BOUNDS FOR THETA SUMS IN HIGHER RANK. II*

By

JENS MARKLOF AND MATTHEW WELSH

To Peter Sarnak on the occasion of his 70th birthday

Abstract. In the first paper of this series we established new upper bounds for multi-variable exponential sums associated with a quadratic form. The present study shows that if one adds a linear term in the exponent, the estimates can be further improved for almost all parameter values. Our results extend the bound for one-variable theta sums obtained by Fedotov and Klopp in 2012.

1 Introduction

For $M > 0$, a real $n \times n$ symmetric matrix X, and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we define a **theta sum** as the exponential sum

$$\theta_f(M, X, \boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^n} f(M^{-1}(\boldsymbol{m} + \boldsymbol{x}))\, \mathrm{e}\Big(\frac{1}{2}\boldsymbol{m} X\, {}^t\boldsymbol{m} + \boldsymbol{m}\, {}^t\boldsymbol{y}\Big), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{C}$ is a rapidly decaying cut-off and $\mathrm{e}(z) = \mathrm{e}^{2\pi \mathrm{i} z}$ for any complex z. If $f = \chi_{\mathcal{B}}$ is the characteristic function of a bounded set $\mathcal{B} \subset \mathbb{R}^n$ we have the finite sum

$$\theta_f(M, X, \boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^n \cap (M\mathcal{B} - \boldsymbol{x})} \mathrm{e}\Big(\frac{1}{2}\boldsymbol{m} X\, {}^t\boldsymbol{m} + \boldsymbol{m}\, {}^t\boldsymbol{y}\Big). \tag{1.2}$$

In this case we will also use the notation $\theta_f = \theta_{\mathcal{B}}$. In this paper we will focus on the case when

$$\mathcal{B} = (0, b_1) \times \cdots \times (0, b_n) \subset \mathbb{R}^n \text{ with } \boldsymbol{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n_{>0}. \tag{1.3}$$

The theorems below remain valid if $f = \chi_{\mathcal{B}}$ is replaced by any function f in the Schwartz class $\mathcal{S}(\mathbb{R}^n)$ (infinitely differentiable, with rapid decay of all derivatives). The results in the latter case follow from a simpler version of the argument for the sharp truncation, so we do not discuss them here.

The principal result of part I [11] in this series is the following.

*Research supported by EPSRC grant EP/S024948/1.

JOURNAL D'ANALYSE MATHÉMATIQUE, Vol. 151 (2023)
DOI 10.1007/s11854-023-0333-9

Theorem 1.1. *Fix a compact subset $\mathcal{K} \subset \mathbb{R}^n_{>0}$, and let $\psi : [0, \infty) \to [1, \infty)$ be an increasing function such that*

$$\int_0^\infty \psi(t)^{-2n-2} dt < \infty. \tag{1.4}$$

Then there exists a subset $\mathcal{X}(\psi) \subset \mathbb{R}^{n\times n}_{\mathrm{sym}}$ of full Lebesgue measure such that

$$\theta_{\mathcal{B}}(M, X, \boldsymbol{x}, \boldsymbol{y}) = O_X(M^{\frac{n}{2}} \psi(\log M)) \tag{1.5}$$

for all $M \geq 1$, $\boldsymbol{b} = (b_1, \ldots, b_n) \in \mathcal{K}$, $X \in \mathcal{X}(\psi)$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. The implied constants are independent of M, $\boldsymbol{b}$, $\boldsymbol{x}$ and $\boldsymbol{y}$.

For example, for any $\epsilon > 0$, the function $\psi(x) = (x + 1)^{\frac{1}{2n+2}+\epsilon}$ satisfies the condition (1.4), which produces the bound $M^{\frac{n}{2}}(\log M)^{\frac{1}{2n+2}+\epsilon}$ for almost every X and any $\boldsymbol{x}$ and $\boldsymbol{y}$. This improved the previously best bound due to Cosentino and Flaminio [3] by a factor of $(\log M)^n$. Moreover, in the case $n = 1$, Theorem 1.1 recovers the optimal result obtained by Fiedler, Jurkat and Körner [5]; cf. also the extension of this result via nilflows by Flaminio and Forni [6].

In what follows we establish a stronger bound than (1.5), for example $M^{\frac{n}{2}}(\log M)^{\frac{1}{2n+4}+\epsilon}$, but now only valid for almost every $\boldsymbol{y}$. In the case $n = 1$, Theorem 1.2 recovers the upper bound in Theorem 0.1 of Fedotov and Klopp [4].

Theorem 1.2. *Fix a compact subset $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2 \subset \mathbb{R}^n_{>0} \times \mathbb{R}^n$, and let $\psi : [0, \infty) \to [1, \infty)$ be an increasing function such that*

$$\int_0^\infty \psi(t)^{-2n-4} dt < \infty. \tag{1.6}$$

Then there exists a subset $\tilde{\mathcal{X}}(\psi) \subset \mathbb{R}^{n\times n}_{\mathrm{sym}} \times \mathbb{R}^n$ of full Lebesgue measure such that

$$\theta_{\mathcal{B}}(M, X, \boldsymbol{x}, \boldsymbol{y}) = O_{X,\boldsymbol{y}}(M^{\frac{n}{2}} \psi(\log M)) \tag{1.7}$$

for all $M \geq 1$, $(\boldsymbol{b}, \boldsymbol{x}) \in \mathcal{K}$, and $(X, \boldsymbol{y}) \in \tilde{\mathcal{X}}(\psi)$. The implied constants are independent of M, $\boldsymbol{b}$ and $\boldsymbol{x}$.

The paper is organized as follows. In Section 2 we review some basic properties of theta functions and the Jacobi group. The Jacobi group is defined as the semi-direct product $H \rtimes G$ of the Heisenberg group H and the symplectic group $G = \mathrm{Sp}(n, \mathbb{R})$, and, following a construction due to Lion and Vergne [9], the theta function associated to a Schwartz function $f \in \mathcal{S}(\mathbb{R}^n)$ is a function $\Theta_f : H \rtimes G \to \mathbb{C}$ that, for appropriate $g \in G$ and $h \in H$, is a simple rescaling of the theta sums θ_f. The theta functions Θ_f satisfy an automorphy equation, Theorem 3.1, under a certain subgroup $\tilde{\Gamma} \subset H \rtimes G$. This subgroup, defined in Section 3, projects to the discrete subgroup $\Gamma = \mathrm{Sp}(n, \mathbb{Z}) \subset G$.

In order to exploit additional savings from the linear term parameterized by $\boldsymbol{y}$, we found it necessary to have a better understanding of the shape of the cusp of $\Gamma\backslash G$ than in the first paper in this series [11]. For this reason we define in Section 3.1 a new fundamental domain for $\Gamma\backslash G$ which has "box-shape" cusps, as explicated in Section 3.2.

Section 4 contains the proof of Theorem 1.2, which is based on a Borel–Cantelli type argument together with a multi-dimensional dyadic decomposition of the characteristic function of the open unit cube $(0, 1)^n$ that is naturally realized as an action of the diagonal subgroup of G. The execution of the Borel–Cantelli argument rests on a kind of "uniform continuity" property of a certain height function on $H \rtimes G$ that controls the theta function Θ_f, see Corollary 4.1. The required property is proved in Section 4.1, see Lemma 4.4, whose proof is the motivation for the creation of the fundamental domain and the study of its cuspidal regions in Sections 3.1 and 3.2. We remark that the interaction of the dyadic decomposition with the H coordinate in the Jacobi group leads to additional complications not seen in [11], see Section 4.2.

2 Theta functions and the Jacobi group

The theta function Θ_f associated to a Schwartz function $f \in \mathcal{S}(\mathbb{R}^n)$ is a complex-valued function defined on the Jacobi group $H \rtimes G$, the semi-direct product of the Heisenberg group H with the rank n symplectic group $G = \mathrm{Sp}(n, \mathbb{R})$. Here H is the set $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ with multiplication given by

$$(\boldsymbol{x}_1, \boldsymbol{y}_1, t_1)(\boldsymbol{x}_2, \boldsymbol{y}_2, t_2) = \Big(\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}_1 + \boldsymbol{y}_2, t_1 + t_2 + \frac{1}{2}(\boldsymbol{y}_1\,{}^t\boldsymbol{x}_2 - \boldsymbol{x}_1\,{}^t\boldsymbol{y}_2)\Big), \tag{2.1}$$

and G is the group of $2n \times 2n$ real matrices g preserving the standard symplectic form:

$$g \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} {}^t g = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \tag{2.2}$$

with I the $n \times n$ identity. Alternatively, writing g in $n \times n$ blocks,

$$G = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} : A\,{}^tB = B\,{}^tA,\ C\,{}^tD = D\,{}^tC,\ A\,{}^tD - B\,{}^tC = I \right\}. \tag{2.3}$$

We note that G acts on H by automorphisms via

$$h^g = (\boldsymbol{x}A + \boldsymbol{y}C, \boldsymbol{x}B + \boldsymbol{y}D, t), \quad \text{where } h = (\boldsymbol{x}, \boldsymbol{y}, t),\ g = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{2.4}$$

so we may define the semi-direct product $H \rtimes G$, the Jacobi group, with multiplication

$$(h_1, g_1)(h_2, g_2) = (h_1 h_2^{g_1^{-1}}, g_1 g_2). \tag{2.5}$$

The theta function is defined by

$$\Theta_f(h, g) = \sum_{\boldsymbol{m} \in \mathbb{Z}^n} (W(h)R(g)f)(\boldsymbol{m}), \tag{2.6}$$

where W is the Schrödinger representation of H and R is the Segal–Shale–Weil (projective) representation of G. We refer the reader to [11] for details regarding these representations, including the slightly non-standard definition of W and the unitary cocycle $\rho : G \times G \to \mathbb{C}$ satisfying $R(g_1 g_2) = \rho(g_1, g_2)R(g_1)R(g_2)$. We recall here that for

$$g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} Y^{\frac{1}{2}} & 0 \\ 0 & {}^tY^{-\frac{1}{2}} \end{pmatrix} \in G, \tag{2.7}$$

we have

$$\begin{aligned} &\Theta_f((\boldsymbol{x}, \boldsymbol{y}, t), g) \\ &= (\det Y)^{\frac{1}{4}}\, \mathrm{e}\Big(-t + \frac{1}{2}\boldsymbol{x}\,{}^t\boldsymbol{y}\Big) \\ &\quad \times \sum_{\boldsymbol{m} \in \mathbb{Z}^n} f((\boldsymbol{m}+\boldsymbol{x})Y^{\frac{1}{2}})\, \mathrm{e}\Big(\frac{1}{2}(\boldsymbol{m}+\boldsymbol{x})X\,{}^t(\boldsymbol{m}+\boldsymbol{x}) + \boldsymbol{m}\,{}^t\boldsymbol{y}\Big). \end{aligned} \tag{2.8}$$

For $f(\boldsymbol{x}) = \exp\big(-\pi \boldsymbol{x}\,{}^t\boldsymbol{x}\big)$ and $h = (0, 0, 0)$, we recover $(\det Y)^{\frac{1}{4}}$ times the classical Siegel theta series that is holomorphic in the complex symmetric matrix $Z = X + \mathrm{i}Y$. Here we choose $Y^{\frac{1}{2}}$ to be the upper-triangular matrix with positive diagonal entries such that $Y^{\frac{1}{2}}\,{}^tY^{\frac{1}{2}} = Y$, and we emphasize that $Y^{-\frac{1}{2}}$ is always interpreted as $(Y^{\frac{1}{2}})^{-1}$ and not $(Y^{-1})^{\frac{1}{2}}$.

For general $g \in G$ we have the Iwasawa decomposition,

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} Y^{\frac{1}{2}} & 0 \\ 0 & {}^tY^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \mathrm{Re}(Q) & -\mathrm{Im}(Q) \\ \mathrm{Im}(Q) & \mathrm{Re}(Q) \end{pmatrix}, \tag{2.9}$$

where X, Y are symmetric and Q is unitary. Explicitly, we have

$$\begin{aligned} Y &= (C\,{}^tC + D\,{}^tD)^{-1} \\ X &= (A\,{}^tC + B\,{}^tD)(C\,{}^tC + D\,{}^tD)^{-1}, \\ Q &= {}^tY^{\frac{1}{2}}(D + \mathrm{i}C). \end{aligned} \tag{2.10}$$

We often further decompose $Y = UV\,{}^tU$ with U upper-triangular unipotent and V positive diagonal, so $Y^{\frac{1}{2}} = UV^{\frac{1}{2}}$. It is easy to express the Haar measure μ on G in these coordinates,

$$\mathrm{d}\mu(g) = \mathrm{d}Q \prod_{1\le i\le j\le n} \mathrm{d}x_{ij} \prod_{1\le i<j\le n} \mathrm{d}u_{ij} \prod_{1\le j\le n} v_j^{-n+j-2}\mathrm{d}v_{jj}, \tag{2.11}$$

where $\mathrm{d}Q$ is Haar measure on $\mathrm{U}(n)$ and $\mathrm{d}x_{ij}$, $\mathrm{d}u_{ij}$, $\mathrm{d}v_{jj}$ are respectively the Lebesgue measures on the entries of X, U, V. We can also express the Haar measure on the open, dense set of g which can be written as

$$g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & {}^tA^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ T & I \end{pmatrix} \tag{2.12}$$

with $A \in \mathrm{GL}(n,\mathbb{R})$ and X and T symmetric. In these coordinates we have

$$\mathrm{d}\mu(g) = c(\det A)^{-2n-1} \prod_{1\le i\le j\le n} \mathrm{d}x_{ij} \prod_{1\le i,j\le n} \mathrm{d}a_{ij} \prod_{1\le i\le j\le n} \mathrm{d}t_{ij} \tag{2.13}$$

where c is a positive constant and $\mathrm{d}x_{ij}$, $\mathrm{d}a_{ij}$, $\mathrm{d}t_{ij}$ are respectively the Lebesgue measure on the entries of X, A, T, see [11]. We note that the Haar measure $\tilde{\mu}$ on the Jacobi group is simply

$$\mathrm{d}\tilde{\mu}(h,g) = \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}t\,\mathrm{d}\mu(g), \tag{2.14}$$

with $h = (\boldsymbol{x},\boldsymbol{y},t)$ and $\mathrm{d}\boldsymbol{x}$, $\mathrm{d}\boldsymbol{y}$, and $\mathrm{d}t$ the Lebesgue measures.

We often make use of the following refinements of the Iwasawa decomposition. For $1 \le l \le n$ and the same Q as in (2.9), we write $g \in G$ as

$$\begin{pmatrix} I & R_l & T_l - S_l\,{}^tR_l & S_l \\ 0 & I & {}^tS_l & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & -{}^tR_l & I \end{pmatrix} \begin{pmatrix} U_lV_l^{\frac{1}{2}} & 0 & 0 & 0 \\ 0 & Y_l^{\frac{1}{2}} & 0 & X_l\,{}^tY_l^{-\frac{1}{2}} \\ 0 & 0 & {}^tU_l^{-1}V_l^{-\frac{1}{2}} & 0 \\ 0 & 0 & 0 & {}^tY_l^{-\frac{1}{2}} \end{pmatrix} \times \begin{pmatrix} \mathrm{Re}(Q) & -\mathrm{Im}(Q) \\ \mathrm{Im}(Q) & \mathrm{Re}(Q) \end{pmatrix}, \tag{2.15}$$

where R_l and S_l are $l \times (n-l)$ matrices, T_l is $l \times l$ symmetric, U_l is $l \times l$ upper-triangular unipotent, V_l is $l \times l$ positive diagonal, X_l is $(n-l)\times(n-l)$ symmetric, and Y_l is $(n-l)\times(n-l)$ positive definite symmetric. We note that for $l = n$ we recover $X = T_l$ and the factorization $Y = U_lV_l\,{}^tU_l$. In what follows we use $g_l = g_l(g) \in \mathrm{Sp}(n-l,\mathbb{R})$ to denote the matrix

$$g_l = \begin{pmatrix} I & X_l \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_l^{\frac{1}{2}} & 0 \\ 0 & {}^tY_l^{-\frac{1}{2}} \end{pmatrix}. \tag{2.16}$$

These decompositions are closely related to the Langlands decompositions of the maximal parabolic subgroups P_l of G. For $1 \leq l < n$, P_l is the subgroup of $g \in G$ which can be written in the form

$$\begin{pmatrix} I & R_l & T_l - S_l\,{}^tR_l & S_l \\ 0 & I & {}^tS_l & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & -{}^tR_l & I \end{pmatrix} \begin{pmatrix} a_l I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & a_l^{-1} I & 0 \\ 0 & 0 & 0 & I \end{pmatrix} \begin{pmatrix} U_l & 0 & 0 & 0 \\ 0 & A_l & 0 & B_l \\ 0 & 0 & {}^tU_l^{-1} & 0 \\ 0 & C_l & 0 & D_l \end{pmatrix} \tag{2.17}$$

where R_l and S_l are $l \times (n-l)$ matrices, T_l is $l \times l$ symmetric, $a_l > 0$, $U_l \in \mathrm{GL}(l, \mathbb{R})$ with $\det U_l = \pm 1$, and $g_l = \left(\begin{smallmatrix} A_l & B_l \\ C_l & D_l \end{smallmatrix}\right) \in \mathrm{Sp}(n-l, \mathbb{R})$. The maximal parabolic P_n is the subgroup of $g \in G$ that can be written as

$$\begin{pmatrix} I & T_n \\ 0 & I \end{pmatrix} \begin{pmatrix} a_n I & 0 \\ 0 & a_n^{-1} I \end{pmatrix} \begin{pmatrix} U_n & 0 \\ 0 & {}^tU_n{}^{-1} \end{pmatrix} \tag{2.18}$$

where T_n is $n \times n$ symmetric, $a_n > 0$, and $U_n \in \mathrm{GL}(n, \mathbb{R})$ with $\det U_n = \pm 1$. The factorizations (2.17), (2.18) are in fact the Langlands decompositions of P_l, P_n. The first paper in this series [11] contains more details on parabolic subgroups and their Langlands decompositions, and we refer the readers to [13], particularly Sections 4.5.3 and 5.1, [8], particularly section 7.7, and the authors' lecture notes [10] for further details.

3 The subgroups Γ and $\tilde{\Gamma}$

We denote by Γ the discrete subgroup $\Gamma = \mathrm{Sp}(n, \mathbb{Z}) \subset G$. Recalling the notation of [11], for

$$\gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma, \tag{3.1}$$

we set $h_\gamma = (\boldsymbol{r}, \boldsymbol{s}, 0) \in H$ where the entries of $\boldsymbol{r}$ are 0 or $\frac{1}{2}$ depending on whether the corresponding diagonal entry of $C\,{}^tD$ is even or odd, and the entries of $\boldsymbol{s}$ are 0 or $\frac{1}{2}$ depending on whether the corresponding diagonal entry of $A\,{}^tB$ is even or odd. As in [11], we now define the group $\tilde{\Gamma} \subset H \rtimes G$ by

$$\tilde{\Gamma} = \{((\boldsymbol{m}, \boldsymbol{n}, t)h_\gamma, \gamma) \in H \rtimes G : \gamma \in \Gamma, \boldsymbol{m} \in \mathbb{Z}^n, \boldsymbol{n} \in \mathbb{Z}^n, t \in \mathbb{R}\}. \tag{3.2}$$

The relevance of the subgroup $\tilde{\Gamma}$ is made apparent by the following theorem, see theorem 4.1 in [11].

Theorem 3.1. *For any* $(uh_\gamma, \gamma) \in \tilde{\Gamma}$ *and* $(h, g) \in H \rtimes G$*, there is a complex number* $\varepsilon(\gamma)$ *with* $|\varepsilon(\gamma)| = 1$ *such that*

$$\Theta_f((uh_\gamma, \gamma)(h, g)) = \varepsilon(\gamma)\rho(\gamma, g)\, \mathrm{e}\Big(-t + \frac{1}{2}\boldsymbol{m}\,{}^t\boldsymbol{n}\Big)\Theta_f(h, g), \tag{3.3}$$

where $u = (\boldsymbol{m}, \boldsymbol{n}, t)$.

A proof of this theorem is found in [9] but with Γ replaced by the finite index subgroup for which $h_\gamma = (0, 0, 0)$. The automorphy under the full $\tilde{\Gamma}$ is proved in [12], but only for the special function $f(\boldsymbol{x}) = \exp\big(-\pi\boldsymbol{x}\,{}^t\boldsymbol{x}\big)$. It is shown in [9] that this f is an eigenfunction for all the operators $R(k(Q))$, with R the Segal–Shale–Weil representation and $Q \in \mathrm{U}(n)$, and it can be seen from the theory built in [9] that the automorphy for any Schwartz function follows from that for $\exp\big(-\pi\boldsymbol{x}\,{}^t\boldsymbol{x}\big)$. A self-contained proof along the lines of [9] is presented in the authors' lecture notes [10].

3.1 Fundamental domains.

We say that a closed set $\mathcal{D} \subset G$ is a fundamental domain for $\Gamma\backslash G$ if

- for all $g \in G$ there exists $\gamma \in \Gamma$ such that $\gamma g \in \mathcal{D}$ and
- if for $g \in \mathcal{D}$ there is a non-identity $\gamma \in \Gamma$ such that $\gamma g \in \mathcal{D}$, then g is contained in the boundary of $\mathcal{D}$.

Similarly a closed set $\tilde{\mathcal{D}} \subset H \rtimes G$ is a fundamental domain for $\tilde{\Gamma}\backslash(H \rtimes G)$ if

- for all $(h, g) \in H \rtimes G$ there exists $\tilde{\gamma} \in \tilde{\Gamma}$ such that $\tilde{\gamma}(h, g) \in \tilde{\mathcal{D}}$ and
- if for $(h, g) \in \tilde{\mathcal{D}}$ there is a non-identity $\tilde{\gamma} \in \tilde{\Gamma}$ such that $\tilde{\gamma}(h, g) \in \tilde{\mathcal{D}}$, then (h, g) is contained in the boundary of $\tilde{\mathcal{D}}$.

We note that if $\mathcal{D}$ is a fundamental domain for $\Gamma\backslash G$, then

$$\tilde{\mathcal{D}} = \Big\{(\boldsymbol{x}, \boldsymbol{y}, 0) \in H : |x_j|, |y_j| \le \frac{1}{2}\Big\} \times \mathcal{D} \tag{3.4}$$

is a fundamental domain for $\tilde{\Gamma}\backslash(H \rtimes G)$.

In contrast to our previous paper [11], here we need to make careful use of the shape of our fundamental domain $\mathcal{D}$ in the cuspidal regions. Drawing inspiration for the fundamental domain for $\mathrm{GL}(n, \mathbb{Z})\backslash\mathrm{GL}(n, \mathbb{R})$ constructed in [7] as well as from the reduction theory developed in [2] (see also [1]), we construct in this section a new fundamental domain $\mathcal{D} = \mathcal{D}_n$ for $\Gamma\backslash G$. In the following section we study the cuspidal region of $\mathcal{D}_n$.

For $n = 1$, we let $\mathcal{D}_1 \subset G$ denote the standard fundamental domain for $\Gamma\backslash G = \mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$. That is,

$$\mathcal{D}_1 = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{\frac12} & 0 \\ 0 & y^{-\frac12} \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} : |x| \le \frac12, x^2+y^2 \ge 1, 0 \le \phi < 2\pi \right\}. \tag{3.5}$$

We now define fundamental domains $\mathcal{D}_n$ inductively using the decomposition (2.15) for $l = 1$. Writing $g \in G$ as

$$g = \begin{pmatrix} 1 & \boldsymbol{r}_1 & t_1 - \boldsymbol{s}_1 {}^t\boldsymbol{r}_1 & \boldsymbol{s}_1 \\ 0 & I & {}^t\boldsymbol{s}_1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -{}^t\boldsymbol{r}_1 & I \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & I & 0 & X_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & I \end{pmatrix} \times \begin{pmatrix} v_1^{\frac12} & 0 & 0 & 0 \\ 0 & Y_1^{\frac12} & 0 & 0 \\ 0 & 0 & v_1^{-\frac12} & 0 \\ 0 & 0 & 0 & {}^tY_1^{-\frac12} \end{pmatrix} k(Q), \tag{3.6}$$

where $\boldsymbol{r} = \boldsymbol{r}(g) \in \mathbb{R}^{n-1}$, $\boldsymbol{s} = \boldsymbol{s}(g) \in \mathbb{R}^{n-1}$, $t_1 = t_1(g) \in \mathbb{R}$, $X_1 = X_1(g)$ is symmetric, $v_1 = v_1(g) > 0$, $Y_1 = Y_1(g)$ is positive definite symmetric, and $Q \in \mathrm{U}(n)$, we define $\mathcal{D}_n$ as the set of all $g \in G$ satisfying

- $v_1(g) \ge v_1(\gamma g)$ for all $\gamma \in \Gamma$,
- $g_1(g) \in \mathcal{D}_{n-1}$, see (2.16), and
- the entries of $\boldsymbol{r}_1(g)$, $\boldsymbol{s}_1(g)$, and $t_1(g)$ are all less than or equal to $\frac12$ in absolute value with the first entry of $\boldsymbol{r}_1$ greater than or equal to 0.

Proposition 3.1. *$\mathcal{D}_n$ is a fundamental domain for $\Gamma\backslash G$.*

Proof. We begin by showing that for $g \in G$, $\sup_{\gamma\in\Gamma} v_1(\gamma g)$ is indeed obtained by some $\gamma \in \Gamma$. From (2.10), we have for

$$\gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma \tag{3.7}$$

that

$$v_1(\gamma g)^{-1} = \boldsymbol{c} Y {}^t\boldsymbol{c} + (\boldsymbol{c}X + \boldsymbol{d}) Y^{-1} {}^t(\boldsymbol{c}X + \boldsymbol{d}) \tag{3.8}$$

where

$$g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} Y^{\frac12} & 0 \\ 0 & {}^tY^{-\frac12} \end{pmatrix} k(Q) \tag{3.9}$$

and $\boldsymbol{c}, \boldsymbol{d}$ are the first rows of C, D. Since Y is positive definite, there are only finitely many $\boldsymbol{c}$ such that $\boldsymbol{c} Y {}^t\boldsymbol{c}$, and hence $v_1(\gamma g)^{-1}$, is below a given bound. Similarly, for a fixed $\boldsymbol{c}$, the positive definiteness of Y^{-1} implies that there are only finitely many $\boldsymbol{d}$ such that $v_1(\gamma g)^{-1}$ is below a given bound. It follows that there are only finitely many $\gamma \in \Gamma_1 \backslash \Gamma$ such that $v_1(\gamma g)$ is larger than a given bound, where $\Gamma_1 = \Gamma \cap P_1$ and we recall P_1 is given by (2.17). As $v_1(\gamma g) = v_1(g)$ for $\gamma \in \Gamma_1$ it follows that $v_1(\gamma g)$ is maximized for some $\gamma \in \Gamma$.

Let γ_0 be so that $v_1(\gamma_0 g)$ is maximal. We now decompose an arbitrary $\gamma \in \Gamma_1$ as in (2.17),

$$\gamma = \begin{pmatrix} 1 & \boldsymbol{r}_1 & t_1 - \boldsymbol{s}_1 {}^t\boldsymbol{r}_1 & \boldsymbol{s}_1 \\ 0 & I & {}^t\boldsymbol{s}_1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -{}^t\boldsymbol{r}_1 & I \end{pmatrix} \begin{pmatrix} \pm 1 & 0 & 0 & 0 \\ 0 & A_1 & 0 & B_1 \\ 0 & 0 & \pm 1 & 0 \\ 0 & C_1 & 0 & D_1 \end{pmatrix} \tag{3.10}$$

with

$$\gamma_1 = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix} \in \mathrm{Sp}(n-1, \mathbb{Z}). \tag{3.11}$$

Proceeding inductively, there exists γ_1 such that $\gamma_1 g_1(\gamma_0 g) = g_1(\gamma\gamma_0 g) \in \mathcal{D}_{n-1}$. Now, we can change $\boldsymbol{r}_1(\gamma), \boldsymbol{s}_1(\gamma), t_1(\gamma)$, and the $\pm$, noting that this does not change $g_1(\gamma\gamma_0 g)$, so that the entries of $\boldsymbol{r}_1(\gamma\gamma_0 g)$, $\boldsymbol{s}_1(\gamma\gamma_0 g)$ and $t_1(\gamma\gamma_0 g)$ are all $\leq \frac{1}{2}$ in absolute value and the first entry of $\boldsymbol{r}_1(\gamma\gamma_0 g)$ is nonnegative. Therefore $\gamma\gamma_1 g \in \mathcal{D}_n$ as required.

We now suppose that $g \in \mathcal{D}_n$ and there is a non-identity $\gamma \in \Gamma$ such that $\gamma g \in \mathcal{D}_n$. We set

$$\gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} Y^{\frac{1}{2}} & 0 \\ 0 & {}^tY^{-\frac{1}{2}} \end{pmatrix} k(Q). \tag{3.12}$$

By the maximality, we have $v_1(g) = v_1(\gamma g)$ and therefore

$$v_1{}^{-1} = \boldsymbol{c} Y {}^t\boldsymbol{c} + (\boldsymbol{c}X + \boldsymbol{d}) Y^{-1}\, {}^t(\boldsymbol{c}X + \boldsymbol{d}) \tag{3.13}$$

where $\boldsymbol{c}$ and $\boldsymbol{d}$ are the first rows of C and D. Let us first consider the case when $\boldsymbol{c} \neq 0$. To show that g is on the boundary of $\mathcal{D}_n$ in this case, we consider

$$g_\epsilon = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} (1-\epsilon)^{\frac{1}{2}} Y^{\frac{1}{2}} & 0 \\ 0 & (1-\epsilon)^{-\frac{1}{2}}\, {}^tY^{-\frac{1}{2}} \end{pmatrix} k(Q) \tag{3.14}$$

for $0 < \epsilon < 1$. We have $v_1(g_\epsilon) = (1-\epsilon) v_1(g)$ and

$$\begin{aligned} v_1(\gamma g_\epsilon)^{-1} &= (1-\epsilon)\boldsymbol{c} Y {}^t\boldsymbol{c} + (1-\epsilon)^{-1} (\boldsymbol{c}X + \boldsymbol{d}) Y^{-1}\, {}^t(\boldsymbol{c}X + \boldsymbol{d}) \\ &= ((1-\epsilon) - (1-\epsilon)^{-1})\, \boldsymbol{c} Y {}^t\boldsymbol{c} + v_1(g_\epsilon)^{-1} \end{aligned} \tag{3.15}$$

by (3.13). Since $v_1(\gamma g_\epsilon) > v_1(g_\epsilon)$, we have that $g_\epsilon \notin \mathcal{D}_n$. As g_ϵ can be made arbitrarily close to g, we conclude that g is on the boundary of $\mathcal{D}_n$.

If $\boldsymbol{c} = 0$, then from (3.13) we have

$$v_1(g)^{-1} = (d^{(1)} - \boldsymbol{d}^{(2)}\,{}^t\boldsymbol{r}_1)^2 v_1(g)^{-1} + \boldsymbol{d}^{(2)} Y_1^{-1}\,{}^t\boldsymbol{d}^{(2)} \tag{3.16}$$

where $\boldsymbol{d} = (d^{(1)} \quad \boldsymbol{d}^{(2)})$ are as above,

$$Y = \begin{pmatrix} 1 & \boldsymbol{r}_1 \\ 0 & I \end{pmatrix} \begin{pmatrix} v_1 & 0 \\ 0 & Y_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ {}^t\boldsymbol{r}_1 & I \end{pmatrix}. \tag{3.17}$$

This time we consider

$$g_\epsilon = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_\epsilon^{\frac{1}{2}} & 0 \\ 0 & {}^tY_\epsilon{}^{-\frac{1}{2}} \end{pmatrix} k(Q) \tag{3.18}$$

with

$$Y_\epsilon = \begin{pmatrix} 1 & \boldsymbol{r}_1 \\ 0 & I \end{pmatrix} \begin{pmatrix} (1-\epsilon)v_1 & 0 \\ 0 & Y_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ {}^t\boldsymbol{r}_1 & I \end{pmatrix}. \tag{3.19}$$

We have $v_1(g_\epsilon) = (1-\epsilon)v_1(g)$ and

$$\begin{aligned} v_1(\gamma g_\epsilon)^{-1} &= (1-\epsilon)^{-1}(d^{(1)} - \boldsymbol{d}^{(2)}\,{}^t\boldsymbol{r}_1)^2 v_1(g)^{-1} + \boldsymbol{d}^{(2)} Y_1^{-1}\,{}^t\boldsymbol{d}^{(2)} \\ &= v_1(g_\epsilon)^{-1} + (1-(1-\epsilon)^{-1})\boldsymbol{d}^{(2)} Y_1\,{}^t\boldsymbol{d}^{(2)} \end{aligned} \tag{3.20}$$

from (3.16). If $\boldsymbol{d}^{(2)} \neq 0$, then $v_1(\gamma g_\epsilon) > v_1(g_\epsilon)$ and we conclude that g is on the boundary of $\mathcal{D}_n$ as before.

When $\boldsymbol{c} = 0$ and $\boldsymbol{d}^{(2)} = 0$ we have $d^{(1)} = \pm 1$, and so $\gamma \in \Gamma_1$. We decompose γ as in (3.10) and define γ_1 as in (3.11). By the construction of $\mathcal{D}_n$, we have $g_1(g) \in \mathcal{D}_{n-1}$ and $g_1(\gamma g) = \gamma_1 g_1(g) \in \mathcal{D}_{n-1}$. By induction, we have that either γ_1 is the identity or $g_1(g)$ is on the boundary of $\mathcal{D}_{n-1}$. In the latter case we have that g is on the boundary of $\mathcal{D}_n$, and so it remains to consider

$$\gamma = \begin{pmatrix} \pm 1 & \boldsymbol{r}_1 & \pm t_1 \mp \boldsymbol{r}_1\,{}^t\boldsymbol{s}_1 & \boldsymbol{s}_1 \\ 0 & I & \pm\,{}^t\boldsymbol{s}_1 & 0 \\ 0 & 0 & \pm 1 & 0 \\ 0 & 0 & \mp\boldsymbol{r}_1 & I \end{pmatrix}. \tag{3.21}$$

If any of the entries of $\boldsymbol{r}_1(\gamma)$ or $\boldsymbol{s}_1(\gamma)$ is not zero, then the corresponding entry of $\boldsymbol{r}_1(g)$ or $\boldsymbol{s}_1(g)$ is $\pm\frac{1}{2}$ and so g is on the boundary of $\mathcal{D}_n$. Similarly if $t_1(\gamma) \neq 0$, we have $t_1(g) = \pm\frac{1}{2}$ and again g is on the boundary of $\mathcal{D}_n$. If all of $\boldsymbol{r}_1, \boldsymbol{s}_1, t_1$ are 0, the sign must be $-$ as γ is not the identity, and it follows that the first entry of $\boldsymbol{r}_1(g)$ is 0 and g is again on the boundary of $\mathcal{D}_n$. $\square$

The following proposition records some useful properties of $\mathcal{D}_n$. It and its proof are very similar to the analogous statement for the different fundamental domain used in [11], see proposition 3.1 there.

Proposition 3.2. *Let $g \in \mathcal{D}_n$ and write*

$$g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} Y^{\frac{1}{2}} & 0 \\ 0 & {}^tY^{-\frac{1}{2}} \end{pmatrix} k(Q), \qquad Y = UV\,{}^tU, \tag{3.22}$$

where X is symmetric, Y is positive definite symmetric, U upper triangular unipotent, V positive diagonal, and $Q \in \mathrm{U}(n)$, and

$$V = \begin{pmatrix} v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_n \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & \boldsymbol{r}_1 \\ 0 & I \end{pmatrix} \begin{pmatrix} v_1 & 0 \\ 0 & Y_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ {}^t\boldsymbol{r}_1 & I \end{pmatrix}. \tag{3.23}$$

Then we have

(1) $v_n \geq \frac{\sqrt{3}}{2}$ *and* $v_j \geq \frac{3}{4} v_{j+1}$ *for* $1 \leq j \leq n-1$,

(2) *for all* $\boldsymbol{x} = (x^{(1)} \quad \boldsymbol{x}^{(2)}) \in \mathbb{R}^n$

$$\boldsymbol{x} Y\, {}^t\boldsymbol{x} \asymp_n v_1 (x^{(1)})^2 + \boldsymbol{x}^{(2)} Y_1\, {}^t\boldsymbol{x}^{(2)}. \tag{3.24}$$

Proof. For the first, we observe that by the inductive construction of $\mathcal{D}_n$, we have that

$$g_{n-1}(g) = \begin{pmatrix} 1 & x_{n-1}(g) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_n^{\frac{1}{2}} & 0 \\ 0 & v_n^{-\frac{1}{2}} \end{pmatrix} \in \mathcal{D}_1. \tag{3.25}$$

As $\mathcal{D}_1$ is the standard fundamental domain for $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$, we conclude that $v_n \geq \frac{\sqrt{3}}{2}$.

To demonstrate that $v_j \geq \frac{3}{4} v_{j+1}$, we note that by the construction of $\mathcal{D}_n$, it suffices to consider only $j = 1$. We start with

$$v_1^{-1} \leq \boldsymbol{c} Y\, {}^t\boldsymbol{c} + (\boldsymbol{c}X + \boldsymbol{d}) Y^{-1}\, {}^t(\boldsymbol{c}X + \boldsymbol{d}) \tag{3.26}$$

for any $(\boldsymbol{c} \quad \boldsymbol{d}) \in \mathbb{Z}^{2n}$ nonzero and primitive. Choosing

$$\boldsymbol{c} = 0 \quad \text{and} \quad \boldsymbol{d} = (0 \quad 1 \quad 0 \quad \cdots \quad 0),$$

we have

$$v_1^{-1} \leq v_1^{-1} (r_1^{(1)})^2 + v_2^{-1}, \tag{3.27}$$

where $r_1^{(1)}$ is the first entry of $\boldsymbol{r}_1$. Since $0 \leq r_1^{(1)} \leq \frac{1}{2}$, we conclude that $v_1 \geq \frac{3}{4} v_2$.

To demonstrate the second part of the proposition, we let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ denote the rows of

$$Y^{\frac{1}{2}} = \begin{pmatrix} 1 & \boldsymbol{r}_1 \\ 0 & I \end{pmatrix} \begin{pmatrix} v_1^{\frac{1}{2}} & 0 \\ 0 & Y_1^{\frac{1}{2}} \end{pmatrix}. \tag{3.28}$$

Setting $\boldsymbol{y} = x_2\boldsymbol{y}_2 + \cdots + x_n\boldsymbol{y}_n$, where the x_j are the entries of $\boldsymbol{x}$, our aim is to prove that for some constants $0 < c_1 < 1 < c_2$ depending only on n,

$$c_1(\|\boldsymbol{y}_1\|^2 x_1^2 + \|\boldsymbol{y}\|^2) \le \|x_1\boldsymbol{y}_1 + \boldsymbol{y}\|^2 \le c_2(\|\boldsymbol{y}_1\|^2 x_1^2 + \|\boldsymbol{y}\|^2), \tag{3.29}$$

from which the lower bound in (3.24) follows as $\|\boldsymbol{y}_1\|^2 \ge v_1$. The upper bound in (3.24) follows from (3.29) and $v_1 \gg \|\boldsymbol{y}_1\|^2$, which is verified below, see (3.35). Expanding the expression in the middle of (3.29), we find that it is enough to show that

$$2|x_1\boldsymbol{y}_1\,{}^t\boldsymbol{y}| \le (1 - c_1)(\|\boldsymbol{y}_1\|^2 x_1^2 + \|\boldsymbol{y}\|^2), \tag{3.30}$$

and

$$2|x_1\boldsymbol{y}_1\,{}^t\boldsymbol{y}| \le (c_2 - 1)(\|\boldsymbol{y}_1\|^2 x_1^2 + \|\boldsymbol{y}\|^2). \tag{3.31}$$

The upper bound (3.31) is trivial if $c_2 = 2$, and the upper bound (3.30) would follow from

$$|\boldsymbol{y}_1\,{}^t\boldsymbol{y}| \le (1 - c_1)\|\boldsymbol{y}_1\|\,\|\boldsymbol{y}\|. \tag{3.32}$$

We let $0 < \phi_1 < \pi$ denote the angle between $\boldsymbol{y}_1$ and $\boldsymbol{y}$ and $0 < \phi_2 < \frac{\pi}{2}$ denote the angle between $\boldsymbol{y}_1$ and the hyperplane $\mathrm{span}(\boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)$. We have $\phi_2 \le \min(\phi_1, \pi - \phi_1)$, and so $|\cos\phi_1| \le |\cos\phi_2|$. We bound $\cos\phi_2$ away from 1 by bounding $\sin\phi_2$ away from 0.

We have

$$|\sin\phi_2| = \frac{\|\boldsymbol{y}_1 \wedge \cdots \wedge \boldsymbol{y}_n\|}{\|\boldsymbol{y}_1\|\,\|\boldsymbol{y}_2 \wedge \cdots \wedge \boldsymbol{y}_n\|} = \frac{v_1^{\frac{1}{2}}}{\|\boldsymbol{y}_1\|}, \tag{3.33}$$

so it suffices to show that $v_1^{\frac{1}{2}} \gg \|\boldsymbol{y}_1\|$. Here $\wedge$ denotes the usual wedge product on $\mathbb{R}^n$ and the norm on $\bigwedge^k \mathbb{R}^n$ is given by

$$\|\boldsymbol{a}_1 \wedge \cdots \wedge \boldsymbol{a}_k\|^2 = \det \begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_k \end{pmatrix} \begin{pmatrix} {}^t\boldsymbol{a}_1 & \cdots & {}^t\boldsymbol{a}_k \end{pmatrix}. \tag{3.34}$$

Using the inductive construction of $\mathcal{D}_n$ and the fact that the entries of $\boldsymbol{r}_1(Y), \boldsymbol{r}_1(Y_1), \ldots$ are at most $\frac{1}{2}$ in absolute value, we observe that U has entries bounded by a constant depending only on n. We find that

$$\|\boldsymbol{y}_1\|^2 \ll v_1 + \cdots + v_n \ll v_1 \tag{3.35}$$

with the implied constant depending on n. □

3.2 Shape of the cusp. As explicated in [1] and [2], the cusp of $\Gamma\backslash G$ can be partitioned into $2^n - 1$ box-shaped regions. These regions are in correspondence with the conjugacy classes of proper parabolic subgroups of G and are formed as K times the product of three subsets, one for each of the components—nilpotent, diagonal, and semisimple—of the Langlands decomposition of P.

In what follows we use the fundamental domain $\mathcal{D}_n$ constructed in Section 3.1 to prove a variation of this fact, although only for the maximal parabolic subgroups (2.17), (2.18). Our main result for this section is Proposition 3.5, which roughly states that if $g \in G$ is close enough the boundary in a precise sense, then g can be brought into $\mathcal{D}_n$ by an element γ in some maximal parabolic subgroup which depends on the way g approaches the boundary.

For $1 \leq l < n$ we denote by $\Gamma_{l,1}$ and $\Gamma_{l,2}$ the subgroups of $\Gamma_l = \Gamma \cap P_l$ given by

$$\Gamma_{l,1} = \left\{ \begin{pmatrix} A & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & {}^t\!A^{-1} & 0 \\ 0 & 0 & 0 & I \end{pmatrix} : A \in \mathrm{GL}(l, \mathbb{Z}) \right\} \tag{3.36}$$

and

$$\Gamma_{l,2} = \left\{ \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & A & 0 & B \\ 0 & 0 & I & 0 \\ 0 & C & 0 & D \end{pmatrix} : \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathrm{Sp}(n-l, \mathbb{Z}) \right\}. \tag{3.37}$$

For $l = n$, we set

$$\Gamma_{n,1} = \left\{ \begin{pmatrix} A & 0 \\ 0 & {}^t\!A^{-1} \end{pmatrix} : A \in \mathrm{GL}(n, \mathbb{Z}) \right\}, \tag{3.38}$$

and we let $\Gamma_{n,2}$ be trivial. We now define, for $g \in G$ and $1 \leq l \leq n$,

$$v_l(\Gamma_l g) := \min_{\gamma \in \Gamma_l} v_l(\gamma g) = \min_{\gamma \in \Gamma_{l,1}} v_l(\gamma g) \tag{3.39}$$

and, for $1 \le l < n$,

$$v_{l+1}(\Gamma_l g) := \max_{\gamma\in\Gamma_l} v_{l+1}(\gamma g) = \max_{\gamma\in\Gamma_{l,2}} v_{l+1}(\gamma g). \tag{3.40}$$

Here $v_l(g)$ denotes the lth diagonal entry of V in the decomposition of g, see (2.9). Abusing notation, we also use $v_l(Y)$ to denote the lth diagonal entry of V, where for a positive definite matrix Y (of any size at least $l \times l$), we decompose $Y = UV\,{}^tU$ as above. The quantities $v_l(g)$, $v_l(Y)$, and $v_l(U_lV_l\,{}^tU_l)$ of course agree when Y comes from the Iwasawa decomposition of g and U_l, V_l come from the refinement (2.15).

We also note that the second equalities in (3.39) and (3.40) follow from the observation that for g as in (2.15) and $\gamma \in \Gamma_l$, we have

$$\begin{aligned}
\gamma g &= \begin{pmatrix} I & M & L-N\,{}^tM & N \\ 0 & I & {}^tN & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & -{}^tM & I \end{pmatrix} \begin{pmatrix} A' & 0 & 0 & 0 \\ 0 & A & 0 & B \\ 0 & 0 & {}^t(A')^{-1} & 0 \\ 0 & C & 0 & D \end{pmatrix} \\
&\quad \times \begin{pmatrix} I & R_l & T_l - S_l\,{}^tR_l & S_l \\ 0 & I & {}^tS_l & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & -{}^tR_l & I \end{pmatrix} \begin{pmatrix} U_lV_l^{\frac12} & 0 & 0 & 0 \\ 0 & Y_l^{\frac12} & 0 & X_l\,{}^tY_l^{-\frac12} \\ 0 & 0 & {}^tU_l^{-1}V_l^{-\frac12} & 0 \\ 0 & 0 & 0 & {}^tY_l^{-\frac12} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \mathrm{Re}(Q) & -\mathrm{Im}(Q) \\ \mathrm{Im}(Q) & \mathrm{Re}(Q) \end{pmatrix} \\
&= \begin{pmatrix} I & * & * & * \\ 0 & I & * & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & * & I \end{pmatrix} \begin{pmatrix} A'U_lV_l^{\frac12} & 0 & 0 & 0 \\ 0 & AY_l^{\frac12} & 0 & (AX_l+B)\,{}^tY_l^{-\frac12} \\ 0 & 0 & {}^t(A'U_lV_l^{\frac12})^{-1} & 0 \\ 0 & CY_l^{\frac12} & 0 & (CX_l+D)\,{}^tY^{-\frac12} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \mathrm{Re}(Q) & -\mathrm{Im}(Q) \\ \mathrm{Im}(Q) & \mathrm{Re}(Q) \end{pmatrix},
\end{aligned} \tag{3.41}$$

and therefore only the part of γ in $\Gamma_{l,1}$ changes $v_l(\gamma g)$ and only the part in $\Gamma_{l,2}$ changes $v_{l+1}(g)$.

Finally, we note that in the proof of Proposition 3.1, we saw that the maximum in (3.40) does exist. As for the minimum in (3.39), we simply note that

$$v_l(AU_lV_l\,{}^tU_l\,{}^tA) = \boldsymbol{a}U_lV_l\,{}^tU_l\,{}^t\boldsymbol{a} \tag{3.42}$$

where $\boldsymbol{a}$ is the last row of $A \in \mathrm{GL}(l, \mathbb{Z})$, so the positive definiteness of $U_lV_l\,{}^tU_l$ implies that there are only finitely many values of $v_l(AU_lV_l\,{}^tU_l\,{}^tA)$ below a given bound.

We now define a fundamental domain $\mathcal{D}'_l$ for the action of $\mathrm{GL}(l,\mathbb{Z})$ on $l \times l$ positive definite symmetric matrices. We set $\mathcal{D}'_1 = \{y > 0\}$ and

$$\mathcal{D}'_2 = \left\{ \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix} : 0 \le r \le \frac{1}{2},\ r^2 + \frac{v_1}{v_2} \ge 1 \right\}, \tag{3.43}$$

the standard fundamental domain for $\mathrm{GL}(2,\mathbb{Z})$ acting on 2×2 positive definite symmetric matrices. The domain $\mathcal{D}'_l$ for $l > 2$ is then defined inductively as the set of all

$$Y = \begin{pmatrix} 1 & \boldsymbol{r} \\ 0 & I \end{pmatrix} \begin{pmatrix} v_1 & 0 \\ 0 & Y_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \boldsymbol{r} & 1 \end{pmatrix} \tag{3.44}$$

such that

(1) $v_1(Y) \ge v_1(AY\,{}^t\!A)$ for all $A \in \mathrm{GL}(l,\mathbb{Z})$,
(2) $Y_1 \in \mathcal{D}'_{l-1}$, and
(3) $|r_j| \le \frac{1}{2}$ and $0 \le r_1 \le \frac{1}{2}$ where r_j are the entries of $\boldsymbol{r}$.

This is in fact the set of Y such that Y^{-1} is in Grenier's fundamental domain, see [7] and [13], so we do not prove that $\mathcal{D}'_l$ is a fundamental domain here. We do however record the following properties of $\mathcal{D}'_l$.

Lemma 3.3. *Let $UV\,{}^tU \in \mathcal{D}'_l$ with*

$$V = \begin{pmatrix} v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_l \end{pmatrix} \tag{3.45}$$

positive diagonal and U upper triangular unipotent. Then we have

(1) $v_j \ge \frac{3}{4} v_{j+1}$ *for* $1 \le j < l$,
(2) *for any* $\boldsymbol{x} \in \mathbb{R}^l$,

$$\boldsymbol{x} UV\,{}^tU\,{}^t\boldsymbol{x} \asymp \boldsymbol{x} V\,{}^t\boldsymbol{x} \tag{3.46}$$

with implied constant depending only on l, and

(3)

$$\min_{A \in \mathrm{GL}(l,\mathbb{Z})} v_l(AUV\,{}^tU\,{}^t\!A) \asymp v_l(UV\,{}^tU) \tag{3.47}$$

with implied constant depending only on l.

Proof. The first and second parts are proved in proposition 3.1 of [11]. To prove the third part, we note that with $\boldsymbol{a}$ the last row of A,

$$v_l(AUV\,{}^tU\,{}^t\!A) = \boldsymbol{a}UV\,{}^tU\,{}^t\boldsymbol{a} \gg \boldsymbol{a}V\,{}^t\boldsymbol{a}, \tag{3.48}$$

by the second part of the lemma. Applying the first part of the lemma we have $\boldsymbol{a}V\,{}^t\boldsymbol{a} \gg v_l\|\boldsymbol{a}\|^2 \ge v_l$, and (3.47) follows. □

As the proof is almost identical to the proof of the third part of Lemma 3.3, we record the following lemma for later use.

Lemma 3.4. *If $g \in \mathcal{D}_n$, then for all $1 \le l < n$,*

$$\upsilon_l(\Gamma_l g) \asymp \upsilon_l(g) \tag{3.49}$$

with the implied constant depending only on n.

Proof. We recall from the second part of Proposition 3.2 that for $\boldsymbol{x} \in \mathbb{R}^l$,

$$\boldsymbol{x} U_l V_l {}^t U_l {}^t\boldsymbol{x} \gg \boldsymbol{x} V_l {}^t\boldsymbol{x}. \tag{3.50}$$

We have

$$\upsilon_l(\Gamma_l g) = \min_{\substack{\boldsymbol{c}\in\mathbb{Z}^l \\ \boldsymbol{c}\neq 0}} \boldsymbol{c} U_l V_l {}^t U_l {}^t\boldsymbol{c} \gg \min_{\substack{\boldsymbol{c}\in\mathbb{Z}^l \\ \boldsymbol{c}\neq 0}} \boldsymbol{c} V_l {}^t\boldsymbol{c}. \tag{3.51}$$

Now as $\boldsymbol{c} \neq 0$, we have $c_j^2 \ge 1$ for some $1 \le j \le l$, and so

$$\upsilon_l(\Gamma_l g) \gg \upsilon_j(g) \gg \upsilon_l(g) \tag{3.52}$$

by the first part of Proposition 3.2. □

We are now ready to prove the main result for this section.

Proposition 3.5. *For $1 \le l \le n$, there are constants $a_l > 0$ such that for $l < n$, if $g \in G$ satisfies $\upsilon_l(\Gamma_l g) \ge a_l \upsilon_{l+1}(\Gamma_l g)$, and for $l = n$ if $g \in G$ satisfies $\upsilon_n(\Gamma_n g) \ge a_n$, then there exists $\gamma \in \Gamma_l$ so that $\gamma g \in \mathcal{D}_n$. Moreover, for this γ we have $\upsilon_l(\Gamma_l g) \asymp \upsilon_l(\gamma g)$ and, for $l < n$, $\upsilon_{l+1}(\Gamma_l g) = \upsilon_{l+1}(\gamma g)$.*

We remark that this proposition can be extended to any of the parabolic subgroups P_L of G by taking intersections of the maximal parabolics. However, some care needs to be taken regarding the possible non-uniqueness of the γ bringing g into $\mathcal{D}_n$. Since it is unnecessary for our goals, we do not discuss this here.

Proof. By multiplying g by

$$\gamma_1 = \begin{pmatrix} A' & 0 & 0 & 0 \\ 0 & A & 0 & B \\ 0 & 0 & {}^t(A')^{-1} & 0 \\ 0 & C & 0 & D \end{pmatrix} \in \Gamma_l, \tag{3.53}$$

which leaves $\upsilon_l(\Gamma_l g)$ and $\upsilon_{l+1}(\Gamma_l g)$ unchanged, see also (3.41), we may assume that $U_l V_l {}^t U_l \in \mathcal{D}'_l$ and

$$\begin{pmatrix} I & X_l \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_l^{\frac{1}{2}} & 0 \\ 0 & {}^tY_l^{-\frac{1}{2}} \end{pmatrix} \in \mathcal{D}_{n-l}. \tag{3.54}$$

We recall that for $\gamma = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$,

$$v_1(\gamma g)^{-1} = \boldsymbol{c} Y {}^t\boldsymbol{c} + (\boldsymbol{c}X + \boldsymbol{d}) Y^{-1}\, {}^t(\boldsymbol{c}X + \boldsymbol{d}) \tag{3.55}$$

where $\boldsymbol{c}, \boldsymbol{d}$ are the first rows of C, D. Now, writing $\boldsymbol{c} = (\boldsymbol{c}^{(1)} \quad \boldsymbol{c}^{(2)})$, $\boldsymbol{d} = (\boldsymbol{d}^{(1)} \quad \boldsymbol{d}^{(2)})$ and

$$X = \begin{pmatrix} T_l + R_l X_l {}^tR_l & S_l + R_l X_l \\ {}^tS_l + X_l {}^tR_l & X_l \end{pmatrix}, \tag{3.56}$$

$$Y = \begin{pmatrix} U_l & R_l \\ 0 & I \end{pmatrix} \begin{pmatrix} V_l & 0 \\ 0 & Y_l \end{pmatrix} \begin{pmatrix} {}^tU_l & 0 \\ {}^tR_l & I \end{pmatrix}, \tag{3.57}$$

see (2.15), we obtain

$$\begin{aligned} v_1(\gamma g)^{-1} &= \boldsymbol{c}^{(1)} U_l V_l {}^tU_l {}^t\boldsymbol{c}^{(1)} + (\boldsymbol{c}^{(1)} R_l + \boldsymbol{c}^{(2)}) Y_l {}^t(\boldsymbol{c}^{(1)} R_l + \boldsymbol{c}^{(2)}) \\ &\quad + (\boldsymbol{c}^{(1)}(T_l - S_l {}^tR_l) + \boldsymbol{c}^{(2)}\, {}^tS_l + \boldsymbol{d}^{(1)} - \boldsymbol{d}^{(2)}\, {}^tR_l)\, {}^tU_l^{-1} V_l^{-1} U_l^{-1} \\ &\qquad \times {}^t(\boldsymbol{c}^{(1)}(T_l - S_l {}^tR_l) + \boldsymbol{c}^{(2)}\, {}^tS_l + \boldsymbol{d}^{(1)} - \boldsymbol{d}^{(2)}\, {}^tR_l) \\ &\quad + (\boldsymbol{c}^{(1)}(S_l + R_l X_l) + \boldsymbol{c}^{(2)} X_l + \boldsymbol{d}^{(2)}) Y_l^{-1} \\ &\qquad \times {}^t(\boldsymbol{c}^{(1)}(S_l + R_l X_l) + \boldsymbol{c}^{(2)} X_l + \boldsymbol{d}^{(2)}). \end{aligned} \tag{3.58}$$

If $\boldsymbol{c}^{(1)} \neq 0$, then, since $U_l V_l {}^tU_l \in \mathcal{D}'_l$, we have

$$v_1(\gamma g)^{-1} \geq \boldsymbol{c}^{(1)} U_l V_l {}^tU_l {}^t\boldsymbol{c}^{(1)} \gg \boldsymbol{c}^{(1)} V_l {}^t\boldsymbol{c}^{(1)} \gg v_l \tag{3.59}$$

by the second part of Lemma 3.3. Since, for $l < n$,

$$\begin{pmatrix} I & X_l \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_l^{\frac{1}{2}} & 0 \\ 0 & {}^tY_l^{-\frac{1}{2}} \end{pmatrix} \in \mathcal{D}_{n-l}, \tag{3.60}$$

we have $v_{l+1} \gg 1$, see Proposition 3.2, and so $v_l \gg a_l$ by the hypothesis. For $l = n$, we directly have $v_n \gg a_n$ by hypothesis. Since also $v_1 \gg v_l$ by Lemma 3.3, we have $v_1 v_l \gg a_l^2$, so by taking a_l to be a sufficiently large constant, it follows that $v_1 \geq v_1(\gamma g)$.

For $l < n$, if $\boldsymbol{c}^{(1)} = 0$ but $(\boldsymbol{c}^{(2)} \quad \boldsymbol{d}^{(2)}) \neq 0$, then we have

$$v_1(\gamma g)^{-1} \geq \boldsymbol{c}^{(2)} Y_l {}^t\boldsymbol{c}^{(2)} + (\boldsymbol{c}^{(2)} X_l + \boldsymbol{d}^{(2)}) Y_l^{-1}\, {}^t(\boldsymbol{c}^{(2)} X_l + \boldsymbol{d}^{(2)}) \geq v_{l+1}(g)^{-1} \tag{3.61}$$

since $g_l(g) \in \mathcal{D}_{n-l}$. We have $v_{l+1}^{-1} \geq a_l {v_l}^{-1} \gg a_l {v_1}^{-1}$, so $v_{l+1}^{-1} \geq {v_1}^{-1}$ for a_l sufficiently large, and it follows that $v_1 \geq v_1(\gamma g)$.

Now, if $l = n$ or if $\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}$, and $\boldsymbol{d}^{(2)}$ are all 0, then we have $\boldsymbol{d}^{(1)} \neq 0$ and

$$v_1(\gamma g)^{-1} = \boldsymbol{d}^{(1)}\, {}^tU_l^{-1} V_l^{-1} U_l^{-1}\, {}^t\boldsymbol{d}^{(1)} \geq {v_1}^{-1} \tag{3.62}$$

as $U_l V_l {}^t U_l \in \mathcal{D}'_l$. We have verified that for any $\gamma \in \Gamma$, $v_1 \le v_1(\gamma g)$, which is the first condition defining the fundamental domain $\mathcal{D}_n$.

Restricting to $\gamma \in \Gamma_1$, which fixes $v_1(g)$, the same argument as above shows that $v_2(g) \ge v_2(\gamma g)$ for all $\gamma \in \Gamma_1$. Continuing this way, we find that the v_j, $1 \le j \le l$ are all maximal (over $\Gamma_{j,2}$), and so, by the construction of $\mathcal{D}_n$, there is a $\gamma \in \Gamma_l$ with the form

$$\gamma = \begin{pmatrix} A & B \\ 0 & {}^t\!A^{-1} \end{pmatrix}, \tag{3.63}$$

where A is upper-triangular unipotent (so $\gamma \in \Gamma_l$ for all l) such that $\gamma g \in \mathcal{D}_n$. $\square$

4 Proof of the main theorem

In the following subsection we gather some technical lemmas regarding the height function needed in the proof of Theorem 1.2, see Section 4.2. This height function is motivated by the following corollary from [11].

Corollary 4.1. *For a Schwartz function $f \in \mathcal{S}(\mathbb{R}^n)$ and $(h, g) \in \tilde{\mathcal{D}}$, and $A > 0$, we have*

$$\Theta_f(h, g) \ll_{f,A} (\det Y)^{\frac{1}{4}} (1 + \boldsymbol{x} Y {}^t\boldsymbol{x})^{-A} \tag{4.1}$$

where

$$g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} Y^{\frac{1}{2}} & 0 \\ 0 & {}^tY^{-\frac{1}{2}} \end{pmatrix} k(Q) \tag{4.2}$$

and $h = (\boldsymbol{x}, \boldsymbol{y}, t)$.

We remark that in [11] this is obtained as a consequence of full asymptotics of the theta function in the various cuspidal regions. We also remark that in [11] we use a slightly different fundamental domain, however an examination of the proof there shows that the fundamental domain can be replaced by any set satisfying the conclusions of Proposition 3.2. Finally, in [11] we use the term $\boldsymbol{x} V {}^t\boldsymbol{x}$ as opposed to $\boldsymbol{x} Y {}^t\boldsymbol{x}$ in (4.1). Up to constants, these are interchangeable by Proposition 3.2 and we only prefer $\boldsymbol{x} Y {}^t\boldsymbol{x}$ for its transformation properties, see the proof of Lemma 4.4.

4.1 Heights and volumes.

For a fixed $A > 0$ sufficiently large depending only on n, we define the function $D : \tilde{\Gamma} \backslash (H \rtimes G) \to \mathbb{R}_{>0}$ by

$$D(\tilde{\Gamma}(h, g)) = \det Y(\gamma g)(1 + \boldsymbol{x}(u h_\gamma h^{\gamma^{-1}}) Y(\gamma g) {}^t\boldsymbol{x}(u h_\gamma h^{\gamma^{-1}}))^{-A} \tag{4.3}$$

where $(uh_\gamma, \gamma) \in \tilde{\Gamma}$ is so that $(uh_\gamma, \gamma)(h, g) \in \tilde{\mathcal{D}}$. Here we write $h \in H$ as

$$h = (\boldsymbol{x}(h), \boldsymbol{y}(h), t(h)).$$

For completeness, in case there are more than one $(uh_\gamma, \gamma) \in \tilde{\Gamma}$ such that $(uh_\gamma, \gamma)(h, g) \in \tilde{\mathcal{D}}$, then we define $D(\tilde{\Gamma}(h, g))$ to be the largest of the finite number of values (4.3). This point is not essential as these values are within constant multiples of each other; see the argument in Lemma 4.4 for how this can be proved.

We begin by analyzing the growth of the height function. We let $\tilde{\mu}$ denote the Haar probability measure on $\tilde{\Gamma}\backslash(H \rtimes G)$, which is μ, the Haar probability measure on $\Gamma\backslash G$, times the Lebesgue measure on the entries of $h = (\boldsymbol{x}, \boldsymbol{y}, t)$.

Lemma 4.2. *For $R \geq 1$ we have*

$$\tilde{\mu}(\{\tilde{\Gamma}(h, g) \in \tilde{\Gamma}\backslash(H \rtimes G) : D(\tilde{\Gamma}(h, g)) \geq R\}) \ll R^{-\frac{n+2}{2}} \tag{4.4}$$

with the implied constant depending only on n.

Proof. We recall that $g \in \mathcal{D}_n$ is written as

$$g = \begin{pmatrix} U & X\,{}^tU^{-1} \\ 0 & {}^tU^{-1} \end{pmatrix} \begin{pmatrix} V^{\frac{1}{2}} & 0 \\ 0 & V^{-\frac{1}{2}} \end{pmatrix} k(Q) \tag{4.5}$$

for U upper-triangular unipotent, X symmetric, $Q \in \mathrm{U}(n)$, and

$$V = V(g) = \begin{pmatrix} v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_n \end{pmatrix} \tag{4.6}$$

positive diagonal. The Haar measure μ on G is then proportional to Lebesgue measure with respect to the entries of X and the off-diagonal entries of U, $\mathrm{U}(n)$-Haar measure on Q, and the measure given by

$$v_1^{-n-1} v_2^{-n} \cdots v_n^{-2} \mathrm{d}v_1 \mathrm{d}v_2 \cdots \mathrm{d}v_n \tag{4.7}$$

on V.

By Proposition 3.2, we observe that the set in (4.4) is contained in the set of (h, g) satisfying $v_j \geq c v_{j+1}$ for all $1 \leq j < n$ and some $c > 0$ in addition to $\det Y \geq R$ and $\boldsymbol{x} Y\,{}^t\boldsymbol{x} \leq R^{-\frac{1}{A}} (\det Y)^{\frac{1}{A}}$. Moreover, the variables $\boldsymbol{x}, \boldsymbol{y}, t$ as well as U, X are constrained to compact sets, and so the measure of the set (4.4) is

$$\ll R^{-\epsilon} \int_{\substack{v_j \geq c v_{j+1} \\ v_1 \cdots v_n \geq R}} \cdots \int v_1^{-n-\frac{3}{2}+\epsilon} v_2^{-n-\frac{1}{2}+\epsilon} \cdots v_n^{-\frac{5}{2}+\epsilon} \mathrm{d}v_1 \mathrm{d}v_2 \cdots \mathrm{d}v_n, \tag{4.8}$$

where $\epsilon = \frac{n}{2A}$.

Changing variables $v_j = \exp(u_j)$, the integral in (4.8) is

$$R^{-\epsilon} \int\cdots\int_{\substack{u_j-u_{j+1}\geq \log c\\ u_1+\cdots+u_n\geq \log R}} \exp\Big(-\Big(n+\frac{1}{2}-\epsilon\Big)u_1 - \Big(n-\frac{1}{2}-\epsilon\Big)u_2 \\ \vdots \\ -\Big(\frac{3}{2}-\epsilon\Big)u_n\Big) \mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_n. \tag{4.9}$$

We now make the linear change of variables $s_j = u_j - u_{j+1}$ for $j < n$ and $s_n = u_1 + \cdots + u_n$. This transformation has determinant n and its inverse is given by

$$u_j = -\frac{1}{n}\sum_{1\leq i<j} is_i + \frac{1}{n}\sum_{j\leq i<n}(n-i)s_i + \frac{1}{n}s_n. \tag{4.10}$$

We find that the exponent in (4.9) is then

$$-\sum_{1\leq j\leq n}(n-j+\frac{3}{2}-\epsilon)u_j = -\Big(\frac{n+2}{2}-\epsilon\Big)s_n - \sum_{1\leq j<n}\frac{j(n-j)}{2}s_j. \tag{4.11}$$

As $\frac{j(n-j)}{2} > 0$ for $j < n$, the bound (4.4) follows. □

Lemma 4.4 below contains a key estimate, establishing a kind of 'uniform continuity' for $\log D$. The proof of this lemma is the primary motivation for defining our new fundamental domain and studying the shape of its cusp in Sections 3.1 and 3.2. For the proof, we first establish a similar kind of 'uniform continuity' for the functions $v_l(\Gamma_l g)$ and $v_{l+1}(\Gamma_l g)$ that are essential to Section 3.2.

Lemma 4.3. *Let $g, g_0 \in G$ with $||g_0 - I|| \leq 1$, then*

$$v_l(g) \asymp v_l(gg_0), \quad v_l(\Gamma_l g) \asymp v_l(\Gamma_l gg_0), \quad v_{l+1}(\Gamma_l g) \asymp v_{l+1}(\Gamma_l gg_0) \tag{4.12}$$

for all $1 \leq l \leq n$ with implied constants depending only on n.

Proof. We first note that we may in fact work with $||I - g_0|| \leq \epsilon$ as then the statement would follow by repeated application of the estimates. In fact, we may assume $||I - g_0^{-1}|| \leq \epsilon$ as well. Now write

$$g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix}\begin{pmatrix} Y^{\frac{1}{2}} & 0 \\ 0 & {}^tY^{-\frac{1}{2}} \end{pmatrix}\begin{pmatrix} R & -S \\ S & R \end{pmatrix}, \tag{4.13}$$

with $R + \mathrm{i}S \in \mathrm{U}(n)$, so in particular $R\,{}^tR + S\,{}^tS = I$. With $g_0 = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, we have from (2.10) that

$$Y(gg_0)^{-1} = {}^tY^{-\frac{1}{2}}\big(SA\,{}^tA\,{}^tS + RC\,{}^tA\,{}^tS + SA\,{}^tC\,{}^tR + RC\,{}^tCR \\ + SB\,{}^tB\,{}^tS + RD\,{}^tB\,{}^tS + SB\,{}^tD\,{}^tR + RD\,{}^tD\,{}^tR\big)Y^{-\frac{1}{2}}. \tag{4.14}$$

As $||g_0 - I|| \le \epsilon$, we have

$$ {}^tY(gg_0)^{-\frac{1}{2}} = {}^tY^{-\frac{1}{2}}(I + O(\epsilon)). \tag{4.15} $$

On the other hand, letting $\boldsymbol{y}_j$ and $\boldsymbol{y}'_J$ denote the rows of ${}^tY^{-\frac{1}{2}}$ and ${}^tY(gg_0)^{-\frac{1}{2}}$, we have

$$ v_1(g)^{-\frac{1}{2}} = ||\boldsymbol{y}_1||, \quad v_1(gg_0)^{-\frac{1}{2}} = ||\boldsymbol{y}'_1|| \tag{4.16} $$

and for $2 \le l \le n$,

$$ v_l(g)^{-\frac{1}{2}} = \frac{||\boldsymbol{y}_1 \wedge \cdots \wedge \boldsymbol{y}_l||}{||\boldsymbol{y}_1 \wedge \cdots \wedge \boldsymbol{y}_{l-1}||}, \quad v_l(gg_0)^{-\frac{1}{2}} = \frac{||\boldsymbol{y}'_1 \wedge \cdots \wedge \boldsymbol{y}'_l||}{||\boldsymbol{y}'_1 \wedge \cdots \wedge \boldsymbol{y}'_{l-1}||}, \tag{4.17} $$

and so $v_l(g) \asymp v_l(gg_0)$ follows.

Now let $\gamma \in \Gamma_l$ be so that $v_l(\Gamma_l g) = v_l(\gamma g)$. We have

$$ v_l(\Gamma_l gg_0) \le v_l(\gamma gg_0) \ll v_l(\gamma g) = v_l(\Gamma_l g), \tag{4.18} $$

and the reverse bound follows by switching the roles of g and gg_0, and using $||g_0^{-1} - I|| \le \epsilon$. The final estimate in (4.12) is proved in the same way. □

Lemma 4.4. *If* $(h, g), (h_0, g_0) \in G$ *with* $||g_0 - I|| \le 1$ *and* $h_0 = (\boldsymbol{x}_0, \boldsymbol{y}_0, t_0)$ *satisfies* $||\boldsymbol{x}_0||, ||\boldsymbol{y}_0|| \le 1$, *then*

$$ D(\tilde{\Gamma}(h, g)) \asymp D(\tilde{\Gamma}(h, g)(h_0, g_0)). \tag{4.19} $$

Proof. We observe as in Lemma 4.3, we may in fact assume

$$ ||g_0 - I|| \le \epsilon, \quad ||\boldsymbol{x}_0|| \le \epsilon, \quad \text{and} \quad ||\boldsymbol{y}_0|| \le \epsilon. \tag{4.20} $$

Moreover, it suffices to show that $D(\tilde{\Gamma}(h, g)(h_0, g_0)) \gg D(\tilde{\Gamma}(h, g))$ as the other inequality follows from switching (h, g) and $(h, g)(h_0, g_0)$ as we may assume in addition that $(h_0, g_0)^{-1} = (h_0^{-g_0}, g_0^{-1})$ also satisfies (4.20).

Now let us suppose that $(h, g) \in \tilde{\mathcal{D}}$ so that

$$ D(\tilde{\Gamma}(h, g)) = (\det Y(g))(1 + \boldsymbol{x}(h)Y(g)\,{}^t\boldsymbol{x}(h))^{-A}. \tag{4.21} $$

Let $1 \le l \le n$ be the largest index such that $v_l(g) \ge a v_{l+1}(g)$ (or $v_n(g) \ge a$ when $l = n$) where a is a constant determined by the constants in Proposition 3.5 and Lemma 4.3. If no such l exists, then we have $v_j(g) \asymp 1$ for all j, and Lemma 4.3 implies that $v_j(gg_0) \asymp 1$ as well. The bounds

$$ D(\tilde{\Gamma}(h, g)(h_0, g_0)) \gg 1 \gg D(\tilde{\Gamma}(h, g)) \tag{4.22} $$

then follow immediately.

Now assuming that such a maximal l exists, we have that $v_j(g) \asymp 1$ for all $j > l$. For these j, Lemma 4.3 then implies that $v_j(gg_0) \asymp 1$, and it follows that $v_j(\gamma gg_0) \asymp 1$ for $\gamma \in \Gamma_l$ such that $g_l(\gamma gg_0) \in \mathcal{D}_{n-l}$, see (2.16). By Lemma 3.4, we have $v_l(\Gamma_l g) \gg v_l(g)$, and so

$$v_l(\Gamma_l g) \gg a v_{l+1}(g) = a v_{l+1}(\Gamma_l g) \tag{4.23}$$

since $g_l(g) \in \mathcal{D}_{n-l}$. Via Lemma 4.3, this implies that $v_l(\Gamma_l gg_0) \gg a v_{l+1}(\Gamma_l gg_0)$, so a can be chosen large enough so that gg_0 satisfies the hypotheses of Proposition 3.5, and we let $\gamma \in \Gamma_l$ be so that $\gamma gg_0 \in \mathcal{D}$.

We write

$$\gamma = \begin{pmatrix} A_1 & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & 0 \\ 0 & * & * & * \end{pmatrix}, \tag{4.24}$$

where $A_1 \in \mathrm{GL}(l, \mathbb{Z})$. From the estimates above, we have

$$\begin{aligned} \det Y(\gamma gg_0) &\asymp \det U_l(\gamma gg_0) V_l(\gamma gg_0)\, {}^t U_l(\gamma gg_0) \\ &= \det U_l(gg_0) V_l(gg_0)\, {}^t U_l(gg_0) \\ &\asymp \det U_l(g) V_l(g)\, {}^t U_l(g) \asymp \det Y(g), \end{aligned} \tag{4.25}$$

where the equality follows from the fact that $\gamma \in \Gamma_l$ normalizes the first matrix in (2.15) and $\det A_1 = \pm 1$.

It now remains to consider the factors $1 + \boldsymbol{x}(*) Y(*)\, {}^t\boldsymbol{x}(*)$ in the definition of the height function D. Let $u = (\boldsymbol{m}, \boldsymbol{n}, 0)$ with $\boldsymbol{m}, \boldsymbol{n} \in \mathbb{Z}^n$ be so that $(uh_\gamma, \gamma)(h, g)(h_0, g_0) \in \tilde{\mathcal{D}}$. Recalling the definition of $h_\gamma = (\boldsymbol{r}, \boldsymbol{s}, 0)$ following (3.1), we have that $\boldsymbol{r}^{(1)} = 0$ where $\boldsymbol{r} = (\boldsymbol{r}^{(1)} \quad \boldsymbol{r}^{(2)})$. Moreover, writing $\boldsymbol{x} = (\boldsymbol{x}^{(1)} \quad \boldsymbol{x}^{(2)})$, we have $\boldsymbol{x}^{(1)}((hh_0^{g^{-1}})^{\gamma^{-1}}) = \boldsymbol{x}^{(1)}(hh_0^{g^{-1}}) A_1^{-1}$. Using Proposition 3.2 together with the fact that u minimizes the absolute values of the entries of $\boldsymbol{x}(uh_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}})$, we have

$$\begin{aligned} &1 + \boldsymbol{x}(uh_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}) Y(\gamma gg_0)\, {}^t\boldsymbol{x}(uh_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}) \\ &\qquad \ll 1 + \boldsymbol{x}(h_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}) Y(\gamma gg_0)\, {}^t\boldsymbol{x}(h_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}), \end{aligned} \tag{4.26}$$

and from the estimates above on the $v_j(\gamma gg_0)$ for $j > l$, we have

$$\begin{aligned} &1 + \boldsymbol{x}(h_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}) Y(\gamma gg_0)\, {}^t\boldsymbol{x}(h_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}) \\ &\quad \asymp 1 + \boldsymbol{x}^{(1)}(h_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}) U_l(\gamma gg_0) V_l(\gamma gg_0)\, {}^t U_l(\gamma gg_0)\, {}^t\boldsymbol{x}(h_\gamma (hh_0^{g^{-1}})^{\gamma^{-1}}). \end{aligned} \tag{4.27}$$

Using the expressions for h_γ, $(hh_0^{g^{-1}})^{\gamma^{-1}}$, and that

$$U_l(\gamma gg_0) V_l(\gamma gg_0)\, {}^t U_l(\gamma gg_0) = A_1 U_l(gg_0) V_l(gg_0)\, {}^t U_l(gg_0)\, {}^t A_1, \tag{4.28}$$

the right side of (4.27) is equal to

$$1 + \boldsymbol{x}^{(1)}(hh_0^{g^{-1}})U_l(gg_0)V_l(gg_0)\,{}^tU_l(gg_0)\,{}^t\boldsymbol{x}^{(1)}(hh_0^{g^{-1}}) \asymp 1 + \boldsymbol{x}(hh_0^{g^{-1}})Y(gg_0)\,{}^t\boldsymbol{x}(hh_0^{g^{-1}}) \tag{4.29}$$

by the above bounds on $v_j(gg_0)$ for $j > l$.

Recalling that

$$g = \begin{pmatrix} I & X(g) \\ 0 & I \end{pmatrix} \begin{pmatrix} Y(g)^{\frac{1}{2}} & 0 \\ 0 & {}^tY(g)^{-\frac{1}{2}} \end{pmatrix} k(g) \tag{4.30}$$

with $k(g) \in K = G \cap \mathrm{SO}(2n, \mathbb{R})$, we set $h_0' = h_0^{k(g)^{-1}}$ and note that

$$||\boldsymbol{x}(h_0')||^2 + ||\boldsymbol{y}(h_0')||^2 = ||\boldsymbol{x}(h_0)||^2 + ||\boldsymbol{y}(h_0)||^2. \tag{4.31}$$

Since $Y(gg_0) = Y(g)^{\frac{1}{2}}Y(k(g)g_0)\,{}^tY(g)^{\frac{1}{2}}$ and $\boldsymbol{x}(hh_0^{g^{-1}}) = \boldsymbol{x}(h) + \boldsymbol{x}(h_0')Y(g)^{-\frac{1}{2}}$, the right side of (4.29) is equal to

$$1 + \boldsymbol{x}(h)Y(g)^{\frac{1}{2}}Y(k(g)g_0)\,{}^tY(g)^{-\frac{1}{2}}\,{}^t\boldsymbol{x}(h) + 2\boldsymbol{x}(h)Y(g)^{\frac{1}{2}}Y(k(g)g_0)\,{}^t\boldsymbol{x}(h_0') + \boldsymbol{x}(h_0')Y(k(g)g_0)\,{}^t\boldsymbol{x}(h_0'). \tag{4.32}$$

We have that $||g_0 - I|| \leq \epsilon$ implies $Y(k(g)g_0) = I + O(\epsilon)$ as in (4.14), so if (4.31) is at most ϵ^2 as well, with ϵ sufficiently small, then (4.32) is

$$\asymp 1 + \boldsymbol{x}(h)Y(g)\,{}^t\boldsymbol{x}(h), \tag{4.33}$$

where we have used

$$2|\boldsymbol{x}(h)Y(g)^{\frac{1}{2}}Y(k(g)g_0)\,{}^t\boldsymbol{x}(h_0')| \leq \sqrt{\boldsymbol{x}(h_0')Y(k(g)g_0)^2\,{}^t\boldsymbol{x}(h_0')}(\boldsymbol{x}(h)Y(g)\,{}^t\boldsymbol{x}(h) + 1) \ll \epsilon(\boldsymbol{x}(h)Y(g)\,{}^t\boldsymbol{x}(h) + 1) \tag{4.34}$$

to bound the third term in (4.32). The bound $D(\tilde{\Gamma}(h, g)(h_0, g_0)) \gg D(\tilde{\Gamma}(h, g)$ now follows. □

4.2 Proof of Theorem 1.2.

We recall the following lemma from [11].

Lemma 4.5. *There exists a smooth, compactly supported function*

$$f_1 : \mathbb{R} \to \mathbb{R}_{\geq 0}$$

such that

$$\chi_1(x) = \sum_{j \geq 0}(f_1(2^jx) + f_1(2^j(1 - x))), \tag{4.35}$$

where χ_1 is the indicator function of the open unit interval $(0, 1)$.

Now, following the method of [11], we define for a subset $S \subset \{1,\ldots,n\}$ and $\boldsymbol{j} = (j_1,\ldots,j_n) \in \mathbb{Z}^n$ with $j_i \geq 0$,

$$g_{\boldsymbol{j},S} = \begin{pmatrix} A_{\boldsymbol{j}}E_S & 0 \\ 0 & A_{\boldsymbol{j}}^{-1}E_S \end{pmatrix} \in G \tag{4.36}$$

where E_S is diagonal with (i,i) entry -1 if $i \in S$, $+1$ if $i \notin S$, and

$$A_{\boldsymbol{j}} = \begin{pmatrix} 2^{j_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 2^{j_n} \end{pmatrix}. \tag{4.37}$$

We also set $h_S = (\boldsymbol{x}_S, 0, 0) \in H$ where $\boldsymbol{x}_S$ has ith entry -1 if $i \in S$ and 0 if $i \notin S$.

As in [11], we have

$$\chi_{\mathcal{B}}(\boldsymbol{x}) = \sum_{\boldsymbol{j}\geq 0} \sum_{S\subset\{1,\ldots,n\}} f_n((\boldsymbol{x}B^{-1} + \boldsymbol{x}_S)A_{\boldsymbol{j}}E_S), \tag{4.38}$$

where $\chi_{\mathcal{B}}$ is the indicator function of the rectangular box $\mathcal{B} = (0,b_1) \times \cdots \times (0,b_n)$, B is the diagonal matrix with entries $b_1,\ldots,b_n$,

$$f_n(x_1,\ldots,x_n) = \prod_{1\leq j\leq n} f_1(x_j), \tag{4.39}$$

and the sums are over $\boldsymbol{j} \in \mathbb{Z}^n$ with nonnegative entries.

Let $\psi : [0,\infty) \to [1,\infty)$ be an increasing function. Then for $C > 0$ we define $\mathcal{G}_{\boldsymbol{j}}(\psi, C)$ to be the set of $\tilde{\Gamma}(h,g) \in \tilde{\Gamma}\backslash(H \rtimes G)$ such that

$$D\left(\tilde{\Gamma}(h,g)\left(1, \begin{pmatrix} \mathrm{e}^{-s}I & 0 \\ 0 & \mathrm{e}^{s}I \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S})\right)^{\frac{1}{4}} \leq C\psi(s) \tag{4.40}$$

for all $S \subset \{1,\ldots,n\}$ and $s \geq 1$.

Lemma 4.6. *Suppose that ψ satisfies*

$$\int_0^\infty \psi(x)^{-(2n+4)}\mathrm{d}x \leq C_\psi \tag{4.41}$$

for some $C_\psi \geq 1$. Then

$$\tilde{\mu}(\tilde{\Gamma}\backslash(H \rtimes G) - \mathcal{G}_{\boldsymbol{j}}(\psi, C)) \ll C_\psi C^{-(2n+4)} 2^{j_1+\cdots+j_n}. \tag{4.42}$$

Proof. Suppose that $\tilde{\Gamma}(h,g) \notin \mathcal{G}_{\boldsymbol{j}}(\psi, C)$, so there exists $S \subset \{1,\ldots,n\}$ and $s \geq 1$ such that

$$D\left(\tilde{\Gamma}(h,g)\left(1, \begin{pmatrix} \mathrm{e}^{-s}I & 0 \\ 0 & \mathrm{e}^{s}I \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S})\right)^{\frac{1}{4}} \geq C\psi(s). \tag{4.43}$$

We let k be a nonnegative integer such that

$$\frac{k}{K_{\boldsymbol{j}}} \le s < \frac{k+1}{K_{\boldsymbol{j}}}, \tag{4.44}$$

where $K_{\boldsymbol{j}} = K2^{j_1+\cdots+j_n}$ with K a constant to be determined. We have

$$\left(1, \begin{pmatrix} \mathrm{e}^{-s}I & 0 \\ 0 & \mathrm{e}^{s}I \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S}) = \left(1, \begin{pmatrix} \mathrm{e}^{-\frac{k}{K_{\boldsymbol{j}}}}I & 0 \\ 0 & \mathrm{e}^{\frac{k}{K_{\boldsymbol{j}}}}I \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S})(h_1, g_1), \tag{4.45}$$

where, with $s' = s - \frac{k}{K_{\boldsymbol{j}}}$,

$$h_1 = ((\mathrm{e}^{s'} - 1)\boldsymbol{x}_S A_{\boldsymbol{j}} E_S, 0, 0), \quad g_1 = \begin{pmatrix} \mathrm{e}^{-s'}I & 0 \\ 0 & \mathrm{e}^{s'}I \end{pmatrix}. \tag{4.46}$$

As $|s'| \le K_{\boldsymbol{j}}^{-1}$, we can make K sufficiently large so that (h_1, g_1) satisfies the conditions of Lemma 4.4. From this and the fact that ψ is increasing, we have that

$$D\left(\tilde{\Gamma}(h,g)\left(1, \begin{pmatrix} \mathrm{e}^{-\frac{k}{K_{\boldsymbol{j}}}}I & 0 \\ 0 & \mathrm{e}^{\frac{k}{K_{\boldsymbol{j}}}}I \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S})\right)^{\frac{1}{4}} \gg C\psi\left(\frac{k}{K_{\boldsymbol{j}}}\right). \tag{4.47}$$

By Lemma 4.2 and the fact that right multiplication is volume preserving, we have that the set of $\tilde{\Gamma}(h,g)$ satisfying (4.47) has $\tilde{\mu}$-volume bounded by a constant times

$$C^{-2n-4}\psi\left(\frac{k}{K_{\boldsymbol{j}}}\right)^{-2n-4}. \tag{4.48}$$

Bounding the volume of the set $\tilde{\Gamma}\backslash(H \rtimes G) - \mathcal{G}_{\boldsymbol{j}}(\psi, C)$ by summing (4.48) over $S \subset \{1, \ldots, n\}$ and nonnegative $k \in \mathbb{Z}$, we obtain the bound

$$C^{-(2n+4)}\sum_{k\ge 0}\psi\left(\frac{k}{K_{\boldsymbol{j}}}\right)^{-(2n+4)} \ll C^{-(2n+4)}\left(\psi(0) + \int_0^\infty \psi\left(\frac{x}{K_{\boldsymbol{j}}}\right)^{-(2n+4)}\mathrm{d}x\right) \tag{4.49}$$

as $\psi(x)$ is increasing. The bound (4.42) follows by changing variables. □

We now proceed to the proof of Theorem 1.2.

Proof of Theorem 1.2. From (4.38) we express $\theta_{\mathcal{B}}(M, X, \boldsymbol{x}, \boldsymbol{y})$ as

$$\sum_{S\subset\{1,\ldots,n\}}\sum_{\boldsymbol{j}\ge 0}\sum_{\boldsymbol{m}\in\mathbb{Z}^n} f_n\left(\frac{1}{M}(\boldsymbol{m} + \boldsymbol{x} + M\boldsymbol{x}_S B)B^{-1}E_S A_{\boldsymbol{j}}\right)\mathrm{e}\left(\frac{1}{2}\boldsymbol{m}X{}^t\boldsymbol{m} + \boldsymbol{m}{}^t\boldsymbol{y}\right). \tag{4.50}$$

We break the sum in (4.50) into terms $\boldsymbol{j}$ such that $2^{j_i}b_{j_i}^{-1} \le M$ for all i and terms $\boldsymbol{j}$ such that $2^{j_i}b_{j_i}^{-1} > M$ for some i. Using (2.8), we write the first part as

$$\mathrm{e}\left(\frac{1}{2}\boldsymbol{x}X{}^t\boldsymbol{x}\right)M^{\frac{n}{2}}(\det B)^{\frac{1}{2}}\sum_{\substack{\boldsymbol{j}\ge 0\\ 2^{j_i}b_{j_i}^{-1}\le M}} 2^{-\frac{1}{2}(j_1+\cdots+j_n)}\Theta_{f_n}((h, g(MB, X))(h_S, g_{\boldsymbol{j},S})), \tag{4.51}$$

where $h = (\boldsymbol{x}, \boldsymbol{y} - \boldsymbol{x}X, 0)$ and

$$g(MB, X) = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} \frac{1}{M}B^{-1} & 0 \\ 0 & MB \end{pmatrix}. \tag{4.52}$$

Bounding this is the main work of the proof, but we first bound the contribution of the terms $\boldsymbol{j}$ with a large index.

Suppose that $L \subset \{1, \ldots, n\}$ is not empty and that $2^{j_l} > b_{j_l}M$ for all $l \in L$. Then the compact support of f_1 implies that the sum over $\boldsymbol{m}^{(L)}$, the vector of entries of $\boldsymbol{m}$ with index in L, has a bounded number of terms. We write

$$\boldsymbol{m}X\,{}^t\boldsymbol{m} = \boldsymbol{m}^{(L)}X^{(L,L)}\,{}^t\boldsymbol{m}^{(L)} + 2\boldsymbol{m}^{(L)}X^{(L,L')}\,{}^t\boldsymbol{m}^{(L')} + \boldsymbol{m}^{(L')}X^{(L',L')}\,{}^t\boldsymbol{m}^{(L')}, \tag{4.53}$$

where L' is the complement of L, and $X^{(L_1,L_2)}$ is the matrix of entries of X with row and column indices in L_1 and L_2 respectively. We have (4.39) that $f_n(\frac{1}{M}(\boldsymbol{m} + \boldsymbol{x} + M\boldsymbol{x}_S B)B^{-1}E_S A_{\boldsymbol{j}})$ factors as

$$\begin{aligned} &f_{\#L}\Big(\frac{1}{M}(\boldsymbol{m}^{(L)} + \boldsymbol{x}^{(L)} + M\boldsymbol{x}_S^{(L)})(B^{(L,L)})^{-1}E_S^{(L,L)}A_{\boldsymbol{j}}^{(L,L)}\Big) \\ &\qquad \times f_{\#L'}\Big(\frac{1}{M}(\boldsymbol{m}^{(L')} + \boldsymbol{x}^{(L')} + M\boldsymbol{x}_S^{(L')})(B^{(L',L')})^{-1}E_S^{(L',L')}A_{\boldsymbol{j}}^{(L',L')}\Big), \end{aligned} \tag{4.54}$$

and so, by inclusion-exclusion and the boundedness of $f_{\#L}$, the term $\boldsymbol{j}$ of (4.50) with $\boldsymbol{j}_l > b_{j_i}M$ for some i is at most a constant times

$$\sum_{\substack{L\subset\{1,\ldots,n\}\\ L\neq\emptyset}} \sum_{S\subset L} \sum_{\boldsymbol{m}^{(L)}} |\theta_{\mathcal{B}^{(L')}}(M, X^{L',L'}, \boldsymbol{x}^{(L')}, \boldsymbol{y}^{(L')} + \boldsymbol{m}^{(L)}X^{(L,L')})|, \tag{4.55}$$

where the sum over $\boldsymbol{m}^{(L)}$ has a bounded number of terms, $\mathcal{B}^{(L')}$ is the edge of $\mathcal{B}$ associated to L', and we have used the decomposition (4.38) to express $\theta_{\mathcal{B}^{(L')}}(M, X^{L',L'}, \boldsymbol{x}^{(L')}, \boldsymbol{y}^{(L')} + \boldsymbol{m}^{(L)}X^{(L,L')})$ as

$$\begin{aligned} &\sum_{S'\subset L'} \sum_{\boldsymbol{j}_{L'}} \sum_{\boldsymbol{m}_{L'}} f_{\#L'}\Big(\frac{1}{M}(\boldsymbol{m}^{(L')} + \boldsymbol{x}^{(L')} + M\boldsymbol{x}_S^{(L')})(B^{(L',L')})^{-1}E_S^{(L',L')}A_{\boldsymbol{j}}^{(L',L')}\Big) \\ &\qquad \times \mathrm{e}\Big(\frac{1}{2}\boldsymbol{m}^{(L')}X^{(L',L')}\,{}^t\boldsymbol{m}^{(L')} + \boldsymbol{m}^{(L')}\,{}^t(\boldsymbol{y}^{(L')} + \boldsymbol{m}^{(L)}X^{(L,L')})\Big). \end{aligned} \tag{4.56}$$

When $L = \{1, \ldots, n\}$, the corresponding part of (4.55) is clearly bounded. For any other L, we may apply Theorem 1.1 (emphasizing the importance of the uniformity in $\boldsymbol{y}$) to conclude for any $\epsilon > 0$, there are full measure sets $\mathcal{X}^{(n-\#L)} = \mathcal{X}^{(n-\#L)}(\epsilon)$ such that if $X^{(L',L')} \in \mathcal{X}^{(n-\#L)}$, the corresponding part of (4.55) is $\ll M^{\frac{n-\#L}{2}+\epsilon}$ for any $\epsilon > 0$. It follows that (4.55) is $\ll M^{\frac{n}{2}}$ assuming that X is such that $X^{(L',L')} \in \mathcal{X}^{(n-\#L)}$ for all nonempty $L \subset \{1, \ldots, n\}$.

We now return to (4.51). We let $\mathcal{X}_{\boldsymbol{j}}(\psi, C)$ be the set of $(X, \boldsymbol{y})$ with all entries in the interval $(-\frac{1}{2}, \frac{1}{2}]$ such that there exist $\boldsymbol{u} \in (-\frac{1}{2}, \frac{1}{2})^n$, $A \in \mathrm{GL}(n, \mathbb{R})$ and $T \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ satisfying

$$\sup_{B\in\mathcal{K}_1} \|((BA)^{-1} - I)A_{\boldsymbol{j}}\| \le \epsilon, \tag{4.57}$$

$\|T\| \le \epsilon$, and

$$\tilde{\Gamma}\left((\boldsymbol{u}, \boldsymbol{y} - \boldsymbol{u}X, 0), \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & {}^t\!A^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ T & I \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S}) \in \mathcal{G}_{\boldsymbol{j}}(\psi, C). \tag{4.58}$$

Here we let $\epsilon > 0$ be a sufficiently small constant, $\mathcal{G}_{\boldsymbol{j}}(\psi, C)$ is defined in (4.40), and $\mathcal{K}_1 \subset \mathbb{R}^n_{>0}$ is the compact subset from the statement of Theorem 1.2 identified with the compact subset of positive diagonal matrices B in the obvious way. We then set $\mathcal{X}(\psi)$ to be the set of $(X, \boldsymbol{y}) \in \mathbb{R}^{n\times n}_{\mathrm{sym}} \times \mathbb{R}^n$ such that

$$\begin{aligned}(X + R, \boldsymbol{y}R + \boldsymbol{s}_R + \boldsymbol{s}) \in & \bigcup_{C>0} \bigcap_{\boldsymbol{j}\ge 0} \mathcal{X}_{\boldsymbol{j}}(\psi, C2^{a(j_1+\cdots+j_n)}) \\ & \cap \bigcap_{\substack{L\subset\{1,\ldots,n\}\\ L\neq\emptyset}} \{(X_1, \boldsymbol{y}_1) \in \mathbb{R}^{n\times n} \times \mathbb{R}^n : X_1^{(L',L')} \in \mathcal{X}^{(n-\#L)}\}\end{aligned} \tag{4.59}$$

for some $(R, \boldsymbol{s}) \in \mathbb{Z}^{n\times n} \times \mathbb{Z}^n$, where $\boldsymbol{s}_R \in \mathbb{R}^n$ has entries 0 or $\frac{1}{2}$ depending on whether the corresponding diagonal entry of R is even or odd, and $a > 0$ is a constant to be determined.

We first verify that $\mathcal{X}(\psi)$ has full measure, noting that it is enough to show that

$$\bigcup_{C>0} \bigcap_{\boldsymbol{j}\ge 0} \mathcal{X}_{\boldsymbol{j}}(\psi, C2^{a(j_1+\cdots+j_n)}) \tag{4.60}$$

has full measure in the subset $\mathcal{X}_0$ of $\mathbb{R}^{n\times n}_{\mathrm{sym}} \times \mathbb{R}^n$ having all entries in the interval $(-\frac{1}{2}, \frac{1}{2}]$. Let us suppose that the Lebesgue measure of the complement of $\mathcal{X}_{\boldsymbol{j}}(\psi, C)$ in $\mathcal{X}_0$ is greater than some $\delta > 0$, which we assume is small. Now, with respect to the measure $(\det A)^{-2n-1} \prod_{i,j} \mathrm{d}a_{ij}$ on $\mathrm{GL}(n, \mathbb{R})$, the volume of the set of $A \in \mathrm{GL}(n, \mathbb{R})$ satisfying (4.57) is within a constant multiple (depending on $\mathcal{K}$) of $2^{-n(j_1+\cdots+j_n)}$. Then, using the expression (2.13), (2.14) for the Haar measure on $H \rtimes G$, we have

$$\tilde{\mu}(\tilde{\Gamma}\backslash(H \rtimes G) - \mathcal{G}_{\boldsymbol{j}}(\psi, C)) \gg \delta 2^{-n(j_1+\cdots+j_n)}, \tag{4.61}$$

with implied constant depending on $\mathcal{K}$. From Lemma 4.6 it follows that

$$\mathrm{meas}(\mathcal{X}_0 - \mathcal{X}_{\boldsymbol{j}}(\psi, C)) \ll C_\psi C^{-2n-4} 2^{(n+1)(j_1+\cdots+j_n)}, \tag{4.62}$$

and we find that

$$\operatorname{meas}\Big(\mathcal{X}_0 - \bigcup_{C>0}\bigcap_{\boldsymbol{j}\geq 0}\mathcal{X}_{\boldsymbol{j}}(\psi, C2^{a(j_1+\cdots+j_n)})\Big) \ll \lim_{C\to\infty} C_\psi C^{-2n-4}\sum_{\boldsymbol{j}\geq 0} 2^{((n+1)-a(2n+4))(j_1+\cdots+j_n)} = 0 \tag{4.63}$$

as long as $a > \frac{n+1}{2n+4}$.

Now let us suppose that $(X, \boldsymbol{y}) \in \mathcal{X}(\psi)$. By Theorem 3.1, the size of the theta functions in (4.51) is invariant under the transformation on the left of (4.59), so we may assume that $X \in \mathcal{X}_0$ as well. In particular, we have that $(X, \boldsymbol{y})$ is in $\mathcal{X}_{\boldsymbol{j}}(\psi, C2^{a(j_1+\cdots+j_n)})$ for some $C > 0$ (independent of $\boldsymbol{j}$) and all $\boldsymbol{j} \geq 0$. We have from Corollary 4.1 and the definition of the height function D that

$$\ll M^{\frac{n}{2}} \sum_{S\subset\{1,\ldots,n\}} \sum_{\substack{\boldsymbol{j}\geq 0\\ 2^{j_i} b_{j_i}^{-1}\leq M}} 2^{-\frac{1}{2}(j_1+\cdots+j_n)} D(\tilde{\Gamma}(h, g(MB, X))(h_S, g_{\boldsymbol{j},S}))^{\frac{1}{4}} \tag{4.64}$$

bounds (4.51). Now for all $\boldsymbol{j} \geq 0$ there is a $\tilde{\Gamma}(h', g) \in \mathcal{G}_{\boldsymbol{j}}(\psi, C2^{a(j_1+\cdots+j_n)})$ with g of the form

$$g = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & {}^t\!A^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ T & I \end{pmatrix} \tag{4.65}$$

satisfying (4.57) and $||T|| \leq \epsilon$, and h' having the form $(\boldsymbol{u}, \boldsymbol{y} - \boldsymbol{u}X, 0)$ for some $\boldsymbol{u} \in (-\frac{1}{2}, \frac{1}{2})^n$. We have

$$(h', g)\left(1, \begin{pmatrix} \frac{1}{M}I & 0 \\ 0 & MI \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S}) = (h, g(MB, X))(h_S, g_{\boldsymbol{j},S})(h_1, g_1), \tag{4.66}$$

where

$$h_1 = \Big(-\boldsymbol{x}_S A_{\boldsymbol{j}} E_S + \boldsymbol{x}_S (BA)^{-1} A_{\boldsymbol{j}} E_S + \frac{1}{M}(\boldsymbol{u} - \boldsymbol{x}) B^{-1} A_{\boldsymbol{j}} E_{\boldsymbol{j}}, 0, 0\Big) \tag{4.67}$$

and

$$g_1 = g_{\boldsymbol{j},S}^{-1} \begin{pmatrix} BA & 0 \\ 0 & {}^t(BA)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{1}{M^2}T & I \end{pmatrix} g_{\boldsymbol{j},S}. \tag{4.68}$$

Recalling that $2^{j_i} \leq M$, the conditions (4.57) and $||T|| \leq \epsilon$ imply that (h_1, g_1) satisfies the conditions of Lemma 4.4 for all M, which then implies

$$D(\tilde{\Gamma}(h, g(MB, X)(h_S, g_{\boldsymbol{j},S}))^{\frac{1}{4}} \asymp D\left(\tilde{\Gamma}(h', g)\left(1, \begin{pmatrix} \frac{1}{M}I & 0 \\ 0 & MI \end{pmatrix}\right)(h_S, g_{\boldsymbol{j},S})\right)^{\frac{1}{4}} \ll C2^{a(j_1+\cdots+j_n)}\psi(\log M) \tag{4.69}$$

since $(h', g) \in \mathcal{G}_{\boldsymbol{j}}(\psi, C2^{a(j_1+\cdots+j_n)})$. Taking $a = \frac{2n+3}{4n+8}$ so that $\frac{n+1}{2n+4} < a < \frac{1}{2}$, it follows that (4.64) is bounded by

$$\ll CM^{\frac{n}{2}}\psi(\log M)\sum_{\boldsymbol{j}\geq 0} 2^{-(\frac{1}{2}-a)(j_1+\cdots+j_n)} \ll CM^{\frac{n}{2}}\psi(\log M), \tag{4.70}$$

and Theorem 1.2 follows. □

Data access statement. No new data were generated or analysed during this study.

REFERENCES

[1] A. Borel. *Introduction aux Groupes Arithmétiques*, Publications de l'Institut de Mathématique de l'Université de Strasbourg, XV, Hermann, Paris, 1969.

[2] A. Borel and L. Ji, *Compactifications of locally symmetric spaces*, J. Differential Geom. **73** (2006), 263–317.

[3] S. Cosentino and L. Flaminio, *Equidistribution for higher-rank Abelian actions on Heisenberg nilmanifolds*, J. Mod. Dyn. **9** (2015), 305–353.

[4] A. Fedotov and F. Klopp, *An exact renormalization formula for Gaussian exponential sums and applications*, Amer. J. Math. **134** (2012), 711–748.

[5] H. Fiedler, W. Jurkat, and O. Körner, *Asymptotic expansions of finite theta series*, Acta Arith. **32** (1977), 129–146.

[6] L. Flaminio and G. Forni, *Equidistribution of nilflows and applications to theta sums*, Ergodic Theory Dynam. Systems **26** (2006), 409–433.

[7] D. Grenier, *Fundamental domains for the general linear group*, Pacific J. Math. **132** (1988), 293–317.

[8] A. W. Knapp, *Lie Groups Beyond an Introduction*, Birkhäuser Boston, Boston, MA, 2002.

[9] G. Lion and M. Vergne, *The Weil Representation, Maslov Index and Theta Series*, Birkhäuser, Boston, MA, 1980.

[10] J. Marklof and M. Welsh, *Segal–Shale–Weil representation, theta functions, and applications*, in preparation.

[11] J. Marklof and M. Welsh, *Bounds for theta sums in higher rank I*, J. Anal. Math. **150** (2023), 325–358.

[12] D. Mumford, *Tata Lectures on Theta. I*, Birkhäuser Boston, Boston, MA, 1983.

[13] A. Terras. *Harmonic Analysis on Symmetric Spaces and Applications. II*, Springer, Berlin, 1988.

Jens Marklof
School of Mathematics
University of Bristol
Bristol BS8 1UG, UK
email: majm@bristol.ac.uk

Matthew Welsh
Department of Mathematics
University of Maryland
College Park, MD 20742, USA
email: mcwelsh@umd.edu

(Received May 1, 2023 and in revised form September 20, 2023)