (\mathbf{x},\mathbf{y}) \in \Lambda_1 \times \Lambda_2\}$. Two ordered pairs $(\mathbf{x}_1,\mathbf{y}_1), (\mathbf{x}_2,\mathbf{y}_2) \in \Lambda_1 \times \Lambda_2$ are considered equivalent if $(\mathbf{x}_1,\mathbf{y}_1) = (\gamma\mathbf{x}_2, \gamma\mathbf{y}_2)$ for some $\gamma \in \Gamma$, and we set $\varphi(c)$ to be the number of equivalence classes of pairs $(\mathbf{x},\mathbf{y}) \in \Lambda_1 \times \Lambda_2$ with $\det(\mathbf{x}|\mathbf{y}) = c$. Our main theorem is the following.

Theorem 1.8. *Let Λ_1, Λ_2 be scaled discrete Γ-orbits, and let f be a bounded semicontinuous function of compact support on $\mathbf{R}^2 \times \mathbf{R}^2$. Then $\Theta_{\Lambda_1,\Lambda_2;f} \in L^\infty(G/\Gamma)$ and*

$$\int_{G/\Gamma} \Theta_{\Lambda_1,\Lambda_2;f}(g)\,d\mu(g) = c_\Gamma \sum_{c\in\mathcal{N}-0} \frac{\varphi(c)}{|c|} \int_{\mathbf{R}}\int_{\mathbf{R}^2} f(\mathbf{x}, t\mathbf{x} + c\mathbf{x}^*)\,d\mathbf{x}\,dt + \delta_{\Lambda_1,\Lambda_2} c_\Gamma \int_{\mathbf{R}^2} (f(\mathbf{x},-\mathbf{x}) + f(\mathbf{x},\mathbf{x}))\,d\mathbf{x},$$

where $\mathbf{x}^*$ is such that $\det(\mathbf{x}\,|\,\mathbf{x}^*) = 1$, and $\delta_{\Lambda_1,\Lambda_2}$ is 1 if Λ_1 and Λ_2 are homothetic and is 0 otherwise.

The right-hand side provides a complete description of the arising Siegel–Veech measures, answering a question of Athreya, Cheung, and Masur [1, Sections 4.4, 4.5.3].

The range of admissible test functions f includes characteristic functions of open and closed measurable sets. However, the evaluation of such integrals is not straightforward. Inspired by a trick used by Schmidt in the geometry of numbers, we rewrite the mean value formula above in terms of the cone measure m introduced earlier.

Theorem 1.8′. *With the same notation as above, we have*

$$\begin{aligned}\int_C \Theta_{\Lambda_1,\Lambda_2;f}(A)\,dm(A) &\\ = c_\Gamma^2 \iint_{\mathbf{R}^2\times\mathbf{R}^2} f(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} &+ \delta_{\Lambda_1,\Lambda_2} c_\Gamma \int_{\mathbf{R}^2} (f(\mathbf{x},-\mathbf{x}) + f(\mathbf{x},\mathbf{x}))\,d\mathbf{x} \\ &+ c_\Gamma \iint_{\mathbf{R}^2\times\mathbf{R}^2} \Psi(|\mathbf{x}\wedge\mathbf{y}|) f(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y},\end{aligned}$$

where $|\mathbf{x}\wedge\mathbf{y}| = |\det(\mathbf{x}\,|\,\mathbf{y})|$, and $\Psi(t) = t\sum_{t\le c\in\mathcal{N}-0} \varphi(c)c^{-3} - c_\Gamma$.

The evaluation of the last double integral therefore depends on counting determinants of pairs of linearly independent vectors in $\Lambda_1 \times \Lambda_2$. Using the spectral theory of automorphic forms, we show the following.

Theorem 1.9. *As $t \to \infty$, $\Psi(t) = O(t^{-\delta})$, where δ is as in (1.2).*

Specializing to $\Lambda_1 = \Lambda_2 = \Lambda$ and $f(\mathbf{x},\mathbf{y}) = h(\mathbf{x})h(\mathbf{y})$, we recover the second moment formula (where Λ does not need to be scaled).

Corollary 1.10.

$$\int_C \Big(\sum_{\mathbf{x}\in\Lambda} h(A\mathbf{x})\Big)^2 \,dm(A) = \Big(c_\Lambda \int_{\mathbf{R}^2} h(\mathbf{x})\,d\mathbf{x}\Big)^2 + c_\Lambda \int_{\mathbf{R}^2} (h(\mathbf{x})h(-\mathbf{x}) + h(\mathbf{x})h(\mathbf{x}))\,d\mathbf{x}$$
$$+ c_\Lambda \iint_{\mathbf{R}^2\times\mathbf{R}^2} \Psi(|\mathbf{x}\wedge\mathbf{y}|)h(\mathbf{x})h(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}.$$

1.3. Discussion of proofs and structure of the paper

Section 2 collects notation and initial definitions. In Section 3 we show that non-uniform lattices yield, up to homothety, only finitely many discrete orbits, in one-to-one correspondence to the set of inequivalent cusps for Γ.

In Section 4 we show that counting problems for discrete lattice orbits can be restricted to scaled discrete orbits. We also recall the state of the art for counting results of discrete lattice orbits, and include a full proof of

$$N_R(\Lambda) = c_\Lambda |B_R| + o(|B_R|^{1/2})$$

when Λ is a discrete orbit of a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$; see Theorem 4.3. (A stronger power-saving can be achieved if we assume the Riemann hypothesis.)

In Section 5 we give a representation theorem for all higher moments of the Siegel–Veech transform over the space of semicontinuous functions. Section 6 reviews Veech's proof of (1.3), filling out some details left to the reader in [41] that help explain our restriction from measurable to semicontinuous functions when working with higher moments.

Concluding the background part of this paper, we give in Section 7 a direct proof of Theorem 1.9 (known to follow from a more general result of Good [17]) based on the meromorphic continuation of the scattering matrix.

Our main Theorem 1.8 (mean value formulas for pairs) and its Corollary 1.10 (second moment formulas) are proved in Section 8. The strategy of proof follows Veech's ergodic approach [41] and its extension by the second author to the second moment of the Siegel–Veech transform for Hecke triangle groups [14, Theorem 1.2]. The key realization is that the geometric totient function introduced in [14] is a special instance of the counting function for double cosets of lattices with respect to maximal parabolic subgroups.

We are now in a position to deduce the applications advertised above. Section 9 collects the needed geometric estimates (using the spectral asymptotic provided by Theorem 1.9) and concludes with the proof of Theorem 1.7 (Poissonian pair correlation on average).

Theorem 1.1 (effective counting) follows directly from Theorem 10.4, which is itself the application of the more general Theorem 10.3 to dilated Borel sets containing the origin. The latter theorem should be seen as the discrete lattice orbit analogue of Schmidt's bound on the discrepancy function for (primitive) lattices in $\mathbf{R}^2$ [37, Theorem 2]. In fact our proof mimicks Schmidt's with input the second moment formula from Corollary 1.10 and a geometric estimate from Section 9. This argument of Schmidt has recently been revisited in various contexts; see [3, 29, 36].

In Section 11, Theorems 1.5 and 1.6 (counting pairs in balls or of bounded determinant) are proved using the full force of the mean value formula for pairs and the well-roundedness of Ecluclidean balls. The paper concludes with the proof of Theorem 1.4 (disjointness of flows) in Appendix A, building on a criterion for disjointness developed in [11].

2. Basic definitions

2.1. Notation

Given a countable set A, we denote by $|A|$ the number of its elements. If A is a Borel set in $\mathbf{R}^2$, we denote its characteristic function by $\mathbf{1}_A$ and its Lebesgue measure by $|A|$ or $\lambda(A)$. We use $\|\cdot\|$ to denote the Euclidean norm, $\|\mathbf{x}\| = \sqrt{\sum x_i^2}$. For the disk $B_R(\mathbf{x}) = \{\mathbf{y} \in \mathbf{R}^2 : \|\mathbf{x} - \mathbf{y}\| \leq R\}$, we have $|B_R(\mathbf{x})| = \pi R^2$ and write $B_R := B_R(\mathbf{0})$.

For two functions $f(x)$ and $g(x)$, we write $f \ll g$ or $f = O(g)$ to say that there exists a constant $C > 0$ such that for x sufficiently large, $|f(x)| \leq C|g(x)|$.

We set $G = \mathrm{SL}_2(\mathbf{R})$ and distinguish the subgroups

$$K = \left\{ k_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbf{R}/2\pi\mathbf{Z} \right\},$$

$$A = \left\{ a_y = \begin{pmatrix} y & 0 \\ 0 & y^{-1} \end{pmatrix} : y > 0 \right\},$$

$$N = \left\{ n_x = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbf{R} \right\}.$$

The Haar measure η on G can be written in coordinates as the Lebesgue measure restricted to the set

$$\mathcal{S} = \{(a, b, s) : a \neq 0, b, s \in \mathbf{R}\} \tag{2.1}$$

with the coordinate transformation

$$(a, b, s) \mapsto g(a, b, s) = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} a & b \\ as & bs + a^{-1} \end{pmatrix}.$$

Let $\Gamma < G$ be a discrete subgroup, with respect to the topology induced from $G \subset \mathbf{R}^4$. We will assume that $\Gamma < G$ is a lattice, that is, the quotient G/Γ carries a finite G-invariant Borel measure, which when normalized to a probability measure we denote by μ. The image of Γ in $\mathrm{PSL}_2(\mathbf{R})$ via the canonical projection is a Fuchsian group, which we also denote by Γ.

The action of Γ by fractional linear transformation on the hyperbolic plane $\mathbf{H}^2$, i.e., $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right).z = \frac{az+b}{cz+d}$ is properly discontinuous and admits a finite-area fundamental domain in $\mathbf{H}^2$ whose (hyperbolic) area we denote by $V = \mathrm{area}(\Gamma \setminus \mathbf{H}^2)$. The area of a geometrically finite Fuchsian group is a numerical invariant that does not depend on the particular choice of fundamental domain; we refer the reader to [27] for more properties of Fuchsian groups.

2.2. *Scaling transformations*

A lattice $\Gamma < G$ is non-uniform if and only if the action of Γ on $\mathbf{H}^2$ has parabolic fixed points, called *cusps*. We say that two cusps $\mathfrak{a}$ and $\mathfrak{b}$ are *equivalent* if their Γ-orbits coincide, i.e., if $\Gamma.\mathfrak{a} = \Gamma.\mathfrak{b}$. Since Γ is a lattice, there are at most finitely many inequivalent cusps. For each cusp $\mathfrak{a}$, we set $\Gamma_{\mathfrak{a}} = \{\gamma \in \Gamma : \gamma.\mathfrak{a} = \mathfrak{a}\}$ and note that if $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent cusps with $\mathfrak{b} = \gamma.\mathfrak{a}$, then $\Gamma_{\mathfrak{b}} = \gamma \Gamma_{\mathfrak{a}} \gamma^{-1}$.

Computations are most convenient at the cusp at ∞; up to conjugation by some $\sigma_{\mathfrak{a}} \in G$ such that $\sigma_{\mathfrak{a}}.\infty = \mathfrak{a}$, we may assume that Γ has a cusp at ∞ (up to replacing Γ by $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}$). If $-I \notin \Gamma$, then Γ_∞ is isomorphic to $\mathbf{Z}$ and generated by a matrix of the form $\pm\left(\begin{smallmatrix} 1 & \omega \\ 0 & 1 \end{smallmatrix}\right)$, where we call $\omega > 0$ the *cusp width*, while if $-I \in \Gamma$, then Γ_∞ is isomorphic to $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}$. The image of Γ_∞ in $\mathrm{PSL}_2(\mathbf{R})$ is always isomorphic to $\mathbf{Z}$. The following criterion is well known; see [34, Lemma 1.7.3].

Lemma 2.1. *Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. If $|c|\omega < 1$, then $\gamma \in \Gamma_\infty$.*

Let $\mathfrak{a}$ be a cusp for Γ and choose $\sigma_{\mathfrak{a}} \in G$ such that

$$\sigma_{\mathfrak{a}}.\infty = \mathfrak{a} \quad \text{and} \quad \sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\sigma_{\mathfrak{a}} = \begin{pmatrix} 1 & \mathbf{Z} \\ 0 & 1 \end{pmatrix},$$

that is, we fix a cusp at ∞ whose width ω is scaled to 1. We call $\sigma_{\mathfrak{a}}$ a *scaling transformation*. Notice that the above conditions only determine $\sigma_{\mathfrak{a}}$ up to right translation by elements $n \in N$. One has the following convenient double coset decomposition, which can be seen as a local expression of the Bruhat decomposition; see [25, Theorem 2.7].

Theorem 2.2. *Let $\mathfrak{a}$, $\mathfrak{b}$ be cusps for $\Gamma < \mathrm{PSL}_2(\mathbf{R})$. We have a disjoint union*

$$\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}} = \delta_{\mathfrak{a}\mathfrak{b}}\begin{pmatrix} 1 & \mathbf{Z} \\ 0 & 1 \end{pmatrix} \cup \bigcup_{0 < a \le c} \begin{pmatrix} 1 & \mathbf{Z} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & * \\ c & * \end{pmatrix}\begin{pmatrix} 1 & \mathbf{Z} \\ 0 & 1 \end{pmatrix}, \tag{2.2}$$

where $\delta_{\mathfrak{a}\mathfrak{b}} = 1$ if $\mathfrak{a}$, $\mathfrak{b}$ are Γ-equivalent and $\delta_{\mathfrak{a}\mathfrak{b}} = 0$ otherwise, and a, c run over real numbers such that $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ contains $\left(\begin{smallmatrix} a & * \\ c & * \end{smallmatrix}\right)$.

3. A dichotomy for lattice orbits in the plane

Given a non-zero vector $\mathbf{x} = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ in the Euclidean plane, we say that *the direction of* $\mathbf{x}$ *is fixed by a parabolic motion in* $G = \mathrm{SL}_2(\mathbf{R})$ if there is a parabolic element $g \in G$ such that $g.\frac{x}{y} = \frac{x}{y}$. We have the following (now classical) dichotomy; see [12, Theorem 3.2].

Theorem 3.1. *Let $\mathbf{x} \in \mathbf{R}^2 \setminus \{\mathbf{0}\}$ and let $\Gamma < G$ be a lattice acting linearly on the plane and containing $-I$. The (linear) orbit $\Gamma\mathbf{x}$ is either discrete or dense in $\mathbf{R}^2$. More precisely, $\Gamma\mathbf{x}$ is discrete if and only if the direction of $\mathbf{x}$ is fixed by a parabolic motion in Γ.*

In particular, it follows that if the underlying lattice is uniform, each lattice orbit is dense in the plane. The distribution of dense lattice orbits was studied by many authors;

see [18–20, 28, 31, 33]. We henceforth assume that Γ is non-uniform. Note that the existence of discrete orbits for Γ still holds if $-I \notin \Gamma$; if $\Gamma^{(\pm)}$ denotes the degree 2 central extension generated by $-I$ and Γ, then $\Gamma^{(\pm)}\mathbf{x} = \Gamma\mathbf{x} \cup -\Gamma\mathbf{x}$ is either a disjoint union or collapses to $\Gamma\mathbf{x}$ and the previous result applies. The following proposition characterizes further the discrete orbits that arise.

Proposition 3.2. *Let $\Gamma < G$ be a non-uniform lattice acting linearly on the plane. Up to homothety, there are only finitely many discrete Γ-orbits, which are pairwise disjoint and in one-to-one correspondence with the finitely many inequivalent cusps of Γ.*

To see this, we record the following preliminary lemma.

Lemma 3.3. *For each $t \in \mathbf{R} \cup \{\infty\}$, set*

$$\mathbf{x}_t = \begin{cases} \mathbf{e}_1 & \text{if } t = \infty, \\ \left(\begin{smallmatrix} t \\ 1 \end{smallmatrix}\right) & \text{if } t \in \mathbf{R}. \end{cases}$$

Let $\mathbf{x} = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \in \mathbf{R}^2 \setminus \{\mathbf{0}\}$ and let $\Gamma < \mathrm{SL}_2(\mathbf{R})$ be a lattice acting linearly on $\mathbf{R}^2$. For $\gamma \in \Gamma$ we set $\gamma.t$ to be the image of γ under the extension of Möbius transformations to the extended real line. Then:

(1) *There is a unique $t \in \mathbf{R} \cup \{\infty\}$ such that $\mathbf{x}$ and $\mathbf{x}_t$ are collinear.*

(2) *For every $\gamma \in \Gamma$ and $t \in \mathbf{R} \cup \{\infty\}$, the vectors $\gamma\mathbf{x}_t$ and $\mathbf{x}_{\gamma.t}$ are collinear.*

(3) *For all $s, t \in \mathbf{R} \cup \{\infty\}$ and $\gamma \in \Gamma$, $\gamma.s = t$ if and only if $\mathbf{x}_{\gamma.s}$ and $\mathbf{x}_t$ are collinear.*

Proof of Lemma 3.3. (1) Clearly, each vector $\mathbf{x}$ is collinear to $\mathbf{x}_t$ for $t = \frac{x}{y}$ if $y \neq 0$ and to $\mathbf{e}_1$ otherwise. Moreover, if $\mathbf{x}_s$ and $\mathbf{x}_t$ are collinear, then $s = t$.

(2) Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $t \neq \infty$. If $ct + d \neq 0$, then $\gamma\mathbf{x}_t$ and $\mathbf{x}_{\gamma.t}$ are collinear. If $ct + d = 0$, then $\gamma.t = \infty$, and $\gamma\mathbf{x}_t$ is collinear to $\mathbf{e}_1 = \mathbf{x}_{\gamma.t}$. Suppose now that $t = \infty$. Then $\gamma\mathbf{e}_1 = \left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right)$ while $\mathbf{x}_{\gamma.\infty} = \mathbf{x}_{a/c}$.

The "only if" part of (3) is immediate. For the converse, we know by (2) that $\mathbf{x}_t$ is collinear to $\gamma\mathbf{x}_s$. Suppose first that $s, t \neq \infty$; we have a vector identity of the form $\left(\begin{smallmatrix} t \\ 1 \end{smallmatrix}\right) = \lambda\left(\begin{smallmatrix} as+b \\ cs+d \end{smallmatrix}\right)$ for some $\lambda \neq 0$. We deduce that $\gamma.s = t$. If instead, $s = t = \infty$, we want to show that $\gamma \in \Gamma_\infty$. This then follows from the vector identity $\mathbf{e}_1 = \lambda\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right)$; since $c = 0$, we have $\gamma \in \Gamma_\infty$. The remaining cases follow along the same lines. ■

Proof of Proposition 3.2. If the direction of $\mathbf{x} = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ is fixed by a parabolic element $\gamma \in \Gamma$, then $\mathbf{x}$ is collinear to $\mathbf{x}_\mathfrak{a}$, whereby $\mathfrak{a} = \frac{x}{y}$ is a cusp for Γ and thus the lattice orbits $\Gamma\mathbf{x}$ and $\Gamma\mathbf{x}_\mathfrak{a}$ are homothetic. Moreover, if $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent cusps, then by Lemma 3.3, the orbits $\Gamma\mathbf{x}_\mathfrak{a}$ and $\Gamma\mathbf{x}_\mathfrak{b}$ are homothetic. Hence up to homothety, there are only finitely many discrete Γ-orbits, in one-to-one correspondence with the finitely many inequivalent cusps of Γ.

Suppose that $\mathbf{x} \in \Gamma\mathbf{x}_\mathfrak{a} \cap \Gamma\mathbf{x}_\mathfrak{b}$, say $\mathbf{x} = \gamma_1\mathbf{x}_\mathfrak{a} = \gamma_2\mathbf{x}_\mathfrak{b}$. Then $\mathbf{x}_\mathfrak{a} = \gamma_1^{-1}\gamma_2\mathbf{x}_\mathfrak{b}$ is collinear to $\mathbf{x}_{\gamma_1^{-1}\gamma_2\mathfrak{b}}$ and this implies that $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent. ■

4. Counting results for discrete lattice orbits

Given a discrete orbit $\Lambda = \Gamma\mathbf{x}$ as above, we wish to understand its distribution in the plane, starting with a precise estimate for

$$N_R(\Lambda) = |\Lambda \cap B_R|$$

as $R \to \infty$, where $B_R = \{\mathbf{x} \in \mathbf{R}^2 : \|\mathbf{x}\| < R\}$.

We will first choose preferred representatives in each homothety class of discrete Γ-orbits. Then, up to computing a scaling factor, it suffices to restrict the counting problem to what we will call scaled discrete lattice orbits. Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_h \in \mathbf{R} \cup \{\infty\}$ be a set of representatives for the inequivalent cusps of Γ. Then for each cusp representative $\mathfrak{a}_i$, we set

$$\Lambda_{\mathfrak{a}_i} := \Gamma\sigma_{\mathfrak{a}_i}\mathbf{e}_1, \tag{4.1}$$

where $\sigma_{\mathfrak{a}_i}$ is a scaling transformation for the cusp $\mathfrak{a}_i$ as defined in Section 2.2. By Proposition 3.2, each discrete Γ-orbit is homothetic to one of $\Lambda_{\mathfrak{a}_1}, \ldots, \Lambda_{\mathfrak{a}_h}$. We call $\Lambda_{\mathfrak{a}}$ as in (4.1) the *scaled* Γ*-orbit attached to* $\mathfrak{a}$. Observe that $\Lambda_{\mathfrak{a}}$ does not depend on the particular choice of scaling $\sigma_{\mathfrak{a}}$ or of cusp representative. Now if $\Lambda = \lambda\Lambda_{\mathfrak{a}}$, the counting function scales by $N_R(\Lambda) = N_{R/\lambda}(\Lambda_{\mathfrak{a}})$. The next proposition provides an algebraic formula to compute the scaling λ. (For a geometric interpretation of λ, see [21, Theorem 6.1].)

Proposition 4.1. *Given a discrete orbit* $\Gamma\mathbf{x}$ *associated to cusp* $\mathfrak{a}$ *of* Γ*, we have* $\Gamma\mathbf{x} = \lambda\Lambda_{\mathfrak{a}}$ *with*

$$\lambda = \frac{\|\mathbf{x}\|}{\sqrt{\operatorname{tr}(S\gamma_{\mathfrak{a}})}},$$

where $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ *and* $\gamma_{\mathfrak{a}}$ *denotes the cyclic generator of* $\Gamma_{\mathfrak{a}}$.

Proof. There exists some $g \in G$ such that $\mathbf{x} = g\mathbf{e}_1$. Then $g^{-1}\gamma_{\mathfrak{a}}g = \left(\begin{smallmatrix} 1 & \omega \\ 0 & 1 \end{smallmatrix}\right)$ for some $\omega > 0$. We may choose $\sigma_{\mathfrak{a}} = ga_{\sqrt{\omega}}$ as scaling transformation; it then follows that $\Gamma\mathbf{x} = \frac{1}{\sqrt{\omega}}\Lambda_{\mathfrak{a}}$. It remains to compute the cusp width ω. Write $\mathbf{x} = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Then $g = \left(\begin{smallmatrix} x & * \\ y & * \end{smallmatrix}\right)$ and a direct computation shows that $\gamma_{\mathfrak{a}} = g\left(\begin{smallmatrix} 1 & \omega \\ 0 & 1 \end{smallmatrix}\right)g^{-1} = \left(\begin{smallmatrix} * & x^2\omega \\ -y^2\omega & * \end{smallmatrix}\right)$, $S\gamma_{\mathfrak{a}} = \left(\begin{smallmatrix} y^2\omega & * \\ * & x^2\omega \end{smallmatrix}\right)$ and thus $\omega = \frac{\operatorname{tr}(S\gamma_{\mathfrak{a}})}{\|\mathbf{x}\|^2}$. ∎

The current best counting result for $N_R(\Lambda)$ is given by the following theorem.

Theorem 4.2 ([6, Theorem 4.1]). *Let* Λ *be a scaled discrete* Γ*-orbit, and let* $g \in G$*. There are numbers* $1 = \rho_0 > \rho_1 > \cdots > \rho_k > 1/2$ *and constants* $c_1, \ldots, c_k$ *such that we have the asymptotic expansion*

$$N_R(g\Lambda) = |g\Lambda \cap B_R| = c_\Gamma|B_R| + c_1|B_R|^{\rho_1} + \cdots + c_k|B_R|^{\rho_k} + O(|B_R|^{2/3}), \tag{4.2}$$

where the leading constant is given by

$$c_\Gamma = \begin{cases} \dfrac{2}{\pi V} & \text{if } -I \in \Gamma, \\[2mm] \dfrac{1}{\pi V} & \text{if } -I \notin \Gamma. \end{cases} \tag{4.3}$$

The constants $\rho_i \in (1/2, 1)$ are eigenparameters of the finitely many possible small residual eigenvalues of the Laplacian, with $\lambda_i = \rho_i(1-\rho_i) \in (0, 1/4)$, where λ_i belongs to the residual spectrum. Set

$$\delta = \min\{2(1-\rho_1), 2/3\}. \tag{4.4}$$

Then $N_R(g\Lambda) = c_\Gamma \pi R^2 + O(R^{2-\delta})$. There are known constructions of non-uniform lattices for which ρ_1 is arbitrarily close to 1 [38].

The small eigenvalues λ_i, when they arise, are not explicit. It is known that $\mathrm{SL}_2(\mathbf{Z})$ and its congruence subgroups have no small residual eigenvalues [25] and this is also true for triangle groups [22, p. 583]. In such cases (4.2) reduces to $N_R(g\Lambda) = c_\Gamma |B_R| + O(|B_R|^{2/3})$, where the exponent $2/3$ is an artefact of the method of proof. For specific lattices, a better result is non-etheless possible. This is easy to see for congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$. For the convenience of the reader, we include a full proof here.

Theorem 4.3. *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$, and let Λ be a scaled discrete Γ-orbit. Then*

$$N_R(\Lambda) = c_\Gamma |B_R| + o(|B_R|^{1/2}).$$

Assuming the Riemann hypothesis (RH), the error term $o(|B_R|^{1/2})$ can be replaced by $O(|B_R|^{5/12+\varepsilon})$.

Proof. Recall that Γ is a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$ if it contains a principal congruence subgroup $\Gamma(N) = \ker(\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}))$ for some $N \geq 1$. We have a finite disjoint decomposition $\Gamma = \bigcup \tau\Gamma(N)$ for a choice of coset representatives $\tau \in \Gamma/\Gamma(N)$. This implies

$$N_R(\Gamma\mathbf{x}) = \sum_\tau N_R(\tau\Gamma(N)\mathbf{x}).$$

We will consider each summand separately.

Since Λ is assumed to be scaled, we have $\mathbf{x} = \sigma_{\mathfrak{a}}\mathbf{e}_1$ for some choice of scaling transformation $\sigma_{\mathfrak{a}}$. Each cusp of Γ is a cusp of $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ and hence we may choose $\sigma \in \Gamma(1)$ such that $\sigma(\infty) = \mathfrak{a}$ and set $\sigma_{\mathfrak{a}} = \sigma a_{\sqrt{\omega}}$, where ω is the cusp width of $\sigma^{-1}\Gamma\sigma$. Then

$$N_R(\tau\Gamma(N)\mathbf{x}) = N_R(\tau\Gamma(N)\sigma a_{\sqrt{\omega}}\mathbf{e}_1) = N_R(\tau\sigma\Gamma(N)a_{\sqrt{\omega}}\mathbf{e}_1),$$

where we use the fact that $\Gamma(N)$ is a normal subgroup of $\Gamma(1)$. For readability we write $A = \tau\sigma$. Then

$$\begin{aligned} N_R(A\Gamma(N)a_{\sqrt{\omega}}\mathbf{e}_1) &= |\{\gamma \in \Gamma(N)/a_{\sqrt{\omega}}\Gamma(N)_\infty a_{\sqrt{\omega}}^{-1} : \gamma\mathbf{e}_1 \in \omega^{-1/2}A^{-1}B_R\}| \\ &= \frac{\omega}{N}|\{\xi \in \mathbf{Z}^2_{\mathrm{prim}} : \xi \equiv \mathbf{e}_1 \ (\mathrm{mod}\ N),\ \xi \in \omega^{-1/2}A^{-1}B_R\}|. \end{aligned} \tag{4.5}$$

Here $\mathbf{Z}^2_{\mathrm{prim}}$ denotes the set of all vectors $\xi = \binom{m}{n}$ with coprime integer coordinates m, n, and $\omega^{-1/2}A^{-1}B_R$ is an ellipsoid centered at the origin with area $\frac{\pi}{\omega}R^2$.

We have

$$\frac{N}{\omega}\cdot(4.5)$$
$$= \sum_{\substack{(m,n)\in\mathbf{Z}^2_{\mathrm{prim}}\\ m\equiv 1\,(N)\\ N\mid n}} \mathbf{1}_{\omega^{-1/2}A^{-1}B_R}((m,n)) = \sum_{\substack{d\geq 1\\ (d,N)=1}} \mu(d) \sum_{\substack{(m,n)\in\mathbf{Z}^2\\ m\equiv 1\,(N)\\ d\mid m,\, dN\mid n}} \mathbf{1}_{\omega^{-1/2}A^{-1}B_R}((m,n)).$$

where $\mu(n)$ is the Möbius function and we have used the fact that $\sum_{d\mid(m,n)}\mu(d)=1$ iff $(m,n)=1$ (and is 0 otherwise). By the Chinese remainder theorem, the system of congruence equations for m has a unique solution m_0 in $[0,dN)$ and thus

$$\begin{aligned}\frac{N}{\omega}\cdot(4.5) &= \sum_{\substack{d\geq 1\\ (d,N)=1}} \mu(d) \sum_{\substack{\xi\in\mathbf{Z}^2\\ \xi\equiv\binom{m_0}{0}\,(dN)}} \mathbf{1}_{\omega^{-1/2}A^{-1}B_R}(\xi)\\ &= \sum_{\substack{d\geq 1\\ (d,N)=1}} \mu(d) \sum_{\xi\in\mathbf{Z}^2} \mathbf{1}_{\frac{1}{dN}(\omega^{-1/2}A^{-1}B_R-m_0\mathbf{e}_1)}(\xi).\end{aligned}$$

For R sufficiently large, $\Omega=\frac{1}{dN}(\omega^{-1/2}A^{-1}B_R-m_0\mathbf{e}_1)$ is a compact convex planar set of area

$$|\Omega|=\frac{|B_R|}{\omega(dN)^2}$$

that contains the origin as an inner point. A simple geometric argument shows that

$$\sum_{\xi\in\mathbf{Z}^2}\mathbf{1}_\Omega(\xi)=\frac{|B_R|}{\omega(dN)^2}+O_{\omega,N}\Big(\frac{R}{d}\Big). \tag{4.6}$$

Using available estimates towards the Gauss circle problem in convex planar domains, the error term can further be replaced by $O(R^{46/73}\log^{315/146}(R))$ [23, Theorem 5]. Let $\rho>1/2$. To summarize, we have, formally,

$$(4.5)=\frac{1}{N^3}\sum_{(d,N)=1}\frac{\mu(d)}{d^2}|B_R|+O\Big(\sum_{d\geq 1}\frac{\mu(d)}{d^\rho}|B_R|^{\rho/2}\Big)$$

with convergence guaranteed when $\rho>1$. Via the prime number theorem, the error term can be replaced with $o(|B_R|^{1/2})$ and under RH by $O(|B_R|^{5/12+\varepsilon})$; see [24]. We conclude that

$$N_R(\Lambda)=\frac{[\Gamma:\Gamma(N)]}{N^3}\sum_{(d,N)=1}\frac{\mu(d)}{d^2}|B_R|+o(|B_R|^{1/2})$$

(unconditionally on RH or with $O(|B_R|^{5/12+\varepsilon})$ under RH). The leading constant can be shown to equal c_Γ; see [34, Theorem 4.2.5]. ∎

5. Theta transforms

Let $k > 0$ and let $\Lambda_1, \ldots, \Lambda_k$ be k discrete Γ-orbits. Let f be a Borel measurable function on $(\mathbf{R}^2)^k = \mathbf{R}^2 \times \cdots \times \mathbf{R}^2$ (k factors). We define the corresponding theta/Siegel–Veech transform to be

$$\Theta = \Theta_{\Lambda_1,\ldots,\Lambda_k;f} : G/\Gamma \to \mathbf{R}, \quad \Theta(g\Gamma) = \sum_{(\mathbf{x}_1,\ldots,\mathbf{x}_k)\in\Lambda_1\times\cdots\times\Lambda_k} f(g\mathbf{x}_1, \ldots, g\mathbf{x}_k).$$

Lemma 5.1. *If f is bounded with compact support, then $\Theta \in L^\infty(G/\Gamma)$.*

Proof. Since f is bounded and has compact support, for all $(\mathbf{x}_1, \ldots, \mathbf{x}_k) \in (\mathbf{R}^2)^k$, and for some constant c_f, $R > 0$, we have $f(\mathbf{x}_1, \ldots, \mathbf{x}_k) \le c_f \mathbf{1}_{B_R}(\mathbf{x}_1) \cdots \mathbf{1}_{B_R}(\mathbf{x}_k)$. By [41, Lemma 16.10], since $\mathbf{1}_{B_R}$ is bounded with compact support, $\Theta_{\Lambda_i;\mathbf{1}_{B_R}}$ is uniformly bounded over G/Γ. Thus since μ is a probability measure, we conclude that $\Theta_{\Lambda_i;\mathbf{1}_{B_R}} \in L^\infty(G/\Gamma)$. We then notice that $\Theta_{\Lambda_1,\ldots,\Lambda_k;f} \le c_f \prod_{i=1}^k \Theta_{\Lambda_i;\mathbf{1}_{B_R}} \in L^\infty(G/\Gamma)$ since G/Γ is a probability space. ■

Recall that a function $f : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ is *lower* (resp. *upper*) *semicontinuous at* $\mathbf{x}_0$ if $\liminf_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) \ge f(\mathbf{x}_0)$, respectively $\limsup_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) \le f(\mathbf{x}_0)$. Characteristic functions of open (resp. closed) sets are lower (resp. upper) semicontinuous. Semicontinuous functions behave nicely with respect to monotone approximation; see Baire's theorem or Remark 5.4 below.

Definition 5.2. We denote by $SC((\mathbf{R}^2)^k)$ the space of all lower semicontinuous functions bounded below and all upper semicontinuous functions bounded above.

Lemma 5.3. *There exists a unit regular G-invariant Borel measure ν on $(\mathbf{R}^2)^k$ such that for any $f \in SC((\mathbf{R}^2)^k)$, we have*

$$\int_{G/\Gamma} \Theta_{\Lambda_1,\ldots,\Lambda_k;f}(g)\, d\mu(g) = \int_{(\mathbf{R}^2)^k} f(\mathbf{x})\, d\nu(\mathbf{x}). \tag{5.1}$$

Proof. By Lemma 5.1, for $f \in C_c((\mathbf{R}^2)^k)$, we have $\Theta_{\Lambda_1,\ldots,\Lambda_k;f} \in L^\infty(G/\Gamma)$. Thus the assignment $f \mapsto \int_{G/\Gamma} \Theta_{\Lambda_1,\ldots,\Lambda_k;f}(g)\, d\mu(g)$ defines a G-invariant positive linear functional on $C_c((\mathbf{R}^2)^k)$ where G acts diagonally on $(\mathbf{R}^2)^k$. Hence by the Riesz–Markov–Kakutani representation theorem, there is a unit regular G-invariant Borel measure ν on $\mathbf{R}^2$ such that (5.1) holds for $f \in C_c((\mathbf{R}^2)^k)$.

Every lower semicontinuous function f bounded below can be approximated by a non-decreasing sequence $(f_j) \subset C_c((\mathbf{R}^2)^k)$ that converges pointwise to f. Since f is bounded below, it suffices to consider non-negative semicontinuous functions. By Baire's thereom, $\tilde{f}$ is the pointwise limit of continuous non-increasing functions h_k. Choosing a continuous bump function $\tilde{h}_k$ to be 1 on $B(0, k)$ and 0 outside $B(0, k + 1)$, the sequence $h_k\tilde{h}_k$ is the desired sequence of continuous functions of compact support. If f is upper semicontinuous and bounded above, then $-f$ is lower semicontinuous and bounded below, reducing the proof to the case of lower semicontinuous functions.

Moreover, pointwise monotone convergence implies $\Theta_{\Lambda_1,\ldots,\Lambda_k;f_j}$ also monotonically increases to $\Theta_{\Lambda_1,\ldots,\Lambda_k;f}$ pointwise for each $g \in G/\Gamma$. We can then apply the monotone convergence theorem. We emphasize here that we use pointwise convergence (in $(\mathbf{R}^2)^k$) in order to automatically guarantee pointwise convergence (in G/Γ) under the transformation $f \mapsto \Theta_{\Lambda_1,\ldots,\Lambda_k;f}$. ∎

Remark 5.4. The proof makes use of the fact that the dual of functions $f \in C_c((\mathbf{R}^2)^k)$ which are continuous with compact support are exactly the Radon measures on $(\mathbf{R}^2)^k$. It is not straightforward to extend this representation from $f \in C_c((\mathbf{R}^2)^k)$ to all integrable Borel functions. For example, if $A = \{x : 1 \leq |x| < 2\}$, the characteristic function $\mathbf{1}_A$ cannot be pointwise approximated from above or below by sequences of continuous functions. But lower (or upper) semicontinuous functions precisely have this property.

6. Veech's mean value theorem

For simplicity of exposition, we will from now on state our results for scaled discrete lattice orbits. Recall that each discrete Γ-orbit Λ_i is homothetic to some scaled orbit Λ_{α_i}; explicitly, we write $\Lambda_i = \lambda_i \Lambda_{\alpha_i}$ and set $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k) \in \mathbf{R}^k$. Then

$$\Theta_{\Lambda_1,\ldots,\Lambda_k;f} = \Theta_{\Lambda_{\alpha_1},\ldots,\Lambda_{\alpha_k};f\circ\boldsymbol{\lambda}},$$

where $f \circ \boldsymbol{\lambda}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = f(\lambda_1\mathbf{x}_1, \ldots, \lambda_k\mathbf{x}_k)$.

Let Λ be a scaled discrete Γ-orbit. The following mean value theorem, extending the classical theorem of Siegel [39] in the geometry of numbers, is due to Veech [41].

Theorem 6.1 ([41, Theorem 6.5]). *Let Λ be a scaled discrete Γ-orbit. For each integrable Borel function $f : \mathbf{R}^2 \to \mathbf{R}$, the Siegel–Veech transform $\Theta_{\Lambda;f}$ satisfies*

$$\int_{G/\Gamma} \Theta_{\Lambda;f}(g)\, d\mu(g) = c_\Gamma \int_{\mathbf{R}^2} f(\mathbf{x})\, d\mathbf{x}, \tag{6.1}$$

with c_Γ as defined in (4.3).

Remark 6.2. For a general (i.e., not scaled) discrete Γ-orbit Λ, the constant c_Γ is simply replaced by c_Γ/λ^2 with λ as defined in Proposition 4.1.

For completeness, we include a proof of Veech's theorem. The statement is fairly immediate for semicontinuous functions, but the extension to integrable Borel functions stated by Veech actually requires the following technical lemma. Here λ_n denotes the Lebesgue measure on $\mathbf{R}^n$.

Lemma 6.3. *If $\{f_k\}_{k=1}^{\infty}$, f are Borel functions in $L^1(\mathbf{R}^2, \lambda_2)$ with $f_k(\mathbf{y}) \to f(\mathbf{y})$ for λ-a.e. $\mathbf{y}$, then $\Theta_{\Lambda;f_k}(g) \to \Theta_{\Lambda;f}(g)$ for μ-a.e. $g \in G/\Gamma$.*

Proof. Recall the definition (2.1) of the set $\mathcal{S}$ and let $\mathcal{Y} \subset \mathcal{S}$ correspond to $\mathcal{G} = \{g \in G : \Theta_{\Lambda;f_k}(g) \nrightarrow \Theta_{\Lambda;f}(g)\}$ up to a set of measure zero. Let $Y = \{\mathbf{y} \in \mathbf{R}^2 : f_k(\mathbf{y}) \nrightarrow f(\mathbf{y})\}$.

The set $\mathcal{Y}$ is contained in

$$\begin{aligned}\bigcup_{\mathbf{x}\in\Lambda}&\{g(a,b,s)\mathbf{x}\in Y\}\\ &=\bigcup_{\mathbf{x}\in\Lambda}\bigcup_{k\geq 1}\{g(a,b,s)\mathbf{x}\in Y : a\in[-2^k,-2^{-k}]\cup[2^{-k},2^k],\ (b,s)\in[-2^k,2^k]^2\}.\end{aligned}$$

We denote the latter set by $\mathcal{X}_{\mathbf{x}}^k\subseteq\mathbf{R}^3$. Fix $\mathbf{x}\in\mathbf{R}^2\setminus\{\mathbf{0}\}$ and $k\in\mathbf{N}$. We will show that

$$\lambda_3(\mathcal{X}_{\mathbf{x}}^k)=0.$$

By the coarea formula for Lipschitz functions [15] and our hypothesis $\lambda_2(Y)=0$, for J the Jacobian of the map $(a,b,c)\mapsto g(a,b,c)$ and $\mathcal{H}^1$ the Hausdorff 1-measure, we have

$$\int_{\mathcal{X}_{\mathbf{x}}^k} J(g(a,b,s)\mathbf{x})\,d\lambda_3(a,b,s)=\int_Y \mathcal{H}^1(\{a,b,s) : g(a,b,s)\mathbf{x}=\mathbf{y}\})\,d\lambda_2(\mathbf{y})=0.$$

Hence $\lambda_3(\mathcal{X}_{\mathbf{x}}^k)=0$ if and only if $\{J(g\mathbf{x})=0\}\cap\mathcal{X}_{\mathbf{x}}^k$ has measure zero. In fact, we can directly compute that

$$D(g(\cdot,\cdot,\cdot)\mathbf{x})(a,b,s)=\begin{pmatrix} x_1 & x_2 & 0\\ sx_1-a^{-2}x_2 & sx_2 & ax_1+bx_2\end{pmatrix}.$$

Since $\mathbf{x}\neq\mathbf{0}$ and $a\neq 0$, in fact $D(g\mathbf{x})$ has rank 2, so $D(g\mathbf{x})[D(g\mathbf{x})]^T$ has full rank and hence the Jacobian satisfies $(J(g\mathbf{x}))^2=\det(D(g\mathbf{x})[D(g\mathbf{x})]^T)\neq 0$.

We have now shown $\lambda_3(\mathcal{X}_{\mathbf{x}}^k)=0$, which implies that $\eta(\mathcal{G})=0$. To conclude the proof, we note that since μ is the probability measure when η is restricted to a finite volume fundamental domain associated to G/Γ, we in fact have convergence $\Theta_{\Lambda;f_k}(g)\to\Theta_{\Lambda;f}(g)$ for μ-a.e. $g\in G/\Gamma$. ■

Proof of Theorem 6.1. Let $f\in SC(\mathbf{R}^2)$. Then Lemma 5.3 applies, and we seek to determine ν. The plane $\mathbf{R}^2$ decomposes into the disjoint G-orbits $\{\mathbf{0}\}$ and $\mathbf{R}^2\setminus\{\mathbf{0}\}$ and hence the only G-invariant measures on $\mathbf{R}^2$ (up to scaling) are δ_0, the point mass at the origin, and the Lebesgue measure λ. Since the Haar measure on G is unique up to scaling, we conclude that $\nu=\alpha\delta_0+\beta\lambda$ for some real constants $\alpha,\beta\geq 0$.

We first show that $\alpha=0$. Indeed, since $f=\mathbf{1_0}$ is the characteristic function of a closed set, $\Lambda\subset\mathbf{R}^2\setminus\{\mathbf{0}\}$ implies that $\alpha=0$.

Let f be the characteristic function of the open disk of radius R centered at the origin, and recall that characteristic functions of open sets are lower semicontinuous. Hence for $R>0$, we are left with

$$\beta=\int_{G/\Gamma}\frac{|g\Lambda\cap B_R|}{|B_R|}\,d\mu(g).$$

Letting $R\to\infty$, we conclude with Theorem 4.2 that $\beta=c_\Gamma$.

To extend the statement from semicontinuous functions to integrable Borel functions, we first consider the class of bounded Borel functions with compact support. With

Lemma 6.3 in hand, we can use the dominated convergence theorem (where pointwise almost everywhere convergence is needed on both the left and right hand side of (6.1)) to bootstrap our way to any integrable Borel function. By Lemmas 5.1 and 6.3, we apply dominated convergence on both sides of (6.1) to a sequence of semicontinuous functions converging a.e. to any bounded Borel function with compact support.

We now extend to all non-negative Borel functions. Consider $A_k = \{|\mathbf{x}| \le 2^k\} \cap \{f(\mathbf{x}) \le 2^k\}$ and define $f_k(\mathbf{x}) := f(\mathbf{x})\mathbf{1}_{A_k}(\mathbf{x})$. Monotone convergence implies that

$$\lim_{k\to\infty} \int_{\mathbf{R}^2} f_k(\mathbf{x})\, d\mathbf{x} = \int_{\mathbf{R}^2} f(\mathbf{x})\, d\mathbf{x}$$

(even when both sides are infinite). Each $f_k(\mathbf{x})$ is a bounded Borel function with compact support, hence L^1 so that (6.1) holds for each k. Clearly $f_k \le f_{k+1}$ implies that $\Theta_{\Lambda;f_k} \le \Theta_{\Lambda;f_{k+1}}$ for each k. Hence, by Lemma 6.3 and monotone convergence, we have

$$\begin{aligned}\int_{G/\Gamma} \Theta_{\Lambda;f}(g)\, d\mu(g) = \lim_{k\to\infty} \int_{G/\Gamma} \Theta_{\Lambda;f_k}(g)\, d\mu(g) \ &= \ c_\Gamma \lim_{k\to\infty} \int_{\mathbf{R}^2} f_k(\mathbf{x})\, d\mathbf{x} \\ &= c_\Gamma \int_{\mathbf{R}^2} f(\mathbf{x})\, d\mathbf{x}.\end{aligned}$$

Finally, for any integrable Borel function f, we decompose f into its positive and negative parts to complete the proof of Theorem 6.1. ∎

7. Admissible determinants

Our goal is to extend Theorem 6.1 to an integral formula for pairs of discrete lattice orbits. Similarly to the proof of Veech, our formula will build on the decomposition of the space $\mathbf{R}^2 \times \mathbf{R}^2$ into the union of disjoint G-orbits given by

$$\begin{aligned}&\mathbf{R}^2 \times \mathbf{R}^2 \\ &\quad = \{\mathbf{0}\} \cup (\{\mathbf{0}\} \times \mathbf{R}^2) \cup (\mathbf{R}^2 \times \{\mathbf{0}\}) \cup \{(\mathbf{x},\mathbf{y}) : \det(\mathbf{x} \mid \mathbf{y}) = 0\} \cup \{(\mathbf{x},\mathbf{y}) : \det(\mathbf{x} \mid \mathbf{y}) \ne 0\} \\ &\quad = G(\mathbf{0},\mathbf{0}) \cup G(\mathbf{0},\mathbf{e}_1) \cup G(\mathbf{e}_1,\mathbf{0}) \cup \{G(\mathbf{e}_1, t\mathbf{e}_1) : t \in \mathbf{R}\} \cup \{G(\mathbf{e}_1, c\mathbf{e}_2) : c \in \mathbf{R}^*\}.\end{aligned}$$

where we write $\det(\mathbf{x} \mid \mathbf{y})$ for the determinant of the matrix with ordered columns $\mathbf{x}$, $\mathbf{y}$. In particular, we will need to understand the set of determinants $\det(\mathbf{x} \mid \mathbf{y})$ arising from ordered pairs $(\mathbf{x},\mathbf{y}) \in \Lambda_1 \times \Lambda_2$. In this section we record the necessary preliminary results. The mean value formula for pairs of discrete lattice orbits will be derived in the next section.

In this section, we fix two (not necessarily distinct) scaled discrete lattice orbits $\Lambda_{\mathfrak{a}}$, $\Lambda_{\mathfrak{b}}$. Let

$$\mathcal{N}_{\mathfrak{ab}} = \{c \in \mathbf{R} : \text{there are } \mathbf{x} \in \Lambda_{\mathfrak{a}}, \mathbf{y} \in \Lambda_{\mathfrak{b}} \text{ such that } \det(\mathbf{x} \mid \mathbf{y}) = c\}$$

be the set of *admissible determinants* for vector pairs in $\Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}}$. Under the assumption that $-I \in \Gamma$, we have $c \in \mathcal{N}_{\mathfrak{ab}}$ if and only if $-c \in \mathcal{N}_{\mathfrak{ab}}$.

Lemma 7.1. *The following statements are equivalent:*

(1) $0 \in \mathcal{N}_{\mathfrak{a}\mathfrak{b}}$;

(2) $\mathfrak{a}$ *and* $\mathfrak{b}$ *are equivalent cusps;*

(3) $\Lambda_{\mathfrak{a}} = \Lambda_{\mathfrak{b}}$.

Proof. If $0 \in \mathcal{N}_{\mathfrak{a}\mathfrak{b}}$, there is $\gamma \in \Gamma$ such that $\gamma \mathbf{x}_{\mathfrak{a}}$ and $\mathbf{x}_{\mathfrak{b}}$ are collinear, and hence by Lemma 3.3, the cusps $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent. Then $\Lambda_{\mathfrak{a}} = \Lambda_{\mathfrak{b}}$ (we may take $\sigma_{\mathfrak{b}} = \gamma\sigma_{\mathfrak{a}}$) and this in turns implies immediately that $0 \in \mathcal{N}_{\mathfrak{a}\mathfrak{b}}$. ■

We introduce an equivalence relation on the set of ordered pairs $(\mathbf{x}, \mathbf{y}) \in \Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}}$ by declaring $(\mathbf{x}_1, \mathbf{y}_1) \sim (\mathbf{x}_2, \mathbf{y}_2)$ iff $(\mathbf{x}_1, \mathbf{y}_1) = (\gamma\mathbf{x}_1, \gamma\mathbf{x}_2)$ for some $\gamma \in \Gamma$. The determinant $\det(\mathbf{x} \mid \mathbf{y})$ is preserved by the equivalence relation. We set

$$\varphi_{\mathfrak{a}\mathfrak{b}}(c) = |\{[(\mathbf{x}, \mathbf{y})] : (\mathbf{x}, \mathbf{y}) \in \Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}}, \det(\mathbf{x} \mid \mathbf{y}) = c\}|$$

to be the counting function of equivalence pairs with determinant c.

Lemma 7.2. *Suppose that* $-I \in \Gamma$ *and let* $\mathcal{N}^*_{\mathfrak{a}\mathfrak{b}} := \mathcal{N}_{\mathfrak{a}\mathfrak{b}} \setminus \{0\}$. *Then* $c \in \mathcal{N}^*_{\mathfrak{a}\mathfrak{b}}$ *if and only if there exists a vector* $\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right) \in (\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}})\mathbf{e}_1$ *with* $0 < a \le |c|$. *In particular,*

$$\varphi_{\mathfrak{a}\mathfrak{b}}(c) = \left|\left\{0 < a \le |c| : \begin{pmatrix} a \\ c \end{pmatrix} \in (\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}})\mathbf{e}_1\right\}\right|.$$

Proof. Let $(\mathbf{x}, \mathbf{y})$ be an ordered pair in $\Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}}$ such that $\det(\mathbf{x} \mid \mathbf{y}) = c$. We can choose $\mathbf{x}'$ such that $(\mathbf{x} \mid \mathbf{x}') \in \Gamma\sigma_{\mathfrak{a}}$. Then

$$(\mathbf{x} \mid \mathbf{x}')^{-1}(\mathbf{x} \mid \mathbf{y}) = \begin{pmatrix} 1 & a \\ 0 & c \end{pmatrix}.$$

This shows that $\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right) \in \sigma_{\mathfrak{a}}^{-1}\Gamma\mathbf{y} = (\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}})\mathbf{e}_1$. Under the assumption that $-I \in \Gamma$, we have $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}$. By matrix multiplication on the left with an appropriate power $\left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right)$, we have

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} a + mc \\ c \end{pmatrix}$$

with $0 < a + mc \le |c|$. Hence we may always choose a with this property. Moreover, we recover the same vector for any equivalent ordered pair $(\mathbf{x}_1, \mathbf{y}_1)$. This assignment is easily seen to be surjective. Indeed, given $\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right) \in (\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}})\mathbf{e}_1$, choose $\mathbf{x} = \sigma_{\mathfrak{a}}\mathbf{e}_1$ and $\mathbf{y} = \sigma_{\mathfrak{a}}\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right)$. To prove that it is also injective, suppose that $(\mathbf{x}_1 \mid \mathbf{y}_1) = (\mathbf{x}_1 \mid \mathbf{x}'_1)(\mathbf{x}_2 \mid \mathbf{x}'_2)^{-1}(\mathbf{x}_2 \mid \mathbf{y}_2)$. Then $(\mathbf{x}_1 \mid \mathbf{y}_1)$ and $(\mathbf{x}_2 \mid \mathbf{y}_2)$ are equivalent ordered pairs of vectors. ■

Remark 7.3. Observe that $\varphi_{\mathfrak{a}\mathfrak{b}}(c)$ counts the number of double coset representatives with lower left entry given by c; compare to Theorem 2.2.

When $\Gamma = \mathrm{SL}_2(\mathbf{Z})$, we may take $\sigma_\infty = I$ and $\Lambda_\infty = \mathrm{SL}_2(\mathbf{Z})\mathbf{e}_1$ for the unique cusp at infinity. Then each integer $n \in \mathbf{Z}$ is realized as an admissible determinant, and $\varphi_\infty(n) = \varphi(|n|)$, Euler's totient function, which is trivially bounded above by $\varphi(n) \leq n-1$ for each $n \in \mathbf{N}$. For more general groups, and in particular groups that are not finite-index subgroups of $\mathrm{SL}_2(\mathbf{Z})$, one only has the estimate $|\varphi_{\mathfrak{a}\mathfrak{b}}(c)| \ll |c|^2$ as a consequence of the pigeonhole principle; see [25, Proposition 2.8]. However, we have stronger bounds on average via the scattering matrix for Γ, which we quickly recall. If h is the number of inequivalent cusps for Γ, the scattering matrix $(\phi_{\mathfrak{a}\mathfrak{b}}(s))_{\mathfrak{a},\mathfrak{b}}$ for Γ is the $h \times h$ matrix with entries given by

$$\phi_{\mathfrak{a}\mathfrak{b}}(s) = \sqrt{\pi}\,\frac{\Gamma(s-1/2)}{\Gamma(s)} \sum_{\substack{c \in \mathcal{N}^*_{\mathfrak{a}\mathfrak{b}} \\ c>0}} \frac{\varphi_{\mathfrak{a}\mathfrak{b}}(c)}{c^{2s}},$$

where $s \in \mathbf{C}$ and $\Gamma(s)$ is Euler's gamma-function. Each entry converges absolutely and uniformly on compact subsets in the half-plane $\Re e(s) > 1$ and admits a meromorphic continuation to the whole complex plane. (See [25] for these facts as well as the ones used in the proof below.)

Theorem 7.4. *Let $\Gamma < G$ be a non-uniform lattice containing $-I$ with cusps $\mathfrak{a}, \mathfrak{b}$. There are constants $1 > \rho_1 > \cdots > \rho_k > 1/2$ and $c_1, \ldots, c_k$ such that*

$$\sum_{\substack{c \in \mathcal{N}^*_{\mathfrak{a}\mathfrak{b}} \\ 0<c<T}} \varphi_{\mathfrak{a}\mathfrak{b}}(c) = \frac{c_\Gamma}{2} T^2 + c_1 T^{2\rho_1} + \cdots + c_k T^{2\rho_k} + O(T^{4/3}) \quad \text{as } T \to \infty.$$

Proof. The asymptotic is a special case of a more general theorem of Good (see [17, Theorem 4, p. 116]) on counting functions over double cosets of Fuchsian groups. We include a much simpler and direct proof for the convenience of the reader that recovers the weaker error growth rate $O(T^{4/3+\varepsilon})$.

We will need the following facts pertaining to the meromorphic continuation of the entries of the scattering matrix, which we will also denote by $\phi_{\mathfrak{a}\mathfrak{b}}(s)$. We record that $\phi_{\mathfrak{a}\mathfrak{b}}(s)$ is holomorphic for $\Re e(s) \geq 1/2$ except for at most finitely many poles $\rho_0, \rho_1, \ldots, \rho_k$, all simple, and all lying on the real segment $(1/2, 1]$, with $\rho_0 = 1$ whose residue is

$$\operatorname*{Res}_{s=1} \phi_{\mathfrak{a}\mathfrak{b}}(s) = \frac{1}{V}.$$

We will also rely on the following standard complex analysis computation; uniformly for $y, \sigma > 0$ and $X \geq 1$ we have

$$\frac{1}{2\pi i}\int_{\sigma-iX}^{\sigma+iX} y^s \frac{ds}{s} = \begin{cases} 1 + O\big(\frac{y^\sigma}{X \log y}\big) & \text{if } y > 1, \\ O\big(\frac{y^\sigma}{X|\log y|}\big) & \text{if } y < 1. \end{cases}$$

We may assume that $T \notin \mathcal{N}_{\mathfrak{a}\mathfrak{b}}$ and let $\sigma > 1$. Then

$$\frac{\pi^{-1/2}}{2\pi i}\int_{\sigma-iX}^{\sigma+iX} \frac{\Gamma(s)}{\Gamma(s-1/2)} \phi_{\mathfrak{a}\mathfrak{b}}(s) \frac{T^{2s}}{s}\, ds = \sum_{\substack{c \in \mathcal{N}_{\mathfrak{a}\mathfrak{b}} \\ 0<c<T}} \varphi_{\mathfrak{a}\mathfrak{b}}(c) + O\left(\frac{T^{2\sigma}}{X}\right).$$

Applying Stirling's estimate for the gamma-function and the Phrágmen–Lindelöf principle we see that the integrand on the LHS is of order $O(T^{2\sigma}/|t|^{1/2})$ for all $s \in \mathbf{C}$ with $\Re e(s) \in [1/2, \sigma]$ and $|t| \geq 1$. To compute the integral on the LHS, we apply Cauchy's residue theorem to the rectangle with sides $[\sigma - iX, \sigma + iX]$, $[\sigma + iX, 1/2 + iX]$, $[1/2 + iX, 1/2 - iX]$, $[1/2 - iX, \sigma - iX]$, and find that the LHS is

$$\sum_{i=0}^{k} \frac{\Gamma(\rho_i)\pi^{-1/2}}{\rho_i \Gamma(\rho_i - 1/2)} \operatorname*{Res}_{s=\rho_i} \phi_{\mathfrak{a}\mathfrak{b}}(s) T^{2\rho_i} + O\left(TX^{1/2} + X^{-1/2}\frac{T^{2\sigma}}{\log T}\right).$$

Combining these results, we have

$$\sum_{\substack{c \in \mathcal{N}_{\mathfrak{a}\mathfrak{b}} \\ 0<c<T}} \varphi_{\mathfrak{a}\mathfrak{b}}(c) = \frac{T^2}{\pi V} + c_1 T^{2\rho_1} + \cdots + c_k T^{2\rho_k} + O\left(\frac{T^{2\sigma}}{X} + TX^{1/2} + X^{-1/2}\frac{T^{2\sigma}}{\log T}\right),$$

for an appropriate choice of $c_1, \ldots, c_k$. The error term is optimized by choosing $X = T^{\sigma/3}$ and σ can be chosen as small as $\sigma = 1 + \varepsilon$ for any $\varepsilon > 0$. ■

We will later require asymptotics for the following modification of this average. For $t > 0$, let

$$\Phi_{\mathfrak{a}\mathfrak{b}}(t) := t \sum_{\substack{c \in \mathcal{N}^*_{\mathfrak{a}\mathfrak{b}} \\ c \geq t}} \frac{\varphi_{\mathfrak{a}\mathfrak{b}}(c)}{c^3}. \tag{7.1}$$

Corollary 7.5. *Let* $\delta = \min\{2(1-\rho_1), 2/3\} \in (0, 2/3]$ *and let* c_Γ *be as in* (4.3). *Then*

$$\Phi_{\mathfrak{a}\mathfrak{b}}(t) = \begin{cases} c_\Gamma + O(t^{-\delta}) & \text{as } t \to \infty, \\ O(t) & \text{as } t \to 0. \end{cases}$$

Proof. The first statement follows by Abel summation on Theorem 7.4; indeed, for any $T \gg t$,

$$\sum_{\substack{c \in \mathcal{N}_{\mathfrak{a}\mathfrak{b}} \\ t<c\leq T}} \frac{\varphi_{\mathfrak{a}\mathfrak{b}}(c)}{c^3} = \sum_{0<c\leq T} \varphi_{\mathfrak{a}\mathfrak{b}}(c) T^{-3} - \sum_{0<c\leq t} \varphi_{\mathfrak{a}\mathfrak{b}}(c) t^{-3} + 3\int_t^T \sum_{0<c\leq u} \varphi_{\mathfrak{a}\mathfrak{b}}(c) u^{-4}\, du.$$

Letting $T \to \infty$ yields

$$\sum_{\substack{c \in \mathcal{N}_{\mathfrak{a}\mathfrak{b}} \\ t\leq c}} \frac{\varphi_{\mathfrak{a}\mathfrak{b}}(c)}{c^3} = c_\Gamma t^{-1} + O(t^{-\delta}) \quad \text{as } t \to \infty.$$

For the second statement, we use Theorem 7.4 and positivity to see $\varphi_{\mathfrak{a}\mathfrak{b}}(c) \ll c^{2-\varepsilon}$ for some $\varepsilon > 0$. Hence the sum defining $\Phi_{\mathfrak{a}\mathfrak{b}}(t)$ is absolutely convergent. ■

8. Mean value formulas for pairs

A formula for the second moment of the Siegel–Veech transform is derived in [14] when the lattice $\Gamma = H_q$ is a Hecke triangle group. In this setting there is only one cusp class, represented by the cusp at ∞, and we have $\Lambda := \Lambda_\infty = \Gamma \mathbf{e}_1$. Then for $f \in SC(\mathbf{R}^2 \times \mathbf{R}^2)$ we have

$$\int_{G/H_q} \Theta_{\Lambda,\Lambda;f}(g)\, d\mu(g) = c_{H_q} \sum_{c \in \mathcal{N}^*} \varphi(c) \int_G f(g\mathbf{e}_1, cg\mathbf{e}_2)\, d\eta(g) \\ + c_{H_q} \int_{\mathbf{R}^2} (f(\mathbf{x}, -\mathbf{x}) + f(\mathbf{x}, \mathbf{x}))\, d\mathbf{x}$$

where $\mathcal{N} = \mathcal{N}_{\infty\infty}$, $\varphi = \varphi_{\infty\infty}$, and η is the Haar measure on G normalized so that $\eta_*(G/H_q) = c_{H_q}$ [14, Theorem 1.2].

Opening up the measure $d\eta = da\, db\, ds$ under the coordinates in Section 2.1, the inner integral becomes

$$\iiint_{\substack{a,b,s \in \mathbf{R} \\ a \neq 0}} f\Big(\big(\begin{smallmatrix} a \\ as \end{smallmatrix}\big), c\big(\begin{smallmatrix} b \\ bs + a^{-1} \end{smallmatrix}\big)\Big)\, da\, db\, ds \\ = \frac{1}{|c|} \iiint_{\substack{u,v,x \in \mathbf{R} \\ u \neq 0}} f\Big(\big(\begin{smallmatrix} u \\ v \end{smallmatrix}\big), x\big(\begin{smallmatrix} u \\ v \end{smallmatrix}\big) + c\big(\begin{smallmatrix} 0 \\ u^{-1} \end{smallmatrix}\big)\Big)\, dx\, dv\, du \\ = \frac{1}{|c|} \int_{\mathbf{R}} \int_{\mathbf{R}^2} f(\mathbf{x}, t\mathbf{x} + c\mathbf{x}^*)\, d\mathbf{x}\, dt.$$

under the change of coordinates $a = u$, $s = vu^{-1}$, $ux = cb$ which has Jacobian $1/|c|$, and for $\mathbf{x}^*$ chosen such that $\det(\mathbf{x} \,|\, \mathbf{x}^*) = 1$.

We can extend this identity from Hecke triangle groups to all non-uniform lattices and pairs of not necessarily homothetic discrete lattice orbits. The mean value formula in Theorem 1.8 is a consequence of the following result.

Theorem 8.1. *Given two scaled discrete Γ-orbits $\Lambda_{\mathfrak{a}}, \Lambda_{\mathfrak{b}}$, $f \in SC(\mathbf{R}^2 \times \mathbf{R}^2)$, we have*

$$\int_{G/\Gamma} \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(g)\, d\mu(g) = c_\Gamma \sum_{c \in \mathcal{N}^*_{\mathfrak{ab}}} \varphi_{\mathfrak{ab}}(c) \int_G f(g\mathbf{e}_1, cg\mathbf{e}_2)\, d\eta(g) \\ + \delta_{\mathfrak{ab}} c_\Gamma \int_{\mathbf{R}^2} (f(\mathbf{x}, -\mathbf{x}) + f(\mathbf{x}, \mathbf{x}))\, d\mathbf{x},$$

where $\delta_{\mathfrak{ab}} = 1$ if and only if $\Lambda_{\mathfrak{a}}$ and $\Lambda_{\mathfrak{b}}$ are homothetic.

Remark 8.2. In order to obtain an extension from $SC(\mathbf{R}^2 \times \mathbf{R}^2)$ to a larger class of functions, one can follow the techniques of Theorem 6.1, where the co-area formula is replaced by the area formula. This technique does not unconditionally yield the set of all integrable Borel functions.

Proof of Theorem 8.1. Following the discussion in [14, Section 3.1], we decompose $\mathbf{R}^2 \times \mathbf{R}^2$ into G-invariant subsets on which μ must be supported; we write

$$\int_{G/\Gamma} \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(g)\,d\mu(g) = a \cdot \delta_{\mathbf{0}}(f) + b_\infty \int_{\mathbf{R}^2} f(\mathbf{0},\mathbf{x})\,d\mathbf{x} + \sum_{t\in\mathbf{R}} b_t \int_{\mathbf{R}^2} f(\mathbf{x},t\mathbf{x})\,d\mathbf{x} + \sum_{t\neq 0} c_t \int_G f(g\mathbf{e}_1, tg\mathbf{e}_2)\,d\eta(g). \tag{8.1}$$

In the following we use the fact that we can choose any $f \in SC(\mathbf{R}^2 \times \mathbf{R}^2)$. In each case we will define $f = \mathbf{1}_S$ for a specific closed set S, and use it to determine the coefficients for each term of (8.1).

Set $S = \{(\mathbf{0},\mathbf{0})\}$. Then since no lattice orbit passes through the origin, we have $a = 0$.

Set $S = \{\mathbf{0}\} \times \overline{B_R}$. Since no discrete lattice orbit contains the origin, $\Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f} = 0$, and we have $b_\infty = 0$. The same argument yields $b_0 = 0$.

Set $S = \{(\mathbf{x},t\mathbf{x}) \in \overline{B_R} \times \overline{B_R}\}$. By Lemma 7.1, if $\delta_{\mathfrak{ab}} = 0$, then $S \cap (\Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}}) = \emptyset$. Then (8.1) becomes

$$0 = b_t \int_{\mathbf{R}^2} \mathbf{1}_S(\mathbf{x},t\mathbf{x})\,d\mathbf{x} = b_t |B_{R/t}|$$

and hence $b_t = 0$ for all $t \in \mathbf{R}^*$ when $\delta_{\mathfrak{ab}} = 0$.

Suppose now $\delta_{\mathfrak{ab}} = 1$. Let $\mathbf{x} \in \Lambda_{\mathfrak{a}}$, $t\mathbf{x} \in \Lambda_{\mathfrak{a}}$. Choose $g \in \Gamma\sigma_{\mathfrak{a}}$ and $h \in \Gamma\sigma_{\mathfrak{a}}$ such that the first column vectors of g and h are respectively $\mathbf{x}$ and $t\mathbf{x}$. Then

$$h^{-1}g = \begin{pmatrix} 1/t & * \\ 0 & t \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}.$$

By Lemma 2.1, we conclude that $t \in \{\pm 1\}$. Set $f_+ = \mathbf{1}_S$ when $S = \{(\mathbf{x},\mathbf{x}) \in \overline{B_R} \times \overline{B_R}\}$ and $f_- = \mathbf{1}_S$ when $S = \{(\mathbf{x},-\mathbf{x}) \in \overline{B_R} \times \overline{B_R}\}$. Hence $\Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f_\pm} \neq 0$ and (8.1) yields

$$\int_{G/\Gamma} \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f_\pm}(g)\,d\mu(g) = b_{\pm 1} \int_{\mathbf{R}^2} f_\pm(\mathbf{x},\pm\mathbf{x})\,d\mathbf{x}.$$

On the other hand, by Theorem 6.1, we have

$$\int_{G/\Gamma} \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f_\pm}(g)\,d\mu(g) = c_\Gamma \int_{\mathbf{R}^2} f_\pm(\mathbf{x},\pm\mathbf{x})\,d\mathbf{x}.$$

Since this is true for all R we conclude that when $\delta_{\mathfrak{ab}} = 1$ we have $b_t = 0$ when $t \neq \pm 1$ and $b_{\pm 1} = c_\Gamma$.

Set $S = \{(\mathbf{x},\mathbf{y}) : \mathbf{x},\mathbf{y} \in \overline{B_R},\ \det(\mathbf{x} \mid \mathbf{y}) = c\}$. Then $\Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(g) = |\{(\mathbf{x},\mathbf{y}) \in (\Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}}) \cap g^{-1}\overline{B_R} : \det(\mathbf{xy}) = c\}| \neq 0$ if and only if $c \in \mathcal{N}^*_{\mathfrak{ab}}$. Now suppose $c \in \mathcal{N}^*_{\mathfrak{ab}}$. By Lemma 7.2, there exists $0 < a \leq |c|$ such that $\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right) \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$, so that by definition $\varphi_{\mathfrak{ab}}(c) \neq 0$.

We want to show that (cf. [14, Lemma 3.6]) for all R, the support of $\Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}$ reduces to $\varphi_{\mathfrak{a}\mathfrak{b}}(c)$ total Γ-orbits. More precisely, we show that

$$D_c^{\mathfrak{a}\mathfrak{b}} := \{(\mathbf{x}_1, \mathbf{x}_2) \in \Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}} : \det(\mathbf{x}_1|\mathbf{x}_2) = c\} = \bigsqcup_{\substack{\binom{a}{c} \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}\mathbf{e}_1 \\ 0<a\leq|c|}} \Gamma\sigma_{\mathfrak{a}} \begin{pmatrix} 1 & a \\ 0 & c \end{pmatrix}.$$

Indeed, by the proof of Lemma 7.2, there is some $h \in \Gamma\sigma_{\mathfrak{a}}$ so that $h^{-1}(\mathbf{x}_1|\mathbf{x}_2) = \left(\begin{smallmatrix} 1 & a \\ 0 & c \end{smallmatrix}\right)$ for some $0 < a \leq |c|$ and $\binom{a}{c} \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}\mathbf{e}_1$. Moreover, each of these orbits is distinct. Namely if $|c| \geq a_2 > a_1 > 0$ with

$$\Gamma\sigma_{\mathfrak{a}} \begin{pmatrix} 1 & a_1 \\ 0 & c \end{pmatrix} = \Gamma\sigma_{\mathfrak{a}} \begin{pmatrix} 1 & a_2 \\ 0 & c \end{pmatrix},$$

then since $\sigma_{\mathfrak{a}}$ is a scaling transformation, we must have $a_2 - a_1 \in c\mathbf{Z}$, which is impossible as $0 < a_2 - a_1 < c$.

Now that we know $D_c^{\mathfrak{a}\mathfrak{b}}$ is decomposed into $\varphi_{\mathfrak{a}\mathfrak{b}}(|c|)$ distinct orbits, we want to show we have the same contribution for each orbit. Namely (cf. [14, Lemma 3.7]) if we fix $\binom{a}{c} \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}\mathbf{e}_1$, with $0 < a \leq |c|$, then by our normalization of the Haar measure, we have

$$\begin{aligned} \int_{G/\Gamma} \sum_{\gamma\in\Gamma} f\left(g\gamma\sigma_{\mathfrak{a}} \begin{pmatrix} 1 & a \\ 0 & c \end{pmatrix}\right) d\mu(g) &= c_\Gamma \int_G f\left(g\sigma_{\mathfrak{a}} \begin{pmatrix} 1 & a \\ 0 & c \end{pmatrix}\right) d\eta(g) \\ &= c_\Gamma \int_G f\left(g \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}\right) d\eta(g) \end{aligned}$$

by multiplying g on the right by $\left(\begin{smallmatrix} 1 & -a/c \\ 0 & 1 \end{smallmatrix}\right)\sigma_{\mathfrak{a}}^{-1}$ and using the fact that η is invariant under the action of G. Hence we get the same contribution from each of the $\varphi_{\mathfrak{a}\mathfrak{b}}(|c|)$ orbits. Since this is true for all R, we obtain the desired coefficients, thus concluding the proof of the first equality. ∎

Let C be the cone of 2×2 matrices A such that $\lambda A \in G/\Gamma$ for some $\lambda \geq 1$, and let dA be the Euclidean volume element. There exists some constant c_0 such that

$$\int_C f(A)\, dA = c_0 \int_0^1 \nu \int_{G/\Gamma} f(\nu^{1/2} g)\, d\mu(g)\, d\nu.$$

Inspired by Schmidt [37], we introduce the equivalent[4] cone measure

$$\int_C f(A)\, dm(A) = \int_0^1 \int_{G/\Gamma} f(\nu^{1/2} g)\, d\mu(g)\, d\nu.$$

[4]That is, the measures $dm(A)$ and dA have the same null measure sets.

Let $\Lambda_{\mathfrak{a}}$, $\Lambda_{\mathfrak{b}}$ be scaled discrete Γ-orbits. With respect to the cone measure dm, we have $m(C) = 1$,

$$\int_C \Theta_{\Lambda_{\mathfrak{a}};f}(A)\det(A)\,dm(A) = \int_0^1 \nu \int_{G/\Gamma} \Theta_{\Lambda_{\mathfrak{a}};f}(\nu^{1/2}g)\,d\mu(g)\ d\nu = c_\Gamma \int_{\mathbf{R}^2} f(\mathbf{x})\,d\mathbf{x} \tag{8.2}$$

and

$$\int_C \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(A)\det(A)^2\,dm(A) = \int_0^1 \nu^2 \int_{G/\Gamma} \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(\nu^{1/2}g)\,d\mu(g)\ d\nu.$$

The second mean value formula given by Theorem 1.8′ is a consequence of the following theorem.

Theorem 8.3. *Let $\Gamma < G$ be a non-uniform lattice containing $-I$, and let $\Lambda_{\mathfrak{a}}$, $\Lambda_{\mathfrak{b}}$ be two scaled discrete Γ-orbits. For each $f \in SC(\mathbf{R}^2 \times \mathbf{R}^2)$, we have*

$$\begin{aligned}\int_C \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(A)\det(A)^2\,dm(A) &= c_\Gamma \iint_{\mathbf{R}^2\times\mathbf{R}^2} \Phi_{\mathfrak{a}\mathfrak{b}}(|\mathbf{x}\wedge\mathbf{y}|)f(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}\\ &\quad + \frac{\delta_{\mathfrak{a}\mathfrak{b}}c_\Gamma}{2}\int_{\mathbf{R}^2}(f(\mathbf{x},\mathbf{x}) + f(\mathbf{x},-\mathbf{x}))\,d\mathbf{x},\end{aligned} \tag{8.3}$$

where $\Phi_{\mathfrak{a}\mathfrak{b}}$ is the function given by (7.1)*.*

Proof. We will obtain this formula from Theorem 8.1 by a change of coordinates. Define $f_\nu(\mathbf{x},\mathbf{y}) = f(\nu^{1/2}\mathbf{x},\nu^{1/2}\mathbf{y})$, and apply Theorem 8.1 to $\Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f_\nu}$. In the linearly dependent subsets, we have via substitution

$$\begin{aligned}&\delta_{\mathfrak{a}\mathfrak{b}}c_\Gamma \int_0^1 \nu^2 \int_{\mathbf{R}^2}(f_\nu(\mathbf{x},\mathbf{x}) + f_\nu(\mathbf{x},-\mathbf{x}))\,d\mathbf{x}\ d\nu\\ &\qquad = \delta_{\mathfrak{a}\mathfrak{b}}c_\Gamma \int_0^1 \nu\ d\nu \int_{\mathbf{R}^2}(f(\mathbf{x},\mathbf{x}) + f(\mathbf{x},-\mathbf{x}))\,d\mathbf{x}.\end{aligned}$$

It remains to show that

$$\begin{aligned}&c_\Gamma \int_0^1 \nu^2 \sum_{c\in\mathcal{N}^*_{\mathfrak{a}\mathfrak{b}}} \varphi_{\mathfrak{a}\mathfrak{b}}(c)\int_G f(\nu^{1/2}gJ_c)\,d\eta(g)\,d\nu\\ &\qquad = c_\Gamma \iint \Phi_{\mathfrak{a}\mathfrak{b}}(|\mathbf{x}\wedge\mathbf{y}|)f(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y},\end{aligned} \tag{8.4}$$

where we set $J_c = \left(\begin{smallmatrix}1&0\\0&c\end{smallmatrix}\right)$. We write $t = \nu c$ and note that

(1) $\mathcal{N}_{\mathfrak{a}\mathfrak{b}}$ and $\varphi_{\mathfrak{a}\mathfrak{b}}(c)$ are symmetric about 0 if $-I \in \Gamma$;

(2) $\mathcal{N}_{\mathfrak{a}\mathfrak{b}}$ is unbounded, so the parameter t ranges over $\mathbf{R} \setminus \{0\}$.

Since $\Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}$ is integrable, Fubini applies and we may exchange the order of integration and summation in what follows. The LHS of (8.4) becomes

$$c_\Gamma \sum_{c\in\mathcal{N}^*_{\mathfrak{ab}}} \varphi_{\mathfrak{ab}}(c) \int_0^1 \nu^2 \int_G f(\nu^{1/2} g J_c)\, d\eta(g)\, d\nu$$

(by a simple change of coordinates $g = g'\mathrm{diag}(\nu^{-1/2}, \nu^{1/2})$ and the G-invariance of η)

$$= c_\Gamma \int_G \sum_{c\in\mathcal{N}^*_{\mathfrak{ab}}} \int_0^1 \nu^2 \varphi_{\mathfrak{ab}}(c) f\big(g\big(\begin{smallmatrix}1 & 0\\ 0 & c\nu\end{smallmatrix}\big)\big)\, d\nu\, d\eta(g)$$

$$\overset{(1)}{=} c_\Gamma \int_G \sum_{c\in(\mathcal{N}^*_{\mathfrak{ab}})_{>0}} \left[\int_0^c t^2 \frac{\varphi_{\mathfrak{ab}}(c)}{c^3} f\big(g\big(\begin{smallmatrix}1 & 0\\ 0 & t\end{smallmatrix}\big)\big) + \int_{-c}^0 t^2 \frac{\varphi_{\mathfrak{ab}}(c)}{c^3} f\big(g\big(\begin{smallmatrix}1 & 0\\ 0 & t\end{smallmatrix}\big)\big)\right] dt\ d\eta(g)$$

$$\overset{(2)}{=} c_\Gamma \int_G \left[\int_0^\infty t^2 f\big(g\big(\begin{smallmatrix}1 & 0\\ 0 & t\end{smallmatrix}\big)\big) \sum_{\substack{c\in\mathcal{N}^*_{\mathfrak{ab}}\\ c\geq t}} \frac{\varphi_{\mathfrak{ab}}(c)}{c^3}\, dt + \int_{-\infty}^0 t^2 f\big(g\big(\begin{smallmatrix}1 & 0\\ 0 & t\end{smallmatrix}\big)\big) \sum_{\substack{c\in\mathcal{N}^*_{\mathfrak{ab}}\\ -c\leq t}} \frac{\varphi_{\mathfrak{ab}}(c)}{c^3}\right] d\eta(g)$$

$$\overset{(1)}{=} c_\Gamma \int_{\mathbf{R}} |t|\Phi_{\mathfrak{ab}}(|t|) \int_G f\big(g\big(\begin{smallmatrix}1 & 0\\ 0 & t\end{smallmatrix}\big)\big)\, d\eta(g)\, dt. \qquad \blacksquare$$

9. Geometric estimates

We may rewrite the mean value formula as

$$\begin{aligned}\int_C \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(A) \det(A)^2\, dm(A) &= c_\Gamma^2 \iint_{\mathbf{R}^2\times\mathbf{R}^2} f(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} + \frac{\delta_{\mathfrak{ab}} c_\Gamma}{2} \int_{\mathbf{R}^2} (f(\mathbf{x},\mathbf{x}) + f(\mathbf{x},-\mathbf{x}))\, d\mathbf{x}\\ &\quad + c_\Gamma \iint_{\mathbf{R}^2\times\mathbf{R}^2} (\Phi_{\mathfrak{ab}}(|\mathbf{x}\wedge\mathbf{y}|) - c_\Gamma) f(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}. \qquad (9.1)\end{aligned}$$

In this section, we use integration by parts to compute estimates for the last double integral,

$$\iint_{\mathbf{R}^2\times\mathbf{R}^2} (\Phi_{\mathfrak{ab}}(|\mathbf{x}\wedge\mathbf{y}|) - c_\Gamma) f(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}. \qquad (9.2)$$

Namely, we will obtain an estimate in (9.4), and the results from this section to be used later are given in Examples 9.1–9.3.

For $t, u > 0$ consider the rectangle

$$P_{t,u}(\mathbf{x}) = \{\mathbf{y}\in\mathbf{R}^2 : |\mathbf{x}^\perp\wedge\mathbf{y}| \leq t,\ |\mathbf{x}\wedge\mathbf{y}| \leq u\}$$

and set $P(\mathbf{x}) = P_{1,1}(\mathbf{x})$. Then $tP(\mathbf{x}) = P_{t,t}(\mathbf{x})$. If we write $\mathbf{x} = \|\mathbf{x}\| k_\theta \mathbf{e}_1$ and $\mathbf{x}^\perp = k_{\pi/2}\|\mathbf{x}\| k_\theta \mathbf{e}_1 = \|\mathbf{x}\| k_\theta \mathbf{e}_2$, we have

$$P_{t,u}(\mathbf{x}) = k_\theta P_{t/\|\mathbf{x}\|, u/\|\mathbf{x}\|}(\mathbf{e}_1).$$

In particular, $P_{t,u}(\mathbf{x})$ is a rotated rectangle, centered at the origin, with sides of length $2t/\|\mathbf{x}\|$ and $2u/\|\mathbf{x}\|$. We consider the inner integral of (9.2). For t sufficiently large, writing $\mathbf{x} = \|\mathbf{x}\|k_\theta \mathbf{e}_1$, we have

$$\begin{aligned}\int_{\mathbf{R}^2} (\Phi_{\mathfrak{ab}}(|\mathbf{x}\wedge\mathbf{y}|) - c_\Gamma \delta_{\mathfrak{ab}}) f(\mathbf{x},\mathbf{y})\, d\mathbf{y} &= \int_{tP(\mathbf{x})} (\Phi_{\mathfrak{ab}}(|\mathbf{e}_1 \wedge \|\mathbf{x}\| k_{-\theta}\mathbf{y}|) - c_\Gamma) f(\mathbf{x},\mathbf{y})\, d\mathbf{y} \\ &= \frac{1}{\|\mathbf{x}\|^2}\int_{tP(\mathbf{e}_1)} (\Phi_{\mathfrak{ab}}(|\mathbf{e}_1\wedge\mathbf{y}|) - c_\Gamma) f\left(\mathbf{x}, \tfrac{1}{\|\mathbf{x}\|}k_\theta \mathbf{y}\right) d\mathbf{y} \\ &= \frac{1}{\|\mathbf{x}\|^2}\int_0^t (\Phi_{\mathfrak{ab}}(y_2) - c_\Gamma)\left(\int_0^t \sum f\left(\mathbf{x}, \tfrac{1}{\|\mathbf{x}\|}k_\theta \left(\begin{smallmatrix}\pm y_1\\ \pm y_2\end{smallmatrix}\right)\right) dy_1\right) dy_2,\end{aligned}$$

where the inner sum contains a summand for each of the four $\left(\begin{smallmatrix}\pm y_1\\ \pm y_2\end{smallmatrix}\right)$. We call the inner integral $g_t(y_2)$ and compute that its primitive is

$$\begin{aligned}G_t(u) &= \int_0^u \int_0^t \sum f\left(\mathbf{x}, \tfrac{1}{\|\mathbf{x}\|}k_\theta\left(\begin{smallmatrix}\pm r\\ \pm s\end{smallmatrix}\right)\right) dr\, ds \\ &= \int_{P_{t,u}(\mathbf{e}_1)} f\left(\mathbf{x}, \tfrac{1}{\|\mathbf{x}\|}k_\theta \mathbf{y}\right) d\mathbf{y} \\ &= \|\mathbf{x}\|^2 \int_{P_{t,u}(\mathbf{x})} f(\mathbf{x},\mathbf{y})\, d\mathbf{y}.\end{aligned}$$

Recall that by Corollary 7.5, there is some $t_0 > 0$ and constant M such that

$$\Phi_{\mathfrak{ab}}(t) \le \tilde{\Phi}_{\mathfrak{ab}}(t) := \begin{cases} c_\Gamma + Mt^{-\delta}, & t \ge t_0, \\ Mt, & 0 \le t < t_0. \end{cases} \tag{9.3}$$

Integration by parts yields

$$\begin{aligned}&\int_{\mathbf{R}^2} (\Phi_{\mathfrak{ab}}(|\mathbf{x}\wedge\mathbf{y}|) - c_\Gamma) f(\mathbf{x},\mathbf{y})\, d\mathbf{y} \\ &\quad = \frac{1}{\|\mathbf{x}\|^2}\int_0^t (\Phi_{\mathfrak{ab}}(y_2) - c_\Gamma)\left(\int_0^t \sum f\left(\mathbf{x}, \tfrac{1}{\|\mathbf{x}\|}k_\theta\left(\begin{smallmatrix}\pm y_1\\ \pm y_2\end{smallmatrix}\right)\right) dy_1\right) dy_2 \\ &\quad \ll (\tilde{\Phi}_{\mathfrak{ab}}(t) - c_\Gamma)\int_{tP(\mathbf{x})} f(\mathbf{x},\mathbf{y})\, d\mathbf{y} - \int_0^t \tilde{\Phi}'_{\mathfrak{ab}}(u)\int_{P_{t,u}(\mathbf{x})} f(\mathbf{x},\mathbf{y})\, d\mathbf{y}\, du \\ &\quad \ll t^{-\delta}\int_{tP(\mathbf{x})} f(\mathbf{x},\mathbf{y})\, d\mathbf{y} + \delta\int_{t_0}^\infty \left(\int_{P_{t,u}(\mathbf{x})} f(\mathbf{x},\mathbf{y})\, d\mathbf{y}\right) u^{-1-\delta}\, du.\end{aligned} \tag{9.4}$$

Example 9.1. If $f(\mathbf{x},\mathbf{y}) = 1_{B_R}(\mathbf{x})1_{B_R}(\mathbf{y})$, letting $t\to\infty$ in (9.4), we have

$$(9.2) \ll \int_{t_0}^\infty \int_{B_R} |P_{\infty,u}(\mathbf{x}) \cap B_R|\, d\mathbf{x}\, u^{-1-\delta}\, du$$

with $|P_{\infty,u}(\mathbf{x})\cap B_R| = |B_R|$ if $u \ge R\|\mathbf{x}\|$, and $|P_{\infty,u}(\mathbf{x})\cap B_R| \ll Ru/\|\mathbf{x}\|$ otherwise. Hence

$$(9.2) \ll \int_{B_R}\left(\int_0^{R\|\mathbf{x}\|} \frac{uR}{\|\mathbf{x}\|} u^{-1-\delta}\, du + \int_{R\|\mathbf{x}\|}^\infty R^2 u^{-1-\delta}\, du\right) d\mathbf{x} \ll R^{2(2-\delta)}.$$

Example 9.2. Let $f(\mathbf{x},\mathbf{y}) = 1_{B_R}(\mathbf{x})1_{B^*_\varepsilon(\mathbf{x})}(\mathbf{y})$. Once again, $|P_{\infty,u}(\mathbf{x}) \cap B^*_\varepsilon(\mathbf{x})| = |B_\varepsilon|$ if $u \geq \varepsilon\|\mathbf{x}\|$ and $|P_{\infty,u}(\mathbf{x}) \cap B^*_\varepsilon(\mathbf{x})| \ll \varepsilon u/\|\mathbf{x}\|$ otherwise. Hence we have the estimate

$$(9.2) \ll \int_{B_R}\left(\int_0^{\varepsilon\|\mathbf{x}\|} \frac{\varepsilon}{\|\mathbf{x}\|}u^{-\delta}\,du + \int_{\varepsilon\|\mathbf{x}\|}^{\infty} \varepsilon^2 u^{-\delta}\,\frac{du}{u}\right)d\mathbf{x} \ll (\varepsilon R)^{2-\delta}.$$

Proof of Theorem 1.7. Apply (9.1) to the characteristic function f supported on $\{\mathbf{x} \in \mathbf{R}^2, \mathbf{y} \in B^*_{s/\sqrt{c}}(\mathbf{x})\}$ and use the estimate from Example 9.2 to bound the remaining term. ■

Example 9.3. For $D, s > 0$, let $f(\mathbf{x},\mathbf{y}) = 1_{B_R}(\mathbf{x})1_{\mathcal{D}_{D,s}(\mathbf{x})}(\mathbf{y})$, where

$$\mathcal{D}_{D,s}(\mathbf{x}) = \{\mathbf{y} \in \mathbf{R}^2 : |\mathbf{y}| \leq s|\mathbf{x}| \text{ and } |\mathbf{x}\wedge\mathbf{y}| \leq D\}.$$

Then $|P_{\infty,u}(\mathbf{x}) \cap \mathcal{D}_{D,s}(\mathbf{x})| \leq \frac{u}{\|\mathbf{x}\|}\cdot s\|\mathbf{x}\| = su$ if $u \leq D$, and $|P_{\infty,u}(\mathbf{x}) \cap \mathcal{D}_{D,s}(\mathbf{x})| \leq sD$ if $u \geq D$. Therefore

$$(9.2) \ll s\int_{B_R}\left(\int_0^D u^{-\delta}du + D\int_D^\infty u^{-1-\delta}\,du\right)d\mathbf{x} \ll sDR^2.$$

10. Effective counting

10.1. Second moment estimate

We will derive Theorem 1.1 (see Theorem 10.3 for a full statement) by a combination of Borel–Cantelli and interpolation from the following "on average" power-saving. Our strategy is inspired by [37].

Proposition 10.1. *Let Λ be a scaled discrete lattice orbit, and let B be a Lebesgue measurable set in the plane with finite volume $|B| > C_\Lambda$. There is a constant M_Λ such that*

$$\int_C \left(\det(A)\Theta_{\Lambda;\mathbf{1}_B}(A) - c_\Gamma|B|\right)^2 dm(A) \leq M_\Lambda|B|^{2-\delta},$$

where δ is given by (4.4).

To prove Proposition 10.1, we use the following result of Schmidt for $n = 2$.

Lemma 10.2 ([37, Theorem 3]). *Let S be a Lebesgue mesurable set in $\mathbf{R}^2$ with characteristic function $\mathbf{1}_S$ and volume $|S|$. Then*

$$\iint \chi(|\mathbf{x}\wedge\mathbf{y}|)\mathbf{1}_S(\mathbf{x})\mathbf{1}_S(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} \leq 8|S|\int_0^\infty \chi(t)\,dt$$

for every non-negative, non-increasing function $\chi(t)$ defined for $t \geq 0$ whose integral $\int_0^\infty \chi(t)\,dt$ converges.

Proof of Proposition 10.1. Let $\alpha(|B|)$ be a function increasing with $|B|$ that we will choose later. Choose $|B|$ large enough so that for $t_0 < \alpha(|B|)$ we can use Corollary 7.5

(cf. (9.3)) to bound

$$|\Phi(t) - c_\Gamma| \leq \chi(t) + \begin{cases} 0, & t \leq \alpha(|B|), \\ M\alpha(|B|)^{-\delta}, & t > \alpha(|B|), \end{cases}$$

where for

$$\chi(t) = \begin{cases} Mt_0, & 0 \leq t < t_0, \\ Mt^{-\delta}, & t_0 \leq t < \alpha(|B|), \\ 0, & t \geq \alpha(|B|). \end{cases}$$

Notice $\chi(t)$ satisfies the assumptions of Lemma 10.2 with

$$\int_0^\infty \chi(t)\, dt = M\left(t_0^2 - \frac{t_0^{1-\delta}}{1-\delta}\right) + \frac{M}{1-\delta}\alpha(|B|)^{1-\delta}.$$

Combining (8.2), (9.1), and Lemma 10.2 we compute

$$\begin{aligned} \int_C \left(\det(A)\Theta_{\Lambda;\mathbf{1}_B}(A) - c_\Gamma|B|\right)^2 dm(A) &= \int_C \Theta_{\Lambda,\Lambda;\mathbf{1}_B}(A)\det(A)^2\, dm(A) - c_\Gamma^2|B|^2 \\ &\leq c_\Gamma|B| + c_\Gamma \iint_{\mathbf{R}^2\times\mathbf{R}^2} (\Phi_{\mathfrak{ab}}(|\mathbf{x}\wedge\mathbf{y}|) - c_\Gamma)\mathbf{1}_B(\mathbf{x})\mathbf{1}_B(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \\ &\ll |B|(1+\alpha(|B|)^{1-\delta}) + |B|^2\alpha(|B|)^{-\delta}. \end{aligned}$$

Choosing $\alpha(|B|) = |B|$ and noting that the implied constants only depend on Λ completes the proof. ∎

10.2. Effective count via interpolation

We can now derive the main result of this section. The proof follows the same lines as that of [37, Theorem 2] and [29, Theorem 6.1]. We include it for the convenience of the reader.

Theorem 10.3. *Let $\Gamma < G$ be a lattice containing $-I$ with scaled discrete orbit Λ. Let $\mathcal{B}$ be a linearly ordered family of Borel sets of finite volume in $\mathbf{R}^2$. Let ψ be a positive non-increasing function such that $e^{t(2-\delta)}\psi(t)$ is eventually non-decreasing and $\int_1^\infty \psi(t)\, dt < \infty$. Then for almost every linear transformation A,*

$$|A(\Lambda)\cap B|\,|\det(A)| = c_\Gamma|B| + O\left(\frac{|B|^{1-\delta/2}\log^{1/2}(|B|)}{\psi^{1/2}(\log(|B|))}\right)$$

Proof. Without loss of generality, we may assume $\mathcal{B}$ has sets of arbitrarily large volumes. By [37, Lemma 1] we may assume without loss of generality that $\{|B| : B \in \mathcal{B}\} = \mathbf{R}^+$. Thus for all $N \in \mathbf{N}_{\geq 1}$ there exists $B_N \in \mathcal{B}$ such that $|B_N| = N$.

Define

$$S_N(A) = |A\Lambda \cap B_N| - c_\Gamma N|\det(A^{-1})|,$$

and for $0 \le N_1 < N_2$,

$$ {}_{N_1}S_{N_2}(A) = |A\Lambda \cap (B_{N_2} \setminus B_{N_1})| - c_\Gamma(N_2 - N_1)|\det(A^{-1})|. $$

For $T \ge 3$ set

$$ \mathcal{K}_T = \{(N_1, N_2) \in \mathbf{Z}^2 : 0 \le N_1 < N_2 \le 2^T, \\ N_1 = \ell 2^t \text{ and } N_2 = (\ell+1)2^t \text{ for some } \ell, t \in \mathbf{N}_{\ge 0}\}. $$

By Proposition 10.1 with $B = B_{N_2} \setminus B_{N_1}$, notice that each value of $N_2 - N_1 = 2^t$ occurs 2^{T-t} times so we have

$$ \sum_{(N_1,N_2)\in\mathcal{K}_T} \int_C |{}_{N_1}S_{N_2}(A)|^2 |\det(A)|^2 \, dm(A) \le M_\Lambda \sum_{t=0}^{T} 2^{T-t} 2^{t(2-\delta)} \le 8M_\Lambda 2^{(2-\delta)T}, \tag{10.1} $$

where the last inequality follows from the geometric series and $\frac{1}{4} \le 2^{1/3} \le 2^{1-\delta} \le 2$.

Next let $\mathcal{E}_T \subseteq C$ be the set of all $A \in C$ such that

$$ \sum_{(N_1,N_2)\in\mathcal{K}_T} |{}_{N_1}S_{N_2}(A)|^2 |\det(A)|^2 > \frac{2^{(2-\delta)T}}{48\psi((T-1)\log(2))}. \tag{10.2} $$

Applying Chebyshev's inequality and (10.1) implies

$$ \begin{aligned} m(\mathcal{E}_T) &< \frac{48\psi((T-1)\log(2))}{2^{(2-\delta)T}} \int_{\mathcal{E}_T} \sum_{(N_1,N_2)\in\mathcal{K}_T} |{}_{N_1}S_{N_2}(A)|^2 |\det(A)|^2 \, dm(A) \\ &\le 48 \cdot 8M_\Lambda \psi((T-1)\log(2)). \end{aligned} \tag{10.3} $$

By the Borel–Cantelli lemma and integrability of ψ, $m(\mathcal{E}_\infty) = 0$ for $\mathcal{E}_\infty = \limsup \mathcal{E}_T$. Define $C \setminus \mathcal{E}_\infty$ to be the full measure set (with respect to dA) of transformations where the discrepancy is small.

For $N \le 2^T$, we can write $[0, N) \subseteq \bigsqcup_{(N_1,N_2)\in\mathcal{J}} [N_1, N_2)$, where $\mathcal{J} \subset \mathcal{K}_T$ is given by

$$ \mathcal{J} = \{(0 \cdot 2^1, 1 \cdot 2^1)\} \cup \{(2^t, 2 \cdot 2^t)\}_{t=1}^{T-1}. $$

Thus

$$ S_N(A) \le \sum_{(N_1,N_2)\in\mathcal{J}} {}_{N_1}S_{N_2}(A). $$

By (10.2), for any $A \notin \mathcal{E}_T$ and any $N < 2^T$,

$$ \begin{aligned} |S_N(A)|^2 |\det(A)|^2 &\le \Big| \sum_{(N_1,N_2)\in\mathcal{J}} {}_{N_1}S_{N_2}(A) \Big|^2 |\det(A)|^2 \\ &\le |\mathcal{J}| \sum_{(N_1,N_2)\in\mathcal{K}_T} |{}_{N_1}S_{N_2}(A)|^2 |\det(A)|^2 \\ &\le T \frac{2^{(2-\delta)T}}{48\psi((T-1)\log(2))}. \end{aligned} $$

Consider $A \in C \setminus \mathcal{E}_\infty$. Choose $T_A \in \mathbf{N}$ so that for all $T \geq T_A$, $A \notin \mathcal{E}_T$. That is, for all $T \geq T_A$ and for any $N < 2^T$,

$$|S_N(A)|^2 |\det(A)|^2 \leq \frac{T 2^{(2-\delta)T}}{48\psi((T-1)\log(2))}.$$

In order to interpolate from B_N to any B, we will set $N_A = \max\{2^{T_A}, N_0\}$, where the assumption that $e^{t(2-\delta)}\psi(t)$ is eventually non-decreasing ensures there exists some N_0 such that for all $N \geq N_0$,

$$\frac{(N+1)^{1-\delta/2}\sqrt{\log(N+1)}}{2\psi^{1/2}(\log(N+1))} + 1 \leq \frac{N^{1-\delta/2}\sqrt{\log(N)}}{\psi^{1/2}(\log(N))}. \tag{10.4}$$

Whenever $N > N_A$, we have $N > 2^{T_A}$ so we can choose an integer $T \geq T_A$ such that $2^{T-1} \leq N \leq 2^T$. For such N we have

$$|S_N(A)|^2 |\det(A)|^2 \leq \left(\frac{\log(N)}{\log(2)} + 1\right) \frac{4N^{2-\delta}}{48\psi(\log(N))} \leq \frac{N^{2-\delta}\log(N)}{4\psi(\log(N))},$$

where we use the fact that ψ is non-increasing, $2^{(2-\delta)T} = 4 \cdot 2^{2(T-1)} 2^{-\delta T}$, and $\frac{\log(N)}{\log(2)} + 1 \leq 3\log(N)$ for all $N \geq 2$.

We have now shown for all $N > 2^{T_A}$ that

$$|S_N(A)|\,|\det(A)| \leq \frac{N^{1-\delta/2}\sqrt{\log(N)}}{2\psi^{1/2}(\log(N))}.$$

For any $B \in \mathcal{B}$ with $|B| \geq N_A$ there exists an integer $N \geq N_A - 1$ such that $N \geq 2^{T_A}$ and $B_N \subseteq B \subseteq B_{N+1}$. We then interpolate and use (10.4) to obtain

$$\begin{aligned} \big||A\Lambda \cap B| - c_\Gamma |B|\,|\det(A^{-1})|\big| &\leq \max\{|S_N(A)|, |S_{N+1}(A)|\} + \frac{c_\Gamma}{|\det(A)|} \\ &\leq \frac{|B|^{1-\delta/2}\sqrt{\log(|B|)}}{\psi^{1/2}(\log(|B|))|\det(A)|}. \end{aligned} \tag{10.5}$$

Since every $\tilde{A}$ with $0 < |\det(\tilde{A})| \leq 1$ is of the form $\tilde{A} = A\gamma$ for some $A \in C$ and $\gamma \in \Gamma$, and since Γ is enumerable, (10.5) holds for almost every linear transformation $\tilde{A}$ with $|\det(\tilde{A})| \leq 1$. Now consider a linear transformation C with $\det(C) = c$. Applying (10.5) to the family of sets $\{C^{-1}B : B \in \mathcal{B}\}$ yields

$$\big||A\Lambda \cap C^{-1}B| - c_\Gamma |C^{-1}B|\,|\det(A^{-1})|\big| = \big||CA\Lambda \cap B| - c_\Gamma |B|\,|\det((CA)^{-1})|\big|.$$

Since $\det(C)$ can be taken arbitrarily large, we conclude (10.5) holds for almost every linear transformation. ■

10.3. Application to dilated sets

We obtain the discrete lattice orbit version of Theorem 1.1 by applying Theorem 10.3 to dilated sets.

Theorem 10.4. *Let Λ be a scaled discrete Γ-orbit and let $\Omega \subset \mathbf{R}^2$ be a bounded Borel set that contains the origin. Consider its dilates $\Omega_R = R \cdot \Omega$. Then for almost every linear transformation A we have*

$$|A(\Lambda) \cap \Omega_R| \, |\det(A)| = c_\Gamma |\Omega| R^2 + O(R^{2-\delta} \log^{3/2}(R)),$$

where δ is as in (4.4).

Proof. Let $\psi(t) = t^{-2}$, which is integrable, non-increasing for $t > 1$, and such that $e^{t(2-\delta)}\psi(t)$ is eventually non-decreasing. Now apply Theorem 10.3 to the family $\mathscr{B} = \{\Omega_R : R \in \mathbf{R}_{\geq 0}\}$ of dilates. ∎

We recall the following known estimates from [6]. For simplicity, suppose that Γ has trivial residual spectrum, i.e., $\delta = 2/3$. By Theorem 10.4, for *almost every* linear transformation $A \in \mathrm{GL}_2(\mathbf{R})$ we find

$$|\det(A)| \, |A(\Lambda) \cap \Omega_R| = c_\Gamma |\Omega| R^2 + O(R^{4/3} \log^{3/2}(R)),$$

while in [6] it is shown that for *every* $A \in G$ (hence with $\det(A) = 1$), we have

$$|A(\Lambda) \cap \Omega_R| = c_\Gamma |\Omega| R^2 + O(R^\alpha)$$

with

$$\alpha = \begin{cases} 4/3 & \text{when } \Omega \text{ is the unit disk,} \\ 8/5 & \text{when } \Omega \text{ is a star-shaped domain with smooth boundary,} \\ 7/4 + \varepsilon & \text{when } \Omega \text{ is a star-shaped domain with piecewise Lipschitz boundary,} \\ 12/7 + \varepsilon & \text{when } \Omega \text{ is a sector.} \end{cases}$$

11. Upper bounds on pair relationships

In this section, we obtain asymptotic upper bounds for the number of ordered pairs $(\mathbf{x}, \mathbf{y}) \in \Lambda_1 \times \Lambda_2$, where

- Λ_1, Λ_2 are scaled discrete Γ-orbits;
- the first vector $\mathbf{x}$ is restricted to $\mathbf{x} \in B_R \cap \Lambda_1$;
- the second vector $\mathbf{y} \in \Lambda_2$ has a prescribed metric relationship to $\mathbf{x}$.

The first relationship we examine is given by those vectors $\mathbf{y} \in \Lambda_2$ distinct from $\mathbf{x}$ that are within distance ε of $\mathbf{x}$. We say that two vectors $\mathbf{x}, \mathbf{y}$ are *ε-friends* if $\|\mathbf{x} - \mathbf{y}\| < \varepsilon$. The second relationship examined is that of a determinant, first studied in [2]. A key aspect of these results is that they hold for every Veech surface. This will be used in Appendix A to extend results previously only known for generic translation surface to every Veech surface.

11.1. ε-friends

The main result of this section is the following.

Theorem 11.1. *Let* $\Lambda_{\mathfrak{a}}$, $\Lambda_{\mathfrak{b}}$ *be two scaled discrete* Γ*-orbits. There exists a constant* C (*depending only on* Γ) *such that*

$$\limsup_{R\to\infty} \frac{|\{(\mathbf{x},\mathbf{y}) \in \Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}} : \mathbf{x} \in B_R,\ \mathbf{y} \in B^*_\eta(\mathbf{x})\}|}{|B_R|} < C|B_\eta|.$$

Proof. Let $f = \mathbf{1}_S$ be the characteristic function of

$$S := S(R,\eta) = \{(\mathbf{x},\mathbf{y}) \in \mathbf{R}^2 \times \mathbf{R}^2 : \mathbf{x} \in B_R,\ \mathbf{y} \in B^*_\eta(\mathbf{x})\}, \tag{11.1}$$

and let

$$\Theta(A) := \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f}(A) = \sum_{\mathbf{x}\in\Lambda_{\mathfrak{a}}} \sum_{\mathbf{y}\in\Lambda_{\mathfrak{b}}} f(A\mathbf{x}, A\mathbf{y})$$

for each $A \in \mathrm{GL}_2(\mathbf{R})$. Fix $\varepsilon > 0$. We will show that we can choose $\eta > 0$ such that

$$\limsup_{R\to\infty} \frac{\Theta(I)}{|B_R|} < \varepsilon.$$

Fix $\alpha > 0$ and consider the open symmetric neighborhood $\mathcal{U}_\alpha = K\left\{\begin{pmatrix} e^{r/2} & \\ & e^{-r/2}\end{pmatrix} : r \in (-\alpha,\alpha)\right\}K \subset G$ (in terms of the Cartan decomposition). If f_α denotes the characteristic function of $S(\sqrt{2}Re^{\alpha/2}, \sqrt{2}\eta e^{\alpha/2})$ and we set $\Theta_\alpha := \Theta_{\Lambda_{\mathfrak{a}},\Lambda_{\mathfrak{b}};f_\alpha}$ then for each $A \in \mathcal{U}_\alpha$,

$$\Theta(I) \le \Theta_\alpha(A). \tag{11.2}$$

Indeed, if $\mathbf{x} \in B_R$, then $\|A\mathbf{x}\| < \sqrt{2}e^{\alpha/2}R$, and if $\mathbf{y} \in B^*_\eta(\mathbf{x})$, then $\|A\mathbf{x} - A\mathbf{y}\| < \sqrt{2}\eta e^{\alpha/2}$.

Let C_α be the matrix cone over $\mathcal{U}_\alpha\Gamma/\Gamma$. Then the above estimate together with the fact that $\Theta_\alpha \ge 0$ yields

$$\Theta(I)\frac{\mu(\mathcal{U}_\alpha\Gamma/\Gamma)}{3} = \int_{C_\alpha} \Theta(I)\det(A)^2\,dm(A) \le \int_C \Theta_\alpha(A)\det(A)^2\,dm(A). \tag{11.3}$$

Expressing the Haar measure via the Cartan decomposition we find

$$\mu(\mathcal{U}_\alpha\Gamma/\Gamma) = \tfrac{1}{2\pi V}\int_{-\alpha}^{\alpha} |\sinh(r)|\,dr = \tfrac{1}{\pi V}(\cosh(\alpha) - 1).$$

By (9.1) and Example 9.2 the RHS of (11.3) is bounded above by

$$\begin{aligned} &c_\Gamma^2 |B_{\sqrt{2}Re^{\alpha/2}}|\,|B_{\sqrt{2}\eta e^{\alpha/2}}| + c_\Gamma \iint_{\mathbf{R}^2\times\mathbf{R}^2} (\Phi_{\mathfrak{ab}}(|\mathbf{x}\wedge\mathbf{y}|) - c_\Gamma) f_\alpha(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} \\ &\qquad = 4c_\Gamma^2 e^{2\alpha}|B_R|\,|B_\eta| + O\big((|B_{\sqrt{2}Re^{\alpha/2}}|\,|B_{\sqrt{2}\eta e^{\alpha/2}}|)^{1-\delta/2}\big). \end{aligned}$$

Hence

$$\limsup_{R\to\infty} \frac{\Theta(I)}{|B_R|} \le 3c_\Gamma|B_\eta|\frac{e^{2\alpha}}{\cosh(\alpha) - 1}.$$

The function in α has a local minimum at $\alpha = \log(3)$ with value $27/2$. ∎

The proof of Theorem 11.1 extends to the set of holonomies of a Veech surface.

Theorem 11.2. *Let M be a Veech surface. Then there exists a constant $C > 0$, depending only on M, such that*

$$\limsup_{R\to\infty} \frac{|\{(\mathbf{x},\mathbf{y}) \in S_M \times S_M : \mathbf{x} \in B_R,\ \mathbf{y} \in B_\eta^*(\mathbf{x})\}|}{|S_M \cap B_R|} < C|B_\eta|.$$

Proof. Recall that if M is a Veech surface, its Veech group Γ_M is a non-uniform lattice in G and the set S_M is a finite disjoint union of discrete Γ_M-orbits, i.e., $S_M = \bigcup \Gamma_M \mathbf{x}_i$. To each orbit corresponds a scaling factor $\lambda_i \neq 0$ given by $\Gamma_M \mathbf{x}_i = \lambda_i \Lambda_i$ – here Λ_i is the associated scaled discrete orbit – and we have

$$|S_M \cap B_\eta^*(\mathbf{x})| = \sum |\Lambda_{\mathfrak{a}} \cap B_{\eta/\lambda_i}^*(\mathbf{x}/\lambda_i)|.$$

Counting the number of pairs that are η-friends we find

$$\sum_{\mathbf{x}\in S_M\cap B_R}\ \sum_{\mathbf{y}\in S_M\cap B_\eta^*(\mathbf{x})} 1 = \sum_{i,j}\ \sum_{\mathbf{x}\in\Lambda_j\cap B_{R/\lambda_j}} \left|\Lambda_i \cap B_{\eta/\lambda_i}^*\left(\frac{\lambda_j}{\lambda_i}\mathbf{x}\right)\right|$$

Let f^{ij} be the characteristic function supported on the set

$$\left\{(\mathbf{x},\mathbf{y}) \in \mathbf{R}^2 \times \mathbf{R}^2 : \mathbf{x} \in B_{R/\lambda_j},\ \mathbf{y} \in B_{\eta/\lambda_i}^*\left(\frac{\lambda_j}{\lambda_i}\mathbf{x}\right)\right\}.$$

Running through the proof of Theorem 11.1 we find that

$$\limsup_{R\to\infty} \sum_{i,j} \frac{\Theta_{\Lambda_j,\Lambda_i;f_\alpha^{ij}}(I)}{|S_M \cap B_R|} \ll_\Gamma \sum_{i,j} \frac{c_\Gamma^2}{\lambda_i^2\lambda_j^2}\frac{|B_\eta|}{c_M} = c_M|B_\eta|. \qquad \blacksquare$$

Theorem 1.3 now follows immediately. We note that this result is trivially true for arithmetic lattice surfaces (e.g., square-tiled surfaces).

Proposition 11.3. *If M is an arithmetic lattice surface, then S_M is uniformly discrete.*

Proof. Let Γ_M be its arithmetic Veech lattice. As such, it is commensurable with $\mathrm{SL}_2(\mathbf{Z})$, and there exists a lattice $L \subset \mathbf{R}^2$ such that any two points in S_M differ by an element of L (see [21, Lemma 5.4, Theorem 5.5]). Let ℓ be the length of a shortest lattice vector in L. Then for any $\mathbf{x} \in \mathbf{R}^2$, $|S_M \cap B_{\ell/2}(\mathbf{x})| \leq 1$. ■

11.2. Determinants

Theorem 11.4. *Fix $D > 0$. Let $\Lambda_{\mathfrak{a}}$ and $\Lambda_{\mathfrak{b}}$ be scaled discrete lattice orbits. There is a constant $C = C_\Gamma$ such that*

$$\limsup_{R\to\infty} \frac{|\{(\mathbf{x},\mathbf{y}) \in \Lambda_{\mathfrak{a}} \times \Lambda_{\mathfrak{b}} : \mathbf{x} \in B_R,\ \mathbf{y} \in \mathfrak{D}_{D,1}(\mathbf{x})|}{|B_R|} \leq DC_\Gamma + 2\delta_{\mathfrak{a}\mathfrak{b}}c_\Gamma.$$

Proof. We follow the proof of Theorem 11.1, where f_α is the characteristic function of

$$\{(\mathbf{x},\mathbf{y}) \in \mathbf{R}^2 \times \mathbf{R}^2 : \mathbf{x} \in B_{Re^\alpha},\ \mathbf{y} \in \mathcal{D}_{D,e^{2\alpha}}(\mathbf{x})\}.$$

The conclusion comes from looking at Example 9.3, and noticing that the linearly dependent terms do indeed contribute to the upper bound since for small enough determinant, we are counting parallel pairs which is determined by the Siegel–Veech constant. ■

The extension to holonomy vectors for Veech surfaces given by Theorem 1.6 is as straightforward as in the case of ε-friends.

Appendix A. An application to flows on translation surfaces (by Jon Chaika)

The first subsection introduces the notion of disjointness, and gives a criterion via spectral arguments for a family of flows to be disjoint. The second subsection verifies the criterion for non-arithmetic Veech surfaces. The main result of this section is Theorem A.7 below. Theorem A.7 implies Theorem 1.4 by Theorem 1.3; see the proof of Corollary A.8. The proof uses [4] and a largely geometric argument based on Theorem 1.3 and [42].

A.1. A criterion for disjointness

Definition A.1. Let $(X, \mathcal{B}, \mu, F_1^t)$ and $(Y, \mathcal{A}, \nu, F_2^t)$ be two ergodic probability measure preserving flows. A *joining* is an $(F_1 \times F_2)^t$-preserved measure with marginals μ and ν. The product measure $\mu \times \nu$ is called the *trivial joining*. If the trivial joining is the only joining between two systems we say they are *disjoint*.

If $(X, \mathcal{B}, \mu, F_1^t)$ and $(Y, \mathcal{A}, \nu, F_2^t)$ are isomorphic with ϕ being the isomorphism then $(\mathrm{Id} \times \phi)_*\mu$ is a joining. Thus if two systems are disjoint then they are not isomorphic (unless they are the one-point system).

Let (X, d) be a compact metric space and (X, μ) be a Borel probability measure space.

Definition A.2. Let F^t be a measurable flow on (X, d) that preserves a probability measure μ. We say that $\{t_j\}_{j\in\mathbf{N}} \subset \mathbf{R}$ is a *c-partial rigidity sequence* if for all j there exists $A_j \subset X$ with $\mu(A_j) \geq c$ so that

$$\lim_{j\to\infty} \int_{A_j} d(F^{t_j}x, x)\, d\mu = 0.$$

We say $\{t_j\}$ is a *partial rigidity sequence* if it is a c-partial rigidity sequence for some $c > 0$.

Definition A.3. Let F^t be a measurable flow on (X, d) that preserves a probability measure μ. We say that $\{t_j\}_{j\in\mathbf{N}} \subset \mathbf{R}$ is a *mixing sequence* for F^t if for all $f, g \in L^2(\mu)$ we have

$$\lim_{j\to\infty} \langle f \circ F^{t_j}, g\rangle = \int f\, d\mu \int g\, d\mu.$$

If $L \subset \mathbf{R}$ has the property that any $\{t_j\}_{j\in\mathbf{N}} \subset L$ with $t_j \to \infty$ is a mixing sequence, then we call L a *mixing set*.

The following lemma is easier than results in the literature and we include a proof for the convenience of the reader.

Lemma A.4. *Let F_1^t and F_2^t be two μ-measure preserving and ergodic flows. If there is a mixing sequence for F_1^t which is a partial rigidity sequence for F_2^t then F_1^t and F_2^t are disjoint.*

The proof uses the following notion: A continuous linear map $P : L^2((X,\mu)) \to L^2((Y,\nu))$ is called a *Markov operator* with adjoint operator P^* if

(1) $P \geq 0$ and $P^* \geq 0$;

(2) $P\mathbf{1}_X = \mathbf{1}_Y$ and $P^*\mathbf{1}_Y = \mathbf{1}_X$;

(3) $PU_{F_1^s} = U_{F_2^s}P$ for $s \in \mathbf{R}$, where $U_{F_j^s}$ is the unitary operator associated to F_j^s.

Let τ be a joining of (F_1^t,μ) and (F_2^t,ν). This defines a Markov operator $P_\tau : L^2(\nu) \to L^2(\mu)$ by $\int_B P_\tau(\mathbf{1}_A)\,d\mu = \tau(B \times A)$. Formally, this Markov operator P_τ is the conditional expectation associated to the disintegration of τ over ν. The set of Markov operators is in a one-to-one correspondence with the set of joinings and given a joining τ we will refer to the corresponding Markov operator P_τ as the *associated Markov operator*. This identification respects the convex structure of preserved measures and so the extreme points come from ergodic joinings. For more on Markov operators, see [16].

Proof of Lemma A.4. Let τ be a joining of F_1^t and F_2^t and P be the associated Markov operator. Choose $c > 0$ and $\{t_j\}_{j\in\mathbf{N}}$ so that $\{t_j\}$ is simultaneously a mixing sequence for F_1^t and a c-partial rigidity sequence for F_2^t.

Since $\{t_j\}$ is a mixing sequence for F_1^t, we see that $PF_1^{t_j} = F_2^{t_j}P$ converge in the weak operator topology to the integral because the same is true for the unitary operators associated with $F_1^{t_j}$. Recall that the weak operator topology is compact on operators with uniformly bounded operator norm. Thus, there is a subsequence of t_{j_k} such that the sequence of unitary operators associated with $F_2^{t_{j_k}}$ converges in the weak operator topology to an operator V. Because t_j is a c-partial rigidity sequence for F_2^t, V necessarily has the form $c\,\mathrm{Id} + V'$ where V' is a positive operator. Set $\int$ to be the bounded linear operator on L^2 given by integrating, so $(c\,\mathrm{Id} + V')P = \int$. Since F_1^t has a mixing sequence, implying F_1^t is weakly mixing, it follows that $(F_1 \times F_2)^t$ is ergodic. Thus $\int$ (the Markov operator associated to the product measure) is an extreme point in the set of Markov operators. The equality $(c\,\mathrm{Id} + V')P = \int$ combined with $\int$ being an extreme point implies $c\,\mathrm{Id}P = c\int$ (and $V'P = (1-c)\int$). Thus P is in fact $\int$, implying the trivial joining is the only joining. ∎

The following disjointness criterion is motivated by the approach of [8] and is somewhat easier than the criterion in [11]. Let $\{F_\theta^t\}_{\theta\in S^1}$ be the 1-parameter family of μ-measure preserving flows on X coming from the straight line flow in each direction on X. Let λ be the Lebesgue measure on S^1.

Proposition A.5. *Suppose that*

- F_θ^t *is weakly mixing for* λ*-almost every* θ*,*
- *for every* $L \subset \mathbf{R}$ *which is a union of intervals of length at least* 1 *and has density* 1*, the flow* F_ϕ^t *has a partial rigidity sequence in* L *for* λ*-almost every* ϕ*.*

Then F_θ^t *and* F_ϕ^t *are disjoint for* λ^2*-almost every* $(\theta, \phi) \in (S^1)^2$*.*

Lemma A.6. *Let* F^t *be a* μ*-weakly mixing flow, and let* $r < \infty$*. We can find* $L \subset \mathbf{R}$ *which is a union of intervals of length at least* r*, has density* 1*, and any sequence in* L *going to infinity is a mixing sequence for* F^t*.*

The condition that there exists such an L of density 1 is well known to be equivalent to weak mixing. The assumption that L can be chosen as a union of intervals of length at least 1 is because the unitary operators for F^t with $t \in [0, r]$ form a compact family in the strong operator topology. So, if L' is a mixing set, its r-neighborhood is too.

Proof of Proposition A.5. It suffices to consider when F_θ^t is weakly mixing. By Lemma A.6 choose a set L satisfying the assumptions of Lemma A.6 for $r = 1$. By assumption there is a full measure set D of directions that have a partial rigidity sequence in L. By Lemma A.4, for all $\phi \in D$, F_θ^t and F_ϕ^t are disjoint, as desired. ∎

A.2. Application to non-arithmetic Veech surfaces

We now set about adapting the criterion of Proposition A.5 to certain translation surfaces. The arguments have more or less been carried out in [7, 10, 32]. We present a slightly different, weaker (in particular avoiding *ubiquity*) argument that is sufficient for our purposes.

Theorem A.7. *Let* M *be a translation surface and* μ *denote the measure coming from the area. Let* F_θ^t *denote the flow in direction* θ *on* M*.*

- *Assume the flow in almost every direction is weakly mixing.*
- *Suppose for every* $\varepsilon > 0$ *there exists* $\eta > 0$ *such that at most an* ε*-proportion of holonomies of cylinders have an* η*-friend.*

Then the family $\{F_\theta^t\}_{\theta \in S^1}$ *with the measure* μ *satisfies the assumption of Proposition* A.5*.*

Corollary A.8. *Let* M *be a non-arithmetic Veech surface. Then for almost every* $\theta, \psi \in S^1$ *the flows* F_θ *and* F_ψ *are disjoint. In particular, the flows are not isomorphic. Moreover, for almost all* $\theta, \psi \in S^1$ *the product flows are uniquely ergodic.*

Proof of Corollary A.8 *assuming Theorem* A.7*.* First, the flow in almost every direction on any non-arithmetic Veech surface is weakly mixing [4]. Theorem 1.3 implies that every non-arithmetic Veech surface has the property that for every $\varepsilon > 0$ there exists $\eta > 0$ such that at most an ε-proportion of holonomies of cylinders have an η-friend. These two results are the assumptions of Theorem A.7 and so by Proposition A.5 the flows in almost every pair of directions on M are disjoint.

We now establish the typical unique ergodicity of the product. Observe that the product of two flows is uniquely ergodic if and only if the flows are uniquely ergodic and disjoint. By [30], for every translation surface the flow in almost every direction is uniquely ergodic. Thus, for almost every θ, ψ the product flow $F_\theta \times F_\psi$ is uniquely ergodic. ∎

In the rest of this section we prove Theorem A.7. The key lemma is the following.

Lemma A.9. *Let M be a Veech surface. If $L \subseteq \mathbf{R}$ is a union of intervals of length at least 1 with density 1, then a density 1 set of holonomies of saddle connections have lengths contained in L. That is,*

$$\lim_{R\to\infty} \frac{|\{(\theta, T) \in S_M : T \in L \cap [0, R)\}|}{|\{(\theta, T) \in S_M : T \in [0, R)\}|} = 1.$$

Proof. Denote an element in S_M by its polar decomposition (θ, T). We want to find a bound for

$$\limsup_{R\to\infty} \frac{|\{(\theta, T) \in S_M : T \in L^c \cap [0, R)\}|}{|\{(\theta, T) \in S_M : T \in [0, R)\}|}.$$

Let $\delta > 0$. We split the set in the numerator into two cases, first those elements with η-friends, and then those without η-friends. Having fixed δ, choose η as given by Theorem 1.3. Then

$$\limsup_{R\to\infty} \frac{|\{(\theta, T) \in S_M : T \in L^c \cap [0, R) \text{ and } (\theta, T) \text{ has an } \eta\text{-friend}\}|}{|\{(\theta, T) \in S_M : T \in [0, R)\}|} < \delta.$$

We now consider the other case, i.e.,

$$\limsup_{R\to\infty} \frac{|\{(\theta, T) \in S_M : T \in L^c \cap [0, R) \text{ and } (\theta, T) \text{ has no } \eta\text{-friend}\}|}{|\{(\theta, T) \in S_M : T \in [0, R)\}|}. \tag{A.1}$$

By quadratic growth, the denominator of (A.1) is bounded below by ρR^2 for some $\rho > 0$ and $R > 0$. By the assumptions on L the complement $L^c \cap [0, R)$ is a finite set of q points, where $q < R$. We thicken L^c by placing an interval I_i of length $|I_i| = \ell$ at each point. Making ℓ small enough we define $L^c_\ell \cap [0, R)$ to be the disjoint union of the intervals: $L^c_\ell \cap [0, R) = \bigsqcup_{i=1}^q I_i$.

So by comparing areas, for any interval I_i we have

$$|\{(\theta, T) \in S_M : T \in I_i \text{ and } (\theta, T) \text{ has no } \eta\text{-friend}\}| \leq \frac{\pi(\ell+\eta)^2}{\pi\eta^2} = \frac{(\ell+\eta)^2}{\eta^2}.$$

Summing over the $q < R$ intervals, we find

$$|\{(\theta, T) \in S_M : T \in L^c_\ell \cap [0, R) \text{ and } (\theta, T) \text{ has no } \eta\text{-friend}\}| \leq R\frac{(\ell+\eta)^2}{\eta^2}. \tag{A.2}$$

Since ℓ can be taken to be non-increasing in R and ρ, η are independent of R, (A.1) yields

$$\limsup_{R\to\infty} \frac{\{(\theta, T) \in S_M : T \in L^c \cap [0, R) \text{ and } (\theta, T) \text{ has no } \eta\text{-friend}\}}{\{(\theta, T) \in S_M : T \in [0, R)\}} \leq \limsup_{R\to\infty} \frac{1}{R}\,\frac{(\ell+\eta)^2}{\rho\eta^2} = 0.$$

We conclude by letting $\delta \to 0$. ∎

Let M be an area 1 translation surface of genus g, and s be the length of the shortest saddle connection on M. Set $\sigma = \frac{1}{2g-2}$. We now fix $L \subset \mathbf{R}$, a union of intervals of length at least 1 with density 1. Let

$$\tilde{S}_M(\sigma, N) = \{(\theta, T) : \theta \text{ is the direction of a periodic cylinder of length } T < N \text{ and volume at least } \sigma\}.$$

Let

$$\tilde{S}'_M(\sigma, N) = \{(\theta, T) \in \tilde{S}_M(\sigma, N) : T \in L\}.$$

Proposition A.10. *It suffices to show that there exists $\hat{c} > 0$ such that for any interval J and for any N large enough (depending on J) we have*

$$\lambda\left(\bigcup_{(\theta,T)\in\tilde{S}'_M(\sigma,N)} B\left(\theta, \frac{1}{TN}\right) \cap J \right) > \hat{c}\lambda(J). \tag{A.3}$$

Proof. We first need a lemma connecting being close to the directions of cylinders which have area at least $c > 0$ with partial rigidity sequences.

Lemma A.11. *If v_i are the holonomies of the circumferences of cylinders on Q with area at least c, θ_i are their directions and*

$$\lim_{i\to\infty} |v_i|^2 |\theta_i - \theta| = 0$$

then $|v_i|$ is a c-partial rigidity sequence for $F_{r_\theta Q}$.

Lemma A.11 follows directly from [11, proof of Lemma 12].

We now show that with Lemma A.11, Proposition A.10 follows from the Lebesgue density theorem because (A.3) for all intervals implies that for any $c' < \hat{c}$, and for all large enough N, we have

$$\lambda\left(\bigcup_{(\theta,T)\in\tilde{S}'_M(\sigma,N)} B\left(\theta, \frac{\varepsilon}{TN}\right) \cap J \right) > \varepsilon c'\lambda(J). \tag{A.4}$$

To see how (A.4) follows from (A.3), note that if $B(\theta, \frac{1}{TN}) \subset J$ then $\lambda(B(\theta, \frac{\varepsilon}{TN}) \cap J) = \varepsilon\lambda(B(\theta, \frac{1}{TN}) \cap J)$. Now

$$\lambda\left(J \cap \bigcup_{(\theta,T)\in\tilde{S}_M(\sigma,N):\, B(\theta,\frac{1}{TN})\not\subset J} B\left(\theta, \frac{1}{TN}\right)\right) \le |\partial J|\frac{1}{sN} = \frac{2}{sN},$$

which clearly goes to zero as $N \to \infty$.

Now (A.4) (and the fact that $N \geq T$) establishes that for any $\varepsilon > 0$, the complement of

$$\bigcup_{(\theta,T)\in\tilde{S}'_M(\sigma,N)} B\left(\theta, \frac{\varepsilon}{T^2}\right)$$

has no Lebesgue density points and so $\bigcup_{(\theta,T)\in\tilde{S}'_M(\sigma,N)} B(\theta,\varepsilon/T^2)$ has full measure. So for almost every $\phi \in S^1$ there exists $(\theta_j, T_j) \in \tilde{S}'_M(\sigma,\infty)$ such that

$$\lim_{j\to\infty} T_j^2|\theta_j - \phi| = 0.$$

If ϕ is not in the countable set of direction of elements of $\tilde{S}'_M(\sigma,\infty)$ then the T_j necessarily tend to infinity. By Lemma A.11, the T_j are a σ-partial rigidity sequence for F^t_ϕ and we have established the theorem. Thus we have shown that (A.3) implies Theorem A.7. ■

To prove (A.3), we use a result of Vorobets.

Theorem A.12 (Vorobets, [42, p. 16]). *Let m be the sum of the multiplicities of singularities. Let $c = 2^{2^{4m}}$. For large enough N we have*

$$\bigcup_{(\theta,T)\in\tilde{S}_M(\sigma,N)} B\left(\theta, \frac{c^2}{TN}\right) = S^1. \tag{A.5}$$

Derivation of Theorem A.12 from Vorobets. Let $(\theta, T) \in S^1 \times \mathbf{R}^+$. For $\alpha > 0$ let $A_{(\theta,T)}(\alpha)$ be the set of directions ϕ such that

$$T|\sin(\theta-\phi)| \le \lambda^{-1}c.$$

Vorobets shows that for any $N \ge c$ and $\lambda = N/c$ we have

$$S^1 = \bigcup_{(\theta,T)\in\tilde{S}_M(\sigma,N)} A_{(\theta,T)}(\lambda).$$

If N is large enough, (depending on the shortest saddle connection in M), then $T\sin(\theta-\phi) \le (N/c)^{-1}c$ implies $|\sin(\theta-\phi)| \ge \frac{1}{2}|\theta-\phi|$. So for large enough N we find that for every ϕ there exists $(\theta, T) \in \tilde{S}_M(\sigma, N)$ such that $T\frac{1}{2}|\theta-\phi| \le T|\sin(\theta-\phi)| \le (N/c)^{-1}c$ or that $|\phi-\theta| \le \frac{c^2}{TN}$. ■

Remark A.13. For the reader's convenience, we briefly connect our notation with the notation in [42]. There T_0 plays the role of c. Because we are assuming our translation surface has area 1, the $\sqrt{S}$ term (which is the area of the translation surface) does not appear. Also note that Vorobets allows for a smaller c when $m = 1, 2$. In [42], T plays the role of R.

Proof of Theorem A.7. By Proposition A.10 it suffices to show that (A.5) implies (A.3). Let J be an interval. Observe that

$$\lambda\left(\bigcup_{(\theta,T)\in\tilde{S}'_M(\sigma,N)} B\left(\theta,\frac{1}{TN}\right)\cap J\right)$$
$$\ge \frac{1}{c^2}\lambda\left(\bigcup_{(\theta,T)\in\tilde{S}_M(\sigma,N)} B\left(\theta,\frac{c^2}{TN}\right)\cap J\right) - \sum_{(\theta,T)\in\tilde{S}_M(\sigma,N)\setminus\tilde{S}'_M(\sigma,N)} 2\frac{c^2}{TN}. \tag{A.6}$$

By Lemma A.9 we may assume $|\tilde{S}_M(\sigma, N) \setminus \tilde{S}'_M(\sigma, N)| < \delta N^2$ for all N large enough. Thus for large enough N, by Abel summation,

$$\sum_{(\theta,T)\in\tilde{S}_M(\sigma,N)\setminus\tilde{S}'_M(\sigma,N)} 2\frac{c^2}{TN} < 4c^2\delta.$$

So for any $\delta > 0$ we can bound the right hand side of (A.6) by $\frac{1}{c^2}\lambda(J) - C\delta$. This establishes (A.3). ∎

Remark A.14. In the previous referenced uses of similar disjointness arguments, [8, 11], one does not assume that the systems are weakly mixing. The author of the appendix considers one of the values of this appendix to be presenting a disjointness criterion of this flavor that is simplified by the presence of density 1 mixing sequences. However, similar to those earlier arguments, one can remove the condition that the flow in almost every direction on M is weakly mixing by using [11, Proposition 2] in place of Lemma A.4. Lemma A.9 provides the necessary input for this criterion as well. Compare with [8, Corollary 5] which uses [8, Corollary 3]. By a similar argument to the one above, one can show that any Veech surface (including arithmetic Veech surfaces) has the property that the flow in almost every pair of directions is disjoint. However, even without the main result of the present paper, one can use similar spectral methods to prove that for any branched translation cover of a torus (a class of surfaces that includes all arithmetic Veech surfaces, which are branched translation covers of tori, branched over one point) the flow in almost every pair of directions is disjoint and so we omit it.

Acknowledgments. We would like to thank MSRI (now SLMath) where this project was started in the Fall of 2019 during the program on Holomorphic Differentials in Mathematics and Physics. The authors would like to thank Jayadev Athreya, Max Goering, Jiyoung Han, and Pedram Safaee for useful discussions, and the anonymous referees for their comments.

Funding. C. B. is supported by the Swiss National Science Foundation Grant No. 201557, and would like to thank the Hausdorff Institute for Mathematics in Bonn, where part of this work was completed in the Summer of 2021. S. F. was partially supported by the Deutsche Forschungsgemeinschaft (DFG), Projektnummer 44546644. J. C. is supported by NSF grants DMS-2055354 and DMS-452762, the Sloan foundation, Poincaré chair, and Warnock chair.

References

[1] Athreya, J. S., Cheung, Y., Masur, H.: Siegel–Veech transforms are in L^2. J. Modern Dynam. **14**, 1–19 (2019) Zbl 1432.32011 MR 3959354

[2] Athreya, J. S., Fairchild, S., Masur, H.: Counting pairs of saddle connections. Adv. Math. **431**, article no. 109233, 55 pp. (2023) Zbl 07741084 MR 4624870

[3] Athreya, J. S., Margulis, G. A.: Logarithm laws for unipotent flows. I. J. Modern Dynam. **3**, 359–378 (2009) Zbl 1184.37007 MR 2538473

[4] Avila, A., Delecroix, V.: Weak mixing directions in non-arithmetic Veech surfaces. J. Amer. Math. Soc. **29**, 1167–1208 (2016) Zbl 1373.37005 MR 3522612

[5] Avila, A., Forni, G.: Weak mixing for interval exchange transformations and translation flows. Ann. of Math. (2) **165**, 637–664 (2007) Zbl 1136.37003 MR 2299743

[6] Burrin, C., Nevo, A., Rühr, R., Weiss, B.: Effective counting for discrete lattice orbits in the plane via Eisenstein series. Enseign. Math. **66**, 259–304 (2020) Zbl 1478.11062 MR 4254262

[7] Chaika, J.: Homogeneous approximation for flows on translation surfaces. arXiv:1110.6167 (2011)

[8] Chaika, J.: Every ergodic transformation is disjoint from almost every interval exchange transformation. Ann. of Math. (2) **175**, 237–253 (2012) Zbl 1243.37002 MR 2874642

[9] Chaika, J., Forni, G.: Weakly mixing polygonal billiards. arXiv:2003.00890 (2020)

[10] Chaika, J., Frączek, K., Kanigowski, A., Ulcigrai, C.: Singularity of the spectrum for smooth area-preserving flows in genus two and translation surfaces well approximated by cylinders. Comm. Math. Phys. **381**, 1369–1407 (2021) Zbl 1480.37050 MR 4218685

[11] Chaika, J., Hubert, P.: Circle averages and disjointness in typical translation surfaces on every Teichmüller disc. Bull. London Math. Soc. **49**, 755–769 (2017) Zbl 1378.37003 MR 3742443

[12] Dal'Bo, F.: Crossroads between hyperbolic geometry and number theory. In: Strasbourg master class on geometry, IRMA Lect. Math. Theor. Phys. 18, European Mathematical Society, Zürich, 183–232 (2012) Zbl 1325.37019 MR 2931887

[13] Eskin, A., Masur, H.: Asymptotic formulas on flat surfaces. Ergodic Theory Dynam. Systems **21**, 443–478 (2001) Zbl 1096.37501 MR 1827113

[14] Fairchild, S.: A higher moment formula for the Siegel–Veech transform over quotients by Hecke triangle groups. Groups Geom. Dynam. **15**, 57–81 (2021) Zbl 1472.11204 MR 4235747

[15] Federer, H.: Geometric measure theory. Grundlehren Math. Wiss. 153, Springer, New York (1969) Zbl 0874.49001 MR 0257325

[16] Glasner, E.: Ergodic theory via joinings. Math. Surveys Monogr. 101, American Mathematical Society, Providence, RI (2003) Zbl 1038.37002 MR 1958753

[17] Good, A.: Local analysis of Selberg's trace formula. Lecture Notes in Math. 1040, Springer, Berlin (1983) Zbl 0525.10013 MR 0727476

[18] Gorodnik, A.: Uniform distribution of orbits of lattices on spaces of frames. Duke Math. J. **122**, 549–589 (2004) Zbl 1067.37004 MR 2057018

[19] Gorodnik, A., Weiss, B.: Distribution of lattice orbits on homogeneous varieties. Geom. Funct. Anal. **17**, 58–115 (2007) Zbl 1112.37001 MR 2306653

[20] Greenberg, L.: Discrete groups with dense orbits. In: Flows on homogeneous spaces, Ann. of Math. Stud. 53, Princeton University Press, 85–104 (1963)

[21] Gutkin, E., Judge, C.: Affine mappings of translation surfaces: geometry and arithmetic. Duke Math. J. **103**, 191–213 (2000) Zbl 0965.30019 MR 1760625

[22] Hejhal, D. A.: The Selberg trace formula for $\mathrm{PSL}(2,\mathbf{R})$. Vol. 2. Lecture Notes in Math. 1001, Springer, Berlin (1983) Zbl 0543.10020 MR 0711197

[23] Huxley, M. N.: Exponential sums and lattice points. II. Proc. London Math. Soc. (3) **66**, 279–301 (1993) Zbl 0820.11060 MR 1199067

[24] Huxley, M. N., Nowak, W. G.: Primitive lattice points in convex planar domains. Acta Arith. **76**, 271–283 (1996) Zbl 0861.11056 MR 1397317

[25] Iwaniec, H.: Introduction to the spectral theory of automorphic forms. Biblioteca de la Revista Matemática Iberoamericana, Revista Matemática Iberoamericana, Madrid (1995) Zbl 0847.11028 MR 1325466

[26] Katok, A.: Interval exchange transformations and some special flows are not mixing. Israel J. Math. **35**, 301–310 (1980) Zbl 0437.28009 MR 0594335

[27] Katok, S.: Fuchsian groups. Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL (1992) Zbl 0753.30001 MR 1177168

[28] Kelmer, D.: Approximation of points in the plane by generic lattice orbits. J. Modern Dynam. **11**, 143–153 (2017) Zbl 1432.11076 MR 3627121

[29] Kelmer, D., Yu, S.: The second moment of the Siegel transform in the space of symplectic lattices. Int. Math. Res. Notices **2021**, 5825–5859 Zbl 1479.11119 MR 4251265

[30] Kerckhoff, S., Masur, H., Smillie, J.: Ergodicity of billiard flows and quadratic differentials. Ann. of Math. (2) **124**, 293–311 (1986) Zbl 0637.58010 MR 0855297

[31] Ledrappier, F.: Distribution des orbites des réseaux sur le plan réel. C. R. Acad. Sci. Paris Sér. I Math. **329**, 61–64 (1999) Zbl 0928.22012 MR 1703338

[32] Marchese, L., Treviño, R., Weil, S.: Diophantine approximations for translation surfaces and planar resonant sets. Comment. Math. Helv. **93**, 225–289 (2018) Zbl 1393.37046 MR 3811752

[33] Maucourant, F., Weiss, B.: Lattice actions on the plane revisited. Geom. Dedicata **157**, 1–21 (2012) Zbl 1252.37004 MR 2893477

[34] Miyake, T.: Modular forms. Springer Monogr. Math., Springer, Berlin (2006) Zbl 0701.11014 MR 2194815

[35] Nevo, A., Rühr, R., Weiss, B.: Effective counting on translation surfaces. Adv. Math. **360**, article no. 106890, 29 pp. (2020) Zbl 1451.37049 MR 4031118

[36] Rühr, R., Smilansky, Y., Weiss, B.: Classification and statistics of cut-and-project sets. J. Eur. Math. Soc. **26**, 3575–3638 (2024) Zbl 07891445 MR 4767499

[37] Schmidt, W. M.: A metrical theorem in geometry of numbers. Trans. Amer. Math. Soc. **95**, 516–529 (1960) Zbl 0101.27904 MR 0117222

[38] Selberg, A.: On the estimation of Fourier coefficients of modular forms. In: Theory of Numbers Proc. Sympos. Pure Math. 8, American Mathematical Society, Providence, RI, 1–15 (1965) Zbl 0142.33903 MR 0182610

[39] Siegel, C. L.: A mean value theorem in geometry of numbers. Ann. of Math. (2) **46**, 340–347 (1945) Zbl 0063.07011 MR 0012093

[40] Veech, W. A.: Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. Invent. Math. **97**, 553–583 (1989) Zbl 0676.32006 MR 1005006

[41] Veech, W. A.: Siegel measures. Ann. of Math. (2) **148**, 895–944 (1998) Zbl 0922.22003 MR 1670061

[42] Vorobets, Y.: Periodic geodesics on generic translation surfaces. In: Algebraic and topological dynamics, Contemp. Math. 385, American Mathematical Society, Providence, RI, 205–258 (2005) Zbl 1130.37015 MR 2180238

[43] Wu, C.: Deloné property of the holonomy vectors of translation surfaces. Israel J. Math. **214**, 733–740 (2016) Zbl 1366.37091 MR 3544700