JOURNAL de Théorie des Nombres de BORDEAUX

anciennement Séminaire de Théorie des Nombres de Bordeaux

Alexander GORODNIK et Amos NEVO

On discrepancy, intrinsic Diophantine approximation, and spectral gaps

Tome 36, n° 1 (2024), p. 127-174.

https://doi.org/10.5802/jtnb.1275

Le Journal de Théorie des Nombres de Bordeaux *est membre du Centre Mersenne pour l'édition scientifique ouverte*
http://www.centre-mersenne.org/
e-ISSN : 2118-8572

Journal de Théorie des Nombres de Bordeaux **36** (2024), 127–174

On discrepancy, intrinsic Diophantine approximation, and spectral gaps

par Alexander GORODNIK et Amos NEVO

Résumé. Dans le présent article, nous établissons des bornes pour la taille de l'écart spectral pour les actions de groupe sur les espaces homogènes. Notre approche est basée sur l'estimation des normes des opérateurs de moyennage appropriés, et nous développons des techniques pour établir des bornes supérieures et inférieures pour de telles normes. Nous montrerons que ce problème analytique est étroitement lié au problème arithmétique de l'établissement de bornes sur la divergence de distribution pour les points rationnels sur les variétés de groupes algébriques. Comme application, nous montrons comment établir une borne effective pour la propriété (τ) des sous-groupes de congruence des treillis arithmétiques dans les groupes algébriques qui sont des formes de SL_2, en utilisant des estimations dans l'approximation diophantienne intrinsèque qui découlent de l'analyse de Heath-Brown des points rationnels sur des variétés quadratiques de dimension 3.

Abstract. In the present paper we establish bounds for the size of the spectral gap for group actions on homogeneous spaces. Our approach is based on estimating operator norms of suitable averaging operators, and we develop techniques for establishing both upper and lower bounds for such norms. We shall show that this analytic problem is closely related to the arithmetic problem of establishing bounds on the discrepancy of distribution for rational points on algebraic group varieties. As an application, we show how to establish an effective bound for property (τ) of congruence subgroups of arithmetic lattices in algebraic groups which are forms of SL_2, using estimates in intrinsic Diophantine approximation which follow from Heath-Brown's analysis of rational points on 3-dimensional quadratic surfaces.

1. Introduction

1.1. Introduction. Let G be a locally compact second countable (lcsc) group. A (strongly continuous) unitary representation π of G on a Hilbert space $\mathcal{H}$ is said to have almost invariant vectors if there exists a sequence

Manuscrit reçu le 28 février 2022, révisé le 24 août 2023, accepté le 15 septembre 2023.

2020 *Mathematics Subject Classification.* 22E46, 11F70, 11F72.

Mots-clefs. Semisimple group, discrepancy, spectral gap, Diophantine approximation, automorphic representation.

A.G. was supported by SNF grant 200021–182089. A.N. was partially supported by ISF Moked Grant 2919-19.

of unit vectors v_n such that $\|\pi(g)v_n - v_n\| \to 0$ uniformly for g varying in compact subsets of G. If π does not have almost invariant vectors, π is said to possess a spectral gap. This fundamental representation-theoretic property expresses the fact that the trivial representation of G is “isolated” from the irreducible representations appearing in the spectral decomposition of the representation π.

The spectral gap property can be also formulated in explicit quantitative terms, via norm estimates of averaging operators. For a finite Borel measure β on G, we define the bounded operator

$$\pi(\beta) : \mathcal{H} \longrightarrow \mathcal{H}\ ,\ v \longmapsto \int_G \pi(g)v\, \mathrm{d}\beta(g).$$

It is well-known that π has the spectral gap property if and only if for some (equivalently, for all) absolutely continuous symmetric probability measure β whose support generates G, one has $\|\pi(\beta)\| < 1$. This leads to the natural problem of establishing explicit estimates on the norm $\|\pi(\beta)\|$ that we aim to address and utilize in the present paper.

We will be especially interested in the spectral gap property for representations of semisimple algebraic groups acting on homogeneous spaces with finite invariant measures, which we now introduce. Let $\mathsf{G} \subset \mathrm{GL}_n$ be an algebraically connected linear algebraic group defined over $\mathbb{Q}$ which is $\mathbb{Q}$-simple. We use the notations:

$$G_\infty := \mathsf{G}(\mathbb{R}) \quad \text{and} \quad G_S := \prod\nolimits_{p\in S} \mathsf{G}(\mathbb{Q}_p)$$

for a finite set of primes S, and

$$\Gamma_S := \mathsf{G}(\mathbb{Z}[S^{-1}]) \quad \text{and} \quad \Gamma_{S,m} := \{\gamma \in \Gamma_S :\ \gamma = e \mod m\}$$

for m coprime to S. We consider the spaces

$$X_{S,m} := (G_\infty \times G_S)/\Gamma_{S,m}$$

and the corresponding unitary representations $\rho_{S,m}$ of the group G_S acting on the spaces $L^2_0(X_{S,m})$ consisting of square-integrable functions with zero integral.

To simplify notation, when the set S consists of a single prime p, we write G_p (instead of $G_{\{p\}}$) and similarly for the other objects involving S in the subscript.

We have a (well-defined and unique) direct integral decomposition

$$\rho_{S,m} = \int_{\widehat{G}_S} \pi^{\oplus k(\pi)}\, \mathrm{d}\Pi_{S,m}(\pi),$$

where $\Pi_{S,m}$ is a Borel measure on the (standard Borel) unitary dual $\widehat{G}_S$ (consisting of equivalence classes of strongly continuous irreducible unitary representations of G_S) and $1 \le k(\pi) \le \infty$ denotes multiplicities. The support of the measure $\Pi_{S,m}$ (w.r.t. the Fell topology on the unitary dual) is called the support of the representation $\rho_{S,m}$. Following [5], we introduce the notion of the automorphic dual of G_S:

$$\widehat{G}_S^{\mathrm{aut},0} := \overline{\bigcup_{(m,S)=1} \mathrm{supp}(\rho_{S,m})} \quad \text{and} \quad \widehat{G}_S^{\mathrm{aut}} = \widehat{G}_S^{\mathrm{aut},0} \cup \{1_{G_S}\},$$

where the closure is taken with respect to the Fell topology on the unitary dual $\widehat{G}_S$, and 1_{G_S} denotes the trivial representation of G_S. A crucially important property (called property (τ) for congruence subgroups or the Ramanujan–Selberg property) of the representations $\rho_{S,m}$ is that they are uniformly isolated from the trivial representation, namely, the sum of the representations $\rho_{S,m}$ has the spectral gap property. This fundamental result was proved in full generality by Clozel [8], but it has a rich history going back in some cases to works of Ramanujan and Selberg [35, 36], Kazhdan [24], Burger–Sarnak [6], and others. We refer, for instance, to [33, 9, 4] for detailed surveys.

We introduce the height function on the group G_S:

$$\mathrm{H}_S(g) := \prod_{p\in S} \max(1, \|g_p\|_p) \quad \text{for } g = (g_p)_{p\in S} \in G_S,$$

where $\|\cdot\|_p$ denotes the maximum p-adic norms, and the corresponding compact subsets

$$B_h^S := \{g \in G_S : \mathrm{H}_S(g) \le h\}. \tag{1.1}$$

We consider the Haar-uniform probability measures β_h^S supported on the sets B_h^S. Then property (τ) amounts to the estimate (when $h \ge h_0(S)$)

$$\sup\{\|\pi(\beta_h^S)\| : \pi \in \widehat{G}_S^{\mathrm{aut},0}\} < 1.$$

In a number of cases, more precise explicit bounds are known. For instance, the Ramanujan–Petersson–Selberg Conjecture amounts to showing that when G is a form of SL_2, then

$$\sup\{\|\pi(\beta_h^S)\| : \pi \in \widehat{G}_S^{\mathrm{aut},0}\} \ll_{\epsilon,S} m_{G_S}(B_h^S)^{-1/2+\epsilon} \quad \text{for all } \epsilon > 0,$$

where m_{G_S} denotes a Haar measure on G_S. It was established by Deligne in the case when G is anisotropic over $\mathbb{R}$ (see, for instance, [32] for a self-contained introduction to these results), but is still open otherwise. On the other hand, one expects only weaker decay bounds for more general algebraic groups.

The goal of the present paper is to develop upper and lower bounds on the norms $\|\pi(\beta_h^S)\|$ for automorphic representations, and to demonstrate their close connection and mutual relationship to problems of intrinsic Diophantine approximation on the corresponding algebraic groups. We note that the norm estimates are interesting only when the group G_S is noncompact, which we assume throughout the paper.

We now turn to state our main results.

1.2. Lower bounds for operator norms in the automorphic representation. Our first result provides an explicit lower bound for the norms of the averaging operators introduced above.

Theorem 1.1. *Let* G *be an algebraically connected* $\mathbb{Q}$*-simple linear algebraic group defined over* $\mathbb{Q}$*. If* L *is a proper reductive algebraic subgroup of* G *defined over* $\mathbb{Q}$*, then there exists* $C > 0$ *such that for all* $h \geq 1$

$$\sup\Big\{\|\pi(\beta_h^S)\| : \pi \in \widehat{G}_S^{\mathrm{aut},0}\Big\} \geq C\, \frac{m_{L_S}(L_S \cap B_h^S)}{m_{G_S}(B_h^S)},$$

where m_{L_S} *denotes a Haar measure on* L_S*, and the constant* C *depends on* $\mathsf{G}, S, \mathsf{L}$ *and the choice of Haar measures. Therefore, for all* $h \geq 1$*,*

$$\sup\Big\{\|\rho_{S,m}(\beta_h^S)\| : m \in \mathbb{N} \text{ with } (m,S) = 1\Big\} \geq C\, \frac{m_{L_S}(L_S \cap B_h^S)}{m_{G_S}(B_h^S)}.$$

We remark that a general, but *weaker*, lower bound valid in the present context is: for all $h \geq 1$

$$\sup\{\|\pi(\beta_h^S)\| : \pi \in \widehat{G}_S^{\mathrm{aut},0}\} \geq C\left(m_{G_S}(B_h^S)\right)^{-1/2}. \tag{1.2}$$

We refer to Remark 3.3 below for a justification of (1.2). Usually, in Theorem 1.1 one can choose the subgroup L, so that $m_{L_S}(L_S \cap B_h^S) \gg m_{G_S}(B_h^S)^\rho$ with $\rho > 1/2$. Then Theorem 1.1 provides a better estimate. We exemplify this by the following explicit estimates for classical groups:

Corollary 1.2.

(1) *For* $\mathsf{G} = \mathrm{SL}_n$ *with* $n \geq 3$*, for every prime* p*, if* h *is sufficiently large, for all* $\epsilon > 0$*.:*

$$\sup\Big\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\Big\} \gg_\epsilon m_{G_p}(B_h^{\{p\}})^{-2/n+\epsilon}.$$

(2) *Let* $\mathsf{G} = \mathrm{SO}_n$ *with* $n \geq 4$ *and* p *is prime satisfying* $p = 1 \pmod 4$*. Then when* n *is even, if* h *is sufficiently large, for all* $\epsilon > 0$

$$\sup\Big\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\Big\} \gg_\epsilon m_{G_p}(B_h^{\{p\}})^{-2/n+\epsilon}, \quad \epsilon > 0,$$

and when n *is odd, if* h *is sufficiently large, for all* $\epsilon > 0$*,*

$$\sup\Big\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\Big\} \gg_\epsilon m_{G_p}(B_h^{\{p\}})^{-2/(n-1)+\epsilon}.$$

(3) *Let* $\mathsf{G} = \mathrm{Sp}_{2n}$ *for* $n \geq 2$*. If* h *is sufficiently large, for all* $\epsilon > 0$*,*

$$\sup\Big\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\Big\} \gg_\epsilon m_{G_p}(B_h^{\{p\}})^{-2/(n+1)+\epsilon}.$$

It is to be expected that the lower bounds in Theorem 1.1 (and in some cases even the best-possible lower bounds) can be deduced from the description of the continuous automorphic spectrum obtained by Langlands [26].

However our approach is different, and uses only relatively elementary considerations.

The above result raises the question of estimating the operator norms for the discrete part of the spectrum. Let $\rho_{S,m}^{\mathrm{disc}}$ denote the subrepresentation of $\rho_{S,m}$ consisting of discretely embedded irreducible subrepresentations of $\rho_{S,m}$, and let $\rho_{S,m}^{\mathrm{cusp}}$ denote its cuspidal subrepresentation, so that $\|\rho_{S,m}^{\mathrm{disc}}(\beta)\| \geq \|\rho_{S,m}^{\mathrm{cusp}}(\beta)\|$. We note that a lower bound for $\|\rho_{S,m}^{\mathrm{cusp}}(\beta)\|$ follows from the remarkable equidistribution results for cuspidal spectrum of automorphic representations that was recently established by Matz and Templier [29] and Finis and Matz [14].

Theorem 1.3. *Let G be a split simple simply connected classical algebraic group defined over $\mathbb{Q}$. Let us assume that G is isotropic and unramified over $\mathbb{Q}_p$ and denote by U_p a hyper-special maximal compact subgroup of G_p. Then for any compactly supported U_p-biinvariant measure β on G_p,*

$$\|\rho_{p,1}^{\mathrm{cusp}}(\beta)\| \geq \|\beta\|_{L^2}.$$

For example, when β is a probability measure given as the normalized characteristic function of a U_p-biinvariant subset B of G_p, we obtain the lower bound

$$\|\rho_{p,1}^{\mathrm{disc}}(\beta)\| \geq m_{G_p}(B)^{-1/2}\,.$$

The bound in Theorem 1.3 implies the bound stated in (1.2) but is established for a more restricted setting, and is far less elementary compared to the discussion in Remark 3.3.

1.3. Spectral gaps and mean-square discrepancy bounds on the group variety. Now we aim to obtain upper bounds on the averaging operators, or equivalently on the size of the spectral gap. We shall show that such bounds can be deduced from the solution of the arithmetic counting problem of estimating the discrepancy of distribution of rational points.

Let now G be an algebraically connected simply connected $\mathbb{Q}$-simple linear algebraic group defined over $\mathbb{Q}$. We use notations as above and additionally assume that G_∞ is non-compact. Then the subgroup $\Gamma_{S,m}$ embedded in G_∞ is dense and, in fact, equidistributed in G_∞ in a suitable sense (see [17, 20]). Let m_{G_∞} and m_{G_S} denote Haar measures on G_∞ and G_S which are normalized so that Γ_S has covolume one in $G_\infty \times G_S$. We fix a left-invariant[1] Riemannian metric ρ on G_∞. Let $B(x,r)$ denote the corresponding balls in G_∞. Let also B be a bounded measurable subset of G_S of positive measure. We consider the problem of counting the points of $\Gamma_{S,m}$ contained in the domains $B(x,r) \times B$. To study the distribution of rational points $\Gamma_{S,m}$ in G_∞, it is natural to consider the following discrepancy

[1]With obvious modifications our argument applies also to a right invariant metric.

function:

$$\mathcal{D}(\Gamma_{S,m}, B(x,r)\times B) := \left| \frac{|\Gamma_{S,m}\cap (B(x,r)\times B)|}{m_{G_S}(B)} - \frac{m_{G_\infty}(B(e,r))}{|\Gamma_S/\Gamma_{S,m}|} \right|$$

where the ball $B(x,r)$ is fixed, and the subsets B eventually exhaust G_S. For instance, one can take B to be the height balls B_h^S with $h\to\infty$.

We consider the discrepancy as a function of x as it varies over G_∞. When the subset B of G_S is left-invariant under a compact open subgroup U_S of G_S, the discrepancy $\mathcal{D}(\Gamma_{S,m}, B(x,r)\times B)$ is left-invariant under the subgroup $\Gamma_{S,m}\cap U_S$, which is a finite index subgroup of $\Gamma_S\cap U_S$. In this case, we define the mean-square discrepancy as

$$E_{S,m}(r,B) := \|\mathcal{D}(\Gamma_{S,m}, B(\,\cdot\,,r)\times B)\|_{L^2\left((\Gamma_{S,m}\cap U_S)\backslash G_\infty\right)},$$

where $(\Gamma_{S,m}\cap U_S)\backslash G_\infty$ is equipped with the invariant probability measure.

Note that computing the discrepancy is a purely arithmetic problem which involves estimating the number of rational points satisfying given Diophantine equations, inequalities and congruence conditions. But we will show that the behaviour of the discrepancy captures the size of the spectral gap for the corresponding automorphic representations, a purely analytic problem. The fact that the mean-square discrepancy can be used to bound the norms of the averaging operators is formulated as follows.

Theorem 1.4. *Let us assume that the set of primes S is unramified and denote by U_S a hyper-special maximal compact subgroup of G_S (cf. Section 2). Let B be any non-empty bounded U_S-bi-invariant subset of G_S of positive measure. Fix an irreducible unitary representation π of G_S which is discretely embedded in $\rho_{S,m}$. Then there exists $r_0(\pi)>0$ such that the averaging operator $\pi(\beta)$ supported on B satisfies the bound*

$$\|\pi(\beta)\| \ll |\Gamma_S/\Gamma_{S,m}|\, r^{-\dim(G_\infty)} E_{S,m}(r,B) \tag{1.3}$$

for all $0<r\le r_0(\pi)$. Here the implied constant depends only on G_∞.

While our approach has little in common with the work of Sarnak and Xue [34], their work was the main inspiration for Theorem 1.4. In fact, [34] develops an approach for establishing the uniform spectral gap property for the Archimedean factor based on lattice point counting estimates.

Let us note the following regarding Theorem 1.4:

- For an upper bound on $\|\pi(\beta)\|$ to be meaningful, it must be strictly less that 1. As we shall see below, in suitable circumstances this is indeed the case for the bound (1.3), provided that the measure of the set B is sufficiently large compared with the inverse of the measure of the ball $B(e,r)$.
- The assumption that π is discretely embedded in $\rho_{S,m}$ is important for our present argument. We expect however that this argument

can be developed further to deal with all representations which are weakly contained in $\rho_{S,m}$. Note that the complete description of continuous spectrum is known due to the work of Langlands [26]. Therefore, understanding the discretely embedded representations is the crucial missing ingredient.

In our previous work [17, 20], we have shown that analysis on the homogeneous spaces $X_{S,m}$ can be used to establish mean, almost everywhere, and pointwise bounds for the discrepancy function. Remarkably, Theorem 1.4 (and more generally Theorem 5.2 below) shows that the converse is also true, and estimates on discrepancy lead to quantitative estimates for the spectral gap. These observations raise an interesting possibility of studying the spectral gap problem via arithmetic consideration. Here we carry this out for forms of SL_2.

1.4. Property (τ) for congruence subgroups. Let G be a linear algebraic group defined over $\mathbb{Q}$ which is a form of SL_2. More explicitly, G can be viewed as the set of norm one elements in a division algebra defined over $\mathbb{Q}$. The integral structure on G is defined with respect to an order of the division algebra. We fix a prime p, and assume that G is isotropic over $\mathbb{R}$ and over $\mathbb{Q}_p$. Moreover, we assume that $\mathsf{G}(\mathbb{Z}_p)$ is a hyper-special maximal compact subgroup of G_p (which is the case for all but finitely many primes). In this setting, we derive a bound on the norms of automorphic representations:

Theorem 1.5. *For every $\ell \in \mathbb{N}$ coprime to a prime p satisfying the foregoing conditions,*

$$\|\rho_{p,\ell}(\beta_h^{\{p\}})\| \ll_{p,\ell} m_{G_p}(B_h^{\{p\}})^{-\kappa},$$

where $\kappa = 1/4$ if G is anisotropic over $\mathbb{Q}$, and $\kappa = 1/16$ if G is isotropic over $\mathbb{Q}$. As a result, for any $\pi \in \widehat{G}_p^{\mathrm{aut},0}$,

$$\|\pi(\beta_h^{\{p\}})\| \ll_{p,\ell} m_{G_p}(B_h^{\{p\}})^{-\kappa}.$$

Our estimate falls short of the best known bound, which corresponds to $\kappa = 25/64$. This bound is due to Kim and Sarnak [25, App. 2] over $\mathbb{Q}$ (and Blomer and Brumley [3] over number fields) and was proved by quite a different argument. Our approach involves establishing a bound on the discrepancy function directly using the refined circle method arguments due to Heath-Brown, and then combining this bound with a version of Theorem 1.4. We remark that for technical reasons, we will work with smooth test-functions on G_∞ rather than the balls as above.

1.5. Organisation of the paper. In the next section we set up notation and review basic facts that will be used throughout the paper. Then in Section 3 we develop a method for proving lower bounds for the norms of averaging operators and establish Theorem 1.1 and Corollary 1.2. In Section 4 we estimate the norms of averaging operators acting on the discrete part of the spectrum and prove Theorem 1.3. In Section 5 we show how to derive norm bounds from estimates on the discrepancy of rational points and prove Theorem 1.4. This approach will be utilized in Section 6 where we use the refined circle method technique developed by Heath-Brown [22] to establish the required estimates on discrepancy and prove Theorem 1.5.

2. Basic notation

Let $\mathsf{G} \subset \mathsf{GL}_n$ be a connected semisimple linear algebraic group defined over a number field K. We denote by K_v, $v \in V$, the completions of K. We fix a finite set S of non-Archimedean completions such that the group G is isotropic for at least one completion from S. We denote by G_∞ the product of $\mathsf{G}(K_v)$ over the Archimedean completions and by G_S the product of $\mathsf{G}(K_v)$ over S. We consider G_∞ and G_S as locally compact groups equipped with the topology arising from the field completions. We fix a choice of Haar measures m_{G_∞} and m_{G_S} on G_∞ and G_S respectively.

Let $O_S := \{x \in K : |x|_v \le 1 \text{ for } v \notin S\}$ denote the subring of K consisting of elements which are K_v-integral for v outside S (also known as the ring of S-integers in K). For non-Archimedean completions, we write O_v for the ring of integers in K_v (note that with this notation $O_v \ne O_{\{v\}}$). We denote by $\Gamma_S := \mathsf{G}(O_S)$ the S-arithmetic subgroup of $\mathsf{G}(K)$. We denote by $\mathcal{O}^S$ the product of $\mathsf{G}(O_v)$ over the non-Archimedean completions v which are not in S. Then Γ_S can be viewed as a subgroup of $\mathcal{O}^S$, and for an open subgroup $\mathcal{O}$ of $\mathcal{O}^S$, we define $\Gamma_{S,\mathcal{O}} := \Gamma_S \cap \mathcal{O}$. Under the natural embedding, we view $\Gamma_{S,\mathcal{O}}$ as a subgroup of $G_\infty \times G_S$. Then it is a discrete subgroup with finite covolume, and we consider the homogeneous spaces $X_{S,\mathcal{O}} := (G_\infty \times G_S)/\Gamma_{S,\mathcal{O}}$ equipped with the invariant probability measures $\mu_{S,\mathcal{O}}$. Let $\rho_{S,\mathcal{O}}$ denote the corresponding unitary representations of the group G_S on the spaces $L^2_0(X_{S,\mathcal{O}})$, the space of L^2-integrable functions with zero integral.

The group G is said to be unramified over a completion K_v if it is quasi-split and split over an unramified extension of K_v. If G is unramified over K_v, there exists a canonical conjugacy class of maximal compact open subgroups of G_v — the so-called hyperspecial maximal compact subgroups (see [39]). We fix a choice U_v of a hyperspecial maximal compact subgroup U_v of G_v. More generally, we write $U_S := \prod_{v\in S} U_v$ when G is unramified over every $v \in S$ (i.e. unramified over S). The pair (G_S, U_S) is known to have the Gelfand–Selberg property, namely the convolution algebra $L^1(U_S\backslash G_S/U_S)$ is commutative. We say that a unitary representation of G_S is spherical if there exists a non-zero U_S-invariant vector. It follows from the Gelfand–Selberg property that if a spherical representation is irreducible, then the dimension of the space of U_S-invariant vectors is at most one.

Let $\widehat{G}_S$ denote the unitary dual of the group G_S equipped with the Fell topology. We recall that given an arbitrary unitary representation σ of G_S on a Hilbert space $\mathcal{H}_\sigma$, we have the direct integral decomposition

$$\sigma = \int_{\widehat{G}_S} \pi^{\oplus k(\pi)}\, \mathrm{d}\Pi_{S,\mathcal{O}}(\pi),$$

where $\Pi_{S,\mathcal{O}}$ is a Borel measure on $\widehat{G}_S$, and $1 \le k(\pi) \le \infty$ denotes the multiplicities. The support of the measure $\Pi_{S,\mathcal{O}}$ is uniquely defined and is called the support of the representation σ, denoted $\mathrm{supp}(\sigma)$. One says that σ is weakly contained in a subset Ω of $\widehat{G}_S$ if $\mathrm{supp}(\sigma) \subset \overline{\Omega}$, where the closure is taken with respect to the Fell topology. Namely, this means that every function of positive type associated to σ can be approximated, uniformly on compact subsets of G_S, by finite sums of functions of positive type associated to $\pi \in \Omega$. More generally, for unitary representations σ_1 and σ_2 of G_S, we say that σ_1 is weakly contained in σ_2 if $\mathrm{supp}(\sigma_1) \subset \mathrm{supp}(\sigma_2)$. We refer to [2, App. F] for basic properties of the notion of weak containment. In particular, we recall that σ_1 is weakly contained in σ_2 is equivalent to

$$\|\sigma_1(F)\| \le \|\sigma_2(F)\| \quad \text{for all } F \in L^1(G_S), \tag{2.1}$$

and that σ is weakly equivalent to the representation $\oplus\{\pi : \pi \in \mathrm{supp}(\sigma)\}$. We say that a representation is tempered if it is weakly contained in the regular representation $L^2(G_S)$.

Following [5], we introduce the notion of automorphic dual of G over S:

$$\widehat{G}_S^{\mathrm{aut}} := \widehat{G}_S^{\mathrm{aut},0} \cup \{1_{G_S}\} \quad \text{with} \quad \widehat{G}_S^{\mathrm{aut},0} := \overline{\bigcup_{\mathcal{O}\subset\mathcal{O}^S} \mathrm{supp}(\rho_{S,\mathcal{O}})},$$

where the union is taken over open subgroups $\mathcal{O}$ of $\mathcal{O}^S$, and the closure is with respect to the Fell topology on $\widehat{G}_S$. Given a subset B of G_S with finite positive measure, we denote by β the Haar-uniform probability measure supported on the set B. For $\pi \in \widehat{G}_S$, we consider the averaging operators $\pi(\beta)$ which are explicitly defined by

$$\pi(\beta) = \frac{1}{m_{G_S}(B)} \int_B \pi(g)\, \mathrm{d}m_{G_S}(g).$$

Recall that for any two finite Borel measures β_1 and β_2 on G_S,

$$\pi(\beta_1 * \beta_2) = \pi(\beta_1)\pi(\beta_2) \quad \text{and} \quad \pi(\beta)^* = \pi(\beta^*).$$

We will be especially interested in the sets

$$B_h^S := \{g \in G_S : \mathrm{H}_S(g) \le h\}$$

defined by the height function

$$\mathrm{H}_S(g) := \prod_{v\in S} \max(1, \|g_v\|_v) \quad \text{for } g = (g_v)_{v\in S} \in G_S.$$

In this case, we denote the corresponding probability measure by β_h^S (and when no ambiguity is present, by β_h).

For some classes of unitary representations, explicit estimates on the operator norms $\|\pi(\beta)\|$ have been established. For instance, according to the Kunze–Stein Phenomenon (proved for p-adic groups in [40] following the method of [10]), for every $q \in [1, 2)$,

$$\|f * \phi\|_2 \ll_q \|f\|_q \|\phi\|_2 \qquad \text{for all } f \in L^q(G_S) \text{ and } \phi \in L^2(G_S).$$

It follows (see [19] for a full account) that for every tempered representation π, and for every subset $B \subset G_S$ of positive finite measure,

$$\|\pi(\beta)\| \ll_\epsilon m_{G_S}(B)^{-1/2+\epsilon} \tag{2.2}$$

for all $\epsilon > 0$. More generally, if the tensor power $\pi^{\otimes n}$ is tempered for some even integer n, then by [30, 18] the following bound holds :

$$\|\pi(\beta)\| \ll_\epsilon m_{G_S}(B)^{-1/(2n)+\epsilon}$$

for all $\epsilon > 0$.

In particular, the tensor power property is known to hold for non-trivial irreducible unitary representations of $\mathrm{SL}_d(K_v)$ (and more generally simple groups with property T). Therefore, for every such representation π,

$$\|\pi(\beta)\| \ll m_{G_S}(B)^{-\kappa(\pi)}$$

for some exponent $\kappa(\pi) > 0$, which is uniform in the representation π when $d > 2$.

When an irreducible representation π is spherical, then

$$\begin{aligned} \|\pi(\beta)\|^2 = \|\pi(\beta^* * \beta)\| &= \left| \frac{1}{m_{G_S}(B)} \int_B \omega_\pi(g)\, \mathrm{d}m_{G_S}(g) \right|^2 \\ &\le \frac{1}{m_{G_S}(B)} \int_B |\omega_\pi(g)|^2 \, \mathrm{d}m_{G_S}(g), \end{aligned} \tag{2.3}$$

where ω_π denotes the spherical functions associated to the representation π. Then $\|\pi(\beta)\|$ can be analysed using estimates on the spherical functions. For instance, considering the complementary series representations of $\mathrm{SL}_2(\mathbb{Q}_p)$, one can observe that the norms $\|\pi(\beta)\|$ may decay arbitrary slowly for irreducible representations π and thus also construct examples of (reducible) representations without invariant vectors such that $\|\pi(\beta)\| = 1$.

3. Lower bounds for norms of automorphic representations

Let G be a connected semisimple linear algebraic group defined over a number field K. We keep the notation and assumptions introduced in Section 2. Our goal is to establish a lower bound for $\|\pi(\beta_h^S)\|$, for the representations $\pi \in \widehat{G}_S^{\mathrm{aut},0}$.

Theorem 3.1. *Let L be any proper reductive algebraic subgroup of G defined over K. If L_S is non-amenable, then for some positive constant C and for all $h \geq 1$,*

$$\sup\left\{\|\pi(\beta_h^S)\| : \pi \in \widehat{G}_S^{\mathrm{aut},0}\right\} \geq C\frac{m_{L_S}(L_S \cap B_h^S)}{m_{G_S}(B_h^S)},$$

where C depends on $\mathsf{G}, S, \mathsf{L}$ and the choice of Haar measures, but not on h.

In particular, for all $h \geq 1$,

$$\sup\{\|\rho_{S,\mathcal{O}}(\beta_h^S)\| : \mathcal{O} \text{ – an open subgroup of } \mathcal{O}^S\} \geq C\frac{m_{L_S}(L_S \cap B_h^S)}{m_{G_S}(B_h^S)}.$$

Proof. The key starting point of our argument is the fundamental results from [5] and [9, §3.3] which assert that if $\pi \in \widehat{L}_S^{\mathrm{aut}}$, then the induced representation $\mathrm{Ind}_{L_S}^{G_S}(\pi)$ is weakly contained in $\widehat{G}_S^{\mathrm{aut}}$. In particular, if we denote by τ the unitary representation of G_S on $L^2(G_S/L_S)$ (namely the representation induced from the trivial representation of L_S), then it follows that τ is weakly contained in $\widehat{G}_S^{\mathrm{aut}}$.

We also note that the representation τ does not weakly contain the trivial representation 1_{G_S}. Indeed, consider any lcsc group G and a closed subgroup L, with G acting on the homogeneous space G/L with infinite invariant measure. Then weak containment of the trivial representation of G in $L^2(G/L)$ is equivalent to the existence of an asymptotically invariant (=Følner) sequence of subsets of G/L. The proof of this statement is due to Eymard [13] and Greenleaf [21, Thm. 4.1]. We refer to [23, Thm. 3.3] for a short proof, and to [1, §3.1, §4.1] for a full account of this topic.

Furthermore, assume that G is the K-points of an algebraically connected semisimple algebraic group over a locally compact non-discrete field K of characteristic zero. Then the existence of an asymptotically invariant sequence on G/L implies that L is Zariski-dense in G. This follows immediately from [23, Thm. 6.1], and thus when L is a proper unimodular algebraic subgroup, we conclude that the trivial representation is not weakly contained in $L^2(G/L)$. This argument also extends to the case of finite products of algebraic groups over local fields of characteristic zero.

We can therefore conclude that τ is in fact weakly contained in $\widehat{G}_S^{\mathrm{aut},0}$. In particular, τ is weakly contained in the direct sum of π in $\widehat{G}_S^{\mathrm{aut},0}$ (see [2, Prop. F.2.7]). Therefore, we deduce that

$$\sup\left\{\|\pi(\beta_h^S)\| : \pi \in \widehat{G}_S^{\mathrm{aut},0}\right\} \geq \|\tau(\beta_h^S)\|. \tag{3.1}$$

Hence, to finish the proof, it remains to establish a suitable lower bound for $\|\tau(\beta_h^S)\|$.

Let ν be a G_S-invariant measure on the homogeneous space G_S/L_S. We denote by $p : G_S \to G_S/L_S$ the factor map. Let ϕ be the characteristic function of a compact neighbourhood U of the identity coset in G_S/L_S. Then

$$\left\langle \tau(\beta_h^S)\phi, \phi \right\rangle = \frac{1}{m_{G_S}(B_h^S)} \int_{G_S/L_S} \int_{B_h^S} \phi(g^{-1}xL_S)\phi(xL_S)\, \mathrm{d}m_{G_S}(g)\mathrm{d}\nu(xL_S).$$

We observe that $\phi(g^{-1}xL_S) = 1$ provided that $g \in xp^{-1}(U)^{-1}$, and it is equal to zero otherwise. Hence,

$$\left\langle \tau(\beta_h^S)\phi, \phi \right\rangle = \frac{1}{m_{G_S}(B_h^S)} \int_U m_{G_S}(B_h^S \cap xp^{-1}(U)^{-1})\, \mathrm{d}\nu(xL_S).$$

If x belongs to a fixed compact subset of G_S, then it follows from properties of p-adic norms that there exists $c_1 > 0$ such that $\mathrm{H}_S(xg) \leq c_1 \mathrm{H}_S(g)$ for all $g \in G_S$. Therefore $x^{-1}B_h^S \supset B_{c_1^{-1}h}^S$, and we deduce that

$$\begin{aligned} m_{G_S}(B_h^S \cap xp^{-1}(U)^{-1}) &= m_{G_S}(x^{-1}B_h^S \cap p^{-1}(U)^{-1}) \\ &\geq m_{G_S}(B_{c_1^{-1}h}^S \cap p^{-1}(U)^{-1}), \end{aligned}$$

and

$$\left\langle \tau(\beta_h^S)\phi, \phi \right\rangle \geq \nu(U) \frac{m_{G_S}(B_{c_1^{-1}h}^S \cap p^{-1}(U)^{-1})}{m_{G_S}(B_h^S)}.$$

If the neighbourhood U is chosen to be sufficiently small, there exists a continuous section $\sigma : U \to p^{-1}(U)$ of the factor map p such that the map $L_S \times U \to p^{-1}(U)^{-1}$ defined by $(u, l) \mapsto l\sigma(u)^{-1}$ is a homeomorphism. Since $\sigma(U)$ is compact, there exists $c_2 > 0$ such that $\mathrm{H}_S(gx) \leq c_2 \mathrm{H}_S(g)$ for all $g \in G_S$ and $x \in \sigma(U)^{-1}$, so that

$$B_{c_1^{-1}h}^S \cap p^{-1}(U)^{-1} = B_{c_1^{-1}h}^S \cap L_S\sigma(U)^{-1} \supset (B_{c_1^{-1}c_2^{-1}h}^S \cap L_S)\sigma(U)^{-1}.$$

We observe that the Haar measure on $p^{-1}(U) \subset G_S$ can be decomposed as

$$\int_{G_S} f \, \mathrm{d}m_{G_S} = \int_U \int_{L_S} f(l\sigma(u)^{-1})\, \mathrm{d}m_{L_S}(l)\, \mathrm{d}\nu(u),$$

so that

$$m_{G_S}((L_S \cap B_{c_1^{-1}c_2^{-1}h}^S)\sigma(U)^{-1}) \geq \nu(U) m_{L_S}(L_S \cap B_{c_1^{-1}c_2^{-1}h}^S).$$

Therefore, combining the above estimates, we conclude that

$$\left\langle \tau(\beta_h^S)\phi, \phi \right\rangle \gg \frac{m_{L_S}(L_S \cap B^S_{c_1^{-1}c_2^{-1}h})}{m_{G_S}(B_h^S)}.$$

Hence, it follows from the definition of the operator norm that

$$\|\tau(\beta_h^S)\| \gg \frac{m_{L_S}(L_S \cap B^S_{c_1^{-1}c_2^{-1}h})}{m_{G_S}(B_h^S)}.$$

To conclude the proof of the first estimate stated in Theorem 3.1, we recall (3.1) and use a property of the volume function formulated and established in Lemma 3.2 immediately below.

To prove the second inequality stated in Theorem 3.1, we observe that every $\pi \in \widehat{G}_S^{\mathrm{aut},0}$ is weakly contained in $\bigoplus_{\mathcal{O}}\{\rho_{S,\mathcal{O}}\}$. Hence, it follows from (2.1) that

$$\|\pi(\beta_h^S)\| \le \sup_{\mathcal{O}} \|\rho_{S,\mathcal{O}}(\beta_h^S)\|,$$

and the second inequality follows directly from the first estimate. □

Lemma 3.2. *Let $\mathsf{L} \subset \mathrm{GL}_n$ be a linear reductive algebraic group defined over a number field K and S a finite set of non-Archimedean completions of K. We denote by m_{L_S} a Haar measure on the group L_S. Then there exists $c > 0$ such that the sets $B_h^S \cap L_S := \{g \in L_S : \mathrm{H}_S(g) \le h\}$ satisfy*

$$m_{L_S}(B^S_{2h} \cap L_S) \le c\, m_{L_S}(B_h^S \cap L_S)$$

for all sufficiently large h.

Proof. First, we consider the case when $S = \{v\}$ consists of a single completion v. For notational simplicity, in the present proof we set $W(K_v) = W_v$ (rather than $W_{\{v\}}$) for any algebraic group W. Let us recall the Cartan decomposition of L_v (see, for instance, [37, Ch. 0]). We take a maximal K_v-split torus T of L and a minimal parabolic subgroup P associated to T which has a decomposition $\mathsf{P} = MU$, where M is the centralizer of T in L, and U is the unipotent radical. We denote by Σ^+ the set of positive roots of A associated to the parabolic subgroup P and $\Pi \subset \Sigma^+$ the set of simple roots. We write $\mathcal{X}(\mathsf{M})$ for the group of algebraic characters of M. Given $\chi \in \mathcal{X}(\mathsf{M})$, we write $|\chi(m)|_v = q_v^{\langle \chi, \omega(m)\rangle}$ for $m \in M_v$, where $\omega : M_v \to \mathrm{Hom}(\mathcal{X}(\mathsf{M}), \mathbb{Z})$. We denote by M_v^0 the kernel of ω. Then M_v/M_v^0 is free abelian group, and moreover $\omega(M_v)$ can be considered as a lattice in $\mathrm{Hom}(\mathcal{X}(\mathsf{M}), \mathbb{R})$. We set

$$M_v^+ := \{m \in M_v : \langle \alpha, \omega(m)\rangle \ge 0 \text{ for all } \alpha \in \Sigma^+\}.$$

Let U_v be a good maximal compact subgroup of L_v associated to A_v. Then the Cartan decomposition holds

$$L_v = \bigsqcup_{m \in M_v^+/M_v^0} U_v m U_v.$$

We also set $A_v^+ := A_v \cap M_v^+$ and $A_v^0 := A_v \cap M_v^0$. Then this decomposition can be rewritten as

$$L_v = \bigsqcup\nolimits_{a \in A_v^+/A_v^0,\, \omega\in\Omega_v} U_v a\omega U_v,$$

where Ω_v is a finite subset of M_v.

Let us consider the representation $L_v \to \mathrm{GL}_n(K_v)$. Since the torus T is K_v-split, the image of T_v is diagonalizable, and we denote by $\Phi \subset \mathcal{X}(\mathsf{A})$ the set of the corresponding weights. We introduce a modified height function $\mathrm{H}'_v : L_v \to \mathbb{R}^+$ defined by

$$\mathrm{H}'_v(g) := \max(1, \|a\|'_v) \quad \text{when } g \in U_v a\omega U_v,$$

where $\|a\|'_v := \max_{\chi\in\Phi} |\chi(a)|_v$. It follows from the basic properties of norms and compactness that there exist $c_1, c_2 > 0$ such that $c_1 \mathrm{H}'_v \le \mathrm{H}_v \le c_2 \mathrm{H}'_v$. Therefore, it will be sufficient to analyse measures of the sets

$$B'_h := \{g \in L_v : \mathrm{H}'_v(g) \le h\}.$$

We obtain

$$m_{L_v}(B'_h) = \sum\nolimits_{a\in A_v^+/A_v^0 : \mathrm{H}'_v(a)\le h,\ \omega\in\Omega_v} m_{L_v}(U_v a\omega U_v).$$

Since $\omega U_v \omega^{-1} \cap U_v$ has finite index in both U_v and $\omega U_v\omega^{-1}$, it is clear that

$$m_{L_v}(U_v a U_v) \ll m_{L_v}(U_v a\omega U_v) \ll m_{L_v}(U_v a U_v).$$

Let $\delta : M_v \to \mathbb{R}^+$ denote the modular function of P_v. Then according to [37, Lem. 4.1.1], there exist $c_1, c_2 > 0$ such that

$$c_1\, \delta(m) \le m_{G_S}(U_v m U_v) \le c_2\, \delta(m) \quad \text{for every } m \in M_v^+.$$

Hence, it remains to investigate the function

$$V(h) := \sum\nolimits_{a\in A_v^+/A_v^0 : \mathrm{H}'_v(a)\le h} \delta(a).$$

Let us fix a basis of A_v/A_v^0 which is dual to the basis Π for simple roots. Then $(A_v/A_v^0)\otimes\mathbb{R}$ can be identified with $\mathbb{R}^r$, so that $\Lambda_v := A_v/A_v^0$ is a lattice in $\mathbb{R}^r$. For $\chi\in\Phi$, $|\chi(a)|_v = q_v^{\sum_{i=1}^r n_i(\chi)t_i}$, where $n_i(\chi)\in\mathbb{Q}$ and $(t_1,\ldots,t_r)$ denote the coordinates of a. Similarly, $\delta(a) = q_v^{\sum_{i=1}^r m_i t_i}$ with $m_i \in \mathbb{N}$. Using this notation, we rewrite $V(h)$ as

$$V(h) := \sum\nolimits_{t\in\Lambda_v\cap D_h} q_v^{\sum_i m_i t_i},$$

where

$$D_h := \left\{ t \in (\mathbb{R}^+)^r : \left(\sum_{i=1}^r n_i(\chi) t_i\right)^+ \le \log_{q_v}(h) \quad \text{for } \chi\in\Phi \right\}.$$

Here we use the notation $x^+ := \max(0, x)$. We shall also consider the integral

$$I(h) := \int_{D_h} q_v^{\sum_i m_i t_i}\, \mathrm{d}t. \tag{3.2}$$

Comparing $I(h)$ with suitable Riemann sums for $I(h)$, we deduce that there exist $c_1, c_2 > 0$ and $d_1, d_2 > 0$ such that

$$c_1\, I(h - d_1) \le V(h) \le c_2\, I(h + d_2)$$

for all sufficiently large h. Finally, it follows by a change of variables that there exists $c > 0$ such that $I(2h) \le c\, I(h)$ for $h \ge 1$. Hence, the similar estimate also holds for $V(h)$, which completes the proof of the lemma when S consists of a single completion.

For general S, we observe that the set $B_h^S \cap L$ is defined by the condition $\sum_{v\in S} \log \mathrm{H}_v(g_v) \le \log h$. Therefore, the required estimate follows from [18, Prop. 7.7]. □

Remark 3.3. *Here we outline a proof of the estimate* (1.2)*. It follows from the proof of Theorem 3.1 (with $L_S = \{e\}$) that the regular representation λ_{G_S} is weakly contained in the automorphic spectrum $\widehat{G}_S^{\mathrm{aut},0}$. Hence, for any absolutely continuous probability measure β,*

$$\|\lambda_{G_S}(\beta)\| \le \sup\left\{\|\pi(\beta)\| : \pi \in \widehat{G}_S^{\mathrm{aut},0}\right\}.$$

Let P_S be a minimal parabolic subgroup of G_S. Since P_S is amenable, the regular representation λ_{P_S} weakly contains trivial representation 1_{P_S}. Therefore, the induced representation $\lambda_{G_S} = \mathrm{Ind}_{P_S}^{G_S}(\lambda_{P_S})$ weakly contains the induced representation $\sigma := \mathrm{Ind}_{P_S}^{G_S}(1_{P_S})$ (see [2, F.3.5]*), and in particular,*

$$\|\sigma(\beta)\| \le \|\lambda_{G_S}(\beta)\|.$$

This suggests that the Harish-Chandra function Ξ_{G_S}, which is the spherical function associated to the representation σ can be used to estimate the norm from below. Indeed, it follows from (2.3) *that whenever the measure β is bi-invariant under U_S,*

$$\|\sigma(\beta)\| = \int_{G_S} \Xi_{G_S}(g)\, \mathrm{d}\beta(g)$$

since $\Xi_{G_S} \ge 0$. We recall furthermore that the Harish-Chandra function satisfies the inequality

$$\Xi_{G_S}(g) \ge C'\, \delta^{1/2}(a(g)),$$

where $\delta(g)$ is the modular character of a minimal parabolic subgroup, and $a(g)$ is the Cartan component of g w.r.t. the associated Cartan decomposition (see e.g. [37] *or* [38, Prop. 2.1] *in the totally disconnected case, and* [15, Thm. 4.6.5] *in the real case). Therefore,*

$$\left\|\sigma(\beta_h^S)\right\| \geq C' \, m_{G_S}(B_h^S)^{-1} \int_{B_h^S} \delta^{1/2}(a(g)) \, \mathrm{d}m_{G_S}(g).$$

The argument of the proof of Lemma 3.2 applied to G_S *and using the analysis of integrals of the form* (3.2)*, gives that*

$$\int_{B_h^S} \delta^{1/2}(a(g)) \, \mathrm{d}m_{G_S}(g) \geq C'' \, m_{G_S}(B_h^S)^{1/2}.$$

This implies (1.2)*.*

We now turn to Corollary 1.2, and note that the proof of Theorem 3.1 also provides a method for estimating the volumes $m_{G_S}(B_h^S)$. We demonstrate the results by executing explicit computations for some classical groups.

Proof of Corollary 1.2. We now fix $S = \{p\}$. Since we know that

$$\sup\left\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\right\} \gg \frac{m_{L_p}(L_p \cap B_h^{\{p\}})}{m_{G_p}(B_h^{\{p\}})}, \tag{3.3}$$

it remains to estimate the relevant volumes. It follows from the proof of Lemma 3.2 that

$$I(h) \ll m_{G_p}(B_h^{\{p\}}) \ll I(h),$$

where $I(h)$ is the integral defined in (3.2). We denote by α the maximum of $\sum_{i=1}^r m_i t_i$ on the domain D_1. We observe that

$$h^\alpha \ll I(h) \ll \mathrm{vol}(D_1) h^\alpha \ll_\epsilon h^{\alpha+\epsilon}$$

for all $\epsilon > 0$. In all the cases considered, the corresponding adjoint representations on the Lie algebra of G are irreducible, and the domain D_1 is given by

$$D_1 := \left\{ t \in (\mathbb{R}^+)^r : \sum_{i=1}^r n_i t_i \leq 1 \right\},$$

where $\sum_{i=1}^r n_i t_i$ is the highest weight of the representation. Then $\alpha = \max(m_i/n_i)$. These considerations can be used to estimate the norms for classical groups.

For $\mathsf{G} = \mathrm{SL}_n$ with $n \geq 3$, we apply the estimate (3.3) with the subgroup $\mathsf{L} = \mathrm{SL}_{n-1}$. We obtain:

$$m_{G_p}(B_h^{\{p\}}) \gg h^{n^2-n} \quad \text{and} \quad m_{L_p}(L_p \cap B_h^{\{p\}}) \ll_\epsilon h^{(n-1)^2-n+1+\epsilon}$$

for all $\epsilon > 0$, and it follows that

$$\sup\left\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\right\} \gg_\epsilon h^{-2(n-1)+\epsilon} \gg m_{G_p}(B_h^{\{p\}})^{-2/n+\epsilon}.$$

Let $\mathsf{G} = \mathrm{SO}_n$ with $n \geq 4$. The assumption that $p = 1 \pmod 4$ implies that G is split over $\mathbb{Q}_p$. We apply the estimate (3.3) with the subgroup $\mathsf{L} = \mathrm{SO}_{n-1}$. When n is even, we have

$$m_{G_p}(B_h^{\{p\}}) \gg h^{n(n-2)/4} \quad \text{and} \quad m_{L_p}(L_p \cap B_h^{\{p\}}) \ll_\epsilon h^{(n-2)^2/4+\epsilon}$$

for all $\epsilon > 0$, and

$$\sup\left\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\right\} \gg_\epsilon h^{-(n-2)/2+\epsilon} \gg m_{G_p}(B_h^{\{p\}})^{-2/n+\epsilon}.$$

Similarly, when n is odd,

$$m_{G_p}(B_h^{\{p\}}) \gg h^{(n-1)^2/4} \quad \text{and} \quad m_{L_p}(L_p \cap B_h^{\{p\}}) \ll_\epsilon h^{(n-1)(n-3)/4+\epsilon}$$

for all $\epsilon > 0$, and

$$\sup\left\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\right\} \gg_\epsilon h^{-(n-1)/2+\epsilon} \gg m_{G_p}(B_h^{\{p\}})^{-2/(n-1)+\epsilon}.$$

For $\mathsf{G} = \mathrm{Sp}_{2n}$ with $n \geq 2$, we apply the estimate (3.3) with the subgroup $\mathsf{L} = \mathrm{Sp}_{2(n-1)}$. We have

$$m_{G_p}(B_h^{\{p\}}) \gg h^{n(n+1)} \quad \text{and} \quad m_{L_p}(L_p \cap B_h^{\{p\}}) \ll_\epsilon h^{n(n-1)+\epsilon}$$

for all $\epsilon > 0$, and

$$\sup\left\{\|\pi(\beta_h^{\{p\}})\| : \pi \in \widehat{G}_p^{\mathrm{aut},0}\right\} \gg_\epsilon h^{-2n+\epsilon} \gg m_{G_p}(B_h^{\{p\}})^{-2/(n+1)+\epsilon}. \qquad \square$$

4. Lower bounds for operator norms in the discrete spectrum

In this section we discuss lower bounds for the norms of averaging operators for the discrete part of the automorphic spectrum and prove Theorem 1.3. We start by reviewing basic facts about the Hecke algebras and representations of p-adic groups (see, for instance, [7]). Let G be a classical simple simply connected algebraic $\mathbb{Q}$-group which is split over $\mathbb{Q}$. We fix a prime p such that G is isotropic and unramified over $\mathbb{Q}_p$. Let U_p be the hyperspecial maximal compact subgroup of G_p. We consider the Hecke algebra $\mathcal{H}(G_p, U_p)$ consisting of compactly supported U_p-biinvariant functions on G_p with the product defined by the convolution. The structure of $\mathcal{H}(G_p, U_p)$ can be explicitly described as follows. We fix a maximal split $\mathbb{Q}$-torus T of G and a Borel subgroup $\mathsf{B} = TN$ such that the Iwasawa decomposition $G_p = U_p T_p N_p$ holds. Let $\Lambda_p := T_p/(T_p \cap U_p) \simeq \mathbb{Z}^{\dim(\mathsf{T})}$ and $W := N_{\mathsf{G}}(\mathsf{T})/Z_{\mathsf{G}}(\mathsf{T})$ be the Weyl group. For $\beta \in \mathcal{H}(G_p, U_p)$, we set

$$\phi_\beta(t) := \Delta_p(t)^{1/2} \int_{N_p} \beta(tn)\, \mathrm{d}m_{N_p}(n), \quad t \in \Lambda_p = T_p/(T_p \cap U_p),$$

where Δ_p denotes the modular function of the group $T_p N_p$ and the invariant measure m_{N_p} on N_p is normalized so that $m_{N_p}(N_p \cap U_p) = 1$. It is known that the map $\beta \mapsto \phi_\beta$ defines an algebra-isomorphism between the Hecke algebra $\mathcal{H}(G_p, U_p)$ and the algebra $\mathbb{C}[\Lambda]^W$ of W-invariant polynomials in

$\mathbb{C}[\Lambda]$. This allows to give a complete description of the spherical functions in terms of the unramified characters $\chi : T_p \to \mathbb{C}^\times$. Given any such character, the corresponding spherical function is defined by

$$\omega_\chi(g) := \int_{U_p} \Delta_p^{1/2}(t(gu))\chi(t(gu))\,\mathrm{d}m_{U_p}(u), \quad g \in G_p,$$

where $t(\,\cdot\,)$ denotes the T_p-component with respect to the Iwasawa decomposition $G_p = U_p T_p N_p$. Moreover, two such spherical functions are equal iff the corresponding characters are conjugate with respect to the Weyl group W. We write $\mathcal{X}_p$ for the sets of unramified characters $\chi : T_p \to \mathbb{C}^\times$ and $\mathcal{X}_p^{temp}$ for the subset of characters with $|\chi| = 1$. We additionally note that for this correspondence

$$\chi(\phi_\beta) = \beta(\omega_\chi) \quad \text{for all } \chi \in \mathcal{X}_p \text{ and } \beta \in \mathcal{H}(G_p, U_p),$$

where $\beta(\omega_\chi) = \int_{G_p} \omega_\chi \,\mathrm{d}m_{G_p}$. On the other hand, the spherical functions naturally arise from irreducible spherical unramified representations of G_p. Given such a representation π_p, the corresponding spherical functions ω_{π_p} are defined by

$$\omega_{\pi_p}(g) := \left\langle \pi_p(g)v_{\pi_p}, v_{\pi_p} \right\rangle, \quad g \in G_p,$$

where v_{π_p} is a unit-norm U_p-invariant vector, which is known to be unique up to scalar multiple. It follows from uniqueness that

$$\pi_p(\beta)v_{\pi_p} = \beta(\omega_{\pi_p})v_{\pi_p}.$$

Under the above identifications, the tempered irreducible spherical unramified representations of G_p correspond to the characters in $\mathcal{X}_p^{temp}$, and the Plancherel formula holds: for all $\beta \in \mathcal{H}(G_p, U_p)$,

$$\int_{G_p} |\beta(g)|^2 \,\mathrm{d}m_{G_p}(g) = \int_{\chi \in \mathcal{X}_p^{temp}/W} |\beta(\omega_\chi)|^2 \,\mathrm{d}\nu_p(\chi),$$

where ν_p denotes the normalised spherical Plancherel measure for the group G_p (see [27, Thm. 2]).

Similarly, the spherical spectrum of G_∞ is parametrized by a subset of $\mathfrak{a}_\mathbb{C}^*$, where $\mathfrak{a}$ is the Lie algebra of the $\mathbb{R}$-split torus in G_∞, and $\mathfrak{a}_\mathbb{C}^*$ denotes the complexificated dual space. Then $i\mathfrak{a}^*$ gives the parametrization of the tempered spherical spectrum. Let Ω be a bounded domain in $i\mathfrak{a}^*$ with rectifiable boundary. The spherical Plancherel density associated to Ω is defined as

$$\Lambda_\Omega(t) := C(G_\infty) \int_{t\Omega} \frac{\mathbf{c}(\rho)}{\mathbf{c}(\lambda)} \,\mathrm{d}\lambda\,,$$

where $C(G_\infty)$ is an explicit positive constant, $\mathbf{c}$ is the Harish-Chandra $\mathbf{c}$-function of G_∞, and ρ is the half the sum of positive roots. It is known that

$$\Lambda_\Omega(t) = C(\Omega)t^d + O(t^{d-1}) \tag{4.1}$$

with explicit $C(\Omega) > 0$ and $d = d(G_\infty) \in \mathbb{N}$.

Let us now consider irreducible spherical unramified automorphic representations π of $G_\infty \times G_p$ discretely embedded in $L^2_0(X_{p,1})$. Such representation splits as a tensor product $\pi = \pi_\infty \otimes \pi_p$, where π_∞ and π_p denote the irreducible spherical representations of G_∞ and G_p respectively. We denote by $\lambda_{\pi_\infty} \in \mathfrak{a}^*_{\mathbb{C}}$ the infinitesimal character of the Archimedean component π_∞ and by $\lambda_{\pi_p} \in \mathcal{X}_p$ the characters corresponding to the representations π_p as described above. The representation $L^2_0(X_{p,1})$ can be viewed as a subrepresentation of $L^2_0(\mathsf{G}(\mathbb{A})/\mathsf{G}(\mathbb{Q}))$ consisting of functions invariant under $\prod_{q\neq p} \mathsf{G}(\mathbb{Z}_q)$. Therefore, with this notation, the equidistribution result of [29, 14] yields, in particular, that for all $\beta \in \mathcal{H}(G_p, U_p)$ and $t \geq 1$,

$$\sum_{\pi:\lambda_{\pi_\infty}\in t\Omega} \chi_{\pi_p}(\phi_\beta) = \Lambda_\Omega(t) \int_{\chi\in\mathcal{X}_p^{temp}/W} \chi(\phi_\beta)\, \mathrm{d}\nu_p(\chi) + O_\Omega\Big(\|\beta\|_{L^1} t^{d-\delta}\Big), \tag{4.2}$$

with explicit $\delta > 0$. Here the sum is taken over irreducible discretely embedded spherical unramified automorphic representations. Let us denote the number of such representations with $\lambda_{\pi_\infty} \in t\Omega$ by $N(t)$. Then taking $\beta = \chi_{U_p}$, we obtain from (4.2) that

$$N(t) = \Lambda_\Omega(t) + O_\Omega(t^{d-\delta}). \tag{4.3}$$

The estimate (4.2) was proved in [29] for SL_d with somewhat weaker error term and in [14] for groups satisfying a technical condition, which is satisfied in particular for all classical split groups. Formula (4.2) underlies our proof of Theorem 1.3, but we note that we will use only the existence of the limit, and not the effective error estimates that (4.2) provides.

Proof of Theorem 1.3. We consider the representation

$$\rho_t := \bigoplus_{\pi:\lambda_{\pi_\infty}\in t\Omega} \pi_p$$

of G_p, where the sum is taken over spherical unramified representations $\pi = \pi_\infty \otimes \pi_p$ of $G_\infty \times G_p$ discretely embedded in $L^2_0(X_{p,1})$. For each π, we denote by f_π the corresponding unit spherical vector and set $f_t := \sum_{\pi:\lambda_{\pi_\infty}\in t\Omega} f_\pi$. Then clearly

$$\|\rho_t(\beta) f_t\| \leq \sum_{\pi:\lambda_{\pi_\infty}\in t\Omega} \|\rho_t(\beta) f_\pi\| \leq N(t)\, \|\rho_t(\beta)\|\ . \tag{4.4}$$

On the other hand,

$$\rho_t(\beta) f_\pi = \pi_p(\beta) f_\pi = \beta(\omega_{\pi_p}) f_\pi.$$

Since the different components in the sum defining f_t are mutually orthogonal, we have

$$\|\rho_t(\beta)f_t\|^2 = \sum_{\pi:\lambda_{\pi_\infty}\in t\Omega} \|\pi_p(\beta)f_\pi\|^2 = \sum_{\pi:\lambda_{\pi_\infty}\in t\Omega} \left|\beta(\omega_{\pi_p})\right|^2$$
$$= \sum_{\pi:\lambda_{\pi_\infty}\in t\Omega} (\beta^* * \beta)(\omega_{\pi_p}) = \sum_{\pi:\lambda_{\pi_\infty}\in t\Omega} \chi_{\pi_p}(\phi_{\beta^* * \beta}).$$

Therefore,

$$\|\rho_t(\beta)f_t\|^2 = \Lambda_\Omega(t)\int_{\chi\in\mathcal{X}_p^{temp}/W} \chi(\phi_{\beta^* * \beta})\,\mathrm{d}\nu_p(\chi) + O_\Omega\Big(\|\beta^* * \beta\|_{L^1} t^{d-\delta}\Big)$$
$$= \Lambda_\Omega(t)\int_{\chi\in\mathcal{X}_p^{temp}/W} |\beta(\omega_\chi)|^2\,\mathrm{d}\nu_p(\chi) + O_\Omega\Big(\|\beta\|_{L^1}^2 t^{d-\delta}\Big).$$

Hence, applying the Plancherel Formula, we deduce that

$$\|\rho_t(\beta)f_t\|^2 = \Lambda_\Omega(t)\|\beta\|_{L^2}^2 + O_\Omega\Big(\|\beta\|_{L^1}^2 t^{d-\delta}\Big).$$

Hence, using (4.1) and (4.3), we conclude that for all $\varepsilon > 0$ and sufficiently large t (depending on β and ε),

$$\|\rho_t(\beta)f_t\| \geq (1-\varepsilon)N(t)\,\|\beta\|_{L^2}.$$

Comparing this estimate with (4.4), we deduce that

$$\left\|\rho_{p,1}^{\mathrm{cusp}}(\beta)\right\| \geq \|\rho_t(\beta)\| \geq (1-\varepsilon)\|\beta\|_{L^2},$$

for all $\varepsilon > 0$. This completes the proof. □

5. From discrepancy estimates to spectral gap

Let G be a simply connected K-simple linear algebraic group defined over a number field K and S a finite set of non-Archimedean completions of K_v. We use the notation introduced in Section 2, and in particular, we recall that $\Gamma_{S,\mathcal{O}}$ denotes the family of congruence lattice subgroups in the product $G_\infty \times G_S$. When S consists of unramified places, we denote by U_S the hyperspecial maximal compact subgroup of G_S. Recall that m_{G_S} and m_{G_∞} denote Haar measures on G_∞ and G_S, respectively. We normalize the Haar measures, so that $m_{G_S}(U_S) = 1$ and Γ_S has covolume one with respect to $m_{G_\infty} \times m_{G_S}$.

We will need the following lemma:

Lemma 5.1. *Assume that G_∞ is not compact. Let U be a compact open subgroup of G_S and $\Gamma := \Gamma_{S,\mathcal{O}} \cap U$. Let $F \subset G_\infty$ be a fundamental domain for Γ in G_∞. Then $F \times U$ is a fundamental domain for $\Gamma_{S,\mathcal{O}}$ in $G_\infty \times G_S$.*

Proof. Since G_∞ is not compact, and G is assumed to be simply connected, it follows from the Strong Approximation Property [31, §7.4] that the image of $\Gamma_{S,\mathcal{O}}$ in G_S is dense. Consider an arbitrary $(g_\infty, g_S) \in G_\infty \times G_S$. Then it follows from density that there exists $\gamma_1 \in \Gamma_{S,\mathcal{O}}$ such that $\gamma_1 \in g_S^{-1} U_S$. Hence, $(g_\infty \gamma_1, g_S \gamma_1) \in G_\infty \times U$. Furthermore, since F is a fundamental domain for Γ in G_∞, there exists $\gamma_2 \in \Gamma$ such that $g_\infty \gamma_1 \gamma_2 \in F$. Then $(g_\infty \gamma_1 \gamma_2, g_S \gamma_1 \gamma_2) \in F \times U$. This proves that $G_\infty \times G_S = (F \times U)\Gamma_{S,\mathcal{O}}$.

Suppose that $(F \times U)\gamma_1 \cap (F \times U)\gamma_2 \neq \emptyset$ for some $\gamma_1, \gamma_2 \in \Gamma_{S,\mathcal{O}}$. Then it follows that $\gamma_1\gamma_2^{-1} \in \Gamma_{S,\mathcal{O}} \cap U = \Gamma$, and $F\gamma_1\gamma_2^{-1} \cap F \neq \emptyset$. Hence, we conclude that $\gamma_1 = \gamma_2$. This completes the proof of the lemma. □

We consider the following counting problem for the congruence lattices $\Gamma_{S,\mathcal{O}}$. Let us fix a left-invariant Riemannian metric on G_∞. For $x \in G_\infty$ and $r > 0$, we denote by $B(x, r)$ the ball centered at x of radius r in G_∞. Let B be a bounded measurable subset of G_S of positive measure. We consider the counting function $|\Gamma_{S,\mathcal{O}} \cap (B(x, r) \times B)|$ and the discrepancy

$$\mathcal{D}(\Gamma_{S,\mathcal{O}}, B(x, r) \times B) := \left| \frac{|\Gamma_{S,\mathcal{O}} \cap (B(x, r) \times B)|}{m_{G_S}(B)} - \frac{m_{G_\infty}(B(e, r))}{|\Gamma_S/\Gamma_{S,\mathcal{O}}|} \right|.$$

We shall show that estimates on the discrepancy can be used to establish norm bounds for $\|\pi(\beta)\|$ for the representations π arising from the G_S-action on the spaces $X_{S,\mathcal{O}}$, where β is the probability measure supported on the set $B \subset G_S$.

We will be interested in utilizing L^2-bounds for the discrepancy function as x varies over G_∞. Let us assume that the subset B is bi-invariant under U_S. Then the discrepancy is left invariant under the subgroup $\Gamma^0_{S,\mathcal{O}} := \Gamma_{S,\mathcal{O}} \cap U_S$ (viewed as a subgroup of G_∞), so that it defines a function on $\Gamma^0_{S,\mathcal{O}} \backslash G_\infty$. We set

$$E_{S,\mathcal{O}}(r, B) := \left\| \mathcal{D}(\Gamma_{S,\mathcal{O}}, B(\,\cdot\,, r) \times B) \right\|_{L^2(\Gamma^0_{S,\mathcal{O}} \backslash G_\infty)},$$

where $\Gamma^0_{S,\mathcal{O}} \backslash G_\infty$ is equipped with the invariant probability measure.

We now formulate and prove a more precise version of Theorem 1.4, as follows.

Theorem 5.2. *Assume that S consists of unramified places, and G_∞ is not compact. Let B be a non-empty bounded U_S-bi-invariant subset of G_S. Fix an irreducible unitary representation π of G_S which is discretely embedded in $\rho_{S,\mathcal{O}}$. Then there exists $r_0(\pi) > 0$ such that for $0 < r \leq r_0(\pi)$, the operator $\pi(\beta)$ satisfies the bound*

$$\|\pi(\beta)\| \leq 2\, |\Gamma_S/\Gamma_{S,\mathcal{O}}|\, m_{G_\infty}(B(e, r))^{-1} E_{S,\mathcal{O}}(r, B).$$

Proof. Consider

$$\chi_r(g_\infty, g_S) := \chi_{B(e,r)}(g_\infty)\chi_{U_S}(g_S), \qquad \text{for } (g_\infty, g_S) \in G_\infty \times G_S,$$

namely the characteristic function of the subset $B(e,r)\times U_S$ of $G_\infty\times G_S$. Then

$$\phi_r(g_\infty,g_S):=\sum\nolimits_{\delta\in\Gamma_{S,\mathcal{O}}}\chi_r(g_\infty\delta,g_S\delta)$$

defines a function on the space $X_{S,\mathcal{O}}=(G_\infty\times G_S)/\Gamma_{S,\mathcal{O}}$. By the Fubini–Tonelli Theorem, for $x\in G_\infty$ and $u\in U_S=U_S^{-1}$,

$$\begin{aligned}\int_B \phi_r(b^{-1}(x^{-1},u))\,\mathrm{d}m_{G_S}(b)&\\ &=\sum\nolimits_{\delta\in\Gamma_{S,\mathcal{O}}}\int_B \chi_r(x^{-1}\delta,b^{-1}u\delta)\,\mathrm{d}m_{G_S}(b)\\ &=\sum\nolimits_{\delta\in\Gamma_{S,\mathcal{O}}}\int_B \chi_{B(e,r)}(x^{-1}\delta)\chi_{U_S}(b^{-1}u\delta)\,\mathrm{d}m_{G_S}(b)\\ &=\sum\nolimits_{\delta\in\Gamma_{S,\mathcal{O}}\cap(B(x,r)\times G_S)} m_{G_S}(u\delta U_S\cap B).\end{aligned}$$

Since the set B is U_S-bi-invariant, if $\delta\in B$, we have $u\delta U_S\subset B$, and if $\delta\notin B$, we have $u\delta U_S\cap B=\emptyset$. Hence, since $m_{G_S}(U_S)=1$, it follows that for every $u\in U_S$,

$$|\Gamma_{S,\mathcal{O}}\cap(B(x,r)\times B)|=\int_B \phi_r(b^{-1}(x^{-1},u))\,\mathrm{d}m_{G_S}(b). \tag{5.1}$$

We also compute:

$$\begin{aligned}\int_{X_{S,\mathcal{O}}}\phi_r\,\mathrm{d}\mu_{S,\mathcal{O}}&=\int_{(G_\infty\times G_S)/\Gamma_{S,\mathcal{O}}}\left(\sum_{\delta\in\Gamma_{S,\mathcal{O}}}\chi_r(g\delta)\right)\mathrm{d}\mu_{S,\mathcal{O}}(g)\\ &=\int_{G_\infty\times G_S}\chi_r(g)\,\frac{d(m_{G_\infty}\times m_{G_S})(g)}{|\Gamma_S/\Gamma_{S,\mathcal{O}}|}\\ &=\frac{m_{G_\infty}(B(e,r))m_{G_S}(U_S)}{|\Gamma_S/\Gamma_{S,\mathcal{O}}|}=\frac{m_{G_\infty}(B(e,r))}{|\Gamma_S/\Gamma_{S,\mathcal{O}}|}.\end{aligned}$$

Therefore, we conclude that for every $u\in U_S$,

$$\begin{aligned}&\mathcal{D}(\Gamma_{S,\mathcal{O}},B(x,r)\times B)\\ &=\left|\frac{1}{m_{G_S}(B)}\int_B\phi_r(b^{-1}(x^{-1},u))\,\mathrm{d}m_{G_S}(b)-\int_{X_{S,\mathcal{O}}}\phi_r\,\mathrm{d}\mu_{S,\mathcal{O}}\right|\\ &=\left|\rho_{S,\mathcal{O}}(\beta)\left(\phi_r-\int_{X_{S,\mathcal{O}}}\phi_r\,\mathrm{d}\mu_{S,\mathcal{O}}\right)(x^{-1},u)\right|.\end{aligned}$$

Let Ω be a measurable fundamental domain for $\Gamma^0_{S,\mathcal{O}}:=\Gamma_{S,\mathcal{O}}\cap U_S$ in G_∞. Then for any bounded measurable subset Q of G_∞ that injects on $G_\infty/\Gamma^0_{S,\mathcal{O}}$ (in particular, for $Q=\Omega$),

$$\left\|\rho_{S,\mathcal{O}}(\beta)\left(\phi_r-\int_{X_{S,\mathcal{O}}}\phi_r\,\mathrm{d}\mu_{S,\mathcal{O}}\right)\right\|_{L^2(Q^{-1}\times U_S)}\le E_{S,\mathcal{O}}(r,B). \tag{5.2}$$

By Lemma 5.1, $\Omega \times U_S$ is a fundamental domain for $\Gamma_{S,\mathcal{O}}$ in $G_\infty \times G_S$. Hence, we deduce from (5.2) that

$$\left\| \rho_{S,\mathcal{O}}(\beta)\phi_r - \int_{X_{S,\mathcal{O}}} \phi_r \, \mathrm{d}\mu_{S,\mathcal{O}} \right\|_{L^2(X_{S,\mathcal{O}})} \le E_{S,\mathcal{O}}(r,B). \tag{5.3}$$

For $x \in G_\infty$ and $r > 0$, let $\chi_{x,r}$ denote the characteristic function of the subset $B(x,r) \times U_S$ of $G_\infty \times G_S$. Because of left-invariance of the metric, $\chi_{x,r}(g) = \chi_r(x^{-1}g)$. We set

$$\phi_{x,\epsilon}(g) := \sum_{\gamma \in \Gamma_{S,\mathcal{O}}} \chi_{x,\epsilon}(g\gamma),$$

which defines a function in $L^2(X_{S,\mathcal{O}})$. For $x \in G_\infty$, let us consider the operators

$$\rho_{\infty,\mathcal{O}}(x) : L^2(X_{S,\mathcal{O}}) \longrightarrow L^2(X_{S,\mathcal{O}}) : \phi \longmapsto \phi \circ x^{-1}.$$

We observe that $\|\rho_{\infty,\mathcal{O}}(x)\| = 1$, $\rho_{\infty,\mathcal{O}}(x)$ commutes with $\rho_{S,\mathcal{O}}(\beta)$, and $\phi_{x,r} = \rho_{\infty,\mathcal{O}}(x)(\phi_r)$. Hence, it follows from (5.3) that for every $x \in G_\infty$,

$$\left\| \rho_{S,\mathcal{O}}(\beta)\phi_{x,r} - \int_{X_{S,\mathcal{O}}} \phi_r \, \mathrm{d}\mu_{S,\mathcal{O}} \right\|_{L^2(X_{S,\mathcal{O}})} \le E_{S,\mathcal{O}}(r,B). \tag{5.4}$$

We use this estimate to conclude the proof of the theorem as follows. Let π be an irreducible unitary representation of G_S which is discretely embedded in $L^2_0(X_{S,\mathcal{O}})$. We observe that since B is U_S-bi-invariant, the image of $\pi(\beta)$ consists of U_S-invariant vectors. Hence, if π is not spherical, then $\pi(\beta) = 0$. Now suppose that π is spherical and denote by F_π the function in $L^2_0(X_{S,\mathcal{O}})$ which is the unique (up to a phase factor) U_S-invariant unit vector of π. Then

$$\omega_\pi(g) := \langle \pi(g)F_\pi, F_\pi \rangle, \qquad \text{with } g \in G_S,$$

is the spherical function associated to the representation π. We have

$$\pi(\beta)F_\pi = \beta(\omega_\pi)F_\pi \quad \text{and} \quad \|\pi(\beta)\|^2 = \|\pi(\beta * \beta^*)\| = |\beta(\omega_\pi)|^2,$$

and $\pi(\beta)$ acts as a projection onto the space of spherical vectors, namely $\pi(\beta)F = \lambda_F F_\pi$. F is orthogonal to the constant functions, and so it follows from (5.4) that when F has unit norm,

$$\begin{aligned}
\left|\langle\rho_{S,\mathcal{O}}(\beta^* * \beta)\phi_{x,r}, F\rangle\right| &= \left|\langle\rho_{S,\mathcal{O}}(\beta)\phi_{x,r}, \pi(\beta)F\rangle\right| \\
&= \left|\left\langle \rho_{S,\mathcal{O}}(\beta)\phi_{x,r} - \int_{X_{S,\mathcal{O}}} \phi_r \,\mathrm{d}\mu_{S,\mathcal{O}}, \pi(\beta)F \right\rangle\right| \\
&\le \left\|\rho_{S,\mathcal{O}}(\beta)\phi_{x,r} - \int_{X_{S,\mathcal{O}}} \phi_r \,\mathrm{d}\mu_{S,\mathcal{O}}\right\|_{L^2(X_{S,\mathcal{O}})} \\
&\qquad \cdot \|\pi(\beta)F\|_{L^2(X_{S,\mathcal{O}})} \\
&\le E_{S,\mathcal{O}}(r,B)\ \|\pi(\beta)\|\,.
\end{aligned}$$

On the other hand, let us choose a sequence of unit vectors ψ_i in the representation π such that

$$\|\pi(\beta^* * \beta)\psi_i\| \longrightarrow \|\pi(\beta^* * \beta)\|.$$

Then $\pi(\beta^* * \beta)\psi_i = \lambda_i F_\pi$ with $\lambda_i \ge 0$ and $\lambda_i \to \|\pi(\beta^* * \beta)\| = \|\pi(\beta)\|^2$, and

$$\left|\langle\rho_{S,\mathcal{O}}(\beta^* * \beta)\phi_{x,r}, \psi_i\rangle\right| = \left|\langle\phi_{x,r}, \rho_{S,\mathcal{O}}(\beta^* * \beta)\psi_i\rangle\right| = \lambda_i\left|\langle\phi_{x,r}, F_\pi\rangle\right|.$$

Hence, we obtain the following norm bound

$$\|\pi(\beta)\| \le \left|\langle\phi_{x,r}, F_\pi\rangle\right|^{-1} E_{S,\mathcal{O}}(r,B) \tag{5.5}$$

provided that $\langle\phi_{x,r}, F_\pi\rangle \neq 0$.

Let us now consider the function

$$f(g_\infty) := F_\pi\left(g_\infty \Gamma_{S,\mathcal{O}}\right) \qquad \text{with } g_\infty \in G_\infty.$$

We note that since F_π is U_S-invariant, we have $f(g_\infty) = F_\pi\left((g_\infty, u)\Gamma_{S,\mathcal{O}}\right)$ for all $u \in U_S$. In particular, f is a well-defined measurable locally L^2-integrable function on G_∞. Since $G_\infty \Gamma_{S,\mathcal{O}}$ is dense in $G_\infty \times G_S$, it is clear that $f \neq 0$. We compute:

$$\begin{aligned}
\langle\phi_{x,r}, F_\pi\rangle &= \int_{X_{S,\mathcal{O}}} \left(\sum_{\gamma\in\Gamma_{S,\mathcal{O}}} \chi_{x,r}(g\gamma)\right) \overline{F_\pi(g\Gamma_{S,\mathcal{O}})}\, \mathrm{d}\mu_{S,\mathcal{O}}(g\Gamma_{S,\mathcal{O}}) \\
&= \int_{G_\infty\times G_S} \chi_{x,r}(g)\overline{F_\pi(g\Gamma_{S,\mathcal{O}})}\, \frac{d(m_{G_\infty}\times m_{G_S})(g)}{|\Gamma_S/\Gamma_{S,\mathcal{O}}|}.
\end{aligned}$$

We recall that $\chi_{x,r}$ is the characteristic function of the set $B(x,r)\times U_S$. Since F_π is U_S-invariant,

$$\langle\phi_{x,r}, F_\pi\rangle = |\Gamma_S/\Gamma_{S,\mathcal{O}}|^{-1} \int_{B(x,r)} f\,\mathrm{d}m_{G_\infty}.$$

It follows from the Local Ergodic Theorem (see, for instance, [28, Cor. 2.14]) that for almost every $x \in G_\infty$,

$$\frac{1}{m_{G_\infty}(B(x,r))} \int_{B(x,r)} f\,\mathrm{d}m_{G_\infty} \longrightarrow f(x) \qquad \text{as } r \longrightarrow 0^+.$$

For a positive measure set of choices of x,

$$|f(x)| \geq \frac{1}{2}\|f\|_\infty = \frac{1}{2}\|F_\pi\|_\infty \geq \frac{1}{2}\|F_\pi\|_2 = \frac{1}{2},$$

so that choosing a point x where the foregoing inequality holds, and in addition convergence in the local ergodic theorem holds, it follows that for $0 < r < r_0(\pi)$

$$|\langle \phi_{x,r}, F_\pi \rangle| \geq \frac{1/2}{|\Gamma_S/\Gamma_{S,\mathcal{O}}|} m_{G_\infty}(B(x,r)).$$

Therefore, the estimate (5.5) implies the claim of the theorem. □

6. Spectral gap for forms of SL_2

Let G be a linear algebraic group defined over $\mathbb{Q}$ which is a form of SL_2. Namely, G can be realised as the group of norm one elements of a quaternion algebra D defined over $\mathbb{Q}$:

$$\mathsf{G} := \{x \in \mathsf{D} : N(x) = 1\}.$$

Throughout this section, we always assume that G is isotropic over $\mathbb{R}$, or equivalently, $\mathsf{D}(\mathbb{R}) \simeq \mathrm{M}_2(\mathbb{R})$. We fix an order Λ of $\mathsf{D}(\mathbb{Q})$ such that $N(\Lambda) \subset \mathbb{Z}$. Then N is an integral quadratic form with respect to this integral structure. The p-adic norm $\|\cdot\|_p$ on $\mathsf{D}(\mathbb{Q}_p)$ is defined with respect to the order Λ.

The group $\Gamma := G_\infty \cap \Lambda$ is an arithmetic lattice in G_∞ corresponding to the integral structure defined by Λ. More generally, for a prime p, we consider the group $\Gamma_p := G_\infty \cap \Lambda[p^{-1}]$ which is a lattice in the product $G_\infty \times G_p$. For $\ell \in \mathbb{N}$ coprime to p, we also consider the congruence subgroups $\Gamma_{p,\ell} := \{\gamma \in \Gamma_p : \gamma = I \pmod{\ell}\}$. The goal of this section is to analyse the unitary representations $\rho_{p,\ell}$ of G_p acting on the space $L_0^2(X_{p,\ell})$, where $X_{p,\ell} := (G_\infty \times G_p)/\Gamma_{p,\ell}$. More specifically, we will be interested in the averaging operators $\rho_{p,\ell}(\beta_h)$ defined with respect to the sets

$$B_h := \{b \in G_p : \|b\|_p \leq h\}.$$

Let m_p be the Haar measure on G_p normalized so that $m_p(\mathsf{G}(\mathbb{Z}_p)) = 1$, and m_∞ the Haar measure on G_∞ normalized so that $m_\infty(G_\infty/\Gamma) = 1$.

Our main result in this section is the following:

Theorem 6.1. *Suppose that G is unramified over $\mathbb{Q}_p$. Then for every $\ell \in \mathbb{N}$ coprime to p,*

$$\|\rho_{p,\ell}(\beta_h)\| \ll_{p,\ell} m_p(B_h)^{-\kappa},$$

where $\kappa = 1/4$ if G is anisotropic over $\mathbb{Q}$, and $\kappa = 1/16$ if G is isotropic over $\mathbb{Q}$, It turns out that the representation-theoretic problem of estimating the norms $\|\rho_{p,\ell}(\beta_h)\|$ is closely related to properties of the distribution of rational points contained in the quadratic surface $N(x)=1$. For $h\in\mathbb{N}$ and a compactly supported function $w:\mathsf{D}(\mathbb{R})\to\mathbb{R}$, we consider the counting function:

$$\mathrm{N}_h(N):=\sum_{x\in\Lambda:N(x)=h^2} w(h^{-1}x).$$

More generally, for $\ell\in\mathbb{N}$ and a coset representative $\xi\in\Lambda/\ell\Lambda$, we define

$$\mathrm{N}_h(N,w;\xi):=\sum_{x\in\xi+\ell\Lambda:N(x)=h^2} w(h^{-1}x).$$

The behaviour of $\mathrm{N}_h(N,w;\xi)$ as $h\to\infty$ captures the distribution of the set of rational points $h^{-1}\Lambda$ with the prescribed congruence condition. In order to state an asymptotic formula for $N_h(N,w;\xi)$, we need to introduce local densities.

For a compactly supported function w on $\mathsf{D}(\mathbb{R})$, we define the *Archimedean local density* as

$$\sigma_\infty(N,w):=\lim_{\epsilon\to 0^+}(2\epsilon)^{-1}\int_{|N(x)-1|\le\epsilon} w(x)\,\mathrm{d}x,$$

where the measure on $D(\mathbb{R})\simeq\mathbb{R}^4$ is normalized so that the lattice Λ has covolume one.

Let $\ell=\prod_q q^{s_q}$ be the prime decomposition of ℓ and $\xi\in\Lambda/\ell\Lambda$. We define, for a general integer $e\in\mathbb{N}$,

$$\mathrm{N}_h(N,q^e,q^{s_q},\xi):=\left|\{x \bmod q^{e+s_q}:\ x=\xi \bmod q^{s_q},\ N(x)=h^2 \bmod q^e\}\right|.$$

Then the *q-adic local density* is defined as

$$\sigma_q(N,\xi,h):=\lim_{e\to\infty}\frac{N_h(q^e,q^{s_q},\xi)}{q^{3e}}.$$

We also set

$$\sigma_f(N,\xi,h):=\prod\nolimits_q \sigma_q(N,\xi,h).$$

With these notations we state:

Theorem 6.2. *For every smooth compactly supported function $w:\mathsf{D}(\mathbb{R})\to\mathbb{R}$ and $\xi\in\Lambda/\ell\Lambda$,*

$$\mathrm{N}_h(N,w;\xi)=\ell^{-4}\sigma_\infty(N,w)\sigma_f(N,\xi,h)h^2+O_{w,\ell,\epsilon}(h^{3/2+\epsilon})\quad \textit{for all } \epsilon>0.$$

To prove Theorem 6.2, we follow closely the refined circle method for quadratic forms developed by Heath-Brown [22], for the case of forms in four variables. This method is particularly suitable for the derivation of estimates which are uniform over families of functions w, which will be crucial for our results, cf. Theorem 6.6 below. We note however that in order to establish uniform bounds on $\|\rho_{p,\ell}(\beta_h)\|$ over ℓ, we have to keep track of additional congruence conditions beyond those considered in [22]. Furthermore, we need explicit bounds over families of functions w. This causes considerable technical complications, which motivates our decision to give a full account of the necessary arguments below.

Making the identification $\Lambda \simeq \mathbb{Z}^4$, we view the reduced norm N as an integral quadratic form in four variable. We denote by $|\cdot|$ the Euclidean norm on $\mathsf{D}(\mathbb{R}) \simeq \mathbb{R}^4$ defined by this identification.

The starting point of our argument is a convenient representation of the Dirac function $\delta_n = 1$ if $n = 0$ and $\delta_n = 0$ for $n \neq 0$. We recall (see [22, Thm. 1] and [12]) that for every $Q > 1$,

$$\delta_n = c_Q Q^{-2} \sum_{k=1}^{\infty} \sum_{a \bmod k}{}^{*} e_k(an) H(Q^{-1}k, Q^{-2}n), \tag{6.1}$$

where $c_Q = 1 + O_N(Q^{-N})$ for any $N > 0$, the sum is taken over a coprime to k, $e_k(x) = \exp(2\pi i x/k)$, and H is a certain explicit smooth function on $(0, \infty) \times \mathbb{R}$. To simplify notation, we set $F_h(x) = N(x) - h^2$. It follows from (6.1) that

$$\mathrm{N}_h(N, w; \xi) = c_h h^{-2} \sum_{x \in \xi + \ell\mathbb{Z}^4} \sum_{k=1}^{\infty} \sum_{a \bmod k}^{*} w(h^{-1}x) e_k(aF_h(x)) H\big(h^{-1}k, h^{-2}F_h(x)\big),$$

and we are required to estimate the following sum

$$\sum_{x \in \xi + \ell\mathbb{Z}^4} w(h^{-1}x) e_k(aF_h(x)) H\big(h^{-1}k, h^{-2}F_h(x)\big) = \sum_{b \bmod k} \sum_{\substack{z \bmod k\ell \\ z = \xi(\ell), z = b(k)}} e_k(aF_h(b)) \left(\sum_{y \in \mathbb{Z}^4} f(y) \right),$$

where

$$f(y) := w(h^{-1}(z + (k\ell)y)) H\big(h^{-1}k, h^{-2}F_h(z + (k\ell)y)\big)$$

is a smooth compactly supported function. By the Poisson Summation Formula,

$$\sum_{y \in \mathbb{Z}^4} f(y) = \sum_{c \in \mathbb{Z}^4} \widehat{f}(c),$$

where

$$\widehat{f}(c) = \int_{\mathbb{R}^4} f(y) e^{-2\pi i (c \cdot y)}\, \mathrm{d}y = (k\ell)^{-4} e_{k\ell}(c \cdot z) I_{k,\ell}(c)$$

with

$$\begin{aligned} I_{k,\ell}(c) &:= \int_{\mathbb{R}^4} w(h^{-1}x) H(h^{-1}k, h^{-2}F_h(x)) e_{k\ell}(-c\cdot x)\,\mathrm{d}x \\ &= h^4 \int_{\mathbb{R}^4} w(x) H(h^{-1}k, N(x)-1) e_{k\ell}(-hc\cdot x)\,\mathrm{d}x. \end{aligned}$$

Hence, we deduce that

$$\mathrm{N}_h(N,w;\xi) = c_h h^{-2} \sum_{c\in\mathbb{Z}^4} \sum_{k=1}^{\infty} (k\ell)^{-4} S_k(c;\xi) I_{k,\ell}(c), \tag{6.2}$$

where

$$\begin{aligned} S_k(c;\xi) &:= \sideset{}{^*}\sum_{a \bmod k} \sum_{b \bmod k} \sum_{\substack{z \bmod k\ell \\ z=\xi(\ell), z=b(k)}} e_{k\ell}(a\ell F_h(b) + c\cdot z) \\ &= \sideset{}{^*}\sum_{a \bmod k} \sum_{\substack{z \bmod k\ell \\ z=\xi(\ell)}} e_{k\ell}(a\ell F_h(z) + c\cdot z) \\ &= \sideset{}{^*}\sum_{a \bmod k} e_k(-ah^2) \sum_{\substack{z \bmod k\ell \\ z=\xi(\ell)}} e_{k\ell}(a\ell N(z) + c\cdot z). \end{aligned}$$

First, we observe that the sum $S_k(c;\xi)$ has the following multiplicative property:

Lemma 6.3. *Let $k = k_1k_2$ and $\ell = \ell_1\ell_2$ such that $k_1\ell_1$ is coprime to $k_2\ell_2$. Choose integers $\bar{k}_1, \bar{k}_2, \bar{\ell}_1, \bar{\ell}_2$ such that*

$$k_1\bar{k}_1 = 1\,(k_2\ell_2),\quad k_2\bar{k}_2 = 1\,(k_1\ell_1),\quad \ell_1\bar{\ell}_1 = 1\,(k_2\ell_2),\quad \ell_2\bar{\ell}_2 = 1\,(k_1\ell_1).$$

Then

$$S_k(c;\xi\,(\ell)) = S_{k_1}(\bar{k}_2\bar{\ell}_2 c;\xi\,(\ell_1))\, S_{k_2}(\bar{k}_1\bar{\ell}_1 c;\xi\,(\ell_2)).$$

Proof. We also choose integers $\bar{\ell}_1$ and $\bar{\ell}_2$ such that $\ell_1\bar{\ell}_1 = 1\,(k_2\ell_2)$ and $k_2\bar{k}_2 = 1\,(k_1\ell_1)$. We write

$$z = k_2\ell_2\bar{k}_2\bar{\ell}_2 z_1 + k_1\ell_1\bar{k}_1\bar{\ell}_1 z_2.$$

If z_1 runs over the integral vectors modulo $k_1\ell_1$ such that $z_1 = \xi\,(\ell_1)$, and z_2 runs over the integral vectors modulo $k_2\ell_2$ such that $z_2 = \xi\,(\ell_2)$, then z runs precisely over the integral vectors modulo $k\ell$ such that $z = \xi\,(\ell)$. Then since the map N is a quadratic form,

$$\begin{aligned} &N(k_2\ell_2\bar{k}_2\bar{\ell}_2 z_1 + k_1\ell_1\bar{k}_1\bar{\ell}_1 z_2) \\ &\qquad = (k_2\ell_2\bar{k}_2\bar{\ell}_2)^2 N(z_1) + (k_1\ell_1\bar{k}_1\bar{\ell}_1)^2 N(z_2) \pmod{k_1k_2\ell}, \end{aligned}$$

and

$$\begin{aligned} e_{k\ell}(a\ell N(z)) &= e_{k_1k_2\ell}\big(a(\ell_1\ell_2(k_2\ell_2\bar{k}_2\bar{\ell}_2)^2N(z_1)+\ell_1\ell_2(k_1\ell_1\bar{k}_1\bar{\ell}_1)^2N(z_2))\big) \\ &= e_{k_1\ell_1}\big(a\ell_1k_2\ell_2^2\bar{k}_2^2\bar{\ell}_2^2N(z_1)\big)e_{k_2\ell_2}\big(a\ell_2k_1\ell_1^2\bar{k}_1^2\bar{\ell}_1^2N(z_2)\big) \\ &= e_{k_1\ell_1}\big(a\bar{k}_2\ell_1N(z_1)\big)e_{k_2\ell_2}\big(a\bar{k}_1\ell_2N(z_2)\big). \end{aligned}$$

Similarly,

$$\begin{aligned} e_{k\ell}(c\cdot z) &= e_{k_1\ell_1k_2\ell_2}\big(c\cdot k_2\ell_2\bar{k}_2\bar{\ell}_2z_1+c\cdot k_1\ell_1\bar{k}_1\bar{\ell}_1z_2\big) \\ &= e_{k_1\ell_1}\big((\bar{k}_2\bar{\ell}_2c)\cdot z_1\big)e_{k_2\ell_2}\big((\bar{k}_1\bar{\ell}_1c)\cdot z_2\big). \end{aligned}$$

Setting

$$S_k(a,k,\xi\,(\ell)) := \sum_{\substack{z \bmod k\ell\\ z=\xi(\ell)}} e_{k\ell}(a\ell N(z)+c\cdot z),$$

we conclude that

$$S_k(a,k,\xi\,(\ell)) = S_{k_1}\big(a\bar{k}_2,\bar{k}_2\bar{\ell}_2c;\xi\,(\ell_1)\big)\,S_{k_2}\big(a\bar{k}_1,\bar{k}_1\bar{\ell}_1c;\xi\,(\ell_2)\big).$$

Then

$$S_k(c;\xi\,(\ell)) = \sum_{a \bmod k}{}^{*}\, e_k(-ah^2)S_{k_1}\big(a\bar{k}_2,\bar{k}_2\bar{\ell}_2c;\xi\,(\ell_1)\big)\,S_{k_2}\big(a\bar{k}_1,\bar{k}_1\bar{\ell}_1c;\xi\,(\ell_2)\big).$$

We observe that every residue a mod k coprime to k can be uniquely written as $k_2a_1+k_1a_2$, where a_i is a residue modulo k_i coprime to k_i. Hence, the above sum can be rewritten as:

$$\begin{aligned} &\sum_{a_1 \bmod k_1}{}^{*}\ \sum_{a_2 \bmod k_2}{}^{*}\, e_k(-(k_2a_1+k_1a_2)h^2)S_{k_1}\big((k_2a_1+k_1a_2)\bar{k}_2,\bar{k}_2\bar{\ell}_2c;\xi\,(\ell_1)\big) \\ &\qquad\qquad \times S_{k_2}\big((k_2a_1+k_1a_2)\bar{k}_1,\bar{k}_1\bar{\ell}_1c;\xi\,(\ell_2)\big) \\ &= \sum_{a_1 \bmod k_1}{}^{*}\ \sum_{a_2 \bmod k_2}{}^{*}\, e_{k_1}(-a_1h^2)e_{k_2}(-a_2h^2)S_{k_1}\big(a_1,\bar{k}_2\bar{\ell}_2c;\xi\,(\ell_1)\big) \\ &\qquad\qquad \times S_{k_2}\big(a_2,\bar{k}_1\bar{\ell}_1c;\xi\,(\ell_2)\big) \\ &= S_{k_1}(\bar{k}_2\bar{\ell}_2c;\xi\,(\ell_1))\,S_{k_2}(\bar{k}_1\bar{\ell}_1c;\xi\,(\ell_2)). \end{aligned}$$

This proves the lemma. □

We record the required properties of the term $S_k(c;\xi)$:

Lemma 6.4. *For every $c\in\mathbb{Z}^4$ and $\xi\in\Lambda/\ell\Lambda$,*

(i) $|S_k(c;\xi)| \ll_\ell k^3$,

(ii) *For all $\epsilon>0$,*

$$\sum_{k\le X}|S_k(c;\xi)| \ll_{\ell,\epsilon} X^{7/2+\epsilon}|c|^\epsilon \quad \textit{when } c\neq 0,$$

and

$$\sum_{k\le X} |S_k(0;\xi)| \ll_{\ell,\epsilon} X^{7/2+\epsilon}h^{\epsilon}.$$

Proof. The proof of (i) proceeds as [22, Lem. 25] with minor modifications. Applying the Cauchy–Schwartz inequality to the sum

$$S_k(c;\xi) = \sum_{a \bmod k}^{*} e_k(-ah^2) \sum_{\substack{z \bmod k\ell \\ z=\xi(\ell)}} e_{k\ell}(a\ell N(z) + c\cdot z),$$

we obtain that

$$(6.3)\quad |S_k(c;\xi)|^2 \le \phi(k) \sum_{a \bmod k}^{*} \left| \sum_{\substack{z \bmod k\ell \\ z=\xi(\ell)}} e_{k\ell}\Big(a\ell N(z) + c\cdot z\Big) \right|^2 \le \phi(k) \sum_{a \bmod k}^{*} \sum_{\substack{z_1,z_2 \bmod k\ell \\ z_1=z_2=\xi(\ell)}} e_{k\ell}\Big(a\ell(N(z_2)-N(z_1)) + c\cdot(z_2-z_1)\Big).$$

We write $N(z) = {}^t z A z$ for a symmetric matrix A and $z_2 = z_1 + v$. Then the last sum can be rewritten as

$$\sum_{\substack{z_1,v \bmod k\ell \\ z_1=\xi(\ell),v=0(\ell)}} e_{k\ell}(a\ell N(v) + c\cdot v) e_k(2a\,{}^t z_1 A v).$$

We consider first the sum over the residues z_1 such that the vector $2\ell\,{}^t z_1 A$ satisfies $2\ell\,{}^t z_1 A = 0 \pmod k$. It is clear that the number of such residues $z_1 \bmod k\ell$ with $z_1 = \xi(\ell)$ is $O_{A,\ell}(1)$. Since the sum over v has k^4 terms, this implies that the contribution of the sum over such z_1 to (6.3) is at most $O_{A,\ell}(k^6)$. Now we consider the sum over the residues z_1 such that $2\ell\,{}^t z_1 A \ne 0 \pmod k$. This implies that

$$\sum_{\substack{v \bmod k\ell \\ v=0(\ell)}} e_k(2a\,{}^t z_1 A v) = \sum_{v' \bmod k} e_k(2a\ell\,{}^t z_1 A v') = 0.$$

Therefore, we conclude that

$$|S_k(c;\xi)| = O_{A,\ell}(k^3),$$

which proves (i).

To prove (ii), let us first consider the case when k is coprime to ℓ. Then

$$S_k(c;\xi) = \sum_{a \bmod k}^{*} \sum_{\substack{z \bmod k\ell \\ z=\xi(\ell)}} e_{k\ell}(a\ell F_h(z) + c\cdot z)$$
$$= \sum_{a \bmod k}^{*} \sum_{z' \bmod k} e_{k\ell}(a\ell F_h(\xi+\ell z') + c\cdot(\xi+\ell z')).$$

Let $\bar{\ell}$ be the residue modulo k such that $\ell\bar{\ell} = 1 \pmod k$. Then it follows that

$$S_k(c;\xi) = \sum_{a \bmod k}^{*} \sum_{z'' \bmod k} e_{k\ell}(a\ell F_h(z'') + c\cdot(\xi + \ell\bar{\ell}(z''-\xi))),$$

and

$$|S_k(c;\xi)| \le |S_k(\bar{\ell}c)|, \tag{6.4}$$

where

$$S_k(c) := \sum_{a \bmod k}^{*} \sum_{z \bmod k} e_k(aF_h(z) + c\cdot z).$$

The quantities $S_k(c)$ were investigated in [22].

We proceed as in [22, Lem. 28]. Let us decompose $k = k_1k_2$ where k_1 is square-free, k_2 is square-full and $(k_1,k_2) = 1$. We also decompose $\ell = \ell_1\ell_2$ so that $k_1\ell_1$ is coprime to $k_2\ell_2$. Then according to Lemma 6.3,

$$S_k(c;\xi\,(\ell)) = S_{k_1}(\bar{k}_2\bar{\ell}_2 c;\xi\,(\ell_1))\,S_{k_2}(\bar{k}_1\bar{\ell}_1 c;\xi\,(\ell_2)).$$

We estimate the second factor using (i):

$$S_{k_2}(\bar{k}_1\bar{\ell}_1 c;\xi\,(\ell_2)) \ll_\ell k_2^3.$$

The first factor can be decomposed further into product over primes $q|k_1$. We use that according to Lemma [22, Lem. 26], for all primes q which do not divide the discriminant of N and do not divide both h^2 and $A^{-1}c$,

$$|S_q(c)| \ll q^{5/2}(q,h^2,A^{-1}c)^{1/2}. \tag{6.5}$$

We note that this bound also holds when q divides both h^2 and $A^{-1}c$ because of the estimate (i). Now for finitely many primes q dividing either k or ℓ, we use the bound from (i), and for the remaining primes we use the estimate (6.5) taking (6.4) into account. This gives the estimate:

$$\sum_{k\le X} |S_k(c;\xi)| \le \sum_{k_2\le X}{}' \sum_{k_1\le X/k_2}{}'' |S_{k_1}(\bar{k}_2\bar{\ell}_2 c;\xi\,(\ell_1))|\,|S_{k_2}(\bar{k}_1\bar{\ell}_1 c;\xi\,(\ell_2))|$$
$$\ll_\ell \sum_{k_2\le X}{}' \sum_{k_1\le X/k_2}{}'' k_1^{5/2}(k_1,h^2,A^{-1}c)^{1/2}\,k_2^3$$
$$\le X^{5/2} \sum_{k_2\le X}{}' k_2^{1/2} \sum_{k_1\le X/k_2}{}'' (k_1,h^2,A^{-1}c)^{1/2}.$$

Here the sums are taken over square-full k_2 and square-free k_1 respectively. We use that for $k \neq 0$,

$$\sum_{n\le Y} (n,k) \le Y d(k) \ll_\epsilon Y k^\epsilon$$

for all $\epsilon > 0$. Hence, we conclude that if $c \neq 0$,

$$\sum_{k\le X} |S_k(c;\xi)| \ll_{\ell,\epsilon} X^{7/2}|c|^\epsilon \sum_{k_2\le X}{}' k_2^{-1/2} \ll_{\ell,\epsilon} X^{7/2+\epsilon}|c|^\epsilon.$$

Similarly, we obtain

$$\sum_{k\le X} |S_k(0;\xi)| \ll_{\ell,\epsilon} X^{7/2+\epsilon}h^\epsilon.$$

This completes the proof of (ii). □

We shall use the following properties of the integral $I_{k,\ell}(c)$:

Lemma 6.5.

(i) $I_{k,\ell}(c) = 0$ *for all* $k \ge h$.

(ii) *For every* $c \neq 0$ *and* $k > 0$,

$$I_{k,\ell}(c) \ll_{w,\ell,k} h^5 k^{-1} |c|^{-k}.$$

(iii) *For* $c \neq 0$ *and* $\epsilon \in (0, 1/2)$,

$$I_{k,\ell}(c) \ll_{w,\ell,\epsilon} \frac{h^{3+2\epsilon}k^{1-2\epsilon}}{|c|^{1-\epsilon}}.$$

(iv) $I_{k,\ell}(0) \ll_{w,\ell} h^4$.

(v) *For every* $k \ll h$ *and* $k > 0$,

$$I_{k,\ell}(0) = h^4\Big(\sigma_\infty(N,w) + O_{w,\ell,k}\big((k/h)^k\big)\Big).$$

Proof. Part (i) follows from properties of the function H (see [22, Lem. 4]). We observe that

$$I_{k,\ell}(c) = (h\ell)^4 \int_{\mathbb{R}^4} w(\ell x) H\big(h^{-1}k, \ell^2 N(x) - 1\big) e_k(-hc\cdot x)\,\mathrm{d}x.$$

The properties of the integrals

$$I_k(c) := h^4 \int_{\mathbb{R}^4} w(x) H\big(h^{-1}k, N(x) - 1\big) e_k(-hc\cdot x)\,\mathrm{d}x.$$

have been studied in [22], and properties (ii)–(iv) can be deduced directly from there. In particular, (ii) follows from [22, Lem. 19], (iii) from [22, Lem. 22], and (iv) from [22, Lem. 22]. To prove (v), we use [22, Lem. 13]. Here we also use that $\sigma_\infty(\ell^2 N, w\circ \ell) = \ell^{-4}\sigma_\infty(N,w)$. □

Proof of Theorem 6.2. Our starting point is the formula

$$\mathrm{N}_h(N, w; \xi) = c_h h^{-2} \sum_{c\in\mathbb{Z}^4} \sum_{k=1}^{\infty} (k\ell)^{-4} S_k(c;\xi) I_{k,\ell}(c).$$

Using Lemmas 6.4 (i) and 6.5 (ii), we deduce that for every $\epsilon > 0$,

$$\left| \sum_{|c|>h^\epsilon} \sum_{k=1}^{\infty} (k\ell)^{-4} S_k(c;\xi) I_{k,\ell}(c) \right| \ll_{w,\ell,k} h^5 \left(\sum_{k=1}^{\infty} k^{-2} \right) \left(\sum_{|c|>h^\epsilon} |c|^{-\theta} \right) = O_{w,\ell,\theta}(1),$$

when θ is chosen to be sufficiently large. Now it is sufficient to consider the sum over $|c| \le h^\epsilon$. It follows from Lemma 6.5 (iii) that when $c \ne 0$ with $|c| \le h^\epsilon$,

$$\left| \sum_{R\le k<2R} (k\ell)^{-4} S_k(c;\xi) I_{k,\ell}(c) \right| \ll_{w,\ell,\epsilon} h^{3+2\epsilon} \sum_{R\le k<2R} k^{-3} |S_k(c;\xi)|,$$

Here we used that $|c| \ge 1$ for $c \in \mathbb{Z}^4 \backslash \{0\}$. Hence, using Lemma 6.4 (ii) and summation by parts, we deduce that

$$\sum_{R\le k<2R} k^{-3} |S_k(c;\xi)| \ll_{\ell,\epsilon} R^{1/2+2\epsilon}.$$

Hence, we deduce that for every $c \ne 0$ with $|c| \le h^\epsilon$,

$$\left| \sum_{k=1}^{h} (k\ell)^{-4} S_k(c;\xi) I_{k,\ell}(c) \right| \ll_{w,\ell,\epsilon} h^{7/2+4\epsilon}.$$

Moreover, according to Lemma 6.5 (i), only the terms with $k \le h$ are non-zero. Therefore, we conclude that for every $\epsilon > 0$,

$$\mathrm{N}_h(N, w; \xi) = c_h h^{-2} \ell^{-4} \sum_{k=1}^{\infty} k^{-4} S_k(0;\xi) I_{k,\ell}(0) + O_{w,\ell,\epsilon}(h^{3/2+\epsilon}). \tag{6.6}$$

Using Lemma 6.4 (ii), Lemma 6.5 (iv), and summation by parts, we deduce that for every $\epsilon > 0$,

$$\sum_{R<k\le 2R} k^{-4} S_k(0;\xi) I_{k,\ell}(0) \ll_{w,\ell,\epsilon} h^{4+\epsilon} R^{-1/2+2\epsilon}.$$

This implies that for every $\epsilon > 0$,

$$\sum_{k>h^{1-\epsilon}} k^{-4} S_k(0;\xi) I_{k,\ell}(0) \ll_{w,\ell,\epsilon} h^{7/2+\epsilon}.$$

It follows from Lemma 6.5 (v) that

$$\sum_{k\le h^{1-\epsilon}} k^{-4} S_k(0;\xi) I_{k,\ell}(0) = h^4 \sigma_\infty(N, w) \left(\sum_{k\le h^{1-\epsilon}} k^{-4} S_k(0;\xi) \right) + O_{w,\ell,\epsilon}(1).$$

Hence, we deduce from (6.6) that

$$\mathrm{N}_h(N,w;\xi) = h^2\sigma_\infty(N,w)\ell^{-4}\left(\sum_{k\le h^{1-\epsilon}} k^{-4}S_k(0;\xi)\right) + O_{w,\ell,\epsilon}(h^{3/2+\epsilon}).$$

The last sum corresponds to the classical singular series. It follows from Lemma 6.4 that it converges absolutely, and moreover,

$$\sum_{k\le X} k^{-4}S_k(0;\xi) = \sum_{k=1}^{\infty} k^{-4}S_k(0;\xi) + O_{\ell,\epsilon}(X^{-1/2+\epsilon})$$

for all $\epsilon > 0$. Finally, we recall that if the singular series converges, it is equal to the product of local densities (see, for instance, [11, Lem. 5.1–5.3]). □

We will also need a uniform version of Theorem 6.2 which applies to families of functions $w_g(x) := w(g^{-1}x)$ with $g \in G_\infty$.

Theorem 6.6. *Let $w : D_\infty \to \mathbb{R}$ be a smooth compactly supported function, $\xi \in \Lambda/\ell\Lambda$, and $g \in G_\infty$. Then for every $\delta > 0$, $\theta > 4$, and $\epsilon > 0$,*

$$\mathrm{N}_h(N,w_g;\xi) = \ell^{-4}\sigma_\infty(N,w)\sigma_f(N,\xi)h^2 + O_{w,\ell,\delta,\theta,\epsilon}\Big(\|g\|^\theta h^{3-(\theta-4)\delta} + \|g\|h^{3/2+3\delta+\epsilon}\Big).$$

Proof. As in the proof of Theorem 6.2,

$$\mathrm{N}_h(N,w_g;\xi) = c_h h^{-2}\sum_{c\in\mathbb{Z}^4}\sum_{k=1}^{\infty}(k\ell)^{-4}S_k(c;\xi)I_{g,k,\ell}(c),$$

where

$$I_{g,k,\ell}(c) := h^4\int_{\mathbb{R}^4} w(g^{-1}x)H(h^{-1}k, N(x)-1)e_{k\ell}(-hc\cdot x)\,\mathrm{d}x.$$

Since the norm N is G_∞-invariant, we obtain

$$I_{g,k,\ell}(c) = h^4\int_{\mathbb{R}^4} w(x)H(h^{-1}k, F(x))e_{k\ell}(-hc\cdot gx)\,\mathrm{d}x = I_{k,\ell}({}^tgc).$$

Our argument proceeds as in the proof of Theorem 6.2 taking the dependences on g into account. Throughout the proof, we will have to deal with the maps $\mathsf{D}(\mathbb{R}) \to \mathsf{D}(\mathbb{R}) : x \mapsto {}^tg^{\pm 1}x$ with $g \in G_\infty$. We use the norm of these maps are estimated in terms of the Euclidean norm $\|g\|_E := \big(\sum_{i,j} g_{ij}^2\big)^{1/2}$ on the group $G_\infty \simeq \mathrm{SL}_2(\mathbb{R})$.

First, we consider the terms with $|c| > h^\delta$. By Lemma 6.5 (ii), for every $\theta > 0$,

$$\left|\sum_{|c|>h^\delta}\sum_{k=1}^{\infty} k^{-4}S_k(c;\xi)I_{k,\ell}({}^tgc)\right| \ll_{w,\ell,\theta} h^5\left(\sum_{k=1}^{\infty} k^{-5}|S_k(c;\xi)|\right)\left(\sum_{|c|>h^\delta} |{}^tgc|^{-\theta}\right).$$

By Lemma 6.4 (i), the first sum is finite. For $\theta > 4$, the second sum is estimated as

$$\sum_{|c|>h^\delta} |{}^t gc|^{-\theta} \ll \|g\|_E^\theta \sum_{|c|>h^\delta} |c|^{-\theta} \ll_\theta \|g\|_E^\theta h^{-(\theta-4)\delta}.$$

Hence,

$$\left|\sum_{|c|>h^\delta} \sum_{k=1}^{\infty} k^{-4} S_k(c;\xi) I_{k,\ell}({}^t gc)\right| \ll_{w,\ell,\theta} \|g\|_E^\theta h^{5-(\theta-4)\delta}. \tag{6.7}$$

Next, we estimate the terms with $0 < |c| \le h^\delta$. By Lemma 6.5 (i)(iii), for every $\epsilon \in (0, 1/2)$,

$$\left|\sum_{0<|c|\le h^\delta} \sum_{k=1}^{\infty} k^{-4} S_k(c;\xi) I_{k,\ell}({}^t gc)\right| \\ \ll_{w,\ell,\epsilon} h^{3+2\epsilon} \left(\sum_{k=1}^{h} k^{-3} |S_k(c;\xi)|\right) \left(\sum_{0<|c|\le h^\delta} |{}^t gc|^{-(1-\epsilon)}\right).$$

It follows from Lemma 6.4 (ii),

$$\sum_{k=1}^{h} k^{-3} |S_k(c;\xi)| \ll_{\ell,\epsilon} h^{1/2+\epsilon},$$

and

$$\sum_{0<|c|\le h^\delta} |{}^t gc|^{-(1-\epsilon)} \ll \|g\|_E \cdot \sum_{0<|c|\le h^\delta} |c|^{-(1-\epsilon)} \ll \|g\|_E h^{(3+\epsilon)\delta}.$$

Hence, we conclude that for every $\epsilon > 0$,

$$\left|\sum_{0<|c|\le h^\delta} \sum_{k=1}^{\infty} k^{-4} S_k(c;\xi) I_{k,\ell}({}^t gc)\right| \ll_{w,\ell,\theta,\epsilon} \|g\|_E h^{7/2+3\delta+\epsilon}. \tag{6.8}$$

Combining (6.7) and (6.8), we deduce that

$$\mathrm{N}_h(N, w_g;\xi) = c_h h^{-2} \ell^{-4} \sum_{k=1}^{\infty} k^{-4} S_k(0;\xi) I_{k,\ell}(0) \\ + O_{w,\ell,\theta,\epsilon}\big(\|g\|_E^\theta h^{3-(\theta-4)\delta} + \|g\|_E h^{3/2+3\delta+\epsilon}\big).$$

The last sum was already estimated as in the proof of Theorem 6.2. □

It will be convenient to interpret the local densities σ_∞ and σ_f group-theoretically, namely, in terms of the Tamagawa measures for the group G. We refer to [42, Ch. 2] or [41, Ch. 5] for basic properties of the Tamagawa measures. Let us fix a nowhere zero regular rational differential form of top degree on G (this form is known to be unique up to a constant factor).

Integration with respect to this form defines Haar measures τ_∞ and τ_q on $G_\infty := \mathsf{G}(\mathbb{R})$ and $G_q := \mathsf{G}(\mathbb{Q}_q)$ respectively. While the local Tamagawa measures are only unique up to constant factors, their product is canonical. In fact, according to the Tamagawa Volume Formula, since G is simply connected,

$$\tau_\infty(G_\infty/\Gamma)\cdot\prod\nolimits_{q\text{-prime}}\tau_q(\mathsf{G}(\mathbb{Z}_q)) = 1. \tag{6.9}$$

It will be convenient to define the Tamagawa measures using that G is a fiber of the norm map N. Let ω_0 be the standard one-form on $\mathbb{A}^1$, and ω be the standard form of top degree on D. Then there exists a form η of degree three on D such that $\eta\wedge N^*(\omega_0)=\omega$. We denote by ω_y the restriction of η to the fiber $N^{-1}(y)$ for $y\neq 0$, and by $\tau_q^{(y)}$ the corresponding measures supported on the fibers $N^{-1}(y)(\mathbb{Q}_q)$. In particular, $\tau_q := \tau_q^{(1)}$ defines a Tamagawa measure on $\mathsf{G}(\mathbb{Q}_q)$. We note that the fiber measures are uniquely defined by the disintegration formula

$$\int_{\mathsf{D}(\mathbb{Q}_q)}\phi(N(x))\psi(x)\,\mathrm{d}x = \int_{\mathbb{Q}_q\setminus\{0\}}\phi(y)\left(\int_{N^{-1}(y)}\psi\,\mathrm{d}\tau_q^{(y)}\right)\mathrm{d}y, \tag{6.10}$$

where ϕ and ψ are compactly supported locally constant function.

Previously, we used the measures m_∞ and m_q that are normalized as

$$m_\infty(G_\infty/\Gamma)=1 \quad\text{and}\quad m_q(\mathsf{G}(\mathbb{Z}_q))=1.$$

In view of (6.9), these measures can be expressed in terms of the Tamagawa measures as

$$m_\infty = \tau_\infty(G_\infty/\Gamma)^{-1}\tau_\infty \quad\text{and}\quad m_q=\tau_q(\mathsf{G}(\mathbb{Z}_q))^{-1}\tau_q. \tag{6.11}$$

Lemma 6.7.

(i) *Let p be a prime, $\ell\in\mathbb{N}$ coprime to p, and $\xi\in\mathsf{G}(\mathbb{Z}/\ell\mathbb{Z})$. Then for $h=p^s$,*

$$\sigma_f(N,\xi,h)=\tau_p(\mathsf{G}(\mathbb{Z}_p))^{-1}h^{-2}\tau_p(B_h)\ell^4\tau_\infty(G_\infty/\Gamma_\ell)^{-1},$$

where $B_h=\{g\in G_p:\ \|g\|_p\le h\}$.

(ii) *For every $w\in C_c(G_\infty)$,*

$$\sigma_\infty(N,w)=\int_{G_\infty} w\,\mathrm{d}\tau_\infty.$$

Proof. The crucial connection between the local densities σ_q and the Tamagawa measures τ_q is given by the following formula:

$$\tau_q(\mathsf{G}(\mathbb{Z}_q)) = \lim_{e\to\infty} q^{-3e}|\mathsf{G}(\mathbb{Z}/q^e\mathbb{Z})|. \tag{6.12}$$

Let q be a prime coprime to both ℓ and p. Then

$$\begin{aligned}\mathrm{N}_h(N, q^e, q^{s_q}, \xi) &= \big|\{x \bmod q^e : N(x) = h^2 \bmod q^e\}\big| \\ &= \big|\{x \bmod q^e : N(x) = 1 \bmod q^e\}\big| \\ &= |\mathsf{G}(\mathbb{Z}/q^e\mathbb{Z})|.\end{aligned}$$

Hence, it follows from (6.12) that in this case

$$\sigma_q(\xi, h) = \tau_q(\mathsf{G}(\mathbb{Z}_q)). \tag{6.13}$$

Let q be a prime dividing $\ell = \prod_{r\text{-prime}} r^{s_r}$ (and coprime to $h = p^s$). We choose a residue $\bar{h}$ (mod p^e) such that $\bar{h}h = 1$ (mod q^e). Then for $e \ge s_q$,

$$\begin{aligned}\mathrm{N}_h(N, q^e, q^{s_q}, \xi) &= \big|\{x \bmod q^{e+s_q} : x = \xi \bmod q^{s_q},\ N(x) = h^2 \bmod q^e\}\big| \\ &= q^{4s_q}\big|\{y \bmod q^e : y = \xi \bmod q^{s_q},\ N(y) = h^2 \bmod q^e\}\big| \\ &= q^{4s_q}\big|\{y \bmod q^e : y = \bar{h}^2\xi \bmod q^{s_q},\ N(y) = 1 \bmod q^e\}\big| \\ &= q^{4s_q}\frac{|\mathsf{G}(\mathbb{Z}/q^e\mathbb{Z})|}{|\mathsf{G}(\mathbb{Z}/q^{s_q}\mathbb{Z})|}.\end{aligned}$$

Hence, by (6.12) as before,

$$\sigma_q(N, \xi, h) = q^{4s_q}|\mathsf{G}(\mathbb{Z}/q^{s_q}\mathbb{Z})|^{-1}\tau_q(\mathsf{G}(\mathbb{Z}_q)). \tag{6.14}$$

Finally, we claim that

$$\sigma_p(N, \xi, h) = h^{-2}\tau_p(B_h). \tag{6.15}$$

For $z \in \mathbb{Q}_p^\times$, let us consider the map $\Phi_z(x) := z^{-1}x$. We observe that the map Φ transforms the fiber $N^{-1}(y)$ to $N^{-1}(z^{-2}y)$ in (6.10), so that it follows from uniqueness of this integral decomposition that

$$(\Phi_z)_*(\tau_p^{(y)}) = |z|_p^2\,\tau_p^{(z^{-2}y)}. \tag{6.16}$$

For $r > 0$ and $y \in \mathbb{Q}_p$, we set

$$B_r(y) := \{x \in \mathsf{D}(\mathbb{Q}_p) : N(x) = y,\ \|x\|_p \le r\}.$$

Then

$$\{x \in \mathsf{D}(\mathbb{Z}_p) : N(x) = h^2 \bmod p^e\} = \bigsqcup\nolimits_{y \in h^2 + p^e\mathbb{Z}_p} B_1(y),$$

and it follows from (6.10) that

$$\int_{h^2+p^e\mathbb{Z}_p} \tau_p^{(y)}(B_1(y))\,\mathrm{d}y = p^{-4e}\big|\{x \in \mathsf{D}(\mathbb{Z}/p^e\mathbb{Z}) : N(x) = h^2 \bmod p^e\}\big|.$$

Hence,

$$\sigma_p(N, \xi, h) = \lim_{e\to\infty} p^e \int_{h^2+p^e\mathbb{Z}_p} \tau_p^{(y)}(B_1(y))\,\mathrm{d}y.$$

If e is sufficiently large, for every $y \in h^2 + p^e\mathbb{Z}_p$ there exists z such that $z^2 = y$. Clearly, $|z|_p = |h|_p = h^{-1}$. Then by (6.16),

$$\tau_p^{(y)}(B_1(y)) = \tau_p^{(y)}\Big(\Phi_z^{-1}(B_{|z|_p^{-2}}(1))\Big) = |z|_p^2\, \tau_p^{(1)}\left(B_{|z|_p^{-2}}(1)\right) = h^{-2}\, \tau_p(B_h).$$

This implies (6.15).

Furthermore, Φ defines a bijection between

$$B_h' := \{x \in \mathsf{D}(\mathbb{Q}_p) : N(x) = h^2,\ \|x\|_p \le 1\} = \{x \in \mathsf{D}(\mathbb{Z}_p) :\ N(x) = h^2\}$$

and

$$B_h = \{x \in \mathsf{D}(\mathbb{Q}_p) : N(x) = 1,\ \|x\|_p \le h\}.$$

Hence, it follows that

$$\tau_p(B_h) = \tau_p^{(1)}(\Phi(B_h')) = |h|_p^{-2}\, \tau_p^{(h^2)}(B_h') = h^2\, \tau_p^{(h^2)}(B_h').$$

Furthermore, it follows from (6.10) that

$$\int_{h^2+p^e\mathbb{Z}_p} \tau_p^{(y)}(B_h')\, \mathrm{d}y = p^{-4e}\big|\{x \in (\mathbb{Z}/p^e\mathbb{Z})^4 : N(x) = h^2 \bmod p^e\}\big|.$$

Hence,

$$\sigma_p(N,\xi,h) = \lim_{e\to\infty} p^e \int_{h^2+p^e\mathbb{Z}_p} \tau_p^{y}(B_h)\, \mathrm{d}y = \tau_p^{h^2}(B_h') = h^{-2}\tau_p(B_h).$$

This proves (6.15).

Combining (6.13), (6.14) and (6.15), we deduce that

$$\sigma_f(N,\xi,h) = \ell^4|\mathsf{G}(\mathbb{Z}/\ell\mathbb{Z})|^{-1}\left(\prod_{q\neq p}\tau_q(\mathsf{G}(\mathbb{Z}_q))\right)h^{-2}\tau_p(B_h).$$

Furthermore, by the Tamagawa Formula (6.9)

$$\begin{aligned}\sigma_f(N,\xi,h) &= \ell^4|G(\mathbb{Z}/\ell\mathbb{Z})|^{-1}\tau_p(\mathsf{G}(\mathbb{Z}_p))^{-1}\tau_\infty(G_\infty/\Gamma)^{-1}h^{-2}\tau_p(B_h)\\ &= \tau_p(\mathsf{G}(\mathbb{Z}_p))^{-1}h^{-2}\tau_p(B_h)\ell^4\tau_\infty(G_\infty/\Gamma_\ell)^{-1}.\end{aligned}$$

This proves the first formula.

It follows from the disintegration formula (6.10) that

$$\begin{aligned}\sigma_\infty(N,w) &= \lim_{\epsilon\to 0^+}(2\epsilon)^{-1}\int_{|N(x)-1|\le\epsilon} w(x)\,\mathrm{d}x\\ &= \lim_{\epsilon\to 0^+}(2\epsilon)^{-1}\int_{1-\epsilon}^{1+\epsilon}\left(\int_{N^{-1}(y)} w\,\mathrm{d}\tau_\infty^{(y)}\right)\mathrm{d}y = \int_{N^{-1}(1)} w\,\mathrm{d}\tau_\infty^{(1)},\end{aligned}$$

which proves the second equality. $\square$

From now on we fix prime p and $\ell \in \mathbb{N}$ coprime to p. We always choose h to be of the form $h = p^s$.

Our next goal is to show that the counting function that we studied can be interpreted in terms of the averaging operators $\rho_{p,\ell}(\beta_h)$. Let $w \in C_c(D_\infty)$, and for $x \in G_\infty$, we also set

$$w_x(g) := w(x^{-1}g).$$

Let χ denote the characteristic function of the subset $\mathsf{G}(\mathbb{Z}_p)$. We introduce a function ϕ_w defined by

$$\phi_w(g_\infty, g_p) := \sum_{\gamma \in \Gamma_{p,\ell}} w(g_\infty \gamma)\chi(g_p \gamma) \quad \text{for } (g_\infty, g_p) \in G_\infty \times G_p. \tag{6.17}$$

This defines the function on $X_{p,\ell} = (G_\infty \times G_p)/\Gamma_{p,\ell}$. The invariant probability measure $\mu_{p,\ell}$ on $X_{p,\ell}$ is defined as

$$\int_{X_{p,\ell}} \left(\sum_{\gamma \in \Gamma_{p,\ell}} f(g\gamma) \right) \mathrm{d}\mu_{p,\ell}(g\Gamma_{p,\ell}) = |\Gamma : \Gamma_\ell|^{-1} \int_{G_\infty \times G_p} f \, d(m_\infty \times m_p)$$

for $f \in C_c(G_\infty \times G_p)$. Indeed, if F is a fundamental domain for Γ_ℓ in G_∞, then $F \times \mathsf{G}(\mathbb{Z}_p)$ is a fundamental domain for $\Gamma_{p,\ell}$ in $G_\infty \times G_p$. Since $m_\infty(G_\infty/\Gamma) = 1$ and $m_p(\mathsf{G}(\mathbb{Z}_p)) = 1$, the above formula indeed defines the invariant probability measure on $X_{p,\ell}$. In particular, it follows that

$$\begin{aligned} \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} &= |\Gamma : \Gamma_\ell|^{-1} \left(\int_{G_\infty} w \, \mathrm{d}m_\infty \right) \left(\int_{G_p} \chi \, \mathrm{d}m_p \right) \\ &= |\mathsf{G}(\mathbb{Z}/\ell\mathbb{Z})|^{-1} \int_{G_\infty} w \, \mathrm{d}m_\infty. \end{aligned}$$

Taking (6.11) and Lemma 6.7 into account, we obtain:

$$\begin{aligned} &\sigma_\infty(N, w)\sigma_f(N, \xi, h) \\ &\qquad = \left(\int_{G_\infty} w \, \mathrm{d}\tau_\infty \right) \tau_p(\mathsf{G}(\mathbb{Z}_p))^{-1} h^{-2} \tau_p(B_h) \ell^4 \tau_\infty(G_\infty/\Gamma_\ell)^{-1} \\ &\qquad = h^{-2}\ell^4 \left(\int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right) m_p(B_h). \end{aligned} \tag{6.18}$$

For $u \in \mathsf{G}(\mathbb{Z}_p)$, we obtain:

$$\begin{aligned}\int_{B_h} \phi_w(b^{-1}(x^{-1},u))\,\mathrm{d}m_p(b) &= \int_{B_h}\left(\sum_{\gamma\in\Gamma_{p,\ell}} w(x^{-1}\gamma)\chi(b^{-1}u\gamma)\right)\mathrm{d}m_p(g)\\ &= \sum_{\gamma\in\Gamma_{p,\ell}} w(x^{-1}\gamma)m_p(u\gamma\mathsf{G}(\mathbb{Z}_p)\cap B_h)\\ &= \sum_{\gamma\in\Gamma_{p,\ell}\cap B_h} w_x(\gamma).\end{aligned}$$

Since

$$\begin{aligned}\Gamma_{p,\ell}\cap B_h &= \{x\in\Lambda[1/p] : N(x)=1,\, x=I \pmod{\ell},\, \|x\|_p\le h\}\\ &= \{h^{-1}y :\, y\in\Lambda,\, N(y)=h^2,\, y=hI \pmod{\ell}\},\end{aligned}$$

we conclude that for any $u \in \mathsf{G}(\mathbb{Z}_p)$,

$$\int_{B_h} \phi_w(b^{-1}(x^{-1},u))\,\mathrm{d}m_p(b) = \mathrm{N}_h\left(N, w_x, hI \bmod \ell\right). \tag{6.19}$$

We will also use the following result about integrability of the function ϕ_w:

Lemma 6.8. *$\phi_w \in L^r(X_{p,\ell})$ for all $r\in[1,\infty)$.*

Proof. Without loss of generality, we may assume that $w \ge 0$.

It is sufficient to show that $\phi_w \in L^r(X_{p,\ell})$ fo every $r\in\mathbb{N}$. We obtain that $\|\phi_w\|^r_{L^r(X_{p,\ell})}$ can be expressed as

$$\begin{aligned}&\int_{X_{p,\ell}} \phi_w(x)^r\,\mathrm{d}\mu_{p,\ell}(x)\\ &= \int_{(G_\infty\times G_S)/\Gamma_{p,\ell}}\left(\sum_{\gamma_1,\dots,\gamma_r\in\Gamma_{p,\ell}} \phi_w(g\gamma_1)\cdots\phi_w(g\gamma_r)\right)\mathrm{d}\mu_{p,\ell}(g\Gamma_{p,\ell})\\ &= \int_{(G_\infty\times G_S)/\Gamma_{p,\ell}}\left(\sum_{\gamma_1,\dots,\gamma_r\in\Gamma_{p,\ell}} \phi_w(g\gamma_1)\phi_w(g\gamma_1\gamma_2)\cdots\phi_w(g\gamma_1\gamma_r)\right)\mathrm{d}\mu_{p,\ell}(g\Gamma_{p,\ell})\\ &= \int_{G_\infty\times G_S}\left(\sum_{\gamma_2,\dots,\gamma_r\in\Gamma_{p,\ell}} \phi_w(g)\phi_w(g\gamma_2)\cdots\phi_w(g\gamma_r)\right)d(m_\infty\times m_p)(g).\end{aligned}$$

We observe that the product is zero unless

$$\gamma_i \in \mathrm{supp}(\phi_w)^{-1}\,\mathrm{supp}(\phi_w) \subset \mathrm{supp}(w)^{-1}\,\mathrm{supp}(w)\times\mathsf{G}(\mathbb{Z}_p),$$

and there are only finitely many such γ_i's. Furthermore, by Hölder inequality,

$$\begin{aligned}\int_{G_\infty \times G_S} &\phi_w(g)\phi_w(g\gamma_2)\cdots\phi_w(g\gamma_r)\, d(m_\infty \times m_p)(g) \\ &\le \|\phi_w\|_{L^r(G_\infty\times G_p)}\|\phi_w\circ\gamma_2\|_{L^r(G_\infty\times G_p)}\cdots\|\phi_w\circ\gamma_r\|_{L^r(G_\infty\times G_p)} \\ &= \|\phi_w\|^r_{L^r(G_\infty\times G_p)} = \|w\|^r_{L^r(G_\infty)}.\end{aligned}$$

Hence, we conclude that

$$\|\phi_w\|_r^r \ll \|w\|^r_{L^r(G_\infty)} < \infty. \qquad \square$$

The following proposition verifies Theorem 6.1 for the class of functions ϕ_w. While this class of functions is quite "sparse", we will eventually show that this can be used to derive this estimate for general functions.

Proposition 6.9. *Let $w \in C_c^\infty(D_\infty)$ be as in Lemma 6.8. Then for every $\epsilon > 0$,*

$$\left\|\rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \,\mathrm{d}\mu_{p,\ell}\right\|_{L^2(X_{p,\ell})} \ll_{w,p,\ell,\epsilon} m_p(B_h)^{-\sigma+\epsilon},$$

where $\sigma = 1/4$ if G is a anisotropic over $\mathbb{Q}$ and $\sigma = 1/16$ otherwise.

Proof. Taking (6.18) and (6.19) into account, Theorem 6.6 can be restated as follows: for every $g \in G_\infty$, $u \in \mathsf{G}(\mathbb{Z}_p)$, $\delta > 0$, $\theta > 4$, and $\epsilon > 0$,

$$\begin{aligned}(6.20)\quad &\rho_{p,\ell}(\beta_h)\phi_w(g^{-1},u) \\ &= \int_{X_{p,\ell}} \phi_w\,\mathrm{d}\mu_{p,\ell} + O_{p,w,\ell,\delta,\theta,\epsilon}\Big(\|g\|_E^\theta h^{1-(\theta-4)\delta} + \|g\|_E h^{-1/2+3\delta+\epsilon}\Big).\end{aligned}$$

We shall use this estimate to prove the proposition.

First, we note that (6.20) implies that for every compact $Q \subset G_\infty$,

$$\left\|\rho_{p,\ell}(\beta_h)\phi_w((\,\cdot\,)^{-1},u) - \int_{X_{p,\ell}} \phi_w\,\mathrm{d}\mu_{p,\ell}\right\|_{L^2(Q)} \ll_{p,w,\ell,Q,\epsilon} h^{-1/2+\epsilon}$$

for all $\epsilon > 0$. If G is anisotropic over $\mathbb{Q}$, then the lattice Γ_ℓ is cocompact in G_∞. Hence, we can choose a compact subset Q such that Q^{-1} surjects onto G_∞/Γ_ℓ, and the second part of the corollary follows. Then $Q^{-1} \times \mathsf{G}(\mathbb{Z}_p)$ surjects onto $(G_\infty \times G_p)/\Gamma_{p,\ell}$. Hence, it follows from the above estimate that

$$\left\|\rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w\,\mathrm{d}\mu_{p,\ell}\right\|_{L^2(X_{p,\ell})} \ll_{p,w,\ell,\epsilon} h^{-1/2+\epsilon} \ll_p m_p(B_h)^{-1/4+\epsilon}.$$

This proves the proposition when G is anisotropic over $\mathbb{Q}$.

Now we consider the case when G is isotropic over $\mathbb{Q}$. We observe that if Q is a subset of G_∞ such that Q^{-1} surjects onto G_∞/Γ_ℓ, then as before

$$\left\|\rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \,\mathrm{d}\mu_{p,\ell}\right\|_{L^2(X_{p,\ell})} \leq \left\|\rho_{p,\ell}(\beta_h)\phi_w((\,\cdot\,)^{-1}, e) - \int_{X_{p,\ell}} \phi_w \,\mathrm{d}\mu_{p,\ell}\right\|_{L^2(Q)}.$$

It follows from the theory of Siegel sets that such Q can be chosen of the form

$$Q = Q_0 \cup Q_1 \cup \cdots \cup Q_s,$$

where Q_0 is compact, and for $i \geq 1$,

$$Q_i^{-1} := \{ka(t)ng_i :\ k \in K, t \geq 0, n \in N_0\}. \tag{6.21}$$

Here $K = \mathrm{SO}(2)$, $a(t) = \mathrm{diag}(e^t, e^{-t})$, N_0 is a compact subset of the upper triangular unipotent group, and $g_i \in \mathsf{G}(\mathbb{Q})$.

Let us consider the case when Q is given by (6.21). We note that the case of the union can be handled by using the triangle inequality. We set

$$Q_{<R} := \{t < \log R\} \quad \text{and} \quad Q_{\geq R} := \{t \geq \log R\}.$$

We observe that for every $\theta > 2$,

$$\int_{Q_{<R}} \|g\|_E^\theta \,\mathrm{d}m_\infty(g) \ll \int_0^{\log R} e^{(\theta-2)t}\,\mathrm{d}t \ll R^{\theta-2},$$

and

$$\int_{Q_{<R}} \|g\|_E \,\mathrm{d}m_\infty(g) \ll 1.$$

Hence, it follows from (6.20) that

$$\left\|\rho_{p,\ell}(\beta_h)\phi_w((\,\cdot\,)^{-1}, e) - \int_{X_{p,\ell}} \phi_w \,\mathrm{d}\mu_{p,\ell}\right\|_{L^2(Q_{<R})} \ll_{p,w,\ell,\delta,\theta,\epsilon} R^{\theta-2}h^{1-(\theta-4)\delta} + h^{-1/2+3\delta+\epsilon}.$$

To estimate the integral over $Q_{\geq R}$, we consider the set

$$\Omega_R := (Q_{\geq R}^{-1} \times \mathsf{G}(\mathbb{Z}_p))\Gamma_{p,\ell} \subset X_{p,\ell},$$

and ω_R denote the characteristic function of this set. Then

$$\left\|\rho_{p,\ell}(\beta_h)\phi_w((\,\cdot\,)^{-1}, e) - \int_{X_{p,\ell}} \phi_w \,\mathrm{d}\mu_{p,\ell}\right\|_{L^2(Q_{\geq R})} \ll \left\|\left(\rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \,\mathrm{d}\mu_{p,\ell}\right)\omega_R\right\|_{L^2(X_{p,\ell})}.$$

Using the Hölder inequality, we obtain that for every $r \geq 1$ and $s = (1-1/r)^{-1}$,

$$\left\| \left(\rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right) \omega_R \right\|^2_{L^2(X_{p,\ell})} \leq \left\| \rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right\|^2_{L^{2r}(X_{p,\ell})} \cdot \|\omega_R\|^2_{L^{2s}(X_{p,\ell})} .$$

It follows from Jensen inequality that the operator $\rho_p(\beta_h) : L^{2r}(X_{p,\ell}) \to L^{2r}(X_{p,\ell})$ is bounded and the corresponding norm satisfies $\|\rho_p(\beta_h)\| \leq 1$. Hence, since $\phi_w \in L^{2r}(X_{p,\ell})$ by Lemma 6.8, we conclude that

$$\begin{aligned} \left\| \rho_{p,\ell}(\beta_h)\phi_w((\,\cdot\,)^{-1}, e) - \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right\|_{L^2(Q_{\geq R})} &\ll_{w,p,\ell} \|\omega_R\|_{L^{2s}(X_{p,\ell})} \\ &= \mu_{p,\ell}(\Omega_R)^{1/(2s)} \\ &\leq m_\infty(Q_{\geq R})^{1/(2s)} \ll_s R^{-1/s}. \end{aligned}$$

This implies that for every $\epsilon > 0$,

$$\left\| \rho_{p,\ell}(\beta_h)\phi_w((\,\cdot\,)^{-1}, e) - \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right\|_{L^2(Q_{\geq R})} \ll_{w,p,\ell,\epsilon} R^{-1+\epsilon}.$$

Ultimately, we conclude that

$$\left\| \rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right\|_{L^2(X_{p,\ell})} \ll_{p,w,\ell,\delta,\theta,\epsilon} R^{\theta-2} h^{1-(\theta-4)\delta} + h^{-1/2+3\delta+\epsilon} + R^{-1+\epsilon}.$$

We choose $R = h^{1/2-3\delta}$. Then for every $\epsilon > 0$,

$$\left\| \rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right\|_{L^2(X_{p,\ell})} \ll_{w,p,\ell,\theta,\delta,\epsilon} h^{-\sigma+\epsilon},$$

where $\sigma := \min\big(-(1/2-3\delta)(\theta-2) + (\theta-4)\delta - 1, 1/2-3\delta\big)$. To optimise the error term, we choose $\delta = (\theta+1)/(8\theta-14)$. Then as $\theta \to \infty$, we get $\sigma \to 1/8$. This implies the theorem. □

Proof of Theorem 6.1. We recall that by Proposition 6.9, for every $w \in C_c^\infty(D_\infty)$,

$$\left\| \rho_{p,\ell}(\beta_h)\phi_w - \int_{X_{p,\ell}} \phi_w \, \mathrm{d}\mu_{p,\ell} \right\|_{L^2(X_{p,\ell})} \ll_{w,p,\ell,\epsilon} m_p(B_h)^{-\sigma+\epsilon}$$

for all $\epsilon > 0$, where $\phi_w \in L^2(X_{p,\ell})$ is defined by (6.17). Let us also consider a family of functions $w_x(y) := w(x^{-1}y)$ with $x \in G_\infty$. We observe that $\phi_{w_x} = \rho_{\infty,\ell}(x)(\phi_w)$ for the operator

$$\rho_{\infty,\ell}(x) : L^2(X_{p,\ell}) \longrightarrow L^2(X_{p,\ell}) : \phi \longmapsto \phi \circ x^{-1}.$$

Since $\rho_{\infty,\ell}(x)$ commutes with $\rho_{p,\ell}(\beta_h)$ and $\|\rho_{\infty,\ell}(x)\| = 1$, we deduce that for every $x \in G_\infty$,

$$\left\| \rho_{p,\ell}(\beta_h)\phi_{w_x} - \int_{X_{p,\ell}} \phi_{w_x} \, \mathrm{d}\mu_{p,\ell} \right\|_{L^2(X_{p,\ell})} \ll_{w,p,\ell,\epsilon} m_p(B_h)^{-\sigma+\epsilon}.$$

Let π be an irreducible unitary representation of G_p which is discretely embedded in $\rho_{p,\ell}$. Since the sets B_h are $\mathsf{G}(\mathbb{Z}_p)$-invariant, $\pi(\beta_h) = 0$ when π has no non-zero $\mathsf{G}(\mathbb{Z}_p)$-invariant vectors. Hence, we may assume that π is spherical and denote by $F_\pi \in L^2(X_{p,\ell})$ the unique unit $\mathsf{G}(\mathbb{Z}_p)$-invariant vector associated to π. Arguing exactly as in the proof of Theorem 5.2 (see (5.5)), we deduce that

$$\|\pi(\beta_h)\| \ll_{w,p,\ell,\epsilon} |\langle \phi_{w_x}, F_\pi \rangle|^{-1} \, m_p(B_h)^{-\sigma+\epsilon},$$

provided that $\langle \phi_{w_x}, F_\pi \rangle \neq 0$. Moreover,

$$\langle \phi_{w_x}, F_\pi \rangle = \int_{G_\infty} w(x^{-1}g) f(g) \, \mathrm{d}m_\infty(g),$$

where $f(g) := F_\pi(g\Gamma_{p,\ell})$. Since f is a non-zero function which is locally L^2-integrable, it follows from the following general version of the Local Ergodic Theorem (Lemma 6.7 below) that there exists $x \in G_\infty$ such that $\langle \phi_{w_x}, F_\pi \rangle \neq 0$. Hence, we conclude that

$$\|\pi(\beta_h)\| \ll_{\pi,w,p,\ell,\epsilon} m_p(B_h)^{-\sigma+\epsilon} \tag{6.22}$$

for all $\epsilon > 0$. We refer to [16, Ch. 2] for the classification of the irreducible unitary representations of $G_p \simeq \mathrm{SL}_2(\mathbb{Q}_p)$. In particular, let us consider the complementary series representations $\pi_s \in \widehat{G}_p$ with $s \in (0,1)$. These representations are spherical, and we recall that the corresponding spherical function are estimated as

$$\|g\|_p^{-(1-s)} \ll_{p,s} |\omega_{\pi_s}(g)| \ll_{p,s} \|g\|_p^{-(1-s)} \quad \text{for } g \in G_p. \tag{6.23}$$

Using this bound, we deduce from (2.3) that

$$h^{-(1-s)} \ll_{p,s} \|\pi_s(\beta_h)\| \ll_{p,s} h^{-(1-s)}.$$

Since also

$$h^2 \ll_p m_p(B_h) \ll_p h^2,$$

we conclude that

$$m_p(B_h)^{-(1-s)/2} \ll_{p,s} \|\pi_s(\beta_h)\| \ll_{p,s} m_p(B_h)^{-(1-s)/2}. \tag{6.24}$$

It will be important for us that the implicit constants in (6.23) and hence in (6.24) are uniformly bounded for $s \le s_0 < 1$.

Comparing (6.22) and (6.24) when $m_p(B_h) \to \infty$, we deduce that if the complementary series representation π_s is discretely embedded in $\rho_{p,\ell}$, then $s \le 1 - 2\sigma$. Hence, we conclude that if $\pi_s \in \widehat{G}_p^{\mathrm{aut},0}$, then $s \le 1/2$ if G is anisotropic over $\mathbb{Q}$, and $s \le 7/8$ if G is anisotropic over $\mathbb{Q}$. The continuous component of the representations $\rho_{p,\ell}$ has been described (in much greater generality) by Langlands [26]. It follows from this description that the continuous component is tempered. Moreover, it follows from the description of the unitary dual of $\mathrm{SL}_2(\mathbb{Q}_p)$ (see, for instance, [16, Ch. 2]) that the only non-tempered irreducible unitary representations are the complementary series π_s. Therefore,

$$\rho_{p,\ell} = \left(\sum_i \pi_{s_i}^{\oplus n_i} \right) \oplus \rho'_{p,\ell},$$

where $\rho'_{p,\ell}$ is a tempered representation and $s_i \le 1 - 2\sigma$ for all i. Since for tempered representations the bound (2.2) holds, we conclude that

$$\|\rho_{p,\ell}(\beta_h)\| \ll_{p,\ell} m_p(B_h)^{-(1-s_{\max})/2},$$

where $s_{\max} := \max(s_i)$. This completes the proof of the theorem modulo Lemma 6.7. ☐

Lemma 6.10. *There exists a collection of smooth non-negative compactly supported functions $w^{(r)}$, $r \in (0, r_0)$, on G_∞ such that for every locally L^2-integrable f on G_∞,*

$$\left(\int_{G_\infty} w^{(r)} \, \mathrm{d}m_\infty \right)^{-1} \int_{G_\infty} w^{(r)}(x^{-1}g) f(g) \, \mathrm{d}m_\infty(g) \longrightarrow f(x^{-1}) \quad \text{as } r \longrightarrow 0^+, \tag{6.25}$$

for almost all $x \in G_\infty$.

Proof. We fix a G_∞-left-invariant Riemannian metric on $\mathsf{D}(\mathbb{R})$ and consider

$$w^{(r)}(g) := \phi\big(r^{-1} d(g, e)\big),$$

where ϕ is a smooth non-negative symmetric bump function at 0. Moreover, we assume that ϕ is non-increasing on $\mathbb{R}^+$. We note that the claim of the lemma clearly holds for continuous functions.

We set

$$A_r f(x) := \left(\int_{G_\infty} w^{(r)} \, \mathrm{d}m_\infty \right)^{-1} \int_{G_\infty} w^{(r)}(x^{-1}g) f(g) \, \mathrm{d}m_\infty(g),$$

and define the corresponding maximal function

$$M_\phi f(x) := \sup_{r \in (0, r_0)} A_r |f|(x).$$

If the maximal function satisfies the bound

$$\|M_\phi f\|_2 \ll \|f\|_2 \tag{6.26}$$

for every L^2-integrable f, then by a standard argument one can extend (6.25) from continuous functions to general L^2-integrable functions. Hence, it remains to prove (6.26). In fact, it can be deduced from the classical maximal inequality for the operators

$$Mf(x) := \sup_{r>0} m_\infty(B(x,r))^{-1} \int_{B(x,r)} |f(g)| \, \mathrm{d}m_\infty(g),$$

where $B(x,r)$ denotes the balls in G_∞ with respect to the metric d. We choose positive parameters $\alpha_i, r_i = O_\phi(1)$ so that

$$\phi \leq \sum\nolimits_i \alpha_i \, \chi_{B(0,r_i)} \qquad \text{and} \qquad \sum\nolimits_i \alpha_i r_i \ll_\phi 1.$$

Then

$$\begin{aligned} \int_{G_\infty} w^{(\epsilon)}(x^{-1}g)|f(g)| \, \mathrm{d}m_\infty(g) &\leq \sum_i \alpha_i \int_{B(x,r_i\epsilon)} |f(g)| \, \mathrm{d}m_\infty(g) \\ &\leq \left(\sum_i \alpha_i m_\infty(B(x,r_i\epsilon)) \right) Mf(x) \\ &\ll_\phi Mf(x). \end{aligned}$$

Hence, (6.26) follows from the classical maximal inequality. □

References

[1] C. ANANTHARAMAN-DELAROCHE, "On spectral characterizations of amenability", *Isr. J. Math.* **137** (2003), p. 1-33.

[2] B. BEKKA, P. DE LA HARPE & A. VALETTE, *Kazhdan's property (T)*, New Mathematical Monographs, vol. 11, Cambridge University Press, 2008, xiii+472 pages.

[3] V. BLOMER & F. BRUMLEY, "On the Ramanujan conjecture over number fields", *Ann. Math.* **174** (2011), no. 1, p. 581-605.

[4] ———, "The role of the Ramanujan conjecture in analytic number theory", *Bull. Am. Math. Soc.* **50** (2013), no. 2, p. 267-320.

[5] M. BURGER, J.-S. LI & P. SARNAK, "Ramanujan duals and automorphic spectrum", *Bull. Am. Math. Soc.* **26** (1992), no. 2, p. 253-257.

[6] M. BURGER & P. SARNAK, "Ramanujan duals. II", *Invent. Math.* **106** (1991), no. 1, p. 1-11.

[7] P. CARTIER, "Representations of $\mathfrak{p}$-adic groups: a survey", in *Automorphic forms, representations and L-functions (Oregon State Univ., Corvallis, 1977)*, Proceedings of Symposia in Pure Mathematics, vol. 33.1, American Mathematical Society, 1979, p. 111-155.

[8] L. CLOZEL, "Démonstration de la conjecture τ", *Invent. Math.* **151** (2003), no. 2, p. 297-328.

[9] ———, "Spectral theory of automorphic forms", in *Automorphic forms and applications*, IAS/Park City Mathematics Series, vol. 12, American Mathematical Society, 2007, p. 43-93.

[10] M. G. COWLING, "The Kunze–Stein phenomenon", *Ann. Math.* (1978), p. 209-234.

[11] H. DAVENPORT, *Analytic methods for Diophantine equations and Diophantine inequalities*, Cambridge Mathematical Library, Cambridge University Press, 2005, xx+140 pages.

[12] W. DUKE, J. FRIEDLANDER & H. IWANIEC, "Bounds for automorphic L-functions", *Invent. Math.* **112** (1993), no. 1, p. 1-8.

[13] P. EYMARD, *Moyennes et représentations unitaires*, Lecture Notes in Mathematics, vol. 300, Springer, 1972, 113 pages.

[14] T. Finis & J. Matz, "On the asymptotics of Hecke operators for reductive groups", *Math. Ann.* **380** (2021), no. 3-4, p. 1037-1104.

[15] R. Gangolli & V. S. Varadarajan, *Harmonic Analysis of Spherical Functions on Real Reductive Groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 101, Springer, 1988, xiv+365 pages.

[16] I. M. Gel'fand, M. I. Graev & I. I. Piatetski-Shapiro, *Representation theory and automorphic functions*, W. B. Saunders Co., 1969.

[17] A. Ghosh, A. Gorodnik & A. Nevo, "Counting intrinsic Diophantine approximations in simple algebraic groups", *Isr. J. Math.* **251** (2022), no. 2, p. 443-466.

[18] A. Gorodnik & A. Nevo, *The ergodic theory of lattice subgroups*, Annals of Mathematics Studies, vol. 172, Princeton University Press, 2009, 136 pages.

[19] ———, "Quantitative ergodic theorems and their number theoretic applications", *Bull. Am. Math. Soc.* **52** (2015), no. 1, p. 65-113.

[20] ———, "Discrepancy of rational points in simple algebraic groups", `https://arxiv.org/abs/2104.06765`, 2021.

[21] F. P. Greenleaf, "Amenable actions of locally compact groups", *J. Funct. Anal.* **4** (1969), p. 295-315.

[22] D. R. Heath-Brown, "A new form of the circle method, and its application to quadratic forms", *J. Reine Angew. Math.* **481** (1996), p. 149-206.

[23] A. Iozzi & A. Nevo, "Algebraic hulls and the Følner property", *Geom. Funct. Anal.* **6** (1996), no. 4, p. 666-688.

[24] D. A. Kazhdan, "Connection of the dual space of a group with the structure of its closed subgroups", *Funct. Anal. Appl.* **1** (1967), p. 63-65.

[25] H. H. Kim & P. Sarnak, "Refined estimates towards the Ramanujan and Selberg conjectures", appendix in H. H. Kim, "Functoriality for the exterior square of GL(4) and the symmetric fourth of GL(2)", *J. Am. Math. Soc.* **16** (2003), no. 1, 139–183.

[26] R. P. Langlands, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Mathematics, vol. 544, Springer, 1976, v+337 pages.

[27] I. G. Macdonald, "Spherical functions on a p-adic Chevalley group", *Bull. Am. Math. Soc.* **74** (1968), no. 3, p. 520-525.

[28] P. Mattila, *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability*, Cambridge Studies in Advanced Mathematics, vol. 44, Cambridge University Press, 1995, xii+343 pages.

[29] J. Matz & N. Templier, "Sato-Tate equidistribution for families of Hecke-Maass forms on $\mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$", *Algebra Number Theory* **15** (2021), no. 6, p. 1343-1428.

[30] A. Nevo, "Spectral transfer and pointwise ergodic theorems for semi-simple Kazhdan groups", *Math. Res. Lett.* **5** (1998), no. 3, p. 305-325.

[31] V. Platonov & A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics, vol. 139, Academic Press Inc., 1994, xi+614 pages.

[32] J. Rogawski, "Modular forms, the Ramanujan conjecture, and the Jacquet-Langlands correspondence", appendix in A. Lubotzky, *Discrete groups, expanding graphs and invariant measures*, Progress in Mathematics, vol. 125, Birkhäuser, 1994.

[33] P. Sarnak, "Notes on the generalized Ramanujan conjectures", in *Harmonic analysis, the trace formula, and Shimura varieties (Toronto, 2003)*, Clay Mathematics Proceedings, vol. 4, American Mathematical Society, 2005, p. 659-685.

[34] P. Sarnak & X. Xue, "Bounds for multiplicities of automorphic representations", *Duke Math. J.* **64** (1991), no. 1, p. 207-227.

[35] A. Selberg, "Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series", *J. Indian Math. Soc., n. Ser.* **20** (1956), p. 47-87.

[36] ———, "On estimation of Fourier coefficients of modular forms", in *Theory of numbers (Pasadena, 1963)*, Proceedings of Symposia in Pure Mathematics, vol. 1965, American Mathematical Society, 1965, p. 1-15.

[37] A. J. Silberger, *Introduction to harmonic analysis on reductive p-adic groups*, Mathematical Notes, vol. 23, Princeton University Press, 1979, iv+372 pages.

[38] M. TADIC, "Harmonic analysis of spherical functions on reductive groups over $\mathfrak{p}$-adic fields", *Pac. J. Math.* **109** (1983), no. 1, p. 215-235.
[39] J. TITS, "Reductive groups over local fields", in *Automorphic forms, representations and L-functions (Oregon State Univ., Corvallis, 1977)*, Proceedings of Symposia in Pure Mathematics, vol. 33.1, American Mathematical Society, 1979, p. 29-69.
[40] A. VECA, "The Kunze–Stein phenomenon", PhD Thesis, Univ. of New South Wales, 2002.
[41] V. E. VOSKRESENSKY, *Algebraic groups and their birational invariants*, Translations of Mathematical Monographs, vol. 179, American Mathematical Society, 1998, xiii+218 pages.
[42] A. WEIL, *Adeles and algebraic groups*, Progress in Mathematics, vol. 23, Birkhäuser, 1982, v+126 pages.

Alexander GORODNIK
Institute für Mathematik, Universität Zürich
Switzerland
E-mail: alexander.gorodnik@math.uzh.ch

Amos NEVO
Department of Mathematics, Technion, Israel
E-mail: anevo@tx.technion.ac.il
Department of Mathematics, University of Chicago, USA
E-mail: nevo@uchicago.edu