Advances in Mathematics 443 (2024) 109613

Contents lists available at ScienceDirect

Advances in Mathematics

journal homepage: www.elsevier.com/locate/aim

Optimal Diophantine exponents for SL(n)

Subhajit Jana [a,*], Amitay Kamber [b]

[a] Queen Mary University of London, Mile End Road, London E14 NS, UK
[b] Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK

ARTICLE INFO

Article history:
Received 20 September 2023
Received in revised form 28 February 2024
Accepted 29 February 2024
Available online xxxx
Communicated by Lars Olsen

MSC:
11F72
11J13

Keywords:
Diophantine approximation
Automorphic spectrum
Density hypothesis

ABSTRACT

The *Diophantine exponent* of an action of a group on a homogeneous space, as defined by Ghosh, Gorodnik, and Nevo, quantifies the complexity of approximating the points of the homogeneous space by the points on an orbit of the group. We show that the Diophantine exponent of the $\mathrm{SL}_n(\mathbb{Z}[1/p])$-action on the generalized upper half-space $\mathrm{SL}_n(\mathbb{R})/\mathrm{SO}_n(\mathbb{R})$, lies in $[1, 1+O(1/n)]$, substantially improving upon Ghosh–Gorodnik–Nevo's method which gives the above range to be $[1, n-1]$. We also show that the exponent is *optimal*, i.e. equals one, under the assumption of *Sarnak's density hypothesis*. The result, in particular, shows that the optimality of Diophantine exponents can be obtained even when the *temperedness* of the underlying representations, the crucial assumption in Ghosh–Gorodnik–Nevo's work, is not satisfied. The proof uses the spectral decomposition of the homogeneous space and bounds on the local L^2-norms of the Eisenstein series.

* Corresponding author.
E-mail addresses: s.jana@qmul.ac.uk (S. Jana), ak2356@dpmms.cam.ac.uk (A. Kamber).

https://doi.org/10.1016/j.aim.2024.109613
0001-8708/

1. Introduction

1.1. Main theorems

Let p be a prime. It is a standard fact that $\mathrm{SL}_n(\mathbb{Z}[1/p])$ is dense in $\mathrm{SL}_n(\mathbb{R})$. Following Ghosh, Gorodnik and Nevo [21,22], we wish to make this density quantitative.

Let $\mathbb{H}^n := \mathrm{SL}_n(\mathbb{R})/\mathrm{SO}_n(\mathbb{R})$ be the symmetric space of $\mathrm{SL}_n(\mathbb{R})$. Fix an $\mathrm{SL}_n(\mathbb{R})$-invariant Riemannian metric dist on $\mathbb{H}^n$.

We define the *height* of $\gamma \in \mathrm{SL}_n(\mathbb{Z}[1/p])$ as

$$\mathrm{ht}(\gamma) = \min\left\{k \in \mathbb{N} \mid p^k\gamma \in \mathrm{Mat}_n\left(\mathbb{Z}\right)\right\},$$

where Mat_n denotes the space of $n \times n$ matrices.

Our work is based on the following definition, motivated by [22] (see Section 2 for comparison).

Definition 1.1. The *Diophantine exponent* $\kappa(x, x_0)$ of $x, x_0 \in \mathbb{H}^n$ is the infimum over $\zeta < \infty$, such that there exists an $\varepsilon_0 = \varepsilon_0(x, x_0, \zeta)$ with the property that for every $\varepsilon < \varepsilon_0$ there is a $\gamma \in \mathrm{SL}_n(\mathbb{Z}[1/p])$ satisfying

$$\mathrm{dist}(\gamma^{-1}x, x_0) \leq \varepsilon \text{ and } \mathrm{ht}(\gamma) \leq \zeta\frac{n+2}{2n}\log_p(\varepsilon^{-1}).$$

The *Diophantine exponent of* $x_0 \in \mathbb{H}^n$ is

$$\kappa(x_0) = \inf\left\{\tau \mid \kappa(x, x_0) \leq \tau \text{ for almost every } x \in \mathbb{H}^n\right\}.$$

The *Diophantine exponent of* $\mathbb{H}^n$ is

$$\kappa = \inf\left\{\tau \mid \kappa(x, x_0) \leq \tau \text{ for almost every } (x, x_0) \in \mathbb{H}^n \times \mathbb{H}^n\right\}.$$

In [22] (in a more generalized context) the following is shown.

Proposition 1.2. *For every $x_0 \in \mathbb{H}^n$ and almost every $x \in \mathbb{H}^n$ we have $\kappa(x, x_0) = \kappa(x_0)$, and for almost every $x, x_0 \in \mathbb{H}^n$ we have $\kappa(x, x_0) = \kappa(x_0) = \kappa$.*

The artificial insertion of the factor $\frac{n+2}{2n}$ in the definition of $\kappa(x, x_0)$ is to ensure that

$$\text{for every } x_0 \in \mathbb{H}^n \quad \kappa(x_0) \geq 1, \quad \text{and} \quad \kappa \geq 1$$

as we explain below. It is thus natural to wonder about a corresponding upper bound for $\kappa(x_0)$, namely whether

$$\text{for every } x_0 \in \mathbb{H}^n \quad \kappa(x_0) = 1,$$

and in particular, whether

$$\kappa = 1.$$

In this case we say that κ is the *optimal Diophantine exponent.*

In [22] the Diophantine exponents are studied in great generality for a lattice Γ in a group G, acting on a homogeneous space G/H, when H is a subgroup of G. In our case $\Gamma = \mathrm{SL}_n(\mathbb{Z}[1/p])$, $G = \mathrm{SL}_n(\mathbb{R}) \times \mathrm{SL}_n(\mathbb{Q}_p)$ and $H = \mathrm{SO}_n(\mathbb{R}) \times \mathrm{SL}_n(\mathbb{Q}_p)$ (see Section 2 for details). The optimality of κ is proved in [22] under certain crucial *temperedness* assumptions of the action of H on $L^2(\Gamma\backslash G)$. Unfortunately, in our particular situation, the temperedness assumption is not satisfied. As we explain in Section 2, the arguments of [22] imply the following non-optimal upper bounds on κ.

Theorem 1 *(Ghosh–Gorodnik–Nevo, [22] and Section 2). Let $n > 1$ be a positive integer.*

(1) *For $n = 2$, assuming the Generalized Ramanujan Conjecture (GRC) for* GL(2)*, for every $x_0 \in \mathbb{H}^2$ we have $\kappa(x_0) = 1$. Unconditionally, for every $x_0 \in \mathbb{H}^2$ we have $\kappa(x_0) \leq 32/25$.*
(2) *For $n \geq 3$, for every $x_0 \in \mathbb{H}^n$ we have $\kappa(x_0) \leq n - 1$.*

In particular, the same bounds also hold for κ.

One of the goals of this paper is to substantially improve the upper bound of κ, in particular, to prove that κ is *essentially* optimal. Our main theorem is as follows.

Theorem 2. *Let $n > 1$ be a positive integer.*

(1) *For $n = 2$ or $n = 3$ we have $\kappa = 1$.*
(2) *For every $n \geq 4$ we have*

$$\kappa \leq 1 + \frac{2\theta_n}{n - 1 - 2\theta_n},$$

where θ_n is the best known bound towards the GRC for GL(n)*.*

Remark 1.3. We refer to Subsection 4.4 for the precise definition of θ_n. From Equation (4.7) and Equation (4.8) we obtain that

$$\kappa \leq \begin{cases} 11/8 & \text{for } n = 4, \\ (n^2+1)/(n^2-n) = 1 + O(1/n) & \text{for } n \geq 5 \end{cases}$$

Notice that the bound on κ gets better as n grows. A non-precise reason is that as n grows, the Hecke operator on the cuspidal spectrum gets closer and closer to having square-root cancellation.

We also show the optimality of κ for any n assuming *Sarnak's Density Hypothesis* for $\mathrm{GL}(n)$ as in Conjecture 2 which is a *much weaker* version of the GRC for $\mathrm{GL}(n)$.

Theorem 3. *For every n, assuming* Sarnak's Density Hypothesis *for* $\mathrm{GL}(n)$ *as in Conjecture 2 below, we have $\kappa = 1$.*

We consider Theorem 3 as a *proof of concept* for the claim that the Diophantine exponent is usually optimal, even without the temperedness assumption. This is in line with *Sarnak's Density Hypothesis* in the theory of automorphic forms, which informally states that the automorphic forms are expected to be tempered *on average* (see discussion in Subsection 4.4). Our result, at least on the assumption of the density hypothesis, also negatively answers a question of Ghosh–Gorodnik–Nevo who asked whether optimal Diophantine exponent implies temperedness; see [22, Remark 3.6] ([29] and [39] also provide answers to this question, in different contexts).

Note that Theorem 1 is about $\kappa(x_0)$ while Theorem 2 and Theorem 3 are about κ. The difference may seem minor but is crucial for the proof. In the general setting of Ghosh–Gorodnik–Nevo, we expect that *usually* $\kappa = 1$ (e.g., when $\mathrm{SL}(n)$ is replaced by another group), but $\kappa(x_0)$ may be larger, because of local obstructions.

An example, based on [19, Section 2.1], is the action of $\mathrm{SO}_{n+1}(\mathbb{Z}[1/p])$ on the sphere S^n, which we discuss shortly in Subsection 2.1. In our situation, we conjecture that for every $x_0 \in \mathbb{H}^n$ we have $\kappa(x_0) = 1$, but do not know how to prove it even assuming the GRC, except for $n = 2$ (as in Theorem 1) and $n = 3$.

Theorem 4. *For $n = 3$, assuming the GRC, we have $\kappa(x_0) = 1$ for every $x_0 \in \mathbb{H}^n$.*

1.2. Almost-covering

Our proofs of Theorem 2 and Theorem 3 use the spectral theory of $L^2(\mathrm{SL}_n(\mathbb{Z})\backslash\mathbb{H}^n)$. It is therefore helpful to understand the problem in an equivalent language that is more suitable for the spectral theory and is of independent interest.

First, it suffices to assume that $x, x_0 \in \mathbb{X} := \mathrm{SL}_n(\mathbb{Z})\backslash\mathbb{H}^n$ as the Riemannian distance dist is left-invariant under $\mathrm{SL}_n(\mathbb{Z})$ and the height function ht is bi-invariant under $\mathrm{SL}_n(\mathbb{Z})$. The set of points of the form $\mathrm{SL}_n(\mathbb{Z})\gamma x_0 \in \mathbb{X}$ for $\gamma \in \mathrm{SL}_n(\mathbb{Z}[1/p])$ with $\mathrm{ht}(\gamma) \le k$ is in bijection with the set $R(1)\backslash R(p^{kn})$, where $R(p^{kn}) := \{A \in \mathrm{Mat}_n(\mathbb{Z}) \mid \det(A) = p^{kn}\}$. More precisely, there is a bijection

$$\mathrm{SL}_n(\mathbb{Z})\backslash\{\gamma \in \mathrm{SL}_n(\mathbb{Z}[1/p]) \mid \mathrm{ht}(\gamma) \le k\} \cong R(1)\backslash R(p^{kn})$$

given by multiplication by p^k, and the bijection above holds for generic x_0. In general, we have a surjection from $R(1)\backslash R(p^{kn})$ to the set on the left hand side above.

It is well known that

$$|R(1)\backslash R(p^{kn})| \asymp p^{kn(n-1)};$$

see Subsection 4.2 (and Subsection 1.6 for the notation $\asymp$).

The parameter $\kappa(x_0)$ measures the *almost-covering* of $\mathbb{X}$ by the set of points above. Consider a sequence of natural numbers $k = k(\varepsilon)$, such that the ε-balls around the $|R(1)\backslash R(p^{kn})|$ points in $\mathbb{X}$ of the form $\mathrm{SL}_n(\mathbb{Z})\gamma x_0$ with $\mathrm{ht}(\gamma) \leq k$ cover all but $o(1)$ of the space $\mathbb{X}$, when $\varepsilon \to 0$ (compare [37, Proposition 3.1]). The number $\kappa(x_0)$ is closely related to the growth of $k(\varepsilon)$ as $\varepsilon \to 0$.

Therefore, it is required that as $\varepsilon \to 0$,

$$m(B_\varepsilon)|R(1)\backslash R(p^{kn})| \geq m(\mathbb{X}) - o(1),$$

where m denotes the $\mathrm{SL}_n(\mathbb{R})$-invariant measure on $\mathbb{X}$ and $m(B_\varepsilon)$ is the volume of a ball of radius ε in $\mathbb{H}^n$. We have $m(B_\varepsilon) \asymp \varepsilon^d$ where $d := \dim \mathbb{H}^n = \frac{(n+2)(n-1)}{2}$. Thus, we deduce that

$$k(\varepsilon) \geq \frac{d}{n(n-1)} \log_p(\varepsilon^{-1})(1 - o(1)).$$

The same argument shows that $\kappa(x_0) \geq 1$. See Fig. 1 for a pictorial description.

We remark that one can also consider the problem of *covering*, where we would like to cover an entire compact region of $\mathbb{X}$ by small balls of radius ε around the Hecke points (unlike Theorem 2 and Theorem 3, which are essentially about *almost-covering*). This is also the difference between part (i) and part (ii) of [19, Theorem 1.3]. See also [14] for the covering problem of Hecke points around $e \in \mathbb{X}$. We will not discuss it further in this work but mention that our methods, and in particular the arguments in Theorem 4, can lead to a better understanding of the covering problem as well, but the results are not expected to be optimal. For example, in the covering problem on the 3-dimensional sphere the spectral approach leads to a covering exponent which is 3/2 times the conjectural value (see [11] and the references therein).

1.3. Outline of the proof

We start by describing the work of Ghosh–Gorodnik–Nevo in [22] in our setting, namely, on $\mathbb{X} := \mathrm{SL}_n(\mathbb{Z})\backslash\mathrm{SL}_n(\mathbb{R})/\mathrm{SO}_n(\mathbb{R})$. Their work crucially relies on the existence and, in fact, an explicit description of the *spectral gap* for a certain averaging operator (i.e., *a quantitative mean ergodic theorem*) on a certain homogeneous space arising from a general reductive group. In our case, the situation is simpler and the relevant operator turns out to be the (adjoint) Hecke operator $T^*(p^{nk})$, for certain k, acting on $L^2(\mathbb{X})$. By a standard duality theorem for Hecke operators [15], the action reduces to an operator from the Hecke algebra of $\mathrm{SL}_n(\mathbb{Q}_p)$ on $L^2(\mathrm{SL}_n(\mathbb{Z}[1/p])\backslash\mathrm{SL}_n(\mathbb{R}) \times \mathrm{SL}_n(\mathbb{Q}_p))$.

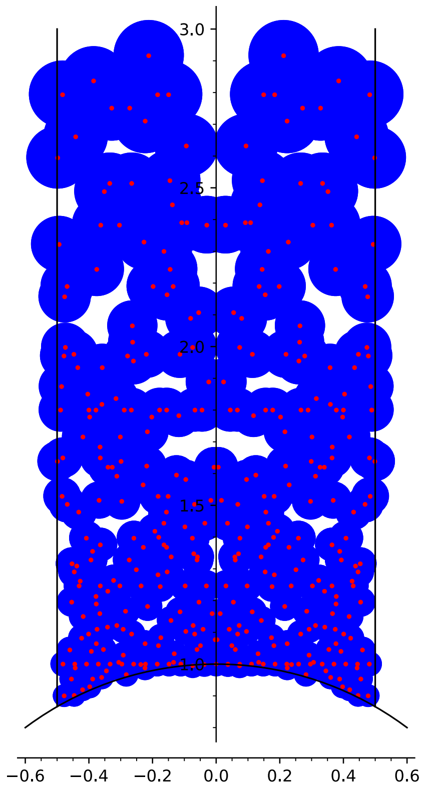

Fig. 1. Covering of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}^2$ by balls around $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{Z}[1/3])$. The height of the points is bounded by 3, and the balls are of radius 3^{-3}.

One can relate the spectral gap of this Hecke operator to the Diophantine exponent (see Theorem 5). The spectral gap, using the theory of spherical functions, can be determined by a certain *integrability exponent*, which is parameterized by a number $2 \le q \le \infty$ (see Proposition 2.4). When the integrability exponent is $q = 2$ (alternatively, an underlying representation is *tempered*, see Section 2) one gets the optimal Diophantine exponent $\kappa = 1$. In general, the method of [22] only shows that $\kappa \le q/2$ (see Theorem 5).

Assuming $n \ge 3$, explicit property (T) implies that $q \le 2(n-1)$ (see the work of Oh [36] for a nice proof). This gives the claimed result that $\kappa \le n-1$. On the other hand, the theory of Eisenstein series implies that $q \ge 2(n-1)$ (see [15, Theorem 1.5]). As the spectral gap cannot be further improved, the method in [22] is limited and we need a different approach to improve the upper bound of κ.

One of the novelties in our work is to use the full spectral decomposition of $L^2(\mathbb{X})$ and treat different types of elements of the spectrum separately. More precisely, we actually analyze the spectral decomposition of $L^2(\mathbb{X}_0)$, where $\mathbb{X}_0 := \mathrm{PGL}_n(\mathbb{Z})\backslash\mathbb{H}^n$. From our experience, this maneuver usually gives more detailed knowledge of the sizes of the

Hecke eigenvalues than that is provided by the spectral gap alone. The basic spectral decomposition uses the *theory of Eisenstein Series* due to Langlands (see [34]) and provides a decomposition of the form $L^2(\mathbb{X}_0) \cong L^2_{\mathrm{cusp}}(\mathbb{X}_0) \oplus L^2_{\mathrm{Eis}}(\mathbb{X}_0)$. The relevant Hecke operator acts on each part of the spectrum and gives rise to certain Hecke eigenvalues.

The cuspidal part $L^2_{\mathrm{cusp}}(\mathbb{X}_0)$ decomposes discretely into irreducible representations. The Generalized Ramanujan Conjecture (GRC) predicts that all such representations are tempered, i.e. the sizes of the Hecke eigenvalues can be bounded optimally. While the GRC is completely open even for $n = 2$, good bounds towards it for each individual representation are known; see [40]. This bound is used in our unconditional results for $n \geq 4$. However, even for $n = 2$ and assuming the best-known bounds, we are unable to reach $\kappa = 1$.

To overcome this problem, we notice that one does not need optimal bounds for individual Hecke eigenvalues, but *optimal bounds on average*. In general, Sarnak's Density Hypothesis (see [42,43]) predicts (in a slightly different setting) that the GRC should hold on average for a nice enough family of automorphic representations. In our conditional result, we assume a certain form of the density hypothesis, namely Conjecture 1 which can be realized as a higher rank analogue of Sarnak's Density Hypothesis (see Proposition 4.12 and discussion there), and apply to our question. This approach was already used in different contexts to deduce results of a similar flavor (see [41,10,24,25]). The version that is relevant for us can be realized as *Density relative to the Weyl's law*. We refer to Subsection 4.4 for a complete discussion. This density property is known for $n = 2$ and $n = 3$ by the work of Blomer [3] and Blomer–Buttcane–Raulf [7], respectively; see Subsection 4.4.

Remark 1.4. Recently, Assing and Blomer in [2, Theorem 1.1] proved Sarnak's density hypothesis in a non-archimedean aspect, namely for the automorphic forms for principal congruence subgroups of square-free level. As an application they solved a related problem in [2, Theorem 1.5] namely, *optimal lifting* for $\mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$ with square-free q, conditional on a hypothesis [2, Hypothesis 1] about certain local L^2-bounds of the Eisenstein series; see also [27, Theorem 4, §8].

Dealing with the Eisenstein part $L^2_{\mathrm{Eis}}(\mathbb{X}_0)$ is less complicated arithmetically than the cuspidal spectrum – the size of the Hecke eigenvalues can be understood inductively using the results of Mœglin and Waldspurger [33]. However, the Eisenstein part is more complicated analytically, because of its growth near the cusp. This problem can be regarded as a Hecke eigenvalue weighted Weyl's law. Similar to the proofs in [32,35] we need to show that the contribution of the Eisenstein part is small compared to the cuspidal spectrum. The problem is non-trivial due to the weights coming from the Hecke eigenvalues, which can be quite large for the non-tempered part of the Eisenstein spectrum. We show that the largeness of the Hecke eigenvalues for the non-tempered automorphic forms is compensated by a *low cardinality* of such forms.

The exact result that we need is estimates on the L^2*-growth of Eisenstein series in compact domains*, see Subsection 4.8 for a formulation. This result was proved by Miller in [32] as the main estimate in his proof of Weyl's law for $\mathrm{SL}_3(\mathbb{Z})$, but was open for $n \geq 4$. In a companion paper [27] we solve this problem; see Subsection 4.8.

1.4. Generalizations and open problems

The questions in this work can be generalized to other groups, by replacing the underlying group $\mathrm{SL}(n)$ by another semisimple simply connected algebraic group. Without giving the full definitions, we expect that the Diophantine exponent κ will be optimal, even without the presence of an optimal spectral gap (compare the discussion in [22, Remark 3.6]). Our proof certainly generalizes to $\mathrm{SL}_m(D)$, when D is a division algebra over $\mathbb{Q}$.

One can also wonder about the Diophantine exponents when $\mathbb{H}^n$ is replaced with $\mathrm{SL}_n(\mathbb{R})$, with some left-invariant Riemannian metric. The methods of this paper can, in principle, be used for this problem as well, but the spectral decomposition, as well as other analytical problems, is more complicated due to the absence of sphericality, and we were not yet able to overcome them. However, one can show that in this situation $\kappa(x_0) = \kappa$, since we have a right $\mathrm{SL}_n(\mathbb{R})$-action (the metric is not right-invariant, so this is not completely trivial).

1.5. Structure of the article

In Section 2 we explain how our question is related to the work of Ghosh–Gorodnik–Nevo, and prove Theorem 1.

In Section 3 we discuss the relevant local groups and their (spherical) representation theory.

In Section 4 we discuss the global preliminaries that we need, and in particular discuss Hecke operators, Langlands spectral decomposition, and the description of the spectrum. In Subsection 4.4 we discuss the Density Conjecture that we require for this work, and in Subsection 4.8 we discuss L^2-bounds on Eisenstein series in compact domains.

In Section 5 we reduce the study of Diophantine exponents to a certain analytic problem (cf. Lemma 5.12), in the spirit of [22].

In Section 6 we apply Langlands spectral decomposition and Proposition 4.20 to reduce the spectral problem to a combinatorial problem.

In Section 7 we complete the proof of Theorem 2 and Theorem 3.

Finally, in Section 8 we prove Theorem 4.

1.6. Notation

The notations $\ll, \asymp, \gg, O, o$ are the usual ones in analytic number theory: for a master parameter $T \to \infty$ and A, B depending on T implicitly we say $A \ll B$ (equivalently,

$A = O(B)$, and $B \gg A$) if there is a constant c such that $A \leq cB$ for T sufficiently large. We write $A \asymp B$ if $A \ll B \ll A$. The implied constants may depend on n and p, without mentioning it explicitly. Also as usual in analytic number theory, δ and η (but not ε) will denote arbitrary small but fixed positive numbers, whose actual values may change from line to line.

1.7. Acknowledgments

The authors thank Amos Nevo for various discussions and detailed comments about a previous version of this work and thank Peter Sarnak for his motivation. The first author thanks Shreyasi Datta for numerous fruitful discussions on various aspects of Diophantine approximation related to this work, and in general. The first author also thanks MPIM Bonn where most of the work was completed during his stay there. The second author is supported by the ERC under the European Union's Horizon 2020 research and innovation program (grant agreement No. 803711). Finally, we thank the anonymous referee for the helpful comments which made the exposition of the paper a lot better.

2. The setting of Ghosh–Gorodnik–Nevo

The goal of this section is to explain how our question fits into the general framework studied by Ghosh–Gorodnik–Nevo in a sequence of works [19–22]. Consequently, we give a sketch of the argument which leads to a proof of Theorem 1. In this section we follow [22, Section 2] and use its notations, which are not the same as the rest of this work, to allow a simple comparison.

Let $G := \mathrm{SL}_n(\mathbb{R}) \times \mathrm{SL}_n(\mathbb{Q}_p)$ and $H := \mathrm{SO}_n(\mathbb{R}) \times \mathrm{SL}_n(\mathbb{Q}_p) < G$. Also let $\Gamma := \mathrm{SL}_n(\mathbb{Z}[1/p])$, which we consider as embedded diagonally in G. Notice that Γ is a lattice in G.

Let $X := G/H \cong \mathrm{SL}_n(\mathbb{R})/\mathrm{SO}_n(\mathbb{R}) \cong \mathbb{H}^n$. Also let dist be the natural Riemannian metric on X, coming from the Killing form on the Lie algebra of $\mathrm{SL}_n(\mathbb{R})$. Notice that the action of G on G/H preserves this metric. We fix natural Haar measures $m_G, m_H, m_X = m_{G/H}$ on G, H, X, respectively.

We define $D\colon \mathrm{SL}_n(\mathbb{Q}_p) \to \mathbb{R}_{\geq 0}$ by

$$D(g_p) = \log(p)\mathrm{ht}(g_p) = \log(p) \min \left\{ k \in \mathbb{Z}_{\geq 0} \mid p^k g_p \in \mathrm{Mat}_n(\mathbb{Z}_p) \right\}.$$

We extend $D\colon G \to \mathbb{R}_{\geq 0}$ by $D(g_\infty, g_p) := D(g_p)$ and denote $|g|_D := e^{D(g)}$, which in our case is simply the p-adic valuation of g_p.

Definition 2.1 *([22], Definition 2.1).* Given $x, x_0 \in X$, the *Diophantine exponent* $\kappa_1(x, x_0)$ is the infimum over ζ, such that there is an $\varepsilon_0 = \varepsilon(x, s_0, \zeta)$ with the property that for every $\varepsilon < \varepsilon_0$ there is a $\gamma \in \Gamma$ satisfying

$$\mathrm{dist}(\gamma^{-1}x, x_0) \leq \varepsilon \text{ and } |\gamma|_D \leq \varepsilon^{-\zeta}.$$

Let us compare Definition 2.1 and Definition 1.1. The inequality $|\gamma|_D = p^{\mathrm{ht}(\gamma)} \le \varepsilon^{-\zeta}$ is equivalent to the inequality $\mathrm{ht}(\gamma) \le \zeta \log_p(\varepsilon^{-1})$. Therefore,

$$\kappa(x, x_0) = \frac{2n}{n+2}\kappa_1(x, x_0).$$

We claim that the choices of G, H, dist, and D satisfy the Assumptions 1 - 4 of [22, Section 2]. First, dist is G-invariant so [22, Assumption 1] holds (in fact, [22] requires a weaker "coarse metric regularity" condition). It also holds that for ε sufficiently small,

$$m_X(\{x \in X \mid \mathrm{dist}(x, x_0) < \varepsilon\}) \asymp \varepsilon^d,$$

where $d := \dim(X) = \dim(\mathrm{SL}_n(\mathbb{R})/\mathrm{SO}_n(\mathbb{R})) = (n+2)(n-1)/2$. Therefore, [22, Assumption 4] holds with local dimension d. Next, it directly follows from the definition that $D(g_1 g_2) \le D(g_1) + D(g_2)$, which means that D is subadditive. While [22, Assumption 2] requires a weaker "coarse norm regularity" condition for D to hold.

Given $t \ge 0$, the set $H_t = \{h \in H \mid D(h) \le t\}$ is of finite Haar measure and moreover, a simple calculation (see Subsection 3.4) shows that

$$m_H(H_t) \asymp e^{n(n-1)t}.$$

Therefore, [22, Assumption 3] is satisfied with an explicit exponent $a := n(n-1)$.

Remark 2.2. Our analysis is slightly different than [22], since the set $G_t = \{g \in G \mid D(g) \le t\}$ is not of finite Haar measure, as assumed in [22]. There are two ways to overcome this minor difference. The first is to ignore it, since this assumption is not used in the proof of [22]. Alternatively, one may define a different metric by

$$D_1(g_\infty, g_p) := D(g_p) + \log \|g_\infty\|,$$

where $\|\cdot\|$ is some submultiplicative matrix norm on $M_n(\mathbb{R})$.

Then it is not hard to show changing D to D_1 will give the same Diophantine exponent, and $G_{t,1} = \{g \in G \mid D_1(g) \le t\}$ will be of finite Haar measure. However, the relation between D_1 and Hecke points is less transparent, so we prefer to work with D instead.

In [22] the authors made the following observation: $\kappa(x, x_0)$ is a $\Gamma \times \Gamma$-invariant function, and Γ acts ergodically on X, and therefore for every $x_0 \in X$ there are constants $\kappa_L(x_0)$ and $\kappa_R(x_0)$ such that $\kappa_L(x_0) = \kappa(x_0, x)$ and $\kappa_R(x_0) = \kappa(x, x_0)$ for almost every x. Similarly, there is a constant κ such that $\kappa = \kappa(x, x_0)$ for almost every $x, x_0 \in X$.

We have the following lower bound of the Diophantine exponent $\kappa_1(x, x_0)$.

Proposition 2.3 *([22], Theorem 3.1). For every $x_0 \in X$ and for almost every $x \in X$, $\kappa_1(x, x_0) \ge d/a = \frac{n+2}{2n}$, or alternatively $\kappa(x, x_0) \ge 1$.*

Therefore, for every $x_0 \in X$, $\kappa_R(x_0) \ge 1$ and $\kappa \ge 1$.

We gave a sketch of the proof of this proposition in the introduction. We remark that [22, Theorem 3.1] actually show that for every $x_0 \in X$ it holds that $\kappa_L(x_0) \geq 1$, but the above statement follows either by the same arguments, or by replacing D with D', where $D'(g) = D(g^{-1})$.

To present upper bounds, we consider the right action of H on $Y := \Gamma\backslash G$, and we endow Y with the natural finite Haar measure coming from G. Let $\beta_t \in C_c(H)$, $\beta_t = \frac{\mathbb{1}_{H_t}}{m_H(H_t)}$ be the normalized characteristic function of H_t. We consider the operator of $\pi_Y(\beta_t)$ on $L^2(Y)$, defined by

$$(\pi_Y(\beta_t)f)(y) := \frac{1}{m_H(H_t)} \int\limits_{H_t} f(yh)\, \mathrm{d}m_H(h).$$

In this case, $\pi_Y(\beta_t)$ can, in fact, be interpreted as a certain spherical Hecke operator for which we have the following mean ergodic theorem.

For any measurable space X we denote

$$L_0^2(X) := \left\{ f \in L^2(X) \mid \int\limits_X f(x)\, \mathrm{d}x = 0 \right\}.$$

Proposition 2.4 *([19], Theorem 4.2). There is an explicit $q = q(n) > 0$ such that as an operator on $L_0^2(Y)$*

$$\|\pi_Y(\beta_t)\|_{\mathrm{op}} \ll_\delta m_H(H_t)^{-q^{-1}+\delta},$$

for every $\delta > 0$.

The value q in the above proposition is the *integrability exponent* for the action of H on $L_0^2(Y)$. The integrability exponent q is the infimum over q' such that the K_H-finite matrix coefficients are in $L^{q'}(H)$, where K_H is a maximal compact subgroup of H. We have the following results on the integrability exponent.

(1) For $n = 2$, using Kim–Sarnak bound towards the Generalized Ramanujan Conjecture (see [40]), one can take $q = 64/25$. Assuming GRC, we have $q = 2$.
(2) For $n \geq 3$, using explicit property (T) from [36] we have $q = 2(n-1)$. Moreover, this choice of q is the best possible.

Using these bounds on the integrability exponents the following result is proved in [22].

Theorem 5 *([22], Theorem 3.3). For every $x_0 \in X$ and almost every $x \in X$,*

$$\kappa_1(x, x_0) \leq \frac{qd}{2a}.$$

Therefore, $\kappa_R(x_0) \leq q/2$, and consequently, $\kappa \leq q/2$.

This recovers Theorem 1 from the results of [22].

2.1. Diophantine exponents on the sphere

This subsection is independent of the rest of the article, and serves to discuss the difference between κ and $\kappa(x_0)$.

The possible difference between κ and $\kappa(x_0)$ has similar origins as the failure of temperedness, and also the failure of optimal L^∞-bounds – embedding of a homogeneous orbit of a subgroup in the space. For example, when $x_0 = I$, there is a homogeneous orbit of $\mathrm{SL}_{n-1}(\mathbb{Z})\backslash\mathrm{SL}_{n-1}(\mathbb{R}) \subset \mathrm{SL}_n(\mathbb{Z})\backslash\mathrm{SL}_n(\mathbb{R})$, and many points of the Hecke orbit of $\mathrm{SL}_n(\mathbb{Z}[1/p])$ around I belong to the image of this homogeneous orbit in $\mathbb{X}$. It seems that this concentration is not dramatic enough to change $\kappa(I)$, but for other groups, this may happen. The goal of this subsection is to give an example with $\mathrm{SL}(n)$ replaced by $\mathrm{SO}(n)$.

Let $n \geq 5$ and $\mathrm{SO}(n)$ be the algebraic group which is the stabilizer of the quadratic from $Q(x_1, \dots, x_n) := x_1^2 + \cdots + x_n^2$.

Replacing $\mathrm{SL}(n)$ by $\mathrm{SO}(n)$, one can study the equidistribution of $\mathrm{SO}_{n+1}(\mathbb{Z}[1/p])$ in $\mathrm{SO}_{n+1}(\mathbb{R})$. This will help up explain the difference between $\kappa(x_0)$ and κ, hinted at in the introduction. For technical reasons, we restrict to $p = 1 \mod 4$.

We let $G := \mathrm{SO}_{n+1}(\mathbb{R}) \times \mathrm{SO}_{n+1}(\mathbb{Q}_p)$, $H = \mathrm{SO}_n(\mathbb{R}) \times \mathrm{SO}_{n+1}(\mathbb{Q}_p)$, and $\Gamma = \mathrm{SO}_{n+1}(\mathbb{Z}[1/p])$ a lattice in G. It holds that $X := G/H \cong S^n$ and we let dist be an $\mathrm{SO}_{n+1}(\mathbb{R})$-invariant Riemannian metric on X. We define D and the Diophantine exponent $\kappa_1(x, x_0)$ as above. The relevant dimension is $d = n$, and the set $H_t = \{h \in H \mid D(h) \leq t\}$ satisfies

$$m_H(H_t) \asymp e^{at},$$

where

$$a = \begin{cases} n^2/4 & n \text{ even} \\ (n+1)(n+3)/4 & n \text{ odd} \end{cases}.$$

We deduce that for every $x_0 \in X$ and almost every $x \in X$ it holds that $\kappa_1(x, x_0) \geq d/a$. The arguments of [22] imply that for every $x_0 \in X$ and almost every $x \in X$, it holds that $\kappa_1(x, x_0) \leq \frac{qd}{2a}$, where

$$q = \begin{cases} n & n \text{ even} \\ n+1 & n \text{ odd} \end{cases}.$$

We *conjecture* that for almost every $x, x_0 \in X$ it holds that

$$\kappa_1(x, x_0) = d/a.$$

We plan to pursue this conjecture in a future work, using the methods of this work.

Now consider the point $x_0 = e = (1, \dots, 0) \in X$. For this specific point, the cardinality of the set of points of the form γe with $D(\gamma) \le t$ is at most the number of solutions to

$$x_1^2 + \cdots + x_{n+1}^2 = p^k,$$

with $x_i \in \mathbb{Z}$ and $p^k \le e^{2t}$. It is standard that the last number is

$$\ll_\varepsilon e^{t(n-1+\varepsilon)}.$$

Therefore, by the same arguments, for almost every $x \in X$ it holds that

$$\kappa_1(x, e) \ge d/(n-1).$$

Notice that this is a lot larger than d/a.

For n odd, Sardari [39, Corollary 1.7] indeed proved that, for almost every $x \in X$ it holds that

$$\kappa_1(x, e) = d/(n-1).$$

The proof uses deep results from automorphic forms to show that the mean ergodic theorem has actually a better spectral gap than given simply by explicit property (T).

The reader is also referred to [29] for calculation of the Diophantine exponents of the $\mathrm{SO}_{n+1}(\mathbb{Q})$-action on the sphere, which is proved by a different method, not directly related to the spectral decomposition.

3. Preliminaries - local theory

In this section we describe some results about spherical representations and the spherical transform of $\mathrm{SL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{Q}_p)$. We mainly follow [17, Section 3] and [30, Section 3] (see also [19, Section 3]).

3.1. Basic set-up

For any ring R the group $\mathrm{GL}_n(R)$ denotes the group defined by the invertible elements of the $n \times n$ matrix algebra over R, which we call $\mathrm{Mat}_n(R)$. Let $\mathrm{PGL}_n(R)$ be the group $\mathrm{GL}_n(R)/R^\times$. We have a map of algebraic groups $\mathrm{GL}_n \to \mathrm{PGL}_n$.

We let $v = \infty$ or $v = p$ a prime, and let $\mathbb{Q}_v$ be the corresponding local field, i.e., $\mathbb{Q}_\infty = \mathbb{R}$ or the p-adic field $\mathbb{Q}_p$. Let $|\cdot|_v$ be the usual valuation, i.e. $|x|_\infty = |x|$ for $x \in \mathbb{R}$, and $|p^l z|_p = p^{-l}$ for $z \in \mathbb{Z}_p^\times$.

Let $G = G_v := \mathrm{PGL}_n(\mathbb{Q}_v)$. If it is clear from the context we will drop v from the notation. Let P be the subgroup of upper triangular matrices, N be the subgroup of upper triangular unipotent matrices, and A be the subgroup of diagonal matrices. We have $P = NA = AN$. Let K be the standard maximal compact subgroup of G, i.e., $K = K_\infty := \mathrm{PO}_n(\mathbb{R})$ when $v = \infty$ and $K = K_p := \mathrm{PGL}_n(\mathbb{Z}_p)$ when $v = p$. We have the Iwasawa decomposition $G = PK$. When $v = \infty$ denote $\dim(G/K)$ by d whose value is $d = \frac{(n-1)(n+2)}{2}$.

We normalize the Haar measure $m = m_G$ on G as in [30]. In particular, we give K Haar measure 1. If $v = p$ this normalization uniquely defines the Haar measure on G. If $v = \infty$, the Killing form on the Lie algebra $\mathfrak{g}$ of G defines an inner product on the tangent space of G/K, and defines a metric and measure on G/K. This uniquely defines the Haar measure on G.

Let $A^+ \subset A$ be the set consisting of the projection to $\mathrm{PGL}(n)$ of the elements of the following form:

- When $v = \infty$, $A^+ := \{\mathrm{diag}(a_1, \dots, a_n) \mid a_1 \geq \cdots \geq a_n > 0\}$.
- When $v = p$, $A^+ := \{\mathrm{diag}(p^{l_1}, \dots, p^{l_n}) \mid l_1 \leq \cdots \leq l_n\}$.

We have the Cartan decomposition $G = KA^+K$.

We let $\mathfrak{a} := \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid \sum x_i = 0\}$ be the coroot space of $\mathrm{PGL}(n)$. There is a natural map $A \to \mathfrak{a}$, given by $\mathrm{diag}(a_1, \dots, a_n) \mapsto (\log(|a_1|_v), \dots, \log(|a_n|_v))$ and further normalized to have sum 0.

Notice that $\mathfrak{a}$ is the same space for all v. We give $\mathfrak{a}$ an inner product using the case $v = \infty$. We identify $\mathfrak{a} \leq \mathfrak{g}$ as the Lie algebra of the connected component of the identity in A and let the inner product on $\mathfrak{a}$ be the restriction of the Killing form to it. Thus when $v = \infty$ the set $B_b := K\{\exp \alpha \mid \alpha \in \mathfrak{a}, \|\alpha\| \leq b\}K$ is the ball of radius b in G/K around the identity. The inner product allows us to identify $\mathfrak{a}$ with its dual $\mathfrak{a}^*$. We let $\mathfrak{a}^*_{\mathbb{C}} = \mathfrak{a}^* \otimes_{\mathbb{R}} \mathbb{C}$, with the natural extension of the inner product.

For every $\mu = (\mu_1, \dots, \mu_n) \in \mathfrak{a}^*_{\mathbb{C}}$, we associate a character $\chi_\mu \colon P \to \mathbb{C}$ by

$$\chi_\mu(na) = \chi_\mu(a) := \prod_{i=1}^{n} |a_i|_v^{\mu_i},$$

for $a = \mathrm{diag}(a_1, \dots, a_n) \in A$ and $n \in N$. Let

$$\rho := ((n-1)/2, (n-3)/2, \dots, -(n-1)/2) \in \mathfrak{a}^*_{\mathbb{C}}$$

be the half sum of the positive roots. It holds that $\Delta = \chi_{2\rho}$ is the modular character of P.

We denote the *Weyl group* of G by W which is isomorphic to the permutation group of S_n.

3.2. Spherical transform

Given $g \in G$, we let $a(g)$ be its A part according to the Iwasawa decomposition $G = NAK$. We define the *spherical function* $\eta_\mu \colon G \to \mathbb{C}$ corresponding to $\mu \in \mathfrak{a}^*_{\mathbb{C}}$ by

$$\eta_\mu(g) := \int_K \chi_{\mu+\rho}(a(kg))\,\mathrm{d}k.$$

Thus η_μ is bi-K-invariant. Moreover, it can be checked, via a change of variable, that η_μ is invariant under the action of W on μ. We can therefore, without loss of generality, assume that for any η_μ the parameter μ is dominant, i.e., $\Re(\mu_1) \geq \cdots \geq \Re(\mu_n)$. In the p-adic case we may also assume that $0 \leq \Im(\mu_i) < 2\pi/\log(p)$.

The *spherical Hecke algebra* of G is the convolution algebra on $C_c^\infty(K\backslash G/K)$, i.e., the convolution algebra of bi-K-invariant compactly supported smooth functions on G. For $h \in C_c^\infty(K\backslash G/K)$, we let $\tilde{h}\colon \mathfrak{a}_{\mathbb{C}}^* \to \mathbb{C}$ be the *spherical transform* of h, defined by[1]

$$\tilde{h}(\mu) := \int_G h(g)\eta_\mu(g)\,\mathrm{d}g.$$

We have the *spherical Plancherel formula* which states that for $h \in C_c^\infty(K\backslash G/K)$,

$$\int_G |h(g)|^2\,\mathrm{d}g = \int_{i\mathfrak{a}^*} |\tilde{h}(\mu)|^2 d(\mu)\,\mathrm{d}\mu,$$

and *spherical inversion formula* which states that

$$h(g) = \int_{i\mathfrak{a}^*} \tilde{h}(\mu)\overline{\eta_\mu(g)}d(\mu)\,\mathrm{d}\mu. \tag{3.1}$$

Here $d(\mu)$ is a smooth function closely related to the *Harish-Chandra's c-function*. For $v = \infty$, we will need the following estimate (see [30, Equation 3.4])

$$d(\mu) \ll (1 + \|\mu\|)^{d-(n-1)} = (1 + \|\mu\|)^{n(n-1)/2}. \tag{3.2}$$

3.3. Spherical representations

We call an irreducible admissible representation π of G *spherical* if π has a non-zero K-invariant vector. It is well known that such a vector is unique up to multiplication by scalar.

We can construct all admissible irreducible spherical representations of G from the unitarily induced principal series representations. Let $\mu \in \mathfrak{a}_{\mathbb{C}}^*$ and $\mathrm{Ind}_P^G \chi_\mu$ denote the normalized parabolic induction of χ_μ from P to G. It is an admissible representation and has a unique irreducible spherical subquotient. Conversely, for any irreducible admissible spherical representation π we can find a $\mu_\pi \in \mathfrak{a}_{\mathbb{C}}^*$ such that π appears as a unique irreducible subquotient of $\mathrm{Ind}_P^G \chi_{\mu_\pi}$. In this case, we call μ_π to be the *Langlands parameter* of π; see [40].

[1] The normalization here is from [17], which is different from [30].

Let π also be unitary. In this case, let $v \in \pi$ be a unit K-invariant vector. Then it follows from the definition of the spherical function that the corresponding matrix coefficient $\langle \pi(g)v, v\rangle$ is equal to $\eta_{\mu_\pi}(g)$. If $h \in C_c(K\backslash G/K)$ then it holds that

$$\pi(h)v = \int_G h(g)\pi(g)v \,\mathrm{d}g = \tilde{h}(\mu)v.$$

This follows from the fact that $\pi(h)v$ is K-invariant and therefore a scalar times v. This scalar may be calculated by evaluating $\langle \pi(h)v, v\rangle$. If μ is Langlands parameter of some irreducible, spherical and unitary representation, in particular if $\mu \in i\mathfrak{a}^*$, then clearly we have $|\eta_\mu(g)| \le 1$.

Let Q be a standard parabolic subgroup of G attached to the partition $n = n_1 + \cdots + n_r$. The Levi subgroup M_Q of Q is isomorphic to $\mathrm{GL}(n_1) \times \cdots \times \mathrm{GL}(n_r)$ modulo $\mathrm{GL}(1)$. We let $\mathfrak{a}_Q^* \cong \{(\lambda_1, \dots, \lambda_r) \in \mathbb{R}^r \mid \sum_{i=1}^r \lambda_i n_i = 0\}$ which is embedded in $\mathfrak{a}^* \subset \mathbb{R}^n$, as

$$(\lambda_1, \dots, \lambda_r) \mapsto (\lambda_1, \dots, \lambda_1, \dots, \lambda_r, \dots, \lambda_r),$$

where λ_i repeats n_i times. Similarly, we have $\mathfrak{a}_{Q,\mathbb{C}}^* = \mathfrak{a}_Q^* \otimes \mathbb{C}$ embedded in $\mathfrak{a}_{\mathbb{C}}^*$. Given $\lambda = (\lambda_1, \dots, \lambda_r) \in \mathfrak{a}_{P,\mathbb{C}}^*$, it defines a character χ_λ of M_Q by

$$\chi_\lambda(\mathrm{diag}(g_1, \dots, g_r)) = \prod |\det(g_i)|^{\lambda_i}, \quad g_i \in \mathrm{GL}(n_i)$$

Given a spherical representation π of M_Q of we construct the representation $\pi_\lambda = \pi \otimes \chi_\lambda$. We realize π_λ as an representation of Q by tensoring with the trivial representation of the unipotent radical of Q. We denote $\mathrm{Ind}_Q^G \pi_\lambda$ to be the normalized parabolic induction. Here the normalization is via the character χ_{ρ_Q}, where ρ_Q is the half sum of the positive roots attached to Q. We express the Langlands parameters of $\mathrm{Ind}_Q^G \pi_\lambda$ in terms of that of π, and λ.

Lemma 3.1. *The Langlands parameters of* $\mathrm{Ind}_Q^G \pi_\lambda$ *are* $\mu_\pi + \lambda$.

Proof. Assume first that Q corresponds to an ordered partition $n = n_1 + n_2$. Therefore, the Levi part M of Q, modulo its center, is equal to $\mathrm{PGL}_{n_1}(\mathbb{Q}_v) \times \mathrm{PGL}_{n_2}(\mathbb{Q}_v)$. Thus, the spherical representation π of M with trivial central character is a tensor product of the representations of $\mathrm{PGL}_{n_1}(\mathbb{Q}_v)$ with Langlands parameters $\mu = (\mu_1, \dots, \mu_{n_1})$ and $\mathrm{PGL}_{n_2}(\mathbb{Q}_v)$ with Langlands parameters $\mu' = (\mu'_1, \dots, \mu'_{n_2})$.

The representation $\mathrm{Ind}_Q^G \pi_\lambda$ has a unique subquotient that is spherical. By the description of the spherical representations above and transitivity of induction, the Langlands parameter of the resulting representation is $\mu'' = (\mu_1, \dots, \mu_{n_1}, \mu'_1, \dots, \mu'_{n_2}) + \lambda$. The general case follows by an inductive argument. $\square$

3.4. Bounds on spherical functions

In this subsection we will give uniform bounds on the spherical transform of some spherical functions. We will only need the case when $v = p$ is a prime and assume it for the rest of the subsection.

As the spherical function η_μ is bi-K-invariant, the value $\eta_\mu(g)$ depends only on the A^+ part of g from the Cartan decomposition. We will therefore focus on elements $g \in A^+$, which we will assume to be of the form

$$g = \mathrm{diag}(p^{l_1}, \dots, p^{l_n}),$$

with $l_1 \leq \cdots \leq l_n$. As described above, we also assume that μ is dominant, i.e., $\Re(\mu_1) \geq \cdots \geq \Re(\mu_n)$.

We record the following bound from [19, Lemma 3.3].

Lemma 3.2. *Let $g \in A^+$, and $\mu \in \mathfrak{a}_{\mathbb{C}}^*$ be dominant. We have*

$$|\eta_\mu(g)| \ll_\delta \chi_{-\rho(1-\delta)+\Re(\mu)}(g),$$

for every $\delta > 0$.

Proof. Since our notations are different, we repeat the proof of [19, Lemma 3.3]. It holds that

$$|\eta_\mu(g)| \leq \int\limits_K |\chi_{\mu+\rho}(a(kg))| \,\mathrm{d}k = \int\limits_K \chi_{\Re(\mu)}(a(kg))\chi_\rho(a(kg)) \,\mathrm{d}k.$$

Since we assume that $g \in A^+$ and μ is dominant, from [12, Proposition 4.4.4(i)], we have

$$\chi_{\Re(\mu)}(a(kg)) \leq \chi_{\Re(\mu)}(g).$$

Therefore,

$$|\eta_\mu(g)| \leq \chi_{\Re(\mu)}(g) \int\limits_K \chi_\rho(a(kg)) \,\mathrm{d}k = \chi_{\Re(\mu)}(g)\eta_0(g).$$

Finally, η_0 is Harish-Chandra's Ξ-function, which is bounded for $g \in A^+$ by

$$\eta_0(g) = \Xi(g) \ll_\delta \chi_{-\rho(1-\delta)}(g),$$

see, e.g., [45, 4.2.1]. $\square$

We will also need an estimate of the measure of double cosets below. This is elementary, a proof can be found in [45, Lemma 4.1.1].

Lemma 3.3. *For every $g \in A^+$ we have $m(KgK) \asymp \chi_{2\rho}(g)$.*

We end this subsection with a discussion of spherical transform of a certain spherical function, which will be needed for latter purposes.

First, we want a measurement of how far are parameters from $i\mathfrak{a}^*$. In our context, the relevant parameter is as follows. For dominant $\mu \in \mathfrak{a}_{\mathbb{C}}^*$ parameterizing a unitary representation, we define

$$\theta(\mu) := \max_i\{|\Re(\mu_i)|\} = \max\{\Re(\mu_1), -\Re(\mu_n)\}. \tag{3.3}$$

We may assume that $0 \le \theta(\mu) \le (n-1)/2$ since it is true for every spherical unitary representation. Notice that $\theta(\mu) = 0$ if and only if $\mu \in i\mathfrak{a}^*$. Such Langlands parameters are called *tempered*.

Remark 3.4. For completeness, we write the relation between θ and the integrability parameter q from Section 2. By [19, Lemma 3.2], given $2 \le q < \infty$, the following are equivalent:

- For every $\varepsilon > 0$ it holds that $\eta_\mu \in L^{q+\varepsilon}(G)$.
- For every $k = 1, \dots, n-1$,

  $$\sum_{i=1}^k \Re(\mu_i) \le (1 - 2/q) \sum_{i=1}^k \rho_i,$$

  where ρ_i is the i-th coordinate of ρ.[2]

Denote the maximal q which satisfies the above equivalent conditions by $q(\mu)$. Then, in general, we have

$$q(\mu) \ge \tilde{q}(\mu) := \frac{2(n-1)}{(n-1) - 2\theta(\mu)},$$

while for $n = 2$ or $n = 3$ it holds that $\tilde{q}(\mu) = q(\mu)$.

Given an integer $l \ge 0$, consider the finite set of tuples $0 \le l_1 \le \cdots \le l_n$ such that $\sum_{i=1}^n l_i = l$. Each such sequence defines a *different* element $\mathrm{diag}(p^{l_1}, p^{l_2}, \dots, p^{l_n}) = \mathrm{diag}(1, p^{l_2 - l_1}, \dots, p^{l_n - l_1}) \in A^+$. We define

$$M(p^l) := \bigsqcup_{\substack{0 \le l_1 \le \cdots \le l_n \\ \sum_{i=1}^n l_i = l}} K\mathrm{diag}(p^{l_1}, \dots, p^{l_n})K = \bigsqcup_{\substack{0 \le l_1 \le \cdots \le l_n \\ \sum_{i=1}^n l_i = l}} K\mathrm{diag}(1, p^{l_2 - l_1}, \dots, p^{l_n - l_1})K. \tag{3.4}$$

[2] We remark that [19, Lemma 3.2] has a typo which is fixed here.

By applying Lemma 3.3 we obtain:

$$m(K\mathrm{diag}(p^{l_1},\dots,p^{l_n})K) \asymp p^{\sum_{i=1}^n l_i(n+1-2i)}. \tag{3.5}$$

Summing over all the possible choices of $l_1 \le \dots \le l_n$ such that $l_1+\dots+l_n = l$, we deduce that

$$m(M(p^l)) \asymp p^{l(n-1)} \tag{3.6}$$

where most of the mass is concentrated on the double coset with $l_1 = \dots = l_{n-1} = 0$, $l_n = l$.

We also define

$$h_{p^l} := \frac{1}{m(M(p^l))} \mathbb{1}_{M(p^l)} \in C_c^\infty(K\backslash G/K) \tag{3.7}$$

which is the normalized characteristic function of $M(p^l)$. This operator will correspond to the usual *Hecke operator* $T^*(p^l)$ which we will define in Subsection 4.2.

Lemma 3.5. *It holds that for* $\mu \in \mathfrak{a}_{\mathbb{C}}^*$

$$|\tilde{h}_{p^l}(\mu)| \ll_\delta p^{l(\theta(\mu)-(n-1)/2+\delta)},$$

for every $\delta > 0$.

Alternatively, if we write

$$\lambda_\mu(p^l) := \tilde{h}_{p^l}(\mu) m(M(p^l)) p^{-l(n-1)/2},$$

then we have

$$|\lambda_\mu(p^l)| \ll_\delta p^{l(\theta(\mu)+\delta)},$$

for every $\delta > 0$.

Proof. Note that using the W-invariance of $\tilde{h}_{p^l}(\mu)$ it suffices to consider μ to be dominant.

From Equation (3.4) and the definition of h_{p^l} we have

$$\tilde{h}_{p^l}(\mu) = \frac{1}{m(M(p^l))} \sum_{l_1,\dots,l_n} m\left(K\mathrm{diag}(p^{l_1},\dots,p^{l_n})K\right) \eta_\mu\left(\mathrm{diag}(p^{l_1},\dots,p^{l_n})\right),$$

where the sum is over $0 \le l_1 \le \dots \le l_n$ with $l_1+\dots+l_n = l$. We use Lemma 3.2, Equation (3.6), and Equation (3.5) to bound the above display equation by

$$\ll_\delta \frac{1}{p^{(n-1)l}} \sum_{l_1,\dots,l_n} \chi_{\rho(1+\delta)+\Re(\mu)}\left(\operatorname{diag}(p^{l_1},\dots,p^{l_n})\right).$$

Each summand above is bounded by

$$p^{l\left(\frac{n-1}{2}+\delta+\theta(\mu)\right)}.$$

Thus we obtain that $\tilde{h}_{p^l}(\mu)$ is bounded by

$$p^{l\left(-\frac{n-1}{2}+\delta+\theta(\mu)\right)} \sum_{l_1,\dots,l_n} 1.$$

Noting that the last sum is bounded by $l^n \ll_\delta p^{l\delta}$ we conclude. □

Remark 3.6. If μ is Langlands' parameter of a unitary representation then trivially we have $|\tilde{h}_{p^l}(\mu)| \le 1$.

3.5. The Paley–Wiener theorem

In this subsection we will assume that $v = \infty$ and discuss the Paley-Wiener theorem for spherical functions. This is a common tool to localize the spectral side of a trace formula (e.g. see proof of Weyl's law in [35]). See [17, Section 3] for details of the results in this subsection.

We define the *Abel–Satake transform* (also known as the *Harish-Chandra transform*) to be the map $C_c(K\backslash G/K) \to C_c(A)$ defined by

$$f \mapsto \mathcal{S}f : a \mapsto \Delta(a)^{1/2} \int_N f(an)\, \mathrm{d}n.$$

Since $\mathcal{S}f$ is left $K \cap A$-invariant, it is actually a map on A^0, the connected component of the identity of A. We have the exponent map $\exp\colon \mathfrak{a} \to A^0$, given by

$$(\alpha_1,\dots,\alpha_n) \to \operatorname{diag}(\exp(\alpha_1),\dots,\exp(\alpha_n)).$$

This gives an identification of $\mathfrak{a}$ with A^0. So we may as well consider

$$\mathcal{S}f \in C_c(\mathfrak{a})$$

after pre-composing with exp map.

It holds that $\mathcal{S}f$ is W-invariant and Gangolli showed that

$$\mathcal{S}\colon C_c^\infty(K\backslash G/K) \to C_c^\infty(\mathfrak{a})^W$$

is an isomorphism of topological algebras (see [17, 3.21]). Harish-Chandra showed that, if we denote the Fourier–Laplace transform $C_c(\mathfrak{a}) \to C(\mathfrak{a}_{\mathbb{C}}^*)$ by the map

$$h \mapsto \hat{h} : \mu \mapsto \int_{\mathfrak{a}} h(\alpha) e^{\mu\alpha} \, \mathrm{d}\alpha,$$

then it holds that

$$\tilde{h} = \widehat{\mathcal{S}(h)}.$$

Gangolli (see [17, eq. 3.22]) also proved the following important result.

Proposition 3.7. *Let $h \in C_c^\infty(\mathfrak{a})^W$ be such that* $\operatorname{supp}(h) \subset \{\alpha \in \mathfrak{a} \mid \|\alpha\| \le b\}$, *then*

$$\operatorname{supp}(\mathcal{S}^{-1}h) \subset B_b := K\{\exp \alpha \mid \alpha \in \mathfrak{a}, \|\alpha\| \le b\}K.$$

We record the classical Paley–Wiener theorem, which asserts that for $h \in C_c^\infty(\mathfrak{a})^W$ with support in $\{\alpha \in \mathfrak{a} \mid \|\alpha\| \le b\}$ we have

$$|\hat{h}(\mu)| \ll_{N,h} \exp(b\|\Re(\mu)\|)(1 + \|\mu\|)^{-N} \tag{3.8}$$

for every $N \ge 0$.

The following lemma is standard (compare, e.g., [17, Subsection 6.2]). We give a proof because the lemma is crucial in our work.

Lemma 3.8. *Let $\varepsilon \to 0$. There exists a function $k_\varepsilon \in C_c(K\backslash G/K)$ that satisfies the following properties:*

(1) *k_ε is supported on $B_\varepsilon := K\{\exp \alpha \mid \alpha \in \mathfrak{a}, \|a\| \le \varepsilon\}K$.*
(2) *We have $\int_G k_\varepsilon(g) \, \mathrm{d}g = 1$.*
(3) *We have $\|k_\varepsilon\|_\infty \ll \varepsilon^{-d}$ and $\|k_\varepsilon\|_2 \ll \varepsilon^{-d/2}$.*
(4) *The spherical transform satisfies for every $\mu \in a_{\mathbb{C}}^*$ with $\theta(\mu) \le (n-1)/2$,*

$$|\tilde{k}_\varepsilon(\mu)| \ll_N (1 + \varepsilon\|\mu\|)^{-N},$$

for all $N > 0$.

(5) *There is a constant $C > 0$ such that for $\mu \in a_{\mathbb{C}}^*$ satisfying $\|\mu\| \le C\varepsilon^{-1}$ it holds that $|\tilde{k}_\varepsilon(\mu)| \gg 1$.*

Proof. Choose a fixed $h \in C_c^\infty(\mathfrak{a})^W$, having the properties that

- h is non-negative,
- $\hat{h}(0) = 1$,

- $\mathrm{supp}(h) \subset \{\alpha \in \mathfrak{a} \mid \|\alpha\| \leq 1/2\}$.

By the Paley–Wiener theorem, as in Equation (3.8), and the third property above we have

$$|\hat{h}(\mu)| \ll_{h,N} \exp(\|\Re(\mu)\|/2)(1+\|\mu\|)^{-N}.$$

In addition, by continuity of h and the second property above we can find a constant $C > 0$ such that $|\hat{h}(\mu)| \geq 1/2$ for $\|\mu\| \leq C$.

We define $h_\varepsilon(\alpha) := \varepsilon^{-(n-1)} h(\alpha/\varepsilon)$. Then we have $\hat{h}_\varepsilon(\mu) = \hat{h}(\varepsilon\mu)$.

Finally, we define

$$k_\varepsilon := C_\varepsilon^{-1} \mathcal{S}^{-1}(h_\varepsilon), \quad \text{where } C_\varepsilon := \int_G \mathcal{S}^{-1}(h_\varepsilon)(g)\,\mathrm{d}g.$$

Hence, $\tilde{k}_\varepsilon = C_\varepsilon^{-1} \hat{h}_\varepsilon$.

Let us now prove the different properties of k_ε. Property (2) follows from the normalization. By Proposition 3.7, k_ε is supported on B_ε, proving property (1).

To estimate C_ε, we first notice that the spherical function $\eta_{-\rho}$ is simply the constant function 1. So we have

$$C_\varepsilon = \int_G \mathcal{S}^{-1}(h_\varepsilon)(g)\,\mathrm{d}g = \int_G \mathcal{S}^{-1}(h_\varepsilon)(g)\overline{\eta_{-\rho}}(g)\,\mathrm{d}g = \hat{h}_\varepsilon(-\rho) = \hat{h}(-\varepsilon\rho).$$

As $\hat{h}(0) = 1$, we have $C_\varepsilon \asymp 1$ as $\varepsilon \to 0$. This implies that

$$|\tilde{k}_\varepsilon(\mu)| = |C_\varepsilon^{-1}\hat{h}_\varepsilon(\mu)| = |C_\varepsilon^{-1}\hat{h}(\varepsilon\mu)| \ll_N \exp(\varepsilon\|\Re(\mu)\|/2)(1+\varepsilon\|\mu\|)^{-N},$$

so property (4) holds.

Similarly, for $\|\mu\| \leq C\varepsilon^{-1}$,

$$|\tilde{k}_\varepsilon(\mu)| = |C_\varepsilon^{-1}\hat{h}(\varepsilon\mu)| \geq C_\varepsilon^{-1}/2 \gg 1,$$

so property (5) holds.

Finally, using the spherical inversion formula as in Equation (3.1), we have

$$|k_\varepsilon(g)| \leq \int_{i\mathfrak{a}^*} |\tilde{k}_\varepsilon(\mu)||\eta_\mu(g)|d(\mu)\,\mathrm{d}\mu.$$

Using the fact that $|\eta_\mu(g)| \leq 1$ for every $g \in G$ and $\mu \in i\mathfrak{a}^*$ unitary, property (4), and Equation (3.2) we obtain that the above is bounded by

$$\ll_N \int_{i\mathfrak{a}^*} (1+\varepsilon\|\mu\|)^{-N}(1+\|\mu\|)^{d-n+1}\,\mathrm{d}\mu.$$

The above integral can be estimated by making N large enough as

$$\ll \int_0^\infty (1+\varepsilon x)^{-N} x^{d-1}\,\mathrm{d}x \ll \varepsilon^{-d}.$$

This shows that $\|k_\varepsilon\|_\infty \ll \varepsilon^{-d}$. The bound on $\|k_\varepsilon\|_2$ follows from the bound on $\|k_\varepsilon\|_\infty$, property (1), and the fact that $\mathrm{vol}(B_\varepsilon) \asymp \varepsilon^d$. $\square$

4. Preliminaries - global theory

4.1. Adelic formulation

In this section, we describe global preliminaries in adelic language that are needed for the proof. Temporarily in this section, p will denote a generic finite prime of $\mathbb{Q}$.

Let $\mathbb{A} := \mathbb{R} \times \prod_p' \mathbb{Q}_p$ be the adele ring of $\mathbb{Q}$, where $\prod'$ means that if $x = (x_\infty, \dots, x_p, \dots) \in \mathbb{A}$ then $x_p \in \mathbb{Z}_p$ for almost all p.

We recall the notations $K_\infty = \mathrm{PO}_n(\mathbb{R})$ and $K_p = \mathrm{PGL}_n(\mathbb{Z}_p)$. We denote $K_\mathbb{A} := K_\infty \times \prod_p K_p$ which is a hyper-special maximal compact subgroup of $\mathrm{PGL}_n(\mathbb{A})$. We also denote

$$\mathbb{X}_\mathbb{A} := \mathrm{PGL}_n(\mathbb{Q})\backslash \mathrm{PGL}_n(\mathbb{A})/K_\mathbb{A}.$$

Recall that $\mathbb{X} := \mathrm{SL}_n(\mathbb{Z})\backslash \mathbb{H}^n \cong \mathrm{SL}_n(\mathbb{Z})\backslash \mathrm{SL}_n(\mathbb{R})/\mathrm{SO}_n(\mathbb{R})$.

Let $\varphi\colon \mathrm{SL}_n(\mathbb{R}) \to \mathrm{PGL}_n(\mathbb{R})$ be the natural quotient map. This map defines an action of $\mathrm{SL}_n(\mathbb{R})$ on $\mathrm{PGL}_n(\mathbb{R})/K_\infty$, which is easily seen to be transitive, and since $\varphi^{-1}(\mathrm{PO}_n(\mathbb{R})) = \mathrm{SO}_n(\mathbb{R})$ we can identify $\mathbb{H}^n = \mathrm{SL}_n(\mathbb{R})/\mathrm{SO}_n(\mathbb{R}) \cong \mathrm{PGL}_n(\mathbb{R})/K_\infty$. By considering the left action of $\mathrm{SL}_n(\mathbb{Z})$ on the two spaces we can identify

$$\mathbb{X} \cong \mathrm{PSL}_n(\mathbb{Z})\backslash \mathrm{PGL}_n(\mathbb{R})/K_\infty.$$

Similarly, by considering the transitive right action of $\mathrm{SL}_n(\mathbb{R})$ on $\mathrm{PSL}_n(\mathbb{Z})\backslash \mathrm{PGL}_n(\mathbb{R})$ we may identify $\mathrm{SL}_n(\mathbb{Z})\backslash \mathrm{SL}_n(\mathbb{R}) \cong \mathrm{PGL}_n(\mathbb{Z})\backslash \mathrm{PGL}_n(\mathbb{R})$ and

$$\mathbb{X} \cong \mathrm{PGL}_n(\mathbb{Z})\backslash \mathrm{PGL}_n(\mathbb{R})/\mathrm{PSO}_n(\mathbb{R}).$$

It is simpler for us to work with the space

$$\mathbb{X}_0 := \mathrm{PGL}_n(\mathbb{Z})\backslash \mathrm{PGL}_n(\mathbb{R})/K_\infty,$$

which is a quotient space of $\mathbb{X}$ by the group $\mathrm{PGL}_n(\mathbb{Z})/\mathrm{PSL}_n(\mathbb{Z})$ of size 2.

The following Lemma 4.1 allows us to identify $\mathbb{X}_0$ with $\mathbb{X}_\mathbb{A}$. The lemma is well known, but essential for this work, so we provide a proof for completeness.

Lemma 4.1. *We have*

$$\mathbb{X}_0 \cong \mathbb{X}_{\mathbb{A}}$$

as topological spaces.

Proof. Let $K_f = \prod_p K_p$ embedded naturally in $\mathrm{GL}_n(\mathbb{A})$. It is enough to prove that

$$\mathrm{PGL}_n(\mathbb{Z})\backslash\mathrm{PGL}_n(\mathbb{R}) \cong \mathrm{PGL}_n(\mathbb{Q})\backslash\mathrm{PGL}_n(\mathbb{A})/K_f.$$

By the fact that GL_n over $\mathbb{Q}$ has class number 1 (see [38, Proposition 8.1]) we have

$$\mathrm{GL}_n(\mathbb{A}) = \mathrm{GL}_n(\mathbb{Q})\mathrm{GL}_n(\mathbb{R})\prod_p \mathrm{GL}_n(\mathbb{Z}_p).$$

Alternatively, this follows from the fact that

$$\mathrm{GL}_n(\mathbb{R}) \times \mathrm{GL}_n(\mathbb{Q}_p) = \mathrm{GL}_n(\mathbb{Z}[1/p])\left(\mathrm{GL}_n(\mathbb{R}) \times \mathrm{GL}_n(\mathbb{Z}_p)\right), \tag{4.1}$$

where $\mathrm{GL}_n(\mathbb{Z}[1/p])$ is embedded diagonally in the left hand side. Using the case $n = 1$ which states

$$\mathbb{A}^\times = \mathbb{Q}^\times\mathbb{R}^\times\prod_p \mathbb{Z}_p^\times,$$

we get

$$\mathrm{PGL}_n(\mathbb{A}) = \mathrm{PGL}_n(\mathbb{Q})(\mathrm{PGL}_n(\mathbb{R}) \times K_f).$$

We deduce that the right action of $\mathrm{PGL}_n(\mathbb{R})$ on $\mathrm{PGL}_n(\mathbb{Q})\backslash\mathrm{PGL}_n(\mathbb{A})/K_f$ is onto. The stabilizer of this action is $\mathrm{PGL}_n(\mathbb{Z})$, so we get the desired homeomorphism. □

Remark 4.2. Alternatively, it holds that

$$\mathbb{X}_0 \cong \mathrm{GL}_n(\mathbb{Z})\mathbb{R}^\times\backslash\mathrm{GL}_n(\mathbb{R})/\mathrm{O}_n(\mathbb{R}) \tag{4.2}$$

and

$$\mathbb{X}_{\mathbb{A}} \cong \mathrm{GL}_n(\mathbb{Q})\mathbb{R}^\times\backslash\mathrm{GL}_n(\mathbb{A})/\mathrm{O}_n(\mathbb{R}) \times \prod_p \mathrm{GL}_n(\mathbb{Z}_p). \tag{4.3}$$

Remark 4.3. A similar proof using the action on $\mathrm{PSL}_n(\mathbb{R}) \subset \mathrm{PGL}_n(\mathbb{R})$ identifies $\mathbb{X}$ with the adelic space $\mathrm{PGL}_n(\mathbb{Q})\backslash\mathrm{PGL}_n(\mathbb{A})/(\mathrm{PSO}_n(\mathbb{R}) \times K_f)$ (i.e., $K_\infty = \mathrm{PO}_n(\mathbb{R})$ is replaced with $\mathrm{PSO}_n(\mathbb{R})$).

4.2. Hecke operators

We want to consider Hecke operators on $L^2(\mathbb{X}_{\mathbb{A}})$ (equivalently, on $L^2(\mathbb{X}_0)$) from a representation theoretic point of view. This is standard (e.g., [15]), but since we work on $\mathrm{PGL}(n)$, which is not simply connected, some modifications are needed.

Let $l \geq 0$ and p be any finite prime. Consider the infinite set

$$R(p^l) = \{x \in \mathrm{Mat}_n(\mathbb{Z}) \mid \det(x) = p^l\}.$$

Recall that for every ring R we have a natural projection $\mathrm{GL}_n(R) \to \mathrm{PGL}_n(R)$, and we will write $\tilde{R}(p^l)$ for the projection of $R(p^l)$ to $\mathrm{PGL}_n(\mathbb{Z}[1/p]) \subset \mathrm{PGL}_n(\mathbb{Q})$. Notice that $\tilde{R}(p^l)$ is both left and right $\tilde{R}(1) = \mathrm{PSL}_n(\mathbb{Z})$-invariant. Let $A(p^l)$ be a finite set of representatives for $\tilde{R}(p^l)/\tilde{R}(1)$. It is possible to explicitly describe the set $A(p^l)$ (*e.g.*, see [23, Lemma 9.3.2] for *right* cosets), but we will not need this explicit description.

Lemma 4.4. *It holds that $K_p\tilde{R}(p^l)K_p = M(p^l)$, where $M(p^l) \subset \mathrm{PGL}_n(\mathbb{Q}_p)$ is (as in Subsection 3.4) the disjoint union of the double cosets of the form*

$$K_p \mathrm{diag}(p^{l_1}, \dots, p^{l_n}) K_p,$$

with $0 \leq l_1 \leq \cdots \leq l_n$ and $\sum l_i = l$.

Moreover, the elements of $A(p^l)$ can be taken as representatives of the left K_p-cosets of $M(p^l)$.

Proof. For $0 \leq l_1 \leq \cdots \leq l_n$ and $\sum l_i = l$ it holds that $\mathrm{diag}(p^{l_1}, \dots, p^{l_n}) \in R(p^l)$. By projecting to $\mathrm{PGL}_n(\mathbb{Q}_p)$ it follows that $M(p^l) \subset K_p\tilde{R}(p^l)K_p$.

We check the other direction. For each element $\gamma \in R(p^l)$ (i.e., $\gamma \in \mathrm{Mat}_n(\mathbb{Z})$ with $\det(\gamma) = p^l$) we can find, using the Cartan decomposition in $\mathrm{GL}_n(\mathbb{Q}_p)$, two elements $k_1, k_2 \in \mathrm{GL}_n(\mathbb{Z}_p)$ and $a = \mathrm{diag}(p^{l_1}, \dots, p^{l_n})$ with $l_1 \leq \cdots \leq l_n$ such that $k_1\gamma k_2 = a$. Therefore, $a \in \mathrm{Mat}_n(\mathbb{Z}_p)$, hence $l_1 \geq 0$. Moreover, by comparing the determinants we have $\sum l_i = l$. By mapping to $\mathrm{PGL}_n(\mathbb{Q}_p)$ we deduce that $\tilde{R}(p^l) \subset M(p^l)$ and $K_p\tilde{R}(p^l)K_p \subset M(p^l)$.

Now, the natural embedding $\tilde{R}(p^l) \to M(p^l)$ extends to a natural embedding

$$\tilde{R}(p^l)/\tilde{R}(1) \to M(p^l)/K_p.$$

We need to show that this map is surjective. By Equation (4.1), the action of $\mathrm{PGL}_n(\mathbb{Z}[1/p])$ on $\mathrm{PGL}_n(\mathbb{Q}_p)/K_p$ is transitive. Therefore, each left coset $M(p^l)/K_p$ has a representative $\gamma \in \mathrm{PGL}_n(\mathbb{Z}[1/p])$. Lifting to $\mathrm{GL}_n(\mathbb{Q}_p)$ and applying the Cartan decomposition, we can write $\gamma = k\mathrm{diag}(p^{l_1}, \dots, p^{l_n})k'$, for some $k, k' \in \mathrm{GL}_n(\mathbb{Z}_p)$ and $0 \leq l_1 \leq \cdots \leq l_n$ with $\sum l_i = l$. Therefore $\gamma \in \mathrm{GL}_n(\mathbb{Z}[1/p]) \cap \mathrm{GL}_n(\mathbb{Z}_p) \subset \mathrm{Mat}_n(\mathbb{Z})$. Finally, $\det(\gamma) = p^l$ implies that $\gamma \in R(p^l)$, as needed. $\square$

Definition 4.5. Let $l \geq 0$. The *Hecke operator* $T^*(p^l)$ acting on $L^2(\mathbb{X}_0)$ is defined as

$$\left(T^*(p^l)\varphi\right)(x) := \frac{1}{|A(p^l)|} \sum_{\gamma \in A(p^l)} \varphi(\gamma^{-1}x).$$

By the facts that $A(p^l)$ are representatives for left $\mathrm{PGL}_n(\mathbb{Z})$-cosets of $\mathrm{PGL}_n(\mathbb{Z})\tilde{R}(p^l)$ which is right and left $\mathrm{PGL}_n(\mathbb{Z})$-invariant, and φ is left $\mathrm{PGL}_n(\mathbb{Z})$-invariant, the operator $T^*(p^l)$ is well-defined, and does not depend on the choice of $A(p^l)$.

Remark 4.6. In analytic number theory, one usually defines the Hecke operator $\tilde{T}(p^l)$ as

$$\left(\tilde{T}(p^l)\varphi\right)(x) = \frac{1}{p^{l(n-1)/2}} \sum_{\gamma \in \tilde{R}(1)\backslash \tilde{R}(p^l)} \varphi(\gamma x),$$

see *e.g.*, [23, §9.3.5]. If we define $T(p^l) := \frac{p^{l(n-1)/2}}{|\tilde{R}(1)\backslash \tilde{R}(p^l)|}\tilde{T}(p^l)$ then $T^*(p^l)$ is indeed the adjoint of $T(p^l)$.

Remark 4.7. Alternatively, using Equation (4.2), we could have defined a Hecke operator $T^*_{\mathrm{GL}}(p^l)$ on the space $L^2(\mathrm{GL}_n(\mathbb{Z})\mathbb{R}^\times\backslash\mathrm{GL}_n(\mathbb{R})/O_n(\mathbb{R}))$ by

$$\left(T^*_{\mathrm{GL}}(p^l)\varphi\right)(x) = \frac{1}{|R(p^l)/R(1)|} \sum_{\gamma \in R(p^l)/R(1)} \varphi(\gamma^{-1}x)$$

This definition agrees with the other definition under the equivalence Equation (4.2).

Using Lemma 4.1 we may lift $\varphi \in L^2(\mathbb{X}_0)$ to a function $\varphi_{\mathbb{A}} \in L^2(\mathbb{X}_{\mathbb{A}})$, by

$$\varphi_{\mathbb{A}}(g_\infty, (e)_p) := \varphi(g_\infty)$$

where $(e)_p := (1, 1, \dots) \in \prod_p \mathrm{GL}_n(\mathbb{Q}_p)$. We extend it to be left $\mathrm{PGL}_n(\mathbb{Q})$-invariant and right $K_{\mathbb{A}}$-invariant. Using such identification we can define certain averaging operators on $L^2(\mathbb{X}_{\mathbb{A}})$ using local functions $h_v \in C_c(K_v\backslash\mathrm{PGL}_n(\mathbb{Q}_v)/K_v)$, such as

$$(R(h_v)\varphi_{\mathbb{A}})(\dots, x_v, \dots) := \int\limits_{\mathrm{PGL}_n(\mathbb{Q}_v)} h_v(y)\varphi_{\mathbb{A}}(\dots, x_v y, \dots)\, \mathrm{d}y.$$

Recall Equation (3.7)

$$h_{p^l} := \frac{1}{m(M(p^l))}\mathbb{1}_{M(p^l)} \in C_c(K_p\backslash\mathrm{PGL}_n(\mathbb{Q}_p)/K_p).$$

We show that the operators $T^*(p^l)$ and $R(h_{p^l})$ are, in fact, classical and adelic versions, respectively, of one another.

Lemma 4.8. *It holds that* $(T^*(p^l)\varphi)_{\mathbb{A}} = R(h_{p^l})\varphi_{\mathbb{A}}$.

Proof. From the definitions of the Hecke operator and the adelic lift above, we have

$$|A(p^l)|(T^*(p^l)\varphi)_{\mathbb{A}}(g_\infty, (e)_q) = \sum_{\gamma \in A(p^l)} \varphi_{\mathbb{A}}(\gamma^{-1} g_\infty, (e)_q) = \sum_{\gamma \in A(p^l)} \varphi_{\mathbb{A}}(g_\infty, (\gamma)_p, (e)_{q \neq p}),$$

where the second equality follows from the left $\mathrm{PGL}_n(\mathbb{Q})$-invariance and the right $K_{\mathbb{A}}$-invariance of $\varphi_{\mathbb{A}}$. Once again using the right $K_{\mathbb{A}}$-invariance of $\varphi_{\mathbb{A}}$ we can write the above as

$$\int_{\mathrm{PGL}_n(\mathbb{Q}_p)} \varphi_{\mathbb{A}}(g_\infty, (y)_p, (e)_{q \neq p}) \sum_{\gamma \in A(p^l)} \mathbb{1}_{\gamma K_p}(y)\, \mathrm{d}y.$$

According to Lemma 4.4,

$$\sum_{\gamma \in A(p^l)} \mathbb{1}_{\gamma K_p} = \mathbb{1}_{M(p^l)}.$$

In addition, it holds that $|A(p^l)| = m(M(p^l))$. Therefore, for $g \in \mathrm{PGL}_n(\mathbb{A})$ of the form $g = (g_\infty, (e)_q)$, it holds that

$$\begin{aligned}(T^*(p^l)\varphi)_{\mathbb{A}}(g) &= \frac{1}{m(M(p^l))} \int_{y \in \mathrm{PGL}_n(\mathbb{Q}_p)} \varphi_{\mathbb{A}}(gy) \mathbb{1}_{M(p^l)}(y)\, \mathrm{d}y \\ &= (R(h_{p^l})\varphi_{\mathbb{A}})(g),\end{aligned}$$

as needed. □

Remark 4.9. One can similarly define Hecke operators on $\mathbb{X}$, by identifying

$$\mathbb{X} = \mathrm{PGL}_n(\mathbb{Z}) \backslash \mathrm{PGL}_n(\mathbb{R}) / \mathrm{PSO}_n(\mathbb{R}).$$

4.3. Discrete spectrum and weak Weyl law

We will need to use Langlands' spectral decomposition of $L^2(\mathbb{X}_{\mathbb{A}})$. In this subsection we describe the discrete part of the spectrum.

The discrete spectrum $L^2_{\mathrm{disc}}(\mathrm{PGL}_n(\mathbb{Q}) \backslash \mathrm{PGL}_n(\mathbb{A}))$ consists of the irreducible representations π of $\mathrm{PGL}_n(\mathbb{A})$ occurring discretely in $L^2(\mathrm{PGL}_n(\mathbb{Q}) \backslash \mathrm{PGL}_n(\mathbb{A}))$. By abstract representation theory, we may write

$$L^2_{\mathrm{disc}}(\mathrm{PGL}_n(\mathbb{Q}) \backslash \mathrm{PGL}_n(\mathbb{A})) \cong \oplus_\pi V_\pi,$$

where V_π is the π-isotypic component. By the multiplicity one theorem ([44] and [33]), each representation π of $\mathrm{PGL}_n(\mathbb{A})$ appears in the decomposition with multiplicity at most 1, so V_π spans a representation isomorphic to π. Moreover, π is isomorphic to a tensor product of representations π_v of $\mathrm{PGL}_n(\mathbb{Q}_v)$ for $v \leq \infty$ (see [18]).

We restrict our attention to spherical representations π, that is, those having a non-zero $K_{\mathbb{A}}$-invariant vector. Since π is equivalent to a tensor product of local representations π_v, and each local representation has a one dimensional K_v-invariant subspace, the $K_{\mathbb{A}}$-invariant subspace $\pi^{K_{\mathbb{A}}}$ of π has dimension 1. We choose once and for all a unit

$$\varphi_\pi \in V_\pi^{K_{\mathbb{A}}}$$

for each π with $V_\pi \neq \{0\}$.

We denote the set of all such φ_π by $\mathcal{B}_n$. Then formally we have

$$L^2_{\mathrm{disc}}(\mathbb{X}_{\mathbb{A}}) \cong \bigoplus_{\varphi \in \mathcal{B}_n} \mathbb{C}\varphi.$$

The discrete spectrum $\mathcal{B}_n$ decomposes naturally into two parts, the *cuspidal part* $\mathcal{B}_{n,\mathrm{cusp}}$ and the *residual part* $\mathcal{B}_{n,\mathrm{res}}$.

Recalling from Subsection 3.3, for every place v we may attach the Langlands parameter $\mu_{\varphi,v} = \mu_{\pi_v}$ to φ. When $v = \infty$ we denote

$$\nu_\varphi := \|\mu_{\varphi,\infty}\|. \tag{4.4}$$

By [17, Equation 3.17] φ is an eigenfunction of the Laplace–Beltrami operator and its eigenvalue is

$$\|\rho\|^2 - \|\Re(\mu_{\varphi,\infty})\|^2 + \|\Im(\mu_{\varphi,\infty})\|^2.$$

In particular, for $\nu_\varphi \geq 1$ the Laplace eigenvalue of φ is $\asymp \nu_\varphi^2$.

We will also denote

$$\theta_{\varphi,p} := \theta(\mu_{\varphi,p}),$$

according to Equation (3.3).

We define

$$\mathcal{F}_T := \{\varphi \in \mathcal{B}_{n,\mathrm{cusp}} \mid \nu_\varphi \leq T\}. \tag{4.5}$$

We record the statement of *Weak Weyl law* due to Donnelly, which gives an upper bound of the cardinality $\mathcal{F}_T$ as $T \to \infty$.

Proposition 4.10 *([16]). We have $|\mathcal{F}_T| \ll T^d$ as $T \to \infty$.*

As a matter of fact, using the corresponding lower bounds by Müller in [35] and Lindenstrauss–Venkatesh in [30], we know that as $T \to \infty$

$$|\mathcal{F}_T| = CT^d(1 + o(1)), \tag{4.6}$$

for some explicit constant C, but we will not need this stronger result.

4.4. The generalized Ramanujan conjecture and Sarnak's density hypothesis

Let $\varphi \in \mathcal{B}_n$ be a spherical discrete series. For every $h_v \in C_c(K_v\backslash \mathrm{PGL}_n(\mathbb{Q}_v)/K_v)$ one has the operator $R(h_v)$, as defined in Subsection 4.1, and it holds that

$$R(h_v)\varphi = \tilde{h}_v(\mu_{\varphi,v})\varphi,$$

where $\tilde{h}_v$ denotes the spherical transform of h_v. When $v = p$, combining Lemma 4.8 and Lemma 3.5 we obtain that for every $x \in \mathbb{X}$,

$$T^*(p^l)\varphi(x) = \lambda_\varphi(p^l)p^{-l(n-1)/2}\varphi(x)$$

such that for all $\delta > 0$

$$|\lambda_\varphi(p^l)| \ll_\delta p^{l(\theta_{\varphi,p}+\delta)}.$$

Let $\varphi \in \mathcal{B}_{n,\mathrm{cusp}}$ be a spherical cusp form. The *Generalized Ramanujan Conjecture* (GRC) predicts that the Langlands parameter of φ at every place v is tempered, which is equivalent in our notations to

$$\theta_{\varphi,v} = 0, \quad v \leq \infty$$

The GRC at the place $v = p$ implies essentially the sharpest bounds on the Hecke eigenvalues, in the following form. For every p prime, $l \geq 0$, and $\delta > 0$,

$$|\lambda_\varphi(p^l)| \ll_\delta p^{l\delta},$$

as $p^l \to \infty$.

The GRC is out of reach of current technology, even for $n = 2$. However, we have various bounds towards it; see [40] for a detailed discussion.

The bounds of Hecke eigenvalues can be understood in terms of the bounds of $\theta_{\varphi,p}$. For $n = 2$, the best bounds are due to Kim–Sarnak [28], for $n = 3$ and $n = 4$, the same are due to Blomer–Brumley [6, Theorem 1], and for $n \geq 5$, they are due to Luo–Rudnick–Sarnak [31]. For $\mathrm{GL}(n)$, these bounds are given by

$$|\theta_{\varphi,p}| \leq \theta_n \tag{4.7}$$

where

$$\theta_2 = \frac{7}{64}, \theta_3 = \frac{5}{14}, \theta_4 = \frac{9}{22},$$

and

$$\theta_n = \frac{1}{2} - \frac{1}{n^2+1}, \quad n \geq 5. \tag{4.8}$$

Our problem requires estimates of the Hecke eigenvalues which is stronger than Equation (4.8). On the other hand, we do not require a strong bound of individual Hecke eigenvalue, but only the GRC *on average*. In general, one expects *Sarnak's Density Hypothesis* to hold; see [43,42,25,4,26,2] for various aspects of this hypothesis.

The hypothesis asserts that for every $\delta > 0$, for a nice enough finite family $\mathcal{F}$ of cusp forms, one has

$$\sum_{\varphi\in\mathcal{F}} |\lambda_\varphi(p^l)|^2 \ll_\delta \left(p^l|\mathcal{F}|\right)^\delta \left(|\mathcal{F}| + p^{l(n-1)}\right) \tag{4.9}$$

uniformly in p, l and as $|\mathcal{F}| \to \infty$.

Informally, the above says that *larger Hecke eigenvalues occur with smaller density.* Note that the above follows from GRC. On the other hand, the occurrence of the summand $p^{l(n-1)}$ represents as if the trivial eigenfunction appeared in $\mathcal{F}$. The hypothesis above is an interpolation between the two cases. As a matter of fact, we expect that it will hold for natural families of *discrete forms*, not only for *cusp forms.*

In this paper we will work on a specific kind of family $\mathcal{F}$, namely, $\mathcal{F}_T$ as defined in Equation (4.5) whose cardinality is $\asymp T^d$ via Equation (4.6). We will need a hypothesis in the following form.

Conjecture 1 *(Density hypothesis, Hecke eigenvalue version). Let p be a fixed prime. For every $l \geq 0$ and $T \geq 1$ one has*

$$\sum_{\varphi\in\mathcal{F}_T} \left|\lambda_\varphi(p^l)\right|^2 \ll_{p,\delta} \left(Tp^l\right)^\delta \left(T^d + p^{l(n-1)}\right),$$

for every $\delta > 0$

Remark 4.11. Conjecture 1 should be compared to the orthogonality conjecture which asserts that

$$\sum_{\varphi\in\mathcal{F}_T} \lambda_\varphi(p_1^{l_1})\overline{\lambda_\varphi(p_2^{l_2})} \sim \delta_{p_1^{l_1}=p_2^{l_2}} T^d,$$

as $T \to \infty$; see *e.g.*, [26, Theorem 1, Theorem 7], [4, Theorem 2] for more details.

Motivated by the original density hypothesis of Sarnak (see [43]) one can propose an analogous density hypothesis in terms of the Langlands parameters in higher rank; see [4,26]. In fact, Conjecture 1 is nothing but a reformulation of *Sarnak's density hypothesis for higher rank* which we describe below.

Conjecture 2 *(Density Conjecture, Langlands parameter version). For every $0 \leq \theta_0 \leq (n-1)/2$ and $T \geq 1$ one has*

$$|\{\varphi \in \mathcal{F}_T \mid \theta_{\varphi,p} \geq \theta_0\}| \ll_{\delta,p} T^{d\left(1-\frac{2\theta_0}{n-1}\right)+\delta},$$

for every $\delta > 0$

Following [4] here we prove the equivalence of Conjecture 1 and Conjecture 2.

Proposition 4.12. *Conjecture 1 and Conjecture 2 are equivalent.*

Proof. Assume that Conjecture 1 holds. Given T sufficiently large, summing over l such that $p^{l(n-1)} \leq T^d$ we get

$$\sum_{\varphi \in \mathcal{F}_T} \sum_{l: p^{l(n-1)} \leq T^d} \left|\lambda_\varphi(p^l)\right|^2 \ll_{p,\delta} T^{d+\delta}.$$

By [4, Lemma 4], for $k \geq n+1$ we have

$$\sum_{l=0}^{k} \left|\lambda_\varphi(p^l)\right|^2 \gg_p p^{2k\theta_{\varphi,p}}.$$

Therefore, for $T \geq p^{(n-1)(n+1)/d}$,

$$\sum_{l: p^{l(n-1)} \leq T^d} \left|\lambda_\varphi(p^l)\right|^2 \gg_p T^{d\frac{2\theta_{\varphi,p}}{n-1}},$$

and

$$\sum_{\varphi \in \mathcal{F}_T} T^{d\frac{2\theta_{\varphi,p}}{n-1}} \ll_{p,\delta} T^{d+\delta},$$

and this implies Conjecture 2.

For the other direction, assume Conjecture 2. By Lemma 3.5,

$$\sum_{\varphi \in \mathcal{F}_T} \left|\lambda_\varphi(p^l)\right|^2 \ll_\delta \sum_{\varphi \in \mathcal{F}_T} p^{l(2\theta_{\varphi,p}+\delta)}.$$

Using partial summation, this is bounded by the main term

$$\ll_{p,\delta} \int_0^{(n-1)/2} |\{\varphi \in \mathcal{F}_T \mid \theta_{\varphi,p} \geq \theta_0\}| p^{l(2\theta_0+\delta)} \, \mathrm{d}\theta_0$$

$$\ll_{p,\delta} (p^l T)^\delta \int_0^{(n-1)/2} T^{d\left(1-\frac{2\theta_0}{n-1}\right)} p^{2l\theta_0} \, \mathrm{d}\theta_0$$

plus the secondary terms

$$|\mathcal{F}_T| p^{l\delta} + |\{\varphi \in \mathcal{F}_T \mid \theta_{\varphi,p} \geq (n-1)/2\}| p^{l(n-1+\delta)}.$$

For $0 \leq \theta_0 \leq \frac{n-1}{2}$ we have

$$T^{d\left(1-\frac{2\theta_0}{n-1}\right)} p^{2l\theta_0} \ll T^d + p^{l(n-1)}.$$

Therefore, the main term is bounded by $(p^l T)^\delta (T^d + p^{l(n-1)})$. Similarly, using Conjecture 2, the secondary terms are also bounded by the same value. We therefore get Conjecture 1. $\square$

Conjecture 1 is actually a *convexity estimate*, just as Conjecture 2. As a matter of fact, one can replace in it $\mathcal{B}_{n,\mathrm{cusp}}$ by $\mathcal{B}_n$, as we will essentially show in Subsection 7.1 that the residual spectrum also satisfies this bound. However, for the cuspidal spectrum itself one should expect better than what Conjecture 1 asserts, *i.e.* the *subconvex* estimates, which are indeed available for $n = 2$ [7, Lemma 1] and $n = 3$ [13, Theorem 3.3]. For $n \geq 3$ Blomer [4] has proved a subconvex estimate for Hecke congruence subgroups in the level aspect. Recently, Blomer and Man [9] (improving upon the work of Assing and Blomer [2]) proved subconvex estimates for principal congruence subgroups, again in the level aspect.

We describe the results for $n = 2, 3$ here which we will use latter. Although, we only need the convexity estimates for our proofs, we record the strongest known estimates. Let m be of the form p^l for some fixed prime[3] p.

Proposition 4.13. *Let $n = 2$. We have*

$$\sum_{\varphi \in \mathcal{F}_T} |\lambda_\varphi(m)|^2 \ll_\delta (Tm)^\delta \left(T^d + m^{1/2}\right),$$

for every $\delta > 0$.

The result follows from [7, Lemma 1] and the discussion following it for the bound on $L(1, \operatorname{sym}^2 u_j)$. It is stronger than the density estimate given in Conjecture 1. In fact, it

[3] The results below are also true for the extension of λ_φ to a multiplicative function on $\mathbb{N}$.

leads to the following subconvex bound using the same methods as in Proposition 4.12: for every $0 \leq \theta_0 \leq 1/2$ and $T \geq 1$ and we have

$$|\{\varphi \in \mathcal{F}_T \mid \theta_{\varphi,p} \geq \theta_0\}| \ll_{\delta,p} T^{2(1-4\theta_0)+\delta}, \tag{4.10}$$

for every $\delta > 0$. Equation (4.10) is slightly stronger than [7, Proposition 1].

Proposition 4.14. *Let $n = 3$. We have*

$$\sum_{\varphi \in \mathcal{F}_T} |\lambda_\varphi(m)|^2 \ll_\delta (Tm)^\delta \left(T^5 + m^{5/4}\right),$$

for every $\delta > 0$.

The result can be obtained from [13, Theorem 3.3] and upper bound of the adjoint L-value as in [7, Equation (22)].[4] This result can be used to prove a considerably stronger statement than Conjecture 2.

Proposition 4.15. *Let $n = 3$ and p fixed. For every $0 \leq \theta_0 \leq 1$ and $T \geq 1$ we have*

$$|\{\varphi \in \mathcal{F}_T \mid \theta_{\varphi,p} \geq \theta_0\}| \ll_{\delta,p} T^{5(1-8\theta_0/5)+\delta},$$

for every $\delta > 0$.

Proof. We use the same arguments as in Proposition 4.12. For T large enough we sum the estimate in Proposition 4.14 with $m = p^l$ and l such that $p^l \leq T^4$. Then using [4, Lemma 4] we get

$$\sum_{\varphi \in \mathcal{F}_T} T^{8\theta_{\varphi,p}} \ll_{p,\varepsilon} T^{5+\varepsilon},$$

and this implies the proposition. □

This substantially improves both [7, Theorem 1] and [7, Theorem 2].

4.5. Eisenstein series

Let P be a standard parabolic in $G := \mathrm{PGL}(n)$ attached to an ordered partition $n = n_1 + \cdots + n_r$. Let M be the corresponding Levi subgroup and N be the corresponding unipotent radical; see Subsection 3.3, where we denote a general parabolic by Q.

Let T_M be the connected component of the identity of the $\mathbb{R}$-points in a maximal torus in the center of M. Let $M(\mathbb{A})^1$ be the kernel of all algebraic characters of M (see

[4] Conjecture 1 for $n = 3$ can actually be deduced from [3].

[1, Chapter 3]). It holds that $M(\mathbb{A}) = M^1(\mathbb{A}) \times T_M$. Following [1, Chapter 7] we give a brief sketch of the construction of the Eisenstein series on $G(\mathbb{A})$, constructed inductively from the elements of $L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash M(\mathbb{A})^1)$.

We follow the construction of induced representation as in Subsection 3.3. Given an representation π of $M(\mathbb{A})$ occurring in $L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash M(\mathbb{A})^1)$ and $\lambda \in \mathfrak{a}^*_{P,\mathbb{C}}$ we consider the representation $\pi_\lambda := \pi \otimes \chi_\lambda$ of $M(\mathbb{A})$ and extend it to $P(\mathbb{A})$ via the trivial representation on $N(\mathbb{A})$. Then we consider the normalized parabolic induction

$$\mathcal{I}_{P,\pi}(\lambda) := \mathrm{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} \pi_\lambda.$$

The representation will be unitary when $\lambda \in i\mathfrak{a}^*_P$.

When π is the right regular representation then we denote $\mathcal{I}_{P,\pi}$ as $\mathcal{I}_P$. One can realize the $\mathcal{I}_P(\lambda)$ on the Hilbert space $\mathcal{H}_P$ (which is λ-independent) defined by the space of functions

$$\varphi\colon N(\mathbb{A})M(\mathbb{Q})T_M\backslash G(\mathbb{A}) \to \mathbb{C}$$

such that for every $x \in G(\mathbb{A})$ the function $\varphi_x\colon m \to \varphi(mx)$ belongs to $L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash M(\mathbb{A})^1)$, and

$$\|\varphi\|^2 := \int\limits_{M(\mathbb{Q})\backslash M(\mathbb{A})^1} \int\limits_{K_\mathbb{A}} |\varphi(mk)|^2 \,\mathrm{d}k \,\mathrm{d}m < \infty.$$

Let $\mathcal{H}^0_P \subset \mathcal{H}_P$ be the subset of $K_\mathbb{A}$-finite vectors. For each element $\varphi \in \mathcal{H}^0_P$ and $\lambda \in \mathfrak{a}^*_{P,\mathbb{C}}$ we define an Eisenstein series as a function on $G(\mathbb{Q})\backslash G(\mathbb{A})$ by

$$\mathrm{Eis}_P(\varphi,\lambda)(x) := \sum_{\gamma \in P(\mathbb{Q})\backslash G(\mathbb{Q})} \varphi(\gamma x)\chi_{\rho_P+\lambda}(a(\gamma x)), \quad x \in G(\mathbb{A}).$$

The above sum converges absolutely for sufficiently dominant λ and can be meromorphically continued for all $\lambda \in \mathfrak{a}^*_{P,\mathbb{C}}$ by the work of Langlands. Moreover, $\mathcal{I}_{P,\pi}(\lambda)$ intertwines with Eis_P, in the sense that,

$$\mathrm{Eis}_P(\mathcal{I}_P(\lambda)(g)\varphi,\lambda)(x) = \mathrm{Eis}_P(\varphi,\lambda)(xg), \quad x \in \mathbb{X}_\mathbb{A}, g \in G(\mathbb{A}).$$

In particular, if $\mathrm{Eis}_P(\varphi,\lambda)$ is $K_\mathbb{A}$-invariant then $\varphi \in \mathcal{H}_P^{K_\mathbb{A}}$. Consequently, φ can be realized as an element of $L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash M(\mathbb{A})^1)^{K_\mathbb{A}\cap M(\mathbb{A})}$.

We have the decomposition

$$\mathcal{H}_P^{K_\mathbb{A}} = \bigoplus_\pi \mathcal{H}_{P,\pi}^{K_\mathbb{A}},$$

where π runs over the isomorphism classes of irreducible sub-representations of $M(\mathbb{A})$ occurring in the subspace $L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash M(\mathbb{A})^1)$. By the Iwasawa decomposition $G(\mathbb{A}) = M(\mathbb{A})N(\mathbb{A})K_\mathbb{A}$, we have an isomorphism of vector spaces $\mathcal{H}_P^{K_\mathbb{A}} \cong$

$L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash M(\mathbb{A})^1)^{K_{M,\mathbb{A}}} \neq \{0\}$, where $K_{M,\mathbb{A}} = K_{\mathbb{A}} \cap M(\mathbb{A})$. Similarly, we have $\mathcal{H}_{P,\pi}^{K_{\mathbb{A}}} \cong \pi^{K_{M,\mathbb{A}}}$, which is one-dimensional by the multiplicity one theorem. In this case we choose $\varphi_\pi \in \mathcal{H}_{P,\pi}^{K_{\mathbb{A}}}$ to be a vector of norm 1. We let $\mathcal{B}_P$ be the set of such vectors φ, when we go over all the possible representation π of $M(\mathbb{A})$ appearing in $L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash M(\mathbb{A})^1)$ with $\pi^{K_{M,\mathbb{A}}} \neq 0$.

Let us describe the last set more explicitly. Each irreducible representation of $M(\mathbb{A})$ appearing in the decomposition of $L^2_{\mathrm{disc}}(M(\mathbb{Q})\backslash G(\mathbb{A}))$ has a central character χ of the center $Z(M(\mathbb{A}))$, trivial on $Z(M(\mathbb{A})^1) \cap M(\mathbb{Q})$. If the representation has a $K_{M,\mathbb{A}}$-invariant vector then χ must also be trivial on $Z(M(\mathbb{A})^1) \cap K_{M,\mathbb{A}}$. Every central character of $\mathrm{GL}_{n_i}(\mathbb{A})$ which is trivial on $\mathrm{GL}_{n_i}(\mathbb{Q})$ and the maximal open compact subgroup $O_{n_i}(\mathbb{R}) \times \prod_p \mathrm{GL}_{n_i}(\mathbb{Z}_p)$ is trivial. Since M is essentially a product of GL_{n_i}, we deduce that χ is trivial. Therefore, $\mathcal{B}_P$ is in bijection with irreducible subrepresentations of

$$L^2_{\mathrm{disc}}((\mathrm{PGL}_{n_1}(\mathbb{Q}) \times \cdots \times \mathrm{PGL}_{n_r}(\mathbb{Q}))\backslash(\mathrm{PGL}_{n_1}(\mathbb{A}) \times \cdots \times \mathrm{PGL}_{n_r}(\mathbb{A}))),$$

having a $K_{n_1,\mathbb{A}} \times \cdots \times K_{n_r,\mathbb{A}}$-invariant vector, where $K_{n_i,\mathbb{A}}$ is the maximal compact subgroup in $\mathrm{PGL}_{n_i}(\mathbb{A})$.

The last space decomposes into a linear span of $\varphi_1 \otimes \cdots \otimes \varphi_r$, $\varphi_i \in \mathcal{B}_{n_i}$. We conclude that $\mathcal{B}_P$ is in bijection with $\mathcal{B}_{n_1} \times \cdots \times \mathcal{B}_{n_r}$.

If $\mu_{\varphi_i,v}$ is the Langlands parameter of φ_i as the place v then we embed it in $\mathfrak{a}^*_{\mathbb{C}}$ in $(n_1 + \cdots + n_{i-1} + 1, \ldots, n_1 + \cdots + n_i)$-th coordinates. Consequently, the Langlands parameter of $\mathrm{Eis}_P(\varphi, \lambda)$ at the place v is

$$\mu_{\varphi,\lambda,v} := (\mu_{\varphi_i,v}, \ldots, \mu_{\varphi_r,v}) + \lambda \in \mathfrak{a}^*_{\mathbb{C}},$$

which follows from Lemma 3.1. In particular, for $\lambda \in i\mathfrak{a}^*_P$ we have[5]

$$\theta_{\varphi,v} = \theta_{\varphi,\lambda,v} := \theta(\mu_{\varphi,\lambda,p}) = \max_i \theta_{\varphi_i,p}, \tag{4.11}$$

where $\theta_{\varphi,v}$ is as defined in Equation (3.3). We also have

$$\nu^2_{\varphi,\lambda} := \|\mu_{\varphi,\lambda,\infty}\|^2 = \nu^2_{\varphi_1} + \cdots + \nu^2_{\varphi_r} + \|\lambda\|^2, \tag{4.12}$$

where ν_φ is as defined in Equation (4.4). We also abbreviate (and slightly abuse notations) $\nu_\varphi = \nu_{\varphi,0}$.

If $h_v \in C_c^\infty(K_v\backslash\mathrm{PGL}_n(\mathbb{Q}_v)/K_v)$, then

$$R(h_v)\mathrm{Eis}_P(\varphi, \lambda) = \mathrm{Eis}_P(\mathcal{I}_{P,\pi}(\lambda)(h_v)\varphi, \lambda) = \tilde{h}(\mu_{\varphi,\lambda,v})\mathrm{Eis}_P(\varphi, \lambda),$$

where $R(h_v)$ is defined as in Subsection 4.1.

[5] Notice that $\theta_{\varphi,\lambda,p}$ does not depend on λ, so we discard it from the notations.

In particular, if $f \in C_c^\infty(K_\mathbb{A}\backslash \mathrm{PGL}_n(\mathbb{A})/K_\mathbb{A})$ is of the form

$$f(g) = f((g_v)_v) = f_\infty(g_\infty) f_p(g_p) \prod_{q \neq p} \mathbb{1}_{K_q}(g_q), \tag{4.13}$$

then we have

$$R(f)\mathrm{Eis}_P(\varphi,\lambda) = \mathrm{Eis}_P(\mathcal{I}_{P,\pi}(h)\varphi,\lambda) = \tilde{f}_\infty(\mu_{\varphi,\lambda,\infty})\tilde{f}_p(\mu_{\varphi,\lambda,p})\mathrm{Eis}_P(\varphi,\lambda). \tag{4.14}$$

4.6. Spectral decomposition

We can describe Langlands spectral decomposition following [1, Section 7].

We denote by $C_c^\infty(K_\mathbb{A}\backslash \mathrm{PGL}_n(\mathbb{A})/K_\mathbb{A})$ the space that is spanned by functions of the form $f = \prod_{v \leq \infty} f_v$, with $f_v \in C_c^\infty(K_v\backslash \mathrm{PGL}_n(\mathbb{Q}_v)/K_v)$, and $f_v = \mathbb{1}_{K_v}$ for almost every v. Given $f \in C_c^\infty(K_\mathbb{A}\backslash \mathrm{PGL}_n(\mathbb{A})/K_\mathbb{A})$, we will consider the operator

$$R(f)\colon L^2(\mathbb{X}_\mathbb{A}) \to L^2(\mathbb{X}_\mathbb{A}),$$

defined by

$$R(f)\varphi(x) := \int\limits_{\mathrm{PGL}_n(\mathbb{A})} f(y)\varphi(xy)\,\mathrm{d}y = \int\limits_{\mathbb{X}_\mathbb{A}} K_f(x,y)\varphi(y)\,\mathrm{d}y,$$

where

$$K_f(x,y) := \sum_{\gamma \in \mathrm{PGL}_n(\mathbb{Q})} f(x^{-1}\gamma y). \tag{4.15}$$

Note that the compact support of f ensures that the above sum is finite.

Finally, we record the spectral decomposition of the automorphic kernel K_f.

$$K_f(x,y) = \sum_P C_P \sum_{\varphi \in \mathcal{B}_P} \int\limits_{i\mathfrak{a}_P^*} \mathrm{Eis}_P(\mathcal{I}_{P,\pi_\varphi}(\lambda)(f)\varphi,\lambda)(x)\overline{\mathrm{Eis}_P(\varphi,\lambda)(y)}\,\mathrm{d}\lambda. \tag{4.16}$$

Notice that when $P = G$ then $\mathcal{B}_G = \mathcal{B}_n$, there is no integral over λ, and $\mathrm{Eis}_P(\varphi,\lambda)$ is simply φ. The constants C_P are certain explicit constants with $C_G = 1$, and are slightly different than in [1], since we normalize the measure of $\mathfrak{a}_\mathbb{C}$ differently.

We will also need the L^2-spectral expansion in the following form.

Proposition 4.16. *Let $f \in C_c^\infty(K_\mathbb{A}\backslash \mathrm{PGL}_n(\mathbb{A})/K_\mathbb{A})$ be as in Equation (4.13). For $x_0 \in \mathbb{X}_\mathbb{A}$, let $F_{x_0} \in C_c^\infty(X_\mathbb{A})$ be*

$$F_{x_0}(x) := K_f(x_0,x)$$

Then

$$\|F_{x_0}\|_2^2 = \sum_P C_P \sum_{\varphi \in \mathcal{B}_P} \int_{i\mathfrak{a}_P^*} |\tilde{f}_\infty(\mu_{\varphi,\lambda,\infty})|^2 |\tilde{f}_p(\mu_{\varphi,\lambda,p})|^2 |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \,\mathrm{d}\lambda,$$

where the notations are as in Equation (4.16).

Proof. Denote $f^*(g) := \overline{f(g^{-1})}$, and $f_1 := f * f^*$. We notice that for notations as in Equation (4.16), using Equation (4.14)

$$\begin{aligned}\mathrm{Eis}_P(\mathcal{I}_{P,\pi_\varphi}(f_1)\varphi,\lambda) &= \mathrm{Eis}_P(\mathcal{I}_{P,\pi_\varphi}(f)\mathcal{I}_{P,\pi_\varphi}(f^*)\varphi,\lambda) \\ &= |\tilde{f}_\infty(\mu_{\varphi,\lambda,\infty})|^2 |\tilde{f}_p(\mu_{\varphi,\lambda,p})|^2 \mathrm{Eis}_P(\varphi,\lambda). \end{aligned} \tag{4.17}$$

We next claim that

$$\|F_{x_0}\|_2^2 = K_{f_1}(x_0,x_0). \tag{4.18}$$

Then the proof follows from Equation (4.16), Equation (4.17) and Equation (4.18).

To see Equation (4.18) we note that,

$$\|F_{x_0}\|_2^2 = \int_{\mathbb{X}_\mathbb{A}} F_{x_0}(x)\overline{F_{x_0}(x)}\,\mathrm{d}x = \int_{\mathbb{X}_\mathbb{A}} \sum_{\gamma \in \mathrm{PGL}_n(\mathbb{Q})} f(x_0^{-1}\gamma x) \sum_{\gamma_1 \in \mathrm{PGL}_n(\mathbb{Q})} \overline{f(x_0^{-1}\gamma_1 x)}\,\mathrm{d}x$$

Both of the above sums are finite. Exchanging order summation and integration, and unfolding the $\mathbb{X}_\mathbb{A}$ integral we obtain the above equals

$$\int_{\mathrm{PGL}_n(\mathbb{A})} \sum_{\gamma \in \mathrm{PGL}_n(\mathbb{Q})} f(x_0^{-1}\gamma g) f^*(g^{-1}x_0)\,\mathrm{d}g.$$

Once again exchanging finite sum with a compact integral we obtain

$$\sum_{\gamma \in \mathrm{PGL}_n(\mathbb{Q})} (f * f^*)(x_0^{-1}\gamma x_0),$$

which concludes the proof. □

We will now use Proposition 4.16 to prove a version of a *local weak Weyl Law*, which will be needed in our proof.

Proposition 4.17. *Let $\Omega \subset \mathbb{X}_A$ be a compact subset. Then for every $x_0 \in \Omega$ it holds that*

$$\sum_P C_P \sum_{\varphi \in \mathcal{B}_P, \nu_\varphi \le T} \int_{\lambda \in i\mathfrak{a}_P^*, |\lambda| \le T} |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \,\mathrm{d}\lambda \ll_\Omega T^d,$$

as T tends to infinity.

Proof. Let $\varepsilon \asymp T^{-1}$ with a sufficiently small implied constant. Choose $f_\infty = k_\varepsilon$ where k_ε is of the form described in Lemma 3.8, and $f_p = \mathbb{1}_{K_p}$. Construct f_1 and F_{x_0} as in (the proof of) Proposition 4.16. Note that $f_{1,p}$ is again $\mathbb{1}_{K_p}$ and $f_{1,\infty}$ is supported on $K_\infty B_{2\varepsilon} K_\infty$, which follows from property (1) of Lemma 3.8. Then we claim that for $x_0 \in \Omega$

$$\|F_{x_0}\|_2^2 \ll_\Omega \|f_\infty\|_2^2 \ll T^d.$$

Notice that the second estimate follows from property 3 in Lemma 3.8.

First, we choose some fixed liftings of $x_0, x \in \mathrm{PGL}_n(\mathbb{A})$, whose p-coordinates $x_{0,p}, x_p$ are in K_p, and their ∞-coordinates are in a fixed fundamental domain of $\mathrm{PGL}_n(\mathbb{Z})\backslash\mathrm{PGL}_n(\mathbb{R})$. This is possible by Lemma 4.1.

If $f_1(x_0^{-1}\gamma x) \neq 0$ it implies that $\mathbb{1}_{K_p}(x_{0,p}^{-1}\gamma x_p) \neq 0$ for all p. Hence, $\gamma \in K_p$ for all p, which implies that $\gamma \in \mathrm{PGL}_n(\mathbb{Z})$. In addition, $f_{1,\infty}(x_{0,\infty}^{-1}\gamma x_\infty) \neq 0$, which implies that $x_{0,\infty}^{-1}\gamma x_\infty \in K_\infty B_{2\varepsilon} K_\infty$. Clearly, the number of $\gamma \in \mathrm{PGL}_n(\mathbb{Z})$ with $x_{0,\infty}^{-1}\gamma x_\infty \in B_{2\varepsilon}$ is $\ll_\Omega 1$.

Using the proof of Proposition 4.16 we obtain

$$\|F_{x_0}\|_2^2 \ll_\Omega \|f_{1,\infty}\|_\infty.$$

Applying Cauchy–Schwarz we see that the above is bounded by $\|f_\infty\|_2^2$, as needed.

First notice that for $\nu_\varphi \leq T$ and $\|\lambda\| \leq T$ it holds that

$$\|\mu_{\varphi,\lambda,\infty}\| = \nu_{\varphi,\lambda} \ll T \asymp \varepsilon^{-1},$$

with a sufficiently small implied constant. Thus using property (5) of Lemma 3.8 we get that

$$|\tilde{f}_\infty(\mu_{\varphi,\lambda,\infty})| \gg 1.$$

Therefore we have

$$\sum_P C_P \sum_{\varphi \in \mathcal{B}_P, \nu_\varphi \leq T} \int_{\lambda \in i\mathfrak{a}_P^*, |\lambda| \leq T} |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \,\mathrm{d}\lambda$$

$$\ll \sum_P C_P \sum_{\varphi \in \mathcal{B}_P} \int_{\lambda \in i\mathfrak{a}_P^*} |\tilde{f}_\infty(\mu_{\varphi,\lambda,\infty})|^2 |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \,\mathrm{d}\lambda.$$

Applying Proposition 4.16 we conclude. □

4.7. The residual spectrum and shapes

Mœglin and Waldspurger in [33] described the residual spectrum $\mathcal{B}_{n,\mathrm{res}}$ as follows. Let $ab = n$ with $b > 1$ and let $\varphi \in \mathcal{B}_{a,\mathrm{cusp}}$. Consider the standard parabolic subgroup

P corresponding to the ordered partition $(a, \ldots, a)$ of n, and let $\varphi' \in \mathcal{B}_P$ correspond to $(\varphi, \ldots, \varphi) \in \mathcal{B}_a \times \cdots \times \mathcal{B}_a$. Construct the Eisenstein series $\mathrm{Eis}_P(\varphi', \lambda)$ for $\lambda \in \mathfrak{a}^*_{P,\mathbb{C}}$. This is a meromorphic function, and it has a (multiple) residue at the point

$$\lambda = \rho_b = ((b-1)/2, (b-3)/2, \ldots, -(b-1)/2) \in \mathfrak{a}^*_{P,\mathbb{C}},$$

where we temporarily identified $\mathfrak{a}^*_{P,\mathbb{C}}$ with a subset of $\mathbb{C}^b$ and recall that it embeds in $\mathfrak{a}^*_{\mathbb{C}}$ by repeating each value a times. The residue can be calculated as

$$\psi' = \lim_{\lambda \to \rho_b} (\lambda_1 - \lambda_2 - 1) \cdot \cdots \cdot (\lambda_{b-1} - \lambda_b + 1)\mathrm{Eis}_P(\varphi', \lambda).$$

After normalization, $\psi = \psi'/\|\psi'\|$ is an element of $\mathcal{B}_{n,\mathrm{res}}$, and every element of $\mathcal{B}_{n,\mathrm{res}}$ can be constructed this way for some $ab = n$, $b > 1$, $\varphi \in \mathcal{B}_a$. Thus we deduce that

$$\mathcal{B}_n = \bigsqcup_{a|n} \mathcal{B}_{a,\mathrm{cusp}}.$$

From the above description it follows that on the level of Langlands parameters we have

$$\mu_{\psi,v} = \mu_{\varphi',\rho_b,v} = (\mu_{\varphi,v}, \ldots, \mu_{\varphi,v}) + \rho_b.$$

In particular, we have

$$\theta_{\psi,p} = \theta_{\varphi,p} + (b-1)/2,$$

and if $\nu_\varphi \gg 1$ then $\nu_\psi \asymp \nu_\varphi$.

Each $\varphi \in \mathcal{B}_P$ is parameterized by a *shape* $(a_1, b_1), \ldots, (a_r, b_r)$ where $a_i, b_i \geq 1$ with $\sum_{i=1}^r a_i b_i = n$, and P corresponds to the ordered partition $n = \sum_{i=1}^r a_i b_i$. Moreover, if φ corresponds to $(\varphi_1, \ldots, \varphi_r)$, $\varphi_i \in \mathcal{B}_{n_i}$, then φ_i corresponds to a cuspidal representation in $\mathcal{B}_{a_i,\mathrm{cusp}}$.

Given a shape $S = ((a_1, b_1), \ldots, (a_r, b_r))$, we let $\mathcal{B}_S \subset \mathcal{B}_P$ be the set of forms $\varphi \in \mathcal{B}_P$ of shape S. We have the following estimates.

Lemma 4.18. *For every $\varphi \in \mathcal{B}_S$, it holds that*

$$\theta_{\varphi,p} \leq \max_i \{(b_i - 1)/2 + \theta_{a_i,p}\},$$

where $\theta_{a_i,p}$ is the best known bound towards the GRC at the place p for the cuspidal spectrum of $\mathrm{PGL}(a_i)$.

Indeed, the lemma follows from the claim above about the behavior of θ of residual forms and Equation (4.11).

The following estimate combines Weyl's law for a given shape.

Lemma 4.19. *We have*

$$|\{\varphi \in \mathcal{B}_S \mid \nu_\varphi \leq T\}| \ll T^{d_S},$$

where $d_S = \sum_{i=1}^r (a_i + 2)(a_i - 1)/2$.

Proof. As described above, there is a bijection between $\mathcal{B}_S$ and $\mathcal{B}_{a_1,\mathrm{cusp}} \times \cdots \times \mathcal{B}_{a_r,\mathrm{cusp}}$. Moreover, under this bijection, by combining the estimates for ν of residual forms above and Equation (4.12) we have

$$\nu_\varphi + 1 \asymp \max_{i=1}^{r} \nu_{\varphi_i} + 1.$$

The estimate now follows from Proposition 4.10. □

4.8. Local L^2-bounds of Eisenstein series

Let $\varphi \in \mathcal{B}_P$ and $\lambda \in i\mathfrak{a}_P^*$, and let $\mathrm{Eis}_P(\varphi, \lambda) \in C^\infty(\mathbb{X})$ be the corresponding Eisenstein series. It is known that the Eisenstein series grow polynomially near the cusp. It is a natural and challenging problem to find *good* pointwise upper bound of $\mathrm{Eis}_P(\varphi, \lambda)$.

A more tractable approach is to take a compact subset $\Omega \subset \mathbb{X}$ of positive measure and ask the size of $\|\mathrm{Eis}_P(\varphi, \lambda)|_\Omega\|_2$ as $\nu_{\varphi,\lambda} \to \infty$. In this paper we need an upper bound of $\|\mathrm{Eis}_P(\varphi, \lambda)|_\Omega\|_2^2$ on an average over λ in a long interval. One may deduce certain bounds of such an average from the local Weyl law (cf. Proposition 4.17) or via the improved L^∞-bounds in [8], but such bounds are not sufficient for our purposes.

One expects that $\|\mathrm{Eis}_P(\varphi, \lambda)|_\Omega\|_2$ remains essentially bounded in $\nu_{\varphi,\lambda}$, which is an analogue of the Lindelöf hypothesis for the L-functions. More precisely, we expect that

$$\int_\Omega |\mathrm{Eis}_P(\varphi, \lambda)(x)|^2 \,\mathrm{d}x \ll_\Omega \log^{n-1}(1 + \nu_{\varphi,\lambda}). \tag{4.19}$$

Note that for $n = 2$ the above is classically known. We refer to [27] for a detailed discussion.

We remark that a recent result of Assing–Blomer [2, Theorem 1.5] on optimal lifting for $\mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$ also requires such a bound on the local L^2-growth but in a non-archimedean aspect; see [2, Hypothesis 1].

Proving Equation (4.19) seems to be quite difficult. A natural way to approach the problem is via the higher rank *Maass–Selberg relations* due to Langlands, at least when φ is *cuspidal*. Among many complications that one faces through this approach (see [27, §1.3]) the major one involves *standard (in $\nu_{\varphi,\lambda}$ aspect)* zero-free region for various $\mathrm{GL}(n) \times \mathrm{GL}(m)$ Rankin–Selberg L-functions, which are available only in a very few cases.

A relatively easy problem would be to find upper bound of a short λ average of $\|\mathrm{Eis}_P(\varphi, \lambda)|_\Omega\|_2^2$. In fact, we expect

$$\int\limits_{\|\lambda'-\lambda\|\leq 1} \int\limits_{\Omega} |\mathrm{Eis}_P(\varphi,\lambda')(x)|^2 \, \mathrm{d}x \, \mathrm{d}\lambda' \ll_\Omega \log^{n-1}(1+\nu_{\varphi,\lambda}). \tag{4.20}$$

In a companion paper [27] we study these problems in detail for a general reductive groups.

For the current paper we only need to find an upper bound of $\|\mathrm{Eis}_P(\varphi,\lambda)|_\Omega\|_2^2$ on an average over λ over a long interval so that the bound in the φ aspect is only polynomial in ν_φ with very small degree. However, it is important for us that the bound is uniform over all φ, cuspidal or not. We describe the required estimate below.

Proposition 4.20. *For every $\varphi \in \mathcal{B}_P$ we have*

$$\int\limits_{\lambda\in i\mathfrak{a}_P^*, \|\lambda\|\leq T} \int\limits_{\Omega} |\mathrm{Eis}_P(\varphi,\lambda)(x)|^2 \, \mathrm{d}x \, \mathrm{d}\lambda \ll_\Omega \log(1+T+\nu_\varphi)^{n-1} T^{\dim \mathfrak{a}_P},$$

for any $T \geq 1$.

The theorem should be compared to the simple upper bounds using the local Weyl law as in Proposition 4.17), which allows us to deduce a similar statement, but with $T^{\dim \mathfrak{a}_P}$ replaced by T^d, which is insufficient for our purpose.

We remark that for $n = 3$, Proposition 4.20 was proved by Miller in [32], as one of the main ingredients in proving Weyl's law for $\mathrm{SL}_3(\mathbb{Z})\backslash \mathrm{SL}_3(\mathbb{R})/\mathrm{SO}_3(\mathbb{R})$. Therefore, for $n = 3$ (and $n = 2$) the results of this paper are unconditional on the result of the companion paper. Proposition 4.20 generalizes Miller's result to higher rank, and similarly implies Weyl's law in the same way. However, we heavily rely on [35], so this does not lead to a new proof.

Proof of Proposition 4.20. We find $\{\eta_j\}_{j=1}^k \in i\mathfrak{a}_P^*$ with $k \ll T^{\dim \mathfrak{a}_P}$ and $\|\eta_j\| \leq T$ so that

$$\{\lambda \in i\mathfrak{a}_P^* \mid \|\lambda\| \leq T\} \subset \cup_{j=1}^k \{\lambda \in i\mathfrak{a}_P^* \mid \|\lambda - \eta_j\| \leq 1\}.$$

Clearly, we can majorize the integral in the proposition by

$$\sum_{j=1}^k \int\limits_{\substack{\lambda\in i\mathfrak{a}_P^* \\ \|\lambda-\eta_j\|\leq 1}} \int\limits_{\Omega} |\mathrm{Eis}_P(\varphi,\lambda)(x)|^2 \, \mathrm{d}x \, \mathrm{d}\lambda.$$

We apply [27, Theorem 1] to each summand on the right hand side above with $\varphi_0 = \varphi$ and $\lambda_0 = \eta_j$ to conclude that the integral in the proposition is bounded by

$$\ll_\Omega k \max_j \log(1+\nu_\varphi + \|\eta_j\|)^{n-1}.$$

We conclude the proof by employing the bounds on η_j and k. $\square$

5. Reduction to a spectral problem

In this section, we finally begin proving the main results Theorem 2 and Theorem 3, and we will reduce it to a spectral problem. Then we will apply a few different reductions to simplify the problem even further.

Consider the set

$$S(k) := \{\gamma \in \mathrm{SL}_n(\mathbb{Z}[1/p]) \mid \mathrm{ht}(\gamma) \le k\}.$$

Lemma 5.1. *The image of $S(k)$ in $\mathrm{PGL}_n(\mathbb{Q})$ is equal to $\tilde{R}(p^{nk})$.*

Proof. The map $S(k) \to R(p^{nk})$ defined by $\gamma \mapsto p^k\gamma$ is a bijection. Therefore, the images of them in $\mathrm{PGL}_n(\mathbb{Q})$ are the same. $\square$

Below we identify $\mathbb{X} := \mathrm{SL}_n(\mathbb{Z})\backslash\mathbb{H}^n \cong \mathrm{PSL}_n(\mathbb{Z})\backslash\mathrm{PGL}_n(\mathbb{R})/K_\infty$. This implies that the action of $\gamma \in \mathrm{SL}_n(\mathbb{R})$ on $\mathbb{H}^n$ depends only on the image of γ in $\mathrm{PGL}_n(\mathbb{R})$.

Definition 5.2. Let $x, x_0 \in \mathbb{X}$ which we identify with some lifts of them in $\mathbb{H}^n \cong \mathrm{PGL}_n(\mathbb{R})/K_\infty$. Also, let $\varepsilon > 0$ and $k \in \mathbb{Z}_{\ge 0}$. We say that the pair (x, x_0) is (ε, k)-*admissible* if there is a solution $\gamma \in \tilde{R}(p^{nk})$ to $\mathrm{dist}(x, \gamma x_0) \le \varepsilon$.

Notice that since $\tilde{R}(p^{nk})$ is left and right $\mathrm{PSL}_n(\mathbb{Z})$-invariant, the above definition does not depend on the lifts of $x, x_0 \in \mathbb{H}^n$.

Unraveling Definition 1.1 and Definition 5.2 we get the following.

Lemma 5.3. *Let $x, x_0 \in \mathbb{X}$. The Diophantine exponent $\kappa(x, x_0)$ is the infimum over $\zeta < \infty$ such that there exists $\varepsilon_0 = \varepsilon_0(x, x_0, \zeta)$ with the property that for every $\varepsilon < \varepsilon_0$ the pair (x, x_0) is $(\varepsilon, \zeta\frac{n+2}{2n}\log_p(\varepsilon^{-1}))$-admissible.*

Let $k_\varepsilon \in C_c^\infty(K_\infty\backslash\mathrm{PGL}_n(\mathbb{R})/K_\infty)$ be as in Lemma 3.8. For $x_0 \in \mathbb{X}$, let $K^{\mathbb{X}}_{\varepsilon,x_0} \in C_c^\infty(\mathbb{X})$ be the automorphic kernel

$$K^{\mathbb{X}}_{\varepsilon,x_0}(y) := \sum_{\gamma \in \mathrm{PSL}_n(\mathbb{Z})} k_\varepsilon(x_0^{-1}\gamma^{-1}y).$$

It is simple to see that

$$\int_{\mathbb{X}} K^{\mathbb{X}}_{\varepsilon,x_0}(y)\,\mathrm{d}y = \int_{\mathrm{PGL}_n(\mathbb{R})} k_\varepsilon(x_0^{-1}g)\,\mathrm{d}g = 1.$$

Recall the Hecke operator $T^*(p^k)$ from Subsection 4.1, and the fact that it acts on functions on $\mathbb{X}$ by Remark 4.9.

Lemma 5.4. *Assume that*

$$T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0}(x) \neq 0,$$

then the pair (x, x_0) is $(\varepsilon, k_1 + k_2)$-admissible.

Proof. We have

$$0 \neq T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0}(x) = \sum_{\gamma_1 \in A(p^{nk_1})} \sum_{\gamma_2 \in A(p^{nk_2})} \sum_{\gamma \in \mathrm{PSL}_n(\mathbb{Z})} k_\varepsilon(x_0^{-1}\gamma^{-1}\gamma_2^{-1}\gamma_1^{-1}x).$$

So there is a $\gamma' := \gamma_1\gamma_2\gamma$ such that $k_\varepsilon(x_0^{-1}\gamma'^{-1}x) \neq 0$. It holds that $\gamma' \in \tilde{R}(p^{n(k_1+k_2)})$. By the assumption of the support of k_ε, we have

$$\mathrm{dist}(x_0^{-1}\gamma'^{-1}x, e) = \mathrm{dist}(x, \gamma' x_0) \leq \varepsilon,$$

as needed. □

Let $\pi_{\mathbb{X}} = \frac{\mathbb{1}_{\mathbb{X}}}{m(\mathbb{X})}$ be the L^1-normalized characteristic function on $\mathbb{X}$.

Lemma 5.5. *Let $x_0 \in \mathbb{X}$. Assume that*

$$\|T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2 \leq \frac{c}{\sqrt{m(\mathbb{X})}}.$$

Then there is a subset $Y \subset \mathbb{X}$, such that

$$m(Y) \geq m(\mathbb{X})(1 - c^2),$$

such that for all $x \in Y$ the pair (x, x_0) is $(\varepsilon, k_1 + k_2)$-admissible.

Proof. Let $Y := \{x \in \mathbb{X} \mid T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0}(x) \neq 0\}$. By Lemma 5.4, each $x \in Y$ is $(\varepsilon, k_1 + k_2)$-admissible. On the other hand,

$$\|T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2^2 \geq \int_{\mathbb{X}\setminus Y} \pi_{\mathbb{X}}(x)^2 \,\mathrm{d}x = m(\mathbb{X} \setminus Y)/m(\mathbb{X})^2.$$

We deduce that $m(\mathbb{X} \setminus Y) \leq c^2 m(\mathbb{X})$, as needed. □

Lemma 5.6. *Let $x_0 \in \mathbb{X}$ and $\beta \geq 1$. Assume that there is $\alpha > 0$ such that for every $\delta > 0$ there is $\varepsilon_0 > 0$, such that for every $0 < \varepsilon < \varepsilon_0$ there are k_1, k_2 with $k_1 + k_2 \leq (1 + \delta)\beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$, such that we have*

$$\|T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2 \leq \varepsilon^{\alpha\delta}.$$

Then $\kappa(x_0) \leq \beta$.

Proof. Let $\delta > 0$. For ε fixed, let $Z_{\varepsilon,\delta} \subset \mathbb{X}$ be the set of $x \in \mathbb{X}$ such that the pairs (x, x_0) are not (ε, k)-admissible with $k \leq (1+\delta)\beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$. Using Lemma 5.3 it suffices to prove that for almost every $x \in \mathbb{X}$, for ε_0 small enough depending on x, δ, and $\varepsilon < \varepsilon_0$ we have $x \notin Z_{\varepsilon,\delta}$.

Let $\varepsilon_j := e^{-cj}$, for some $c > 0$ sufficiently small relatively to δ. Then for ε small enough, there is ε_j such that $Z_{\varepsilon,\delta} \subset Z_{\varepsilon_j,\delta/2}$. Therefore, it suffices to prove that for almost every $x \in \mathbb{X}$, for $m \in \mathbb{Z}_{\geq 0}$ large enough, $x \notin Z_{\varepsilon_j,\delta/2}$. Using the Borel–Cantelli lemma it is enough to prove that

$$\sum_j m(Z_{\varepsilon_j,\delta/2}) < \infty. \tag{5.1}$$

By the assumption and Lemma 5.5, there is $\varepsilon_0 > 0$ such that for $\varepsilon_j < \varepsilon_0$,

$$m(Z_{\varepsilon_j,\delta/2}) \ll \varepsilon_j^{2\alpha\delta} = e^{-2c\alpha\delta j}.$$

This shows that Equation (5.1) holds, as needed. $\square$

We now add an additional average over x_0. Let $\Omega \subset \mathbb{X}$ be a fixed compact subset of positive measure.

Lemma 5.7. *Let $\beta \geq 1$. Assume that there is $\alpha > 0$ such that for every $\delta > 0$ there is $\varepsilon_0 > 0$, such that for every $0 < \varepsilon < \varepsilon_0$ there are k_1, k_2 with $k_1 + k_2 \leq (1+\delta)\beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$, such that we have*

$$\int_\Omega \|T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2^2 \, \mathrm{d}x_0 \leq \varepsilon^{\alpha\delta}.$$

Then $\kappa \leq \beta$.

Proof. Let $\delta > 0$. Let $Z_{\varepsilon,\delta} \subset \mathbb{X} \times \mathbb{X}$ be the set of $(x, x_0) \in \mathbb{X} \times \Omega$ that are not (ε, k)-admissible, for $k \leq (1+\delta)\beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$.

Since $\kappa = \kappa(x, x_0)$ for almost every x, x_0 (see Section 2), using Lemma 5.3 it suffices to prove that for almost every $x_0 \in \Omega$ and almost every $x \in \mathbb{X}$, there is an ε_0 such that for $\varepsilon < \varepsilon_0$ the pair $(x, x_0) \notin Z_{\varepsilon,\delta}$. Using the same argument as in the proof of Lemma 5.6, we may assume that $\varepsilon = \varepsilon_j = e^{-cj}$, and using Borel–Cantelli it is enough to prove that

$$\sum_j m(Z_{\varepsilon_j,\delta}) < \infty.$$

For $\varepsilon < \varepsilon_0$ small enough, let $Y_{\varepsilon,\delta} \subset \Omega$ be the set of $x_0 \in \Omega$ such that for k as in the assumption of the lemma

$$\|T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2^2 \leq \varepsilon^{\alpha\delta/2}.$$

We claim that $m(\Omega - Y_{\varepsilon,\delta}) \leq \varepsilon^{\alpha\delta/2}$. Indeed,

$$\int\limits_{\Omega - Y_{\varepsilon,\delta}} \|T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2^2 \, \mathrm{d}x_0 \geq \varepsilon^{\alpha\delta/2} m(\Omega - Y_{\varepsilon,\delta}),$$

so $\varepsilon^{\alpha\delta/2} m(\Omega - Y_{\varepsilon,\delta}) \leq \varepsilon^{\alpha\delta}$ giving the desired estimate.

Now, for $x_0 \in Y_{\varepsilon,\delta}$ by Lemma 5.5 we have

$$m(\{x \in \mathbb{X} \mid (x, x_0) \in Z_{\varepsilon,\delta}\}) \ll \varepsilon^{\alpha\delta/2}.$$

Therefore,

$$m(Z_{\varepsilon,\delta}) \leq m(\Omega - Y_{\varepsilon,\delta}) m(\mathbb{X}) + m(Y_{\varepsilon,\delta}) \varepsilon^{\alpha\delta/2} \ll \varepsilon^{\alpha\delta/2}.$$

Using the last estimate we get

$$\sum_j m(Z_{\varepsilon_j,\delta}) < \infty,$$

as needed. □

We now discuss a further reduction, which allows us to prove bounds of κ but with weaker assumptions than that of Lemma 5.6 and Lemma 5.7. First, we will need the following estimates, which play a major role in the work [22], and which we already discussed in Proposition 2.4 using different notations.

Lemma 5.8. *For all $n \geq 2$ there is an $\alpha > 0$ such that as an operator on $L^2_0(\mathbb{X})$*

$$\|T^*(p^l)\|_{\mathrm{op}} \ll p^{-l\alpha}.$$

Moreover, for $n \geq 3$ any $\alpha < 1/2$ and for $n = 2$ (resp. under the GRC) any $\alpha < 25/64$ (resp. $\alpha < 1/2$) work.

Proof. Using Remark 4.9, the proof follows from bounds on the integrability exponents of the action of $\mathrm{PGL}_n(\mathbb{Q}_p)$ on $L^2(\mathrm{PGL}_n(\mathbb{Q})\backslash\mathrm{PGL}_n(\mathbb{A}))$, as in Proposition 2.4 and the discussions after it. □

Some remarks are in order now.

(1) By combining Lemma 5.6 and Lemma 5.8 we may deduce Theorem 1. Indeed, we essentially recovered the arguments in [22] for our specific case.
(2) For $n \geq 3$, by [15, Theorem 1.5], Lemma 5.8 is optimal in the sense that for every $\delta > 0$ there exists $f \in L^2_0(\mathbb{X})$, with

$$\|T^*(p^l)f\|_2 \gg_\delta p^{-l(1/2+\delta)}\|f\|_2.$$

This shows that to prove Theorem 2 one needs stronger tools than spectral gap alone.

Lemma 5.8 allows us to give the following versions of Lemma 5.6 and Lemma 5.7.

Lemma 5.9. *Let $\beta \geq 1$. Assume that there is an $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ there is $k \leq \beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$ such that we have*

$$\|T^*(p^{nk})K^{\mathbb{X}}_{\varepsilon,x_0}\|_2 \ll_\eta \varepsilon^{-\eta},$$

for every $\eta > 0$. Then $\kappa(x_0) \leq \beta$.

Proof. By the assumption, there is an ε_0 such that for $\varepsilon < \varepsilon_0$ and for some $k_2 \leq \beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$, it holds that

$$\|T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0}\|_2 \ll_\eta \varepsilon^{-\eta}.$$

Since $T^*(p^{nk_2})$ is an average operator and $\int_{\mathbb{X}} K^{\mathbb{X}}_{\varepsilon,x_0}(x)\,\mathrm{d}x = 1$ we have

$$T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}} \in L^2_0(\mathbb{X}).$$

Let $\delta > 0$. Let $k_1 = \lfloor \beta\delta\frac{n+2}{2n}\log_p(\varepsilon^{-1})\rfloor$. Notice that $k_1 + k_2 \leq \beta(1+\delta)\frac{n+2}{2n}\log_p(\varepsilon^{-1})$.

Applying Lemma 5.8 we find some $\alpha > 0$ such that

$$\begin{aligned}\|T^*(p^{nk_1})T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2 &= \|T^*(p^{nk_1})(T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}})\|_2 \\ &\ll \varepsilon^{\alpha\delta}\|T^*(p^{nk_2})K^{\mathbb{X}}_{\varepsilon,x_0} - \pi_{\mathbb{X}}\|_2 \ll_\eta \varepsilon^{\alpha\delta-\eta}.\end{aligned}$$

By choosing ε_0', η small enough, for $\varepsilon < \varepsilon_0'$ the above is $\leq \varepsilon^{\alpha\delta/2}$. The lemma now follows from Lemma 5.6. □

Our final reduction will allow us to replace the space $\mathbb{X}$ by the nicer space $\mathbb{X}_0$. For $x_0 \in \mathbb{X}_0$, let $K^{\mathbb{X}_0}_{\varepsilon,x_0} \in C^\infty_c(\mathbb{X}_0)$ be

$$K^{\mathbb{X}_0}_{\varepsilon,x_0}(y) := \sum_{\gamma\in\mathrm{PGL}_n(\mathbb{Z})} k_\varepsilon(x_0^{-1}\gamma^{-1}y).$$

Let $\Phi\colon \mathbb{X} \to \mathbb{X}_0$ be the covering map. Then we can define a push-forward map $\Phi_*\colon L^2(\mathbb{X}) \to L^2(\mathbb{X}_0)$, defined for $f \in L^2(\mathbb{X})$, $y \in \mathbb{X}$ as

$$\Phi_*(f)(\Phi(y)) := \sum_{\gamma\in\mathrm{PGL}_n(\mathbb{Z})/\mathrm{PSL}_n(\mathbb{Z})} f(\gamma y).$$

We have the simple norm estimate on push-forward maps for *non-negative* f,

$$\|f\|_2 \ \leq \|\Phi_* f\|_2, \tag{5.2}$$

where on the left-hand side the norm on $L^2(\mathbb{X})$ and on the right-hand side the norm is on $L^2(\mathbb{X}_0)$.

Lemma 5.10. *Let $x_0 \in \mathbb{X}$. Then it holds that*

$$K^{\mathbb{X}_0}_{\varepsilon,\Phi(x_0)} = \Phi_* K^{\mathbb{X}}_{\varepsilon,x_0}$$

and similarly

$$T^*(p^{nk})K^{\mathbb{X}_0}_{\varepsilon,\Phi(x_0)} = \Phi_*(T^*(p^{nk})K^{\mathbb{X}}_{\varepsilon,x_0})$$

Proof. It is sufficient to prove the second estimate. Indeed, unwinding the definitions we get that

$$T^*(p^{nk})K^{\mathbb{X}}_{\varepsilon,x_0}(y) = \sum_{\gamma\in\tilde{R}(p^{nk})} k_\varepsilon(x_0^{-1}\gamma^{-1}y),$$

and similarly,

$$T^*(p^{nk})K^{\mathbb{X}_0}_{\varepsilon,x_0}(y) = \sum_{\gamma'\in \mathrm{PGL}_n(\mathbb{Z})/\mathrm{PSL}_n(\mathbb{Z})} \sum_{\gamma\in\tilde{R}(p^{nk})} k_\varepsilon(x_0^{-1}\gamma^{-1}\gamma' y).$$

The lemma follows. □

Finally, combining Lemma 5.9, Lemma 5.10, and Equation (5.2), we deduce the following.

Lemma 5.11. *Let $\beta \geq 1$ and $x_0 \in \mathbb{X}$. Assume that there is an $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ and for some $k \leq \beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$ we have*

$$\|T^*(p^{nk})K^{\mathbb{X}_0}_{\varepsilon,\Phi(x_0)}\|_2 \ll_\eta \varepsilon^{-\eta},$$

for every $\eta > 0$. Then $\kappa(x_0) \leq \beta$.

The same set of arguments, with Lemma 5.7 in place of Lemma 5.6 will give:

Lemma 5.12. *Let $\Omega \subset \mathbb{X}_0$ be a compact set of positive measure. Let $\beta \geq 1$. Assume that for every $\delta > 0$ there is $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ there is $k \leq (1+\delta)\beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})$ such that*

$$\int_{\Omega} \|T^*(p^{nk}) K^{\mathbb{X}_0}_{\varepsilon,x_0}\|_2^2 \, \mathrm{d}x_0 \ll_\eta \varepsilon^{-\eta},$$

for every $\eta > 0$*. Then* $\kappa \le \beta$*.*

6. Applying the spectral decomposition

Consider the adelic function $f \in C_c^\infty(K_{\mathbb{A}} \backslash \mathrm{PGL}_n(\mathbb{A})/K_{\mathbb{A}})$ defined by

$$f((g)_v) := k_\varepsilon(g_\infty) h_{p^{nk}}(g_p) \prod_{q \neq p} \mathbb{1}_{K_q}(g_q),$$

where $h_{p^{nk}}$ is as in Equation (3.7) and k_ε is given by Lemma 3.8.

Given $x_0 \in \mathbb{X}_0$, identify it by a slight abuse of notations as an element $x_0 \in \mathbb{X}_{\mathbb{A}}$. Consider the function

$$F_{x_0}(x) = \sum_{\gamma \in \mathrm{PGL}_n(\mathbb{Q})} f_1(x_0^{-1}\gamma^{-1}x),$$

where f_1 is the self-convolution of f, as defined in the proof of Proposition 4.16. Using the discussion in Subsection 4.1, we see that F_{x_0} is the adelic version of the function $T^*(p^{nk}) K^{\mathbb{X}_0}_{\varepsilon,x_0}$. Therefore,

$$\|T^*(p^{nk}) K^{\mathbb{X}_0}_{\varepsilon,x_0}\|_2^2 = \|F_{x_0}\|_2^2,$$

where the underlying space on the left-hand side is $\mathbb{X}_0$ and on the right-hand side is $\mathbb{X}_{\mathbb{A}}$.

We apply Proposition 4.16 to F_{x_0}, and obtain that

$$\begin{aligned} \|T^*(p^{nk}) K^{\mathbb{X}_0}_{\varepsilon,x_0}\|_2^2 &= \sum_P C_P \sum_{\varphi \in \mathcal{B}_P} \int_{i\mathfrak{a}_P^*} |\tilde{k}_\varepsilon(\mu_{\varphi,\lambda,\infty})|^2 |\tilde{h}_{p^{nk}}(\mu_{\varphi,\lambda,p})|^2 |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \, \mathrm{d}\lambda \\ &= \sum_{\varphi \in \mathcal{B}_G} |\tilde{k}_\varepsilon(\mu_{\varphi,\infty})|^2 |\tilde{h}_{p^{nk}}(\mu_{\varphi,p})|^2 |\varphi(x_0)|^2 \\ &+ \sum_{P \neq G} C_P \sum_{\varphi \in \mathcal{B}_P} \int_{i\mathfrak{a}_P^*} |\tilde{k}_\varepsilon(\mu_{\varphi,\lambda,\infty})|^2 |\tilde{h}_{p^{nk}}(\mu_{\varphi,\lambda,p})|^2 |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \, \mathrm{d}\lambda \end{aligned} \tag{6.1}$$

Using Lemma 3.8 we get that for every $N > 0$,

$$\begin{aligned} |\tilde{k}_\varepsilon(\mu_{\varphi,\lambda,\infty})| &\ll_N (1+\varepsilon\nu_{\varphi,\lambda})^{-N} \ll (1+\varepsilon\nu_\varphi)^{-N/2}(1+\varepsilon\|\lambda\|)^{-N/2}, \\ |\tilde{k}_\varepsilon(\mu_{\varphi,\infty})| &\ll_N (1+\varepsilon\nu_\varphi)^{-N}. \end{aligned} \tag{6.2}$$

Also using Lemma 3.5 we get that for every $\eta > 0$,

$$|\tilde{h}_{p^{nk}}(\mu_{\varphi,\lambda,p})| \ll_\eta p^{kn(\theta_{\varphi,p}-(n-1)/2+\eta)}, \quad |\tilde{h}_{p^{nk}}(\mu_{\varphi,p})| \asymp p^{-kn(n-1)/2}|\lambda_\varphi(p^{nk})|. \tag{6.3}$$

Thus we arrive at the following proposition.

Proposition 6.1 *(Truncation). Let $\Omega \subset \mathbb{X}_0$ be a fixed compact set. Then for every $x_0 \in \Omega$ and $\delta, \eta > 0$ we have*

$$\begin{aligned} \|T^*(p^{nk})K^{\mathbb{X}_0}_{\varepsilon,x_0}\|_2^2 \ll_{\Omega,N,\eta,\delta} &\sum_{\varphi\in\mathcal{B}_G,\nu_\varphi\le\varepsilon^{-1-\delta}} p^{-kn(n-1)}|\lambda_\varphi(p^{nk})|^2|\varphi(x_0)|^2 \\ &+ \sum_{P\neq G} C_P \sum_{\varphi\in\mathcal{B}_P,\nu_\varphi\le\varepsilon^{-1-\delta}} p^{kn(2\theta_{\varphi,p}-(n-1)+\eta)} \int\limits_{\lambda\in i\mathfrak{a}_P^*,\|\lambda\|\le\varepsilon^{-1-\delta}} |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \,\mathrm{d}\lambda + \varepsilon^N \end{aligned}$$

for every $N > 0$.

Proof. We notice that the proposition follows from Equation (6.1) and Equation (6.3), if we can show that the contribution of $\varphi \in \mathcal{B}_P$ with $\nu_\varphi \ge \varepsilon^{-1-\delta}$, and the contribution of $\varphi \in \mathcal{B}_p$ for $P \neq G$ and λ with $\varphi \le \varepsilon^{-1-\delta}$ and $\|\lambda\| \ge \varepsilon^{-1-\delta}$ are $O_{\Omega,N,\delta}(\varepsilon^N)$.

We use the estimate

$$|\tilde{h}_{p^{nk}}(\mu_{\varphi,\lambda,p)})| \le 1,$$

which follows from Remark 3.6, throughout the proof.

We first handle the contribution to Equation (6.1) from φ with $S \le \nu_\varphi \le 2S$, with $S \ge \varepsilon^{-1-\delta}$. Notice that in this case $(1+\varepsilon\nu_\varphi) \gg S^{\delta'}$ for some δ' depending on δ. Applying Equation (6.2) we see that the contribution of such φ is bounded by

$$\ll_{N,\delta} \sum_P \sum_{\varphi\in\mathcal{B}_P,S\le\nu_\varphi\le 2S} S^{-N} \int\limits_{i\mathfrak{a}_P^*} (1+\varepsilon\|\lambda\|)^{-N}|\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \,\mathrm{d}\lambda.$$

Take N large enough, use Proposition 4.17, and apply integration by parts. Then the inner integral is bounded by $\ll_\Omega \varepsilon^{-L_1}$ for some absolute L_1. Making N large enough the entire sum is bounded by

$$\ll_{\Omega,N,\delta} \sum_P \sum_{\varphi\in\mathcal{B}_P,S\le\nu_\varphi\le 2S} S^{-N}.$$

Using Proposition 4.10 and Lemma 4.19 we have

$$|\{\varphi \in \mathcal{B}_P, S \le \nu_\varphi \le 2S\}| \le S^{L_2}$$

for some absolute L_2. Once again making N sufficiently large the entire contribution is bounded by $\ll_{\Omega,N,\delta} S^{-N}$. Summing over $S \ge \varepsilon^{-1-\delta}$ in dyadic intervals we deduce the entire contribution from $\nu_\varphi \ge \varepsilon^{-1-\delta}$ is bounded by $\ll_{\Omega,N,\delta} \varepsilon^N$.

We next deal with the case when $\nu_\varphi \leq \varepsilon^{-1-\delta}$ and $\|\lambda\| \geq \varepsilon^{-1-\delta}$. Applying Equation (6.2) we see that the contribution is bounded by

$$\ll_N \sum_P \sum_{\varphi\in\mathcal{B}_P,\nu_\varphi\leq\varepsilon^{-1-\delta}} (1+\varepsilon\nu_\varphi)^{-N} \int\limits_{\lambda\in i\mathfrak{a}_P^*,\|\lambda\|\geq\varepsilon^{-1-\delta}} (1+\varepsilon\|\lambda\|)^{-N}|\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2\,\mathrm{d}\lambda.$$

Using Proposition 4.17, applying integration by parts, and making N large enough, the inner integral is bounded by $\ll_{\Omega,N,\delta} \varepsilon^N$. Therefore, the entire sum is bounded by

$$\ll_{\Omega,N,\delta} \varepsilon^N \sum_P \sum_{\varphi\in\mathcal{B}_P,\nu_\varphi\leq\varepsilon^{-1-\delta}} 1.$$

Applying Proposition 4.10 and Lemma 4.19 again, and making N sufficiently large, this is bounded by $\ll_{\Omega,N,\delta} \varepsilon^N$. $\square$

A similar proposition treats the averaged version.

Proposition 6.2. *Let $\Omega \subset \mathbb{X}_0$ be a fixed compact set. For every $\delta, \eta > 0$ we have*

$$\int\limits_\Omega \|T^*(p^{nk})K_{\varepsilon,x_0}^{\mathbb{X}_0}\|_2^2\,\mathrm{d}x_0 \ll_{\Omega,N,\eta,\delta} \sum_{\varphi\in\mathcal{B}_G,\nu_\varphi\leq\varepsilon^{-1-\delta}} p^{-kn(n-1)}|\lambda_\varphi(p^{nk})|^2$$
$$+\sum_{P\neq G} C_P \sum_{\varphi\in\mathcal{B}_P,\nu_\varphi\leq\varepsilon^{-1-\delta}} \varepsilon^{-(1+\delta+\eta)\dim\mathfrak{a}_P^*} p^{kn(2\theta_{\varphi,p}-(n-1)+\eta)} + \varepsilon^N,$$

for every $N > 0$.

Proof. We integrate both sides of the estimate in Proposition 6.1 over $x_0 \in \Omega$. Apply Proposition 4.20 to bound

$$\int\limits_\Omega \int\limits_{\lambda\in i\mathfrak{a}_P^*,\|\lambda\|\leq\varepsilon^{-1-\delta}} |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2\,\mathrm{d}\lambda\,\mathrm{d}x_0 \ll_{\Omega,\eta} \varepsilon^{-(1+\delta)(1+\eta)\dim a_P^*},$$

for $\nu_\varphi \leq \varepsilon^{-1-\delta}$. The claim follows from the fact that $\varphi \in \mathcal{B}_G$ are L^2-normalized. $\square$

7. Proof of Theorem 2 and Theorem 3

The goal of this section is to prove Theorem 2 and Theorem 3. By Lemma 5.12, to prove that $\kappa \leq \beta := \frac{n-1}{n-1-2\theta_n}$ it is sufficient to prove that for ε_0 small enough, for every $\varepsilon < \varepsilon_0$ and for $k = \lfloor \beta\frac{n+2}{2n}\log_p(\varepsilon^{-1})\rfloor$

$$\int\limits_\Omega \|T^*(p^{nk})K_{\varepsilon,x_0}^{\mathbb{X}_0})\|_2\,\mathrm{d}x_0 \ll_\eta \varepsilon^{-\eta},$$

for every $\eta > 0$.

Using Proposition 6.2 and standard modifications, it is sufficient to prove that under the same conditions that

$$p^{-kn(n-1)} \sum_{\varphi \in \mathcal{B}_G, \nu_\varphi \leq \varepsilon^{-1}} |\lambda_\varphi(p^{nk})|^2 \ll_\eta \varepsilon^{-\eta}, \tag{7.1}$$

and for every standard parabolic $P \neq G$

$$\sum_{\varphi \in \mathcal{B}_P, \nu_\varphi \leq \varepsilon^{-1}} \varepsilon^{-\dim \mathfrak{a}_P} p^{kn(2\theta_{\varphi,p} - (n-1))} \ll_\eta \varepsilon^{-\eta}, \tag{7.2}$$

for every $\eta > 0$.

7.1. The discrete spectrum

We start by handling Equation (7.1). We can further divide $\mathcal{B}_G$ according to shapes, as in Subsection 4.7. Each such shape is of the form $S = ((a, b))$, for $n = ab$. We can uniformly bound

$$\theta_{\varphi,p} \leq (b-1)/2 + \theta_a,$$

where θ_a is the best known bound towards the GRC for GL_a. So by Lemma 3.5 we have

$$|\lambda_\varphi(p^{nk})| \ll_\eta p^{nk((b-1)/2 + \theta_a + \eta)}.$$

We have $\theta_a = 0$ for $a = 1$, and by Equation (4.8) $\theta_a \leq \frac{1}{2} - \frac{1}{a^2+1}$. By Lemma 4.19, we have

$$\#\{\varphi \in \mathcal{B}_S, \nu_\varphi \leq \varepsilon^{-1}\} \ll \varepsilon^{-(a+2)(a-1)/2}.$$

Therefore, applying the above bounds we get

$$p^{-kn(n-1)} \sum_{\varphi \in \mathcal{B}_S, \nu_\varphi \leq \varepsilon^{-1}} |\lambda_\varphi(p^{nk})|^2 \ll \varepsilon^{-(a+2)(a-1)/2} p^{nk(b-1-(n-1)+2\theta_a+\eta)},$$

plugging in $k = \lfloor \beta \frac{n+2}{2n} \log_p(\varepsilon^{-1}) \rfloor$, the last value is

$$\asymp \varepsilon^{-\eta} \varepsilon^{-(a+2)(a-1)/2 - \beta(n+2)(b-1-(n-1)+2\theta_a)/2}.$$

Making η small enough it suffices to show that

$$\beta(n+2)(n-1-(b-1)-2\theta_a) - (a+2)(a-1) \geq 0. \tag{7.3}$$

For $a = 1$, $b = n$ we have $\theta_a = 0$. Hence Equation (7.3) is obvious for any $\beta \geq 1$.

For $1 < a < n$, $b = n/a$ we have $2\theta_a \le 1$. Then $n + 2 \ge a + 2$ and

$$n - b - 2\theta_a > n/2 - 1 \ge a - 1.$$

So Equation (7.3) still holds for any $\beta \ge 1$.

Finally, for $a = n, b = 1$, Equation (7.3) will hold for as long as

$$\beta \ge \frac{n-1}{n-1-2\theta_n}.$$

Now, assuming Conjecture 1 for $n \ge 4$ and using Proposition 4.13 and Proposition 4.14 for $n = 2$ and $n = 3$, respectively we can handle $a = n$, $b = 1$ for $\beta = 1$. Indeed, in this case $\mathcal{B}_S = \mathcal{B}_{n,\mathrm{cusp}}$. We have

$$p^{-kn(n-1)} \sum_{\varphi \in \mathcal{B}_S, \nu_\varphi \le \varepsilon^{-1}} |\lambda_\varphi(p^{nk})|^2 \ll_\eta p^{-kn(n-1)}(\varepsilon^{-1}p^{nk})^\eta(\varepsilon^{-d} + p^{nk(n-1)}).$$

Assuming $\beta = 1$, i.e., $k = \lfloor \frac{n+2}{2n} \log_p(\varepsilon^{-1}) \rfloor$, then

$$p^{nk(n-1)} \asymp \varepsilon^{(n+2)(n-1)/2} = \varepsilon^{-d},$$

so the above is $\ll_\eta \varepsilon^{-\eta}$, as needed.

7.2. The continuous spectrum

In this subsection we prove Equation (7.2) for every $\beta \ge 1$, which is enough for our purpose.

We further divide into shapes, as in Subsection 4.7. Let $S = ((a_1, b_1), \ldots, (a_r, b_r))$ be a shape, and let $\mathcal{B}_S \subset \mathcal{B}_P$. Notice that in this case $\dim \mathfrak{a}_P = r - 1$. Since we assume that $P \ne G$, we assume that $r > 1$.

We need to prove that for every shape S, it holds that

$$\sum_{\varphi \in \mathcal{B}_S, \nu_\varphi \le \varepsilon^{-1}} \varepsilon^{-(r-1)} p^{kn(2\theta_{\varphi,p} - (n-1))} \ll_\eta \varepsilon^{-\eta}.$$

Without loss of generality, we assume that it holds that

$$\max_i \{(b_i - 1)/2 + \theta_{a_i}\} = (b_1 - 1)/2 + \theta_{a_1}.$$

Then by Lemma 4.18 we have $\theta_{\varphi,p} \le (b_1 - 1)/2 + \theta_{a_1}$. Using Lemma 4.19, we deduce

$$\sum_{\varphi \in \mathcal{B}_S, \nu_\varphi \le \varepsilon^{-1}} \varepsilon^{-(r-1)} p^{kn(2\theta_{\varphi,p} - (n-1))} \ll \varepsilon^{-(r-1) - \sum_{i=1}^r (a_i+2)(a_i-1)/2} p^{kn(b_1 - 1 + 2\theta_{a_1} - (n-1))}.$$

We plug in $k = \lfloor \frac{n+2}{2n} \log_p(\varepsilon^{-1}) \rfloor$, and deduce that it is sufficient to prove that for every shape S it holds that

$$-(r-1) - \sum_{i=1}^{r} (a_i+2)(a_i-1)/2 + \frac{n+2}{2}((n-1)-(b_1-1)-2\theta_{a_1}) \geq 0.$$

We start by noticing that

$$r + \sum_{i=1}^{r} (a_i+2)(a_i-1)/2 = \sum_{i=1}^{r} ((a_i+2)(a_i-1)/2+1) = \sum_{i=1}^{r} a_i(a_i+1)/2.$$

We have the following simple lemma

Lemma 7.1. *Assume that* $\sum_{i=2}^{r} a_i \leq M$. *Then*

$$\sum_{i=2}^{r} a_i(a_i+1)/2 \leq M(M+1)/2.$$

Proof. The polynomial $p(x) = x(x+1)/2$ satisfies for $x_1, x_2 \geq 0$ that $p(x_1) + p(x_2) \leq p(x_1 + x_2)$. The lemma follows. $\square$

In our case, we have $\sum_{i=2}^{r} a_i \leq n - a_1 b_1$. So we deduce that

$$\sum_{i=1}^{r} a_i(a_i+1)/2 \leq (n-a_1b_1)(n-a_1b_1+1)/2.$$

We deduce that it is sufficient to prove the following.

Lemma 7.2. *Denote*

$$F(a,b,n) := 2 + (n+2)(n-b-2\theta_a) - a(a+1) - (n-ab)(n-ab+1).$$

Then for every positive integers $n \geq 2$ *and* a, b *such that* $ab < n$ *it holds that* $F(a,b,n) \geq 0$.

Proof. We show by case-by-case analysis. First, it is easy to see that the claim holds for $n = 2$, since then $a = b = 1$ and $\theta_a = 0$.

Now, assume that $a = 1$. Then $\theta_a = 0$, and it holds that

$$F(1,b,n) = 2 + (n+2)(n-b) - 2 - (n-b)(n-b+1) = (n-b)(n+2-(n-b+1)) \geq 0.$$

Next, assume that $a > 1$. We first take care of the case $n = 3$. Then we only need to consider $a = 2, b = 1$ case. It holds that

$$F(2,1,3) = 2 + 5(2 - 2\theta_2) - 6 - 2 = 4 - 10\theta_2.$$

Plugging in the Kim–Sarnak's bound $\theta_2 \leq \frac{7}{64}$ we get the desired result.

Now assume that $a > 1$ and $n \geq 4$. We will use the bound $\theta_a \leq 1/2$ which follows from Equation (4.8). Then

$$F(a,b,n) \geq G(a,b,n) := 2 + (n+2)(n-b-1) - a(a+1) - (n-ab)(n-ab+1).$$

First consider the case $b = 1$. Then

$$G(a,1,n) = 2 + (n+2)(n-2) - a(a+1) - (n-a)(n-a+1).$$

The values of $G(a,1,n)$, when n is fixed and a varies, lie on a parabola with negative leading coefficient. To prove lower bounds in the range $2 \leq a \leq n-1$ it is sufficient to check the extreme values, namely $a = 1$ and $a = n-1$.

We see that

$$G(n-1,1,n) = G(1,1,n) = 2+(n+2)(n-2)-2-(n-1)n = n^2-4-n^2+n = n-4 \geq 0,$$

as we assumed that $n \geq 4$.

Finally, we are left with the case $a \geq 2, b \geq 2$. In this case we have $n \geq ab+1 \geq 5$ and $a \leq (n-1)/2$. Fix a and n. Then $G(a,b,n)$ as a function of b is again a parabola with negative leading coefficient. So it suffices to check the extreme cases $b = 1$ and $b = (n-1)/a$ We already proved that $G(a,1,n) \geq 0$, so we are left to show that for $2 \leq a \leq (n-1)/2$,

$$G(a,(n-1)/a,n) \geq 0.$$

Indeed, we get

$$G(a,(n-1)/a,n) = 2 + (n+2)(n-(n-1)/a-1) - a(a+1) - 2.$$

Using $(n-1)/a \leq (n-1)/2$ and $a \leq (n-1)/2$, we get

$$\begin{aligned} G(a,(n-1)/a,n) &\geq (n+2)((n-1)/2-1) - (n-1)(n+1)/4 \\ &= \frac{2(n+2)(n-3)-(n^2-1)}{4} = \frac{n^2-2n-11}{4}. \end{aligned}$$

The last value is non-negative since $n \geq 5$. □

8. Proof of Theorem 4

We start the proof for n general, assuming the GRC for GL_m for all $m \leq n$. The proof is similar to the last section, but we use Lemma 5.11 instead of Lemma 5.12. It is therefore sufficient to show that for every $x_0 \in \mathbb{X}_0$, for $\varepsilon < \varepsilon_0$, for $k = \lfloor \frac{n+2}{2n} \log_p(\varepsilon^{-1}) \rfloor$, it holds that

$$\|T^*(p^{nk}) K^{\mathbb{X}_0}_{\varepsilon,x_0}\|_2 \ll_\eta \varepsilon^{-\eta}.$$

We choose a compact subset Ω which contains x_0. Using Proposition 6.1, we reduce to proving that

$$\sum_P C_P \sum_{\varphi \in \mathcal{B}_P, \nu_\varphi \leq \varepsilon^{-1}} p^{kn(2\theta_{\varphi,p} - (n-1))} \int\limits_{\lambda \in i\mathfrak{a}^*_P, \|\lambda\| \leq \varepsilon^{-1}} |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \, \mathrm{d}\lambda \ll \varepsilon^{-\eta}. \tag{8.1}$$

We can then further divide the above sum into shapes as done in Section 7. The basic observation is that for a shape $S = ((a_1, b_1), \ldots, (a_r, b_r))$, if $b_i = 1$ for all i (i.e., the shape is *cuspidal*), then by Lemma 4.18 for every $\varphi \in \mathcal{B}_S$ the GRC implies that $\theta_{\varphi,p} = 0$. Therefore, for cuspidal S, assuming GRC,

$$\begin{aligned} &\sum_{\varphi \in \mathcal{B}_S, \nu_\varphi \leq \varepsilon^{-1}} p^{kn(2\theta_{\varphi,p} - (n-1))} \int\limits_{\lambda \in i\mathfrak{a}^*_P, \|\lambda\| \leq \varepsilon^{-1}} |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \, \mathrm{d}\lambda \\ =& p^{-kn(n-1)} \sum_{\varphi \in \mathcal{B}_S, \nu_\varphi \leq \varepsilon^{-1}} \int\limits_{\lambda \in i\mathfrak{a}^*_P, \|\lambda\| \leq \varepsilon^{-1}} |\mathrm{Eis}_P(\varphi,\lambda)(x_0)|^2 \, \mathrm{d}\lambda \end{aligned}$$

Using Proposition 4.17, the last value is

$$\ll_\Omega p^{-kn(n-1)} \varepsilon^{-d},$$

and plugging in the value of k the last value is $\ll 1$.

We, therefore, deduce that assuming the GRC, we are only left with non-cuspidal shapes to deal with.

For $S = ((1,n))$, the only representation $\varphi \in \mathcal{B}_S$ is the constant L^2-normalized function φ_0. In this case $\theta_{\varphi_0,p} = (n-1)/2$, and its contribution is

$$p^{kn(2\theta_{\varphi_0,p} - (n-1))} \varphi_0(x_0) \ll 1.$$

Now, let us fix $n = 3$. In this case, the only non-cuspidal shapes are $S = ((1,3))$ and $S = ((1,2),(1,1))$. Following the above, we are left with the shape $S = ((1,2),(1,1))$. In this case there is exactly one element $\varphi_1 \in \mathcal{B}_S$ with $\theta_{\varphi_1,p} = 1/2$. We have the following result.

Proposition 8.1. *For every $\lambda \in i\mathfrak{a}^*_P$ and every $\eta > 0$, it holds that*

$$|\mathrm{Eis}_P(\varphi_1,\lambda)(x_0)| \ll_{\eta,x_0} \lambda^{3/4+\eta}.$$

The result follows from the functional equation of $\mathrm{Eis}_P(\varphi_1, \lambda)$, standard bounds of the Riemann ξ-function, and the Phragmén–Lindelöf convexity principle. This is explained in [5]. As a matter of fact, [5, Theorem 1] proves a stronger and far deeper result where the exponent is $1/2$ instead of $3/4$.

Now, using Proposition 8.1, we deduce that

$$\sum_{\varphi \in \mathcal{B}_S, \nu_\varphi \leq \varepsilon^{-1}} p^{3k(2\theta_{\varphi,p}-2)} \int_{\lambda \in i\mathfrak{a}_P^*, \|\lambda\| \leq \varepsilon^{-1}} |\mathrm{Eis}_P(\varphi, \lambda)(x_0)|^2 \, \mathrm{d}\lambda \ll_{\eta, x_0} p^{-3k} \varepsilon^{-5/2-\eta}.$$

In this case $p^k \asymp \varepsilon^{-5/6}$, so we conclude.

Remark 8.2. The argument above using the local L^∞-bound of the maximal degenerate Eisenstein series extends to the shapes of the form $S = ((1, n-1), (1, 1))$ for any n. However, for $n = 4$ we do not know how to handle shapes of the form $S = ((1, 2), (1, 2))$ or $S = ((1, 2), (2, 1))$ or $S = ((2, 2))$. In all cases, we need good uniform bounds for $\mathrm{Eis}_P(\varphi, \lambda)(x_0)$.

Remark 8.3. Without assuming the GRC we do not know, even for $n = 2$, whether $\kappa(x_0) = 1$ for all $x_0 \in \mathbb{X}$. The problem reduces to the following local version of Sarnak's density conjecture, which we state for general n: for every $x_0 \in \mathbb{X}$ and $l \geq 0$, $T \geq 1$

$$\sum_{\varphi \in \mathcal{F}_T} \left|\lambda_\varphi(p^l)\right|^2 |\varphi(x_0)|^2 \ll_{\delta, p, x_0} \left(Tp^l\right)^\delta \left(T^d + p^{l(n-1)}\right),$$

for every $\delta > 0$. This version is open even for $n = 2$ but can be proven for some x_0 using different methods. This result will appear elsewhere.

References

[1] J. Arthur, An introduction to the trace formula, in: Harmonic Analysis, the Trace Formula, and Shimura Varieties, in: Clay Math. Proc., vol. 4, Amer. Math. Soc., Providence, RI, 2005, pp. 1–263.

[2] E. Assing, V. Blomer, The density conjecture for principal congruence subgroups, arXiv preprint, arXiv:2204.08868, 2022. To appear in Duke Mathematical Journal.

[3] V. Blomer, Applications of the Kuznetsov formula on GL(3), Invent. Math. 194 (3) (2013) 673–729.

[4] V. Blomer, Density theorems for GL(n), Invent. Math. 232 (2) (2023) 683–711, https://doi.org/10.1007/s00222-022-01172-3.

[5] V. Blomer, Epstein zeta-functions, subconvexity, and the purity conjecture, J. Inst. Math. Jussieu 19 (2) (2020) 581–596.

[6] V. Blomer, F. Brumley, On the Ramanujan conjecture over number fields, Ann. Math. (2) 174 (1) (2011) 581–605.

[7] V. Blomer, J. Buttcane, N. Raulf, A Sato-Tate law for GL(3), Comment. Math. Helv. 89 (4) (2014) 895–919.

[8] V. Blomer, P. Maga, Subconvexity for sup-norms of cusp forms on PGL(n), Sel. Math. New Ser. 22 (3) (2016) 1269–1287.

[9] V. Blomer, S.H. Man, Bounds for Kloosterman sums on GL(n), arXiv preprint, arXiv:2208.01295, 2022. To appear in Mathematische Annalen.

[10] C. Bordenave, H. Lacoin, Cutoff at the entropic time for random walks on covered expander graphs, J. Inst. Math. Jussieu (2021) 1–46.

[11] T.D. Browning, V.V. Kumaraswamy, R.S. Steiner, Twisted Linnik implies optimal covering exponent for S^3, Int. Math. Res. Not. 1 (2019) 140–164.

[12] F. Bruhat, J. Tits, Groupes réductifs sur un corps local, Publ. Math. Inst. Hautes Études Sci. 41 (1972) 5–251.

[13] J. Buttcane, F. Zhou, Plancherel distribution of Satake parameters of Maass cusp forms on GL_3, Int. Math. Res. Not. 5 (2020) 1417–1444.

[14] P. Chiu, Covering with Hecke points, J. Number Theory 53 (1) (1995) 25–44.

[15] L. Clozel, H. Oh, E. Ullmo, Hecke operators and equidistribution of Hecke points, Invent. Math. 144 (2) (2001) 327–351.

[16] H. Donnelly, On the cuspidal spectrum for finite volume symmetric spaces, J. Differ. Geom. 17 (2) (1982) 239–253.

[17] J.J. Duistermaat, J.A.C. Kolk, V.S. Varadarajan, Spectra of compact locally symmetric manifolds of negative curvature, Invent. Math. 52 (1) (1979) 27–93.

[18] D. Flath, Decomposition of representations into tensor products, in: Automorphic Forms, Representations and L-Functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1, in: Proc. Sympos. Pure Math., vol. XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 179–183.

[19] A. Ghosh, A. Gorodnik, A. Nevo, Diophantine approximation and automorphic spectrum, Int. Math. Res. Not. 21 (2013) 5002–5058.

[20] A. Ghosh, A. Gorodnik, A. Nevo, Metric Diophantine approximation on homogeneous varieties, Compos. Math. 150 (8) (2014) 1435–1456.

[21] A. Ghosh, A. Gorodnik, A. Nevo, Diophantine approximation exponents on homogeneous varieties, in: Recent Trends in Ergodic Theory and Dynamical Systems, in: Contemp. Math., vol. 631, Amer. Math. Soc., Providence, RI, 2015, pp. 181–200.

[22] A. Ghosh, A. Gorodnik, A. Nevo, Best possible rates of distribution of dense lattice orbits in homogeneous spaces, J. Reine Angew. Math. 745 (2018) 155–188.

[23] D. Goldfeld, Automorphic Forms and L-Functions for the Group $\mathrm{GL}(n, \mathbb{R})$, Cambridge Studies in Advanced Mathematics, vol. 99, Cambridge University Press, Cambridge, 2006. With an appendix by Kevin A. Broughan.

[24] K. Golubev, A. Kamber, Cutoff on hyperbolic surfaces, Geom. Dedic. 203 (2019) 225–255.

[25] K. Golubev, A. Kamber, On Sarnak's density conjecture and its applications, Forum Math. Sigma 11 (2023) 1–51, Paper No. e48, https://doi.org/10.1017/fms.2023.40.

[26] S. Jana, Applications of analytic newvectors for $\mathrm{GL}(n)$, Math. Ann. 380 (3–4) (2021) 915–952.

[27] S. Jana, A. Kamber, On the local L^2-bound of the Eisenstein series, arXiv preprint, arXiv:2210.16291, 2022.

[28] H.H. Kim, P. Sarnak, Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2, appendix 2, J. Am. Math. Soc. 16 (1) (2003) 139–183.

[29] D. Kleinbock, K. Merrill, Rational approximation on spheres, Isr. J. Math. 209 (1) (2015) 293–322.

[30] E. Lindenstrauss, A. Venkatesh, Existence and Weyl's law for spherical cusp forms, Geom. Funct. Anal. 17 (1) (2007) 220–251.

[31] W. Luo, Z. Rudnick, P. Sarnak, On the generalized Ramanujan conjecture for $\mathrm{GL}(n)$, in: Automorphic Forms, Automorphic Representations, and Arithmetic, Fort Worth, TX, 1996, in: Proc. Sympos. Pure Math., vol. 66, Amer. Math. Soc., Providence, RI, 1999, pp. 301–310.

[32] S.D. Miller, On the existence and temperedness of cusp forms for $\mathrm{SL}_3(\mathbb{Z})$, J. Reine Angew. Math. 533 (2001) 127–169.

[33] C. Mœglin, J.-L. Waldspurger, Le spectre résiduel de $\mathrm{GL}(n)$, Ann. Sci. Éc. Norm. Supér. (4) 22 (4) (1989) 605–674.

[34] C. Mœglin, J.-L. Waldspurger, Spectral Decomposition and Eisenstein Series, Cambridge Tracts in Mathematics, vol. 113, Cambridge University Press, Cambridge, 1995. Une paraphrase de l'Écriture [A paraphrase of Scripture].

[35] W. Müller, Weyl's law for the cuspidal spectrum of SL_n, Ann. Math. (2) 165 (1) (2007) 275–333.

[36] H. Oh, Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants, Duke Math. J. 113 (1) (2002) 133–192.

[37] O. Parzanchevski, P. Sarnak, Super-Golden-Gates for PU(2), Adv. Math. 327 (2018) 869–901.

[38] V. Platonov, A. Rapinchuk, Algebraic Groups and Number Theory, Pure and Applied Mathematics, vol. 139, Academic Press, Inc., Boston, MA, 1994. Translated from the 1991 Russian original by Rachel Rowen.

[39] N.T. Sardari, The Siegel variance formula for quadratic forms, arXiv preprint, arXiv:1904.08041, 2019.
[40] P. Sarnak, Notes on the generalized Ramanujan conjectures, in: Harmonic Analysis, the Trace Formula, and Shimura Varieties, in: Clay Math. Proc., vol. 4, Amer. Math. Soc., Providence, RI, 2005, pp. 659–685.
[41] P. Sarnak, Letter to Stephen D. Miller and Naser Talebizadeh Sardari on optimal strong approximation by integral points on quadrics, available at http://publications.ias.edu/sarnak, 2015.
[42] P. Sarnak, X.X. Xue, Bounds for multiplicities of automorphic representations, Duke Math. J. 64 (1) (1991) 207–227.
[43] P.C. Sarnak, Diophantine problems and linear groups, in: Proceedings of the International Congress of Mathematicians, vol. I, II, Kyoto, 1990, Math. Soc. Japan, Tokyo, 1991, pp. 459–471.
[44] J.A. Shalika, The multiplicity one theorem for GL_n, Ann. Math. (2) 100 (1974) 171–193.
[45] A.J. Silberger, Introduction to Harmonic Analysis on Reductive p-Adic Groups: Based on Lectures by Harish-Chandra at the Institute for Advanced Study, 1971–1973, Mathematical Notes, vol. 23, Princeton University Press/University of Tokyo Press, Princeton, N.J./Tokyo, 1979.