

RESEARCH ARTICLE

# On Sarnak’s Density Conjecture and Its Applications

Konstantin Golubev <sup>1</sup> and Amitay Kamber <sup>2</sup>

<sup>1</sup>ETH, Ramistrasse 101 Zurich, 8092, Switzerland; E-mail: [k.golubev@gmail.com](mailto:k.golubev@gmail.com).

<sup>2</sup>Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK;  
E-mail: [ak2356@dpmms.cam.ac.uk](mailto:ak2356@dpmms.cam.ac.uk).

**Received:** 25 January 2022; **Accepted:** 29 April 2023

**2020 Mathematics Subject Classification:** *Primary* – 11F99; *Secondary* – 22E50

## Abstract

Sarnak’s density conjecture is an explicit bound on the multiplicities of nontempered representations in a sequence of cocompact congruence arithmetic lattices in a semisimple Lie group, which is motivated by the work of Sarnak and Xue ([58]). The goal of this work is to discuss similar hypotheses, their interrelation and their applications. We mainly focus on two properties – the spectral spherical density hypothesis and the geometric Weak injective radius property. Our results are strongest in the  $p$ -adic case, where we show that the two properties are equivalent, and both imply Sarnak’s general density hypothesis. One possible application is that either the spherical density hypothesis or the Weak injective radius property imply Sarnak’s optimal lifting property ([57]). Conjecturally, all those properties should hold in great generality. We hope that this work will motivate their proofs in new cases.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	The relations between the different properties . . . . .	5
<b>2</b>	<b>Applications and open problems</b>	<b>6</b>
2.1	Ramanujan graphs and complexes . . . . .	6
2.2	Adding a parameter to the properties . . . . .	7
2.3	Congruence subgroups of arithmetic groups . . . . .	7
2.4	The Work of Sarnak and Xue and its implications . . . . .	10
2.5	On Some congruence subgroups of $SL_3(\mathbb{Z})$ . . . . .	10
2.6	The weak injective radius of principal congruence subgroups of $SL_n(\mathbb{Z})$ . . . . .	11
<b>3</b>	<b>Main ideas of the proofs</b>	<b>11</b>
<b>4</b>	<b>Preliminaries</b>	<b>14</b>
4.1	Distances and length of elements . . . . .	14
4.2	Growth of matrix coefficients and Harish-Chandra’s bounds . . . . .	17
4.3	Upper bounds on operators . . . . .	20
4.4	Spherical Functions . . . . .	21
4.5	Convolution of Operators . . . . .	23
4.6	Traces of operators on irreducible unitary representations . . . . .	24
4.7	The pretrace formula . . . . .	24
4.8	Spectral decomposition of a characteristic function of a small ball . . . . .	25
<b>5</b>	<b>The weak injective radius property</b>	<b>26</b>

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

<b>6</b>	<b>The weak injective radius property implies the optimal lifting property</b>	<b>31</b>
6.1	Reduction to a spectral argument	31
6.2	Completing the proof of Theorem 1.5	33
<b>7</b>	<b>The spectral to geometric direction</b>	<b>34</b>
7.1	Some technical calculations	34
7.2	Proof of Theorem 1.6	36
7.3	A strong version of Theorem 1.5, assuming the spherical density hypothesis	38
<b>8</b>	<b>The Bernstein theory of nonbacktracking operators</b>	<b>39</b>
8.1	Nonbacktracking operators and temperedness	41
<b>9</b>	<b>Lower Bounds on Matrix Coefficients for a Specific Representation</b>	<b>42</b>
9.1	Leading exponents	44
9.2	Some technical lemmas	45

**1. Introduction**

Let  $k$  be a local field (Archimedean or non-Archimedean), let  $G$  be the  $k$ -rational points of a semisimple algebraic group defined over  $k$ , let  $\Gamma_1 \subset G$  be a lattice and let  $(\Gamma_N)$  be a sequence of finite index subgroups of  $\Gamma_1$  with  $[\Gamma_1 : \Gamma_N] \rightarrow \infty$ . There are various results about the multiplicities in the decomposition of  $L^2(\Gamma_N \backslash G)$  into irreducible representations (e.g., [16, 59, 1]). An extremely strong property is the *very naive Ramanujan property*, stating that if  $\pi$  is nontempered and nontrivial, then it does not appear in the decomposition. However, the *very naive Ramanujan property* is usually not true in high rank (see, e.g., [9]). Notice that we do not make a distinction between cusp forms and noncusp forms – the *naive Ramanujan conjecture* states that cusp forms are tempered, and it is also not true in general ([34]). Moreover, even when the naive Ramanujan conjecture is expected to be true, it is usually out of reach by the existing methods.

Recently, Sarnak made a density conjecture which approximates the very naive Ramanujan property and should serve as a replacement for it for applications. Some instances of this general idea were previously given for hyperbolic surfaces ([57, 28]), and for graphs ([7, 39]). Our goal here is to give a general framework for similar density conjectures and their use in applications.

We give a geometric and somewhat elementary approach to the problem. An alternative approach based on deep results in the Langlands program may be found in an ongoing work of Shai Evra.

To state Sarnak’s density conjecture, we first set some notations. Let  $(\pi, V)$  be a unitary irreducible representation of  $G$ , and let  $0 \neq v_0 \in V$  be a  $K$ -finite vector, where  $K$  is a maximal compact subgroup of  $G$ , which is good in the sense of Bruhat and Tits in the non-Archimedean case ([8]). We let  $1 \leq p(\pi) \leq \infty$  be the infimum over  $p \geq 1$  such that the matrix coefficient  $\beta : G \rightarrow \mathbb{C}, \beta(g) = \langle v_0, \pi(g)v_0 \rangle$  is in  $L^p(G)$ . It is a simple fact that  $p(\pi)$  does not depend on the choice of  $v_0$ . Let  $\Pi(G)$  be the set of isomorphism classes of irreducible unitary representations of  $G$  endowed with the Fell topology. For a cocompact lattice  $\Gamma$  and  $(\pi, V) \in \Pi(G)$ , denote  $m(\pi, \Gamma) = \dim \text{Hom}_G(V, L^2(\Gamma \backslash G))$ , that is, the multiplicity of  $\pi$  in the decomposition of  $L^2(\Gamma \backslash G)$  into irreducible representations. For a subset  $A \subset \Pi(G)$ , denote  $M(A, \Gamma, p) = \sum_{\pi \in A, p(\pi) \geq p} m(\pi, \Gamma)$ .

**Conjecture 1.1** (Sarnak’s density conjecture). *Let  $G$  be a real, semisimple, almost-simple and simply connected Lie group, let  $\Gamma_1$  be a cocompact arithmetic lattice of  $G$  and let  $(\Gamma_N)$  be a sequence of finite index congruence subgroups of  $\Gamma_1$ , with  $[\Gamma_1 : \Gamma_N] \rightarrow \infty$ . Then for every precompact subset  $A \subset \Pi(G)$  and  $\epsilon > 0$  there exists a constant  $C_{\epsilon, A}$  such that for every  $N$  and  $p > 2$ ,*

$$M(A, \Gamma, p) \leq C_{A, \epsilon} [\Gamma_1 : \Gamma_N]^{2/p + \epsilon}.$$

We refer to a sequence of lattices satisfying this multiplicity property as a sequence that satisfies the *general density hypothesis*. A similar conjecture appeared in the work of Sarnak and Xue ([58]), but they only considered the case when  $A = \{\pi\}$  is a singleton. In such a case, we say that the sequence of lattices satisfies the *pointwise multiplicity hypothesis*.

We will prefer to work with a different spectral definition, the *spherical density hypothesis*, which is easier to use for applications, and concerns only spherical representations. Let  $\Pi(G)_{\text{sph}} \subset \Pi(G)$  be the set of isomorphism classes of spherical representations, that is, of irreducible unitary representations with a nonzero  $K$ -invariant vector. In the  $p$ -adic case, or the rank 1 case, the set  $\{\pi \in \Pi(G)_{\text{sph}} : p(\pi) > 2\}$  is precompact, and the spherical density hypothesis is simply the case when  $A = \{\pi \in \Pi(G)_{\text{sph}} : p(\pi) > 2\}$  in the general density hypothesis. When  $G$  is Archimedean of high rank,  $\{\pi \in \Pi(G)_{\text{sph}} : p(\pi) > 2\}$  is not necessarily precompact, so we associate to a spherical representation  $(\pi, V) \in \Pi(G)$  a number  $\lambda(\pi) \in \mathbb{R}_{\geq 0}$ , which is the eigenvalue of the Casimir operator on the  $K$ -invariant subspace of  $\pi$ , and define:

**Definition 1.2.** The sequence  $(\Gamma_N)$  of cocompact lattices satisfies the *spherical density hypothesis* if:

- o In the  $p$ -adic or rank 1 case, for every  $\epsilon > 0$  there exists  $C_\epsilon$  such that for every  $N \geq 1, p > 2$ ,

$$M(\Pi(G)_{\text{sph}}, \Gamma_N, p) \leq C_\epsilon [\Gamma_1 : \Gamma_N]^{2/p+\epsilon}.$$

- o In the general Archimedean case, there exists  $L > 0$  large enough, such that for every  $\epsilon > 0$  there exists  $C_\epsilon$  such that for every  $\lambda \geq 0, N \geq 1, p > 2$ ,

$$M(\{\pi \in \Pi(G)_{\text{sph}} : \lambda(\pi) \leq \lambda\}, \Gamma_N, p) \leq C_\epsilon (1 + \lambda)^L [\Gamma_1 : \Gamma_N]^{2/p+\epsilon}.$$

In the second case, the spherical density hypothesis does not a priori follow from the general density hypothesis.

To state our main geometric definition, we need to set some more notations. We view  $G, \Gamma_1$  and the sequence  $(\Gamma_N)$  as fixed. We use the standard  $O, \Theta, o$  notations, where, for example,  $f(N, \epsilon) = O_\epsilon(g(N, \epsilon))$  says that for every  $\epsilon$  there exists  $C$  depending only on  $\epsilon$  (and on  $G, \Gamma_1, (\Gamma_N)$ ) such that  $f(N, \epsilon) \leq Cg(N, \epsilon)$  for  $N$  large enough. The notation  $f(N, \epsilon) \ll_\epsilon g(N, \epsilon)$  is the same as  $f(N, \epsilon) = O_\epsilon(g(N, \epsilon))$  and  $f(N, \epsilon) \asymp_\epsilon g(N, \epsilon)$  is the same as  $f(N, \epsilon) \ll_\epsilon g(N, \epsilon)$  and  $g(N, \epsilon) \ll_\epsilon f(N, \epsilon)$ .

We fix a Cartan decomposition  $G = KA_+K$  and an Iwasawa decomposition  $G = KP$ . Let  $\delta : P \rightarrow \mathbb{R}_{>0}$  be the left modular character of  $P$  (see Section 4 for more details).

We define a length  $l : G \rightarrow \mathbb{R}_{\geq 0}$  by first denoting for  $a \in A_+, l(a) = \log_q \delta(a)$ , where  $q$  is equal to  $e$  in the Archimedean case and is equal to the size of the quotient field of  $k$  otherwise. Then we extend  $l$  to  $G$  using the Cartan decomposition, that is,  $l(k_1ak_2) = l(a)$ . Finally, we define a metric on  $G/K$  by  $d(x, y) = l(x^{-1}y)$ .

The weak injective radius property is based on the lattice point counting approach of Sarnak and Xue ([58, Conjecture 2]). Given an element  $y = \Gamma_N y \in \Gamma_N \backslash \Gamma_1$ , we denote

$$N(\Gamma_N, d_0, y) = \#\{\gamma \in \Gamma_N : l(y^{-1}\gamma y) \leq d_0\}.$$

**Definition 1.3.** The sequence  $(\Gamma_N)$  satisfies the *weak injective radius property* if for every  $0 \leq d_0 \leq 2 \log_q([\Gamma_1 : \Gamma_N]), \epsilon > 0$ ,

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{y \in \Gamma_N \backslash \Gamma_1} N(\Gamma_N, d_0, y) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon q^{d_0/2}.$$

This definition is somewhat different from [58, Conjecture 2]. For rank 1, it is slightly weaker (see Proposition 5.3), while for higher rank we use a different length. In this form, the weak injective radius property follows from the spherical density hypothesis – see Theorem 1.6 below. On the other hand, the weak injective radius property also makes sense for nonuniform lattices.

We can now state our intended application. First, we say that a sequence of lattices  $(\Gamma_N)$  has a spectral gap if there exists  $p_0 < \infty$  such that  $p(\pi) \leq p_0$  for every nontrivial spherical  $\pi \in \Pi(G)$  weakly contained in  $L^2(\Gamma_N \backslash G)$ . This definition can be applied to the nonuniform case as well, and in the cocompact case we may replace *weakly contained* by  $m(\pi, \Gamma) > 0$ .

We look at the natural action  $\pi_N : \Gamma_1 \rightarrow \text{Aut}(\Gamma_N \backslash \Gamma_1)$ , defined by  $\pi_N(\gamma)(\Gamma_N \gamma') = \Gamma_N \gamma' \gamma^{-1}$ . Given  $x, y \in \Gamma_N \backslash \Gamma_1$ , we look for a *small* element  $\gamma \in \Gamma_1$  such that  $\pi_N(\gamma)x = y$ . A very general way of measuring how small is an element is by the Cartan decomposition,  $G = KA_+K$ . For  $\gamma \in \Gamma \subset G$  we let  $a_\gamma$  be the element in  $A_+$  in the Cartan decomposition of  $\gamma$ . We also fix some norm  $\|\cdot\|_a$  on the underlying coroot space of  $A_+$ . By [19], the number of  $\gamma \in \Gamma_1$  with  $\|a_\gamma - a\|_a < \delta$  is  $\asymp_{\Gamma_1} q^{l(a)}$ . Therefore, the following definition is optimal.

**Definition 1.4.** The sequence  $(\Gamma_N)$  has the *optimal lifting property* if for every  $\epsilon > 0$ , for every  $a \in A_+$  with  $l(a) \geq (1 + \epsilon) \log_q([\Gamma_1 : \Gamma_N])$ ,

$$\#\{(x, y) \in (\Gamma_N \backslash \Gamma_1)^2 : \exists \gamma \in \Gamma_1 \text{ s.t. } \pi_N(\gamma)x = y, \|a_\gamma - a\|_a < \epsilon \|a\|_a\} = (1 - o_\epsilon(1))[\Gamma_1 : \Gamma_N]^2.$$

Conjecturally, every sequence of congruence subgroups of an arithmetic lattice in an almost-simple and simply connected Lie group satisfies the optimal lifting property. We refer to Conjecture 2.4 for a full statement.

The following two theorems show that our two main properties imply the optimal lifting property.

**Theorem 1.5.** *Let  $(\Gamma_N)$  be a sequence of lattices having a spectral gap and satisfying the weak injective radius property. Then the sequence  $(\Gamma_N)$  has the optimal lifting property.*

**Theorem 1.6.** *Let  $(\Gamma_N)$  be a sequence of cocompact lattices satisfying the spherical density hypothesis. Then  $(\Gamma_N)$  satisfies the weak injective radius property. Therefore, assuming also spectral gap, the sequence also has the optimal lifting property.*

The definition of optimal lifting is based on the main result in an influential letter of Sarnak ([57]), who proved the optimal lifting property for principal congruence subgroups of  $\text{SL}_2(\mathbb{Z})$ , by utilizing a version of the spherical density hypothesis proved by Huxley ([35]).

In the cocompact case, one may relate the optimal lifting property to almost-diameter of the quotient space as follows. If we give  $X_N = \Gamma_N \backslash G/K$  the quotient metric  $d$ , the optimal lifting property implies that the distances between the points of  $X_N$  are concentrated at the optimal location  $\log_q(\mu(X_N))$ , that is, for every  $\epsilon > 0$ ,

$$\mu\left(\{(x, y) \in X_N \times X_N : d(x, y) < (1 + \epsilon) \log_q([\Gamma_1 : \Gamma_N])\}\right) = (1 - o_\epsilon(1))\mu^2(X_N).$$

This concentration of distances phenomena was proven for Ramanujan graphs by Sardari ([56]) and Lubetzky–Peres ([48]), who also related it to the cutoff phenomena. In higher dimensions, similar results for Ramanujan complexes appear in [38, 47]. Theorem 1.5 implies that one may get results that are almost as strong, as long as we assume only the far weaker spherical density hypothesis or the injective radius property.

The weak injective radius property is intended as the arithmetic, or geometric, input to our approach, and we discuss it further in Section 2. There are a few cases where it is known, most notably, following the work of Sarnak and Xue, for principal congruence subgroups of arithmetic lattices in  $\text{SL}_2(\mathbb{R})$  and  $\text{SL}_2(\mathbb{C})$  (see Subsection 2.4). In a companion paper by the second named author and Hagai Lavner, the weak injective radius property is proven for some nonprincipal congruence subgroups of  $\text{SL}_3(\mathbb{Z})$ , and this result is closely related to the works [6, 5] (see Subsection 2.5 for a full discussion). If we allow ourselves to relax the definition and add a parameter  $0 < \alpha \leq 1$  to the definition of the weak injective radius property (see Subsection 2.2), then it is quite straightforward to show that *principal* congruence subgroups of arithmetic groups satisfy the weak injective radius property with some explicit parameter  $\alpha > 0$  (see Corollary 2.2). As a matter of fact, recent results in [1, 22] show that every sequence of congruence subgroups satisfies the weak injective radius property with some explicit parameter  $\alpha > 0$  (see Theorem 2.3). However, one must have  $\alpha = 1$  for the optimal lifting application.

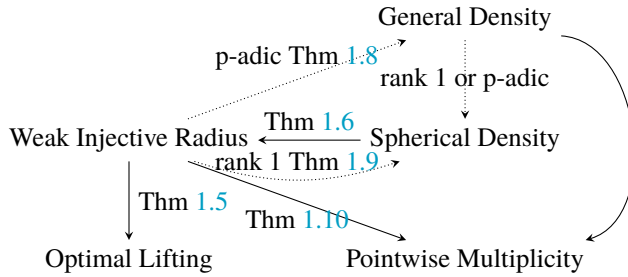


Figure 1.1. The relations between our main properties for a sequence of cocompact lattices.

1.1. The relations between the different properties

We already stated that the spherical density hypothesis implies the weak injective radius property. For the deduction of spectral results from the weak injective radius property, we have partial results in the Archimedean case and full results in the  $p$ -adic case.

We believe that the following is true:

**Conjecture 1.7.** *The weak injective radius property implies both the general density hypothesis and the spherical density hypothesis.*

In the  $p$ -adic case, one can choose larger and larger sets covering  $\Pi(G)$  as follows. For a compact open subgroup  $K'$  of  $G$ , let  $\Pi(G)_{K'-\text{sph}}$  be the set of isomorphism classes of irreducible unitary representations with a nonzero  $K'$ -invariant vector. If we have a sequence  $(K'_m)$  of arbitrarily small compact open subgroups (i.e., they generate the topology of  $G$  near the identity), then we have

$$\Pi(G) = \bigcup_m \Pi(G)_{K'_m-\text{sph}}.$$

We can now state:

**Theorem 1.8.** *Conjecture 1.7 is true when  $G$  is non-Archimedean. More precisely, there exists a sequence  $(K'_m)$  consisting of arbitrarily small compact open subgroups of  $G$  such that if the weak injective radius property holds for a sequence of cocompact lattices  $(\Gamma_N)$ , then for every  $N \geq 1, m \geq 1, p > 2, \epsilon > 0$ ,*

$$M(\Pi(G)_{K'_m-\text{sph}}, \Gamma_N, p) \ll_{K'_m, \epsilon} [\Gamma_1 : \Gamma_N]^{2/p+\epsilon}.$$

In the Archimedean case, we do not know even whether the weak injective radius property implies the spherical density hypothesis. However, for rank 1 it was essentially proven in [58] (see the remark after the statement of Theorem 3 in [58]):

**Theorem 1.9.** *If  $G$  is of rank one and  $(\Gamma_N)$  is a sequence of cocompact lattices, then the weak injective radius property implies the spherical density hypothesis.*

For general representations in the Archimedean case, we have the following theorem. For rank 1 it was proven in [58, Theorem 3].

**Theorem 1.10.** *If the sequence  $(\Gamma_N)$  of cocompact lattices satisfies the weak injective radius property, then  $(\Gamma_N)$  satisfies the pointwise multiplicity hypothesis.*

Figure 1.1 summarizes the different relations between our main properties for a sequence of cocompact lattices.

Let us end this introduction by stating some open problems this work leads to.

The main open problem is to prove either the spherical density hypothesis or the injective radius property for new cases, which would lead to a proof of the optimal lifting property. See Conjecture 2.4 for

a general conjecture for the Archimedean case. Very few cases of this conjecture are known for groups of rank greater than 1. We remark that it seems that the problem is harder for principal congruence subgroups and easier for groups that are far from being normal, such as  $\Gamma_0(N)$  of  $\mathrm{SL}_m(\mathbb{Z})$  (see Subsection 2.5). See, for example, the *density amplification* phenomena for graphs in [29].

On the more technical side, one of the main problems this work does not resolve is Conjecture 1.7, which would show, in particular, that the spherical density hypothesis and the injective radius property are indeed equivalent, also for the Archimedean high-rank case. It relates in particular to the understanding of uniform *lower bounds* on matrix coefficients; see, for example, Conjecture 3.12.

Finally, we only discuss multiplicities for cocompact lattices. We strongly believe that the weak injective radius property has spectral implications for nonuniform lattices as well, for example, for bounds of multiplicities of representations in the discrete spectrum. This problem is related to the concentration of  $L^2$ -mass of nontempered automorphic functions away from the cusp, in a uniform way. In hyperbolic spaces this problem is essentially solved, even for some discrete groups that are not lattices, thanks to the work of Gamburd on hyperbolic surfaces ([23]) and the work of Magee on general hyperbolic spaces ([53]). Therefore, Theorem 1.9 can be generalized to such cases.

### **Structure of the article**

In Section 2, we give some applications of our results and state some open problems.

In Section 3, we present the main ideas behind the proofs.

In Section 4, we collect various results, mainly from representation theory. In particular, we discuss upper bounds for matrix coefficients, which are well understood. Using those upper bounds, we prove the essential Convolution Lemma 4.18.

In Section 5, we discuss the weak injective radius property and the spectral results it implies. We prove Proposition 3.9, which reduces Theorem 1.10, Theorem 1.9, and Theorem 1.8 to finding some explicit and strict lower bounds on matrix coefficients.

In Section 6, we prove Theorem 1.5.

In Section 7, we discuss the spherical density hypothesis and the results it implies. We prove Theorem 1.6 and prove Theorem 7.4, which is a version of Theorem 1.5 which assumes the spherical density hypothesis and has stronger implications.

In Section 8, we discuss Bernstein's description of the Hecke algebra in the non-Archimedean case and prove Theorem 3.11, which implies Theorem 1.8 together with Proposition 3.9.

In Section 9, we discuss the theory of leading coefficients and prove Theorem 3.10, which implies Theorem 1.10 together with Proposition 3.9.

### **Added in proof**

Since the completion of the draft of this paper, some exciting new developments have occurred. The most significant is the verification of the weak injective radius property and the optimal lifting property for the principal congruence subgroups  $\Gamma(q)$  of  $\mathrm{SL}_n(\mathbb{Z})$ , for square-free  $q$ , by Jana and the second named author ([36, Theorem 5 and Theorem 6]). This followed a spectral breakthrough, similar to the spherical density hypothesis, by Assing and Blomer ([2]).

Another development is the extension of the ideas of this paper to the theory of Ghosh–Gorodnik–Nevo diophantine exponents by Jana and the second named author ([37]).

## **2. Applications and open problems**

### **2.1. Ramanujan graphs and complexes**

We shortly note that the results of this paper, and in particular Theorem 1.5 (or the stronger Theorem 7.4) apply to Ramanujan complexes, by which we mean here the situation when  $G$  is  $p$ -adic,  $\Gamma_1$  is cocompact

and no nontempered and nontrivial spherical representation appears in the decomposition of  $L^2(\Gamma_N \backslash G)$ . In this case, the sequence  $(\Gamma_N)$  obviously satisfies the spherical density hypothesis. The Ramanujan complexes themselves are the quotients of the Bruhat–Tits buildings of  $G$  by  $\Gamma_N$ . Such complexes were constructed (with the same definition) by Lubotzky, Samuels and Vishne (see [51, 50]). Similar results to Theorem 7.4 appear in [38, 47, 13].

Note that the definition we use is not the same as in the more modern approach to Ramanujan complexes, where one considers not only spherical representation but also representations with a nontrivial vector fixed by the Iwahori subgroup (see [49, Subsection 2.3] and the references within).

The standard way of proving that a complex is a Ramanujan complex is to use results from the Langlands program (e.g., [45]) and to eventually apply Deligne’s proof of the Weil conjectures ([17]). This approach has obvious limitations, and in particular, it seems that one must assume that  $G$  is  $p$ -adic for it to succeed. We refer to [21] for some recent work on this subject and to an ongoing and yet unpublished work of Shai Evra.

We will avoid diving deeper into this subject, as it is based on spectral input, unlike our approach which is geometric.

### 2.2. Adding a parameter to the properties

It is useful to add a parameter  $0 < \alpha \leq 1$  to the properties, with  $\alpha = 1$  being equivalent to the property without the parameter.

**Definition.** We say that the sequence  $(\Gamma_N)$  satisfies the *weak injective radius property with parameter  $\alpha$* , if for every  $0 \leq d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ ,  $\epsilon > 0$ ,

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{y \in \Gamma_N \backslash \Gamma_1} N(\Gamma_N, d_0, y) \ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{\epsilon} q^{d_0/2}.$$

**Definition.** We say that the sequence  $(\Gamma_N)$  satisfies the *general density hypothesis with parameter  $\alpha$*  if for every precompact subset  $A \subset \Pi(G)$  and every  $\epsilon > 0$ ,  $N$  and  $p \geq 2$ ,

$$M(A, \Gamma, p) \ll_{\epsilon, A} [\Gamma_1 : \Gamma_N]^{1-\alpha(1-2/p)+\epsilon}.$$

One may similarly define the spherical density hypothesis with parameter  $\alpha$  and the pointwise density hypothesis with parameter  $\alpha$ , by changing the exponent  $2/p$  to  $1 - \alpha(1 - 2/p)$ .

Figure 1.1 remains true if we replace the properties with their parameterized version, for the same parameter  $0 < \alpha \leq 1$ , with the exception that the derivation of the optimal lifting property from the spherical density hypothesis requires  $\alpha = 1$ .

In the work itself, we work with the parameterized version of the properties.

### 2.3. Congruence subgroups of arithmetic groups

Let  $\Gamma_1$  be an arithmetic lattice in a semisimple noncompact Lie group  $G$ . Following [58], by arithmetic we mean that  $G$  is defined over  $\mathbb{Q}$ ,  $\Gamma_1 \subset G(\mathbb{Q})$ , there is a  $\mathbb{Q}$ -embedding  $\rho: G \rightarrow \text{GL}_n$  and  $\Gamma_1$  is commensurable with  $\rho^{-1}(\text{GL}_n(\mathbb{Z}))$ .

In this case, we may define a sequence of principal congruence subgroups  $(\Gamma_N)$  of  $\Gamma_1$  by letting

$$\Gamma_N = \Gamma_1(N) = \Gamma_1 \cap \rho^{-1}(\{A \in \text{GL}_n(\mathbb{Z}) : A \equiv I \pmod{N}\}).$$

It is a well-known fact that such subgroups have an injective radius which is logarithmic in the index (e.g., [58, Lemma 1] or [30, Proposition 16]).



Let us shortly give the argument – we assume by moving to a finite index in  $\Gamma_1$  that  $\Gamma_1 \subset \rho^{-1}(\text{GL}_n(\mathbb{Z}))$ . In this case,  $\Gamma_N$  is normal in  $\Gamma_1$  and it is left to verify that for every  $d_0 \leq 2\alpha \ln([\Gamma_1 : \Gamma_N])$  and  $\epsilon > 0$ ,

$$\#\{\gamma \in \Gamma_N : l(\gamma) \leq d_0\} \ll_\epsilon e^{d_0(1/2+\epsilon)}.$$

We note that there exists a constant  $C$  such that for  $g \in G$  outside a compact set the length  $l(g)$  we defined satisfies

$$C^{-1} \ln(\|\rho(g)\|) \leq l(g),$$

where  $\|\cdot\| : \text{GL}_n(\mathbb{R}) \rightarrow \mathbb{R}_{\geq 0}$  is the maximal absolute value of an entry of the matrix. Since each element of  $\rho(\Gamma_N)$  which is not the identity has an entry of size  $N$ , We deduce that for every  $N$  large enough,

$$\{\gamma \in \Gamma_N : l(\gamma) < C \ln(N)\} = \{I\}.$$

On the other hand, there is an injective map of  $\Gamma_1/\Gamma_N$  into  $\text{GL}_n(\mathbb{Z}/N\mathbb{Z})$ , so

$$[\Gamma_1 : \Gamma_N] \ll N^{n^2}$$

(and the exponent  $n^2$  can usually be improved).

Combining the different estimates, for  $d_0 < 2\frac{C}{2n^2} \ln([\Gamma_1 : \Gamma_N]) \leq 2\frac{C}{2n^2} N^{n^2} = C \ln(N)$ , it holds that for  $N$  large enough,

$$\#\{\gamma \in \Gamma_N : l(\gamma) \leq d_0\} = 1.$$

Therefore, the weak injective radius property is satisfied with parameter  $\alpha = \frac{C}{2n^2}$ .

**Remark 2.1.** Pushing this argument further, we may take the parameter to be  $\alpha = \frac{C}{d}$ , where  $d = \dim G$ , at least when  $G$  is split. For example, for the principal congruence subgroups of  $\text{SL}_2(\mathbb{Z})$  one gets this way the parameter  $\alpha = \frac{2}{3}$ , while  $\alpha = 1$  can be reached by a better analysis; see Subsection 2.4 below.

As a direct application of this fact, we have:

**Corollary 2.2.** *Let  $G$  be Archimedean, let  $\Gamma_1$  be an arithmetic lattice, and let  $(\Gamma_N)$  be the sequence of principal congruence subgroups of  $\Gamma_1$ . Then the sequence  $(\Gamma_N)$  satisfies the weak injective radius property with parameter  $\alpha = \alpha(\Gamma_1)$ .*

As a matter of fact, this corollary can be easily extended to principal congruence subgroups of  $S$ -arithmetic groups, once those are properly defined.

For arbitrary congruence subgroups, let us recall some of the results of [22] (the same result can be deduced from [1, Theorem 1.11 and Theorem 5.6]).

First, for  $\gamma \in \Gamma_1$ , we denote

$$c_{\Gamma_N}(\gamma) = \left| \{y \in \Gamma_N \setminus \Gamma_1 : y\gamma y^{-1} \in \Gamma_N\} \right|,$$

which is the number of fixed points of the right action of  $\gamma$  on  $\Gamma_N \setminus \Gamma_1$ . Then the weak injective radius property with parameter  $\alpha$  is equivalent to the fact that for every  $0 \leq d_0 \leq 2\alpha \ln([\Gamma_1 : \Gamma_N])$ ,  $\epsilon > 0$ ,

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{l(\gamma) \leq d_0} c_{\Gamma_N}(\gamma) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon e^{d_0(1/2+\epsilon)}.$$

A congruence subgroup is a subgroup of one of the groups  $\Gamma_1(N)$  as above.

**Theorem 2.3** (Following [22, Corollary 5.9]). *Let  $G$  be Archimedean, semisimple, almost-simple and simply connected; let  $\Gamma_1$  be an arithmetic lattice and let  $(\Gamma_N)$  be a sequence of congruence subgroups*



of  $\Gamma_1$ . Then there exist constants  $\mu > 0, c > 0$  depending only on  $\Gamma_1$  such that for every  $d_0 > 0, N \geq 1$  it holds that

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{\gamma \in \Gamma_1, l(\gamma) \leq d_0} c_{\Gamma_N}(\gamma) \ll 1 + e^{cd_0} [\Gamma_1 : \Gamma_N]^{-\mu}. \tag{2.1}$$

In particular, the injective radius property holds with parameter  $\alpha = \frac{\mu}{c}$ .

*Proof.* By [22, Corollary 5.9], there exists a constant  $c'$  such that for every  $\gamma \in \Gamma_1$  which does not belong to a proper normal subgroup of  $G$  it holds that

$$\frac{1}{[\Gamma_1 : \Gamma_N]} c_{\Gamma_N}(\gamma) \ll e^{c'l(\gamma)} [\Gamma_1 : \Gamma_N]^{-\mu}.$$

We remark that the dependence on  $\gamma$  in [22] is different, but it is obvious that  $e^{c'l(\gamma)}$  for  $c$  large enough is an upper bound on it. By our assumptions on  $G$ , the number of  $\gamma \in \Gamma_1$  belonging to a proper normal subgroup is  $\ll 1$ .

By summing over all  $\gamma \in \Gamma_1$  with  $l(\gamma) \leq d_0$ , we get

$$\begin{aligned} \frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{\gamma \in \Gamma_1, l(\gamma) \leq d_0} c_{\Gamma_N}(\gamma) &\ll \frac{[\Gamma_1 : \Gamma_N]}{[\Gamma_1 : \Gamma_N]} + e^{c'd_0} [\Gamma_1 : \Gamma_N]^{-\mu} \sum_{\gamma \in \Gamma_1, l(\gamma) \leq d_0} 1 \\ &\ll 1 + e^{(1+c')d_0} [\Gamma_1 : \Gamma_N]^{-\mu}, \end{aligned}$$

which implies equation (2.1) with  $c = 1 + c'$ .

When  $d_0 \leq 2\frac{\mu}{c} \ln([\Gamma_1 : \Gamma_N])$  or  $[\Gamma_1 : \Gamma_N]^{-\mu} \leq e^{d_0c/2}$ , we get

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{\gamma \in \Gamma_1, l(\gamma) \leq d_0} c_{\Gamma_N}(\gamma) \ll e^{d_0/2},$$

which implies the weak injective radius property with parameter  $\alpha = \frac{\mu}{c}$ , as needed. □

In the cocompact case, Theorem 1.10 implies the pointwise multiplicity hypothesis with parameter  $\alpha = \frac{\mu}{c}$ , namely

$$m(\pi, \Gamma_N) \ll_{\pi, \epsilon} [\Gamma_1 : \Gamma_N]^{1 - \frac{\mu}{c}(1 - 2/p(\pi)) + \epsilon}.$$

It should be compared with [1, Theorem 7.15], which states

$$m(\pi, \Gamma_N) \ll_{\pi} [\Gamma_1 : \Gamma_N]^{1 - \alpha(\pi)}.$$

Finally, we wish to make the following conjecture:

**Conjecture 2.4.** *Let  $G$  be Archimedean, semisimple, almost-simple and simply connected; let  $\Gamma_1$  be an arithmetic lattice and let  $(\Gamma_N)$  be a sequence of congruence subgroups of  $\Gamma_1$ . Then the sequence  $(\Gamma_N)$  satisfies the weak injective radius property with parameter  $\alpha = 1$ , and if  $\Gamma_1$  is cocompact then the sequence  $(\Gamma_N)$  satisfies the general density hypothesis with parameter  $\alpha = 1$ .*

*As a corollary, the sequence  $(\Gamma_N)$  has the optimal lifting property.*

The conjecture generalizes Conjecture 1 and Conjecture 2 from the work of Sarnak and Xue ([58]). Finally, a similar conjecture should also hold in the  $S$ -arithmetic setting when  $G$  is  $p$ -adic.

**2.4. The Work of Sarnak and Xue and its implications**

Sarnak and Xue proved the weak injective radius property with parameter  $\alpha = 1$  for principal congruence subgroups of cocompact arithmetic lattices in  $SL_2(\mathbb{R})$  and  $SL_2(\mathbb{C})$  and proved the weak injective radius property with parameter  $\alpha = 5/6$  for principal congruence subgroups of cocompact arithmetic lattices of  $SU(2, 1)$ .

We refer to [58, Section 3] for their calculations. Their argument actually works also for principal congruence subgroups of lattices in  $SL_2$  over  $p$ -adic fields coming from division algebras. See, for example, [15, Theorem 4.4.4].

As a simple example, let us prove here the weak injective radius property for the principal congruence subgroups of  $SL_2(\mathbb{Z})$ . In this case, we need to show that the number  $N(T, \Gamma(N))$ , of solutions to  $ad - bc = 1$ , with  $a \equiv d \equiv 1 \pmod N$ ,  $b \equiv c \equiv 0 \pmod N$  and  $\max\{|a|, |b|, |c|, |d|\} \leq T$ , for  $T \leq N^3$ , is bounded by

$$N(T, \Gamma(N)) \ll_{\epsilon} T^{1+\epsilon}.$$

This is done as follows (see also [23, Proposition 5.3]). From the congruence condition, it follows that

$$a + d - 2 = -(a - 1)(d - 1) + bc \equiv 0 \pmod{N^2}.$$

One may therefore choose  $a + d$  in  $(2T/N^2 + 1)$  ways, and choose  $(a, d)$  in  $(2T/N^2 + 1)(2T/N + 1)$  ways.

If  $ad \neq 1$ , from  $bc = 1 - ad$  and bounds on the divisor function, there are  $T^{\epsilon}$  ways of choosing  $bc$ . If  $ad = 1$ , it is also simple to bound the number of possibilities of  $b, c$  by  $4(T/N + 1)$ . In total, we get for  $T \leq N^3$ ,

$$N(T, \Gamma(N)) \ll_{\epsilon} T^{\epsilon} (T/N^2 + 1)(T/N + 1) \ll_{\epsilon} T^{1+\epsilon}.$$

**2.5. On Some congruence subgroups of  $SL_3(\mathbb{Z})$**

Consider the following subgroups of  $\Gamma_1 = SL_3(\mathbb{Z})$ :

$$\Gamma_0(N) = \left\{ \begin{pmatrix} * & * & * \\ * & * & * \\ a & b & * \end{pmatrix} \in SL_3(\mathbb{Z}) : a \equiv b \equiv 0 \pmod N \right\} \subset$$

$$\Gamma_2(N) = \left\{ \begin{pmatrix} * & * & * \\ c & * & * \\ a & b & * \end{pmatrix} \in SL_3(\mathbb{Z}) : a \equiv b \equiv c \equiv 0 \pmod N \right\} \subset \Gamma_1 = SL_3(\mathbb{Z}).$$

In a companion paper by the second named author and Hagai Lavner ([40]), the following is proven:

**Theorem 2.5.** *The sequences of lattices  $(\Gamma_0(N))$  and  $(\Gamma_2(N))$ , for  $N$  prime, have the weak injective radius property (with parameter 1).*

*As a result, the sequences have the optimal lifting property.*

We refer to [40] for an interpretation of this result in terms of the action of  $SL_3(\mathbb{Z})$  on the projective plane over the field with  $N$  elements, or on the corresponding flag space.

The work [40] is strongly influenced by a deep work of Blomer, Buttcane and Maga on  $\Gamma_0(N)$  ([6]), which is based on the Kuznetsov trace formula:

**Theorem 2.6** [6, Theorem 4]. *For  $\pi \in \Pi(SL_3(\mathbb{R}))_{\text{sph}}$ , let  $m_{\text{cusp}}(\pi, \Gamma_0(N))$  be the multiplicity of  $\pi$  in the cuspidal part of  $L^2(\Gamma_0(N) \backslash SL_3(\mathbb{R}))$ . Then for every compact  $A \subset \Pi(SL_3(\mathbb{R}))_{\text{sph}}$ , it holds that for every  $N$  prime,  $p > 2$ ,  $\epsilon > 0$ ,*

$$\sum_{\pi \in A, p(\pi) > 0} m_{\text{cusp}}(\pi, \Gamma_0(N)) \ll_{A, \epsilon} [\Gamma_1 : \Gamma_0(N)]^{1-2(1-2/p)+\epsilon}.$$

The result of the theorem is very similar to the spherical density hypothesis with parameter  $\alpha = 2$ . There are also a number of differences: first,  $SL_3(\mathbb{Z})$  is not cocompact, so our discussion does not apply to it. In particular, we have to deal with the continuous spectrum if we wish to deduce the optimal lifting property. Secondly and less crucially, the dependence on the subset  $A$  is not explicit, as needed in the definition of the spherical density hypotheses, so it is hard to use it for geometric applications.

This result was recently generalized by Blomer ([5]) to general  $SL_M(\mathbb{Z})$ , where the subgroup  $\Gamma_0(N)$  is similarly defined to be the set of matrices with the entries in the last row, except for the  $(M, M)$  entry, equal to 0 modulo  $N$ . We remark that Blomer uses a slightly weaker way to measure temperedness, rather than  $p(\pi)$  (for  $SL_3$  the two ways are equivalent).

In any case, it seems that the methods of [6, 5] are not applicable to the subgroups of the form  $\Gamma_2(N)$ .

### 2.6. The weak injective radius of principal congruence subgroups of $SL_n(\mathbb{Z})$

If the weak injective radius property would hold for the principal congruence subgroups of  $SL_n(\mathbb{Z})$ , then it will imply that

$$\#\{\gamma \in SL_n(\mathbb{Z}) : \gamma \equiv I \pmod N, \|\gamma\| \leq T\} \ll_\epsilon T^\epsilon N^\epsilon \left( \frac{T^{n^2-n}}{N^{n^2-1}} + T^{(n^2-n)/2} \right),$$

where  $\|\cdot\|$  is the maximal absolute value of an entry. One may also try to improve this estimate, in particular, the ‘error’ part  $T^{(n^2-n)/2}$ .

One of the results of [42], which is also cited in [58], states that

$$\#\{\gamma \in SL_n(\mathbb{Z}) : \gamma \equiv I \pmod N, \|\gamma\| \leq T\} \ll \frac{T^{n^2-n}}{N^{n^2-1}} \log T + 1.$$

As was pointed to us by Sarnak, the proof contains an error. As a matter of fact, this naive estimate is actually *false*, for simple reasons. For example, in  $SL_2$ ,

$$\#\left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \in SL_2(\mathbb{Z}) : \gamma \equiv I \pmod N, \|\gamma\| \leq T \right\} \asymp T/N + 1.$$

and this is larger than  $\frac{T^2}{N^3} \log T$  in the range  $N^{1+\epsilon} < T < N^{2(1-\epsilon)}$ . The same argument works for every  $n$ , but there are  $n(n-1)/2$  entries we can use. Therefore, we have the lower bound:

$$\#\{\gamma \in SL_n(\mathbb{Z}) : \gamma \equiv I \pmod N, \|\gamma\| \leq T\} \gg \left( \frac{T^{n^2-n}}{N^{n^2-1}} + \left( \frac{T}{N} \right)^{(n^2-n)/2} + 1 \right).$$

Up to  $T^\epsilon$ , this is also the upper bound for  $n = 2$ . We conjecture that up to  $T^\epsilon$  this is also the upper bound for larger values of  $n$ .

### 3. Main ideas of the proofs

In this section, we discuss some ideas from the proofs of the main theorems and the technical problems they lead to.

All the proofs are based on a reduction of the geometric properties into spectral ones. We first restate the weak injective radius property in terms of traces of operators.

**Definition 3.1.** For  $d_0 \in \mathbb{R}_{\geq 0}$ , let  $\chi_{d_0} \in C_c^\infty(K \backslash G / K)$  be a function that is equal to 1 for  $l(g) \leq d_0$ , equal to 0 for  $l(g) \geq d_0 + 1$  and for  $d_0 \leq l(g) \leq d_0 + 1$  it is defined between 0 and 1 so that  $\chi_{d_0} \in C_c^\infty(K \backslash G / K)$ .

Let  $\psi_{d_0}(g) \in C_c^\infty(K \backslash G / K)$  be  $\psi_{d_0}(g) = q^{(d_0 - l(g))/2} \chi_{d_0}(g)$  for  $l(g) \geq 1$  and for  $0 \leq l(g) \leq 1$   $\psi_{d_0}(g)$  is defined between  $q^{(d_0-1)/2}$  and  $q^{d_0/2}$  so that eventually  $\psi_{d_0}(g) \in C_c^\infty(K \backslash G / K)$ .

As  $\chi_{d_0}, \psi_{d_0} \in C_c^\infty(K \backslash G / K)$ , they act naturally on  $L^2(\Gamma_N \backslash G)$ , and moreover, are trace-class ([25, Chapter 1]; see also Subsection 4.6). The trivial eigenvalues of  $\chi_{d_0}, \psi_{d_0}$  satisfy  $q^{d_0} \ll \int_G \chi_{d_0}(g) dg \ll_\epsilon q^{d_0(1+\epsilon)}$  and  $q^{d_0} \ll \int_G \psi_{d_0}(g) dg \ll_\epsilon q^{d_0(1+\epsilon)}$ . Since  $\chi_{d_0}, \psi_{d_0} \in C_c(K \backslash G / K)$ , their action is actually on  $L^2(\Gamma_N \backslash G / K)$ .

The reason that we look at  $\psi_{d_0}$ , and one of the main reasons we use the length  $l(g)$ , is the following convolution lemma, which replaces the rank 1 case in [58, Lemma 3.1]. The lemma says that  $\psi_{d_0}$  provides an approximated upper bound for the convolution of  $\chi_{d_0/2}$  with itself:

**Lemma 3.2** (See Lemma 4.18). *It holds that  $c_{d_0} = \chi_{d_0} * \chi_{d_0} \in C_c^\infty(K \backslash G / K)$ , and for every  $g \in G$  and  $\epsilon > 0$ , we have*

$$c_{d_0}(g) \ll_\epsilon q^{d_0 \epsilon} \psi_{2d_0}(g).$$

Using the pretrace formula (see Subsection 4.7), we get:

**Proposition 3.3.** *The following are equivalent for a sequence  $(\Gamma_N)$  of cocompact lattices:*

1. *The sequence  $(\Gamma_N)$  satisfies the weak injective radius property with parameter  $\alpha$ .*
2. *For every  $0 \leq d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ ,  $\epsilon > 0$ ,*

$$\text{tr } \chi_{d_0}|_{L^2(\Gamma_N \backslash G)} \ll_\epsilon [\Gamma_1 : \Gamma_N]^{1+\epsilon} q^{d_0(1/2+\epsilon)}. \tag{3.1}$$

3. *For every  $\epsilon > 0$ , for every  $h \in C_c(G)$  satisfying  $h(g) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon \psi_{2\alpha \log_q([\Gamma_1 : \Gamma_N])}(g)$ , it holds that*

$$\text{tr } h|_{L^2(\Gamma_N \backslash G)} \ll_\epsilon [\Gamma_1 : \Gamma_N]^{1+\alpha+\epsilon}. \tag{3.2}$$

The proposition allows one to prove the weak injective radius property using the trace formula and equation (3.1). It also allows proving various multiplicity results using equation (3.2).

By combining the Convolution Lemma 3.2, spectral gap and a version of equation (3.2), we prove Theorem 1.5. See Section 6.

The direction ‘spectral to geometric’ uses carefully Harish-Chandra’s upper bounds on spherical functions, and its generalization to arbitrary matrix coefficients of representations. Our analysis is closely related to the work of Ghosh, Gorodnik and Nevo on Diophantine exponents ([26, 27] and the reference therein). Very generally, the main difference between our analysis and theirs is that we assume density, while they assume spectral gap. The new idea is to note that one applies the spectral estimates to characteristic functions of ‘small sets’, and if there are few ‘bad eigenvectors’ they correlate poorly with such functions.

Let us state some concrete results. Let  $\Xi(g) = \int \delta(gk)^{-1/2} dk$  be Harish-Chandra’s function of  $G$ . This function satisfies  $|\Xi(g)| \ll_\epsilon q^{-l(g)(1/2+\epsilon)}$  (which is another motivation for the definition of  $l$ ). Now, combining the two theorems of [14], we get:

**Theorem 3.4** [14]. *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq 2$ , and let  $v_1, v_2 \in V$  be two  $K$ -finite vectors such that  $\dim \text{span } Kv_1 = d_1$ ,  $\dim \text{span } Kv_2 = d_2$ . Then*

$$|\langle v_1, gv_2 \rangle| \leq \sqrt{d_1 d_2} \|v_1\| \|v_2\| \Xi(g).$$

However, this theorem is not sufficient for our uses since we wish to consider  $p(\pi)$  arbitrary. We remark that [14] does give some results for arbitrary  $\pi$  and  $p$ , but they are not precise enough for us.

Let us state here a general theorem that generalizes the above and gives a bound of the form that we need. For  $2 \leq p \leq \infty$ , let  $\Xi_p(g) = \int \delta(gk)^{-1/p} dk$  be the  $p$ -th version of Harish-Chandra’s function of  $G$ . This function satisfies  $\Xi_p(g) \ll_\epsilon \Xi(g) q^{-l(g)(1/p-1/2)} \ll_\epsilon q^{-l(g)(1/p+\epsilon)}$ . Then we have:

**Theorem 3.5** (See Theorem 4.10). *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq p$ ,  $p \geq 2$ , and let  $v_1, v_2 \in V$  be two  $K$ -finite vectors such that  $\dim \text{span } Kv_1 = d_1$ ,  $\dim \text{span } Kv_2 = d_2$ . Then*

$$|\langle v_1, gv_2 \rangle| \leq \sqrt{d_1 d_2} \|v_1\| \|v_2\| \Xi_p(g).$$

We discuss such bounds further in Subsection 4.2. Using Theorem 3.5 one may deduce various *upper bounds* on matrix coefficients, norms of operators and traces of operators. In particular, a useful bound is:

**Theorem 3.6** (See Corollary 4.14). *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) = p$ ,  $p \geq 2$ . Then for every  $d_0 > 0$  and  $\epsilon > 0$*

$$\|\pi(\chi_{d_0})\| \ll_{\epsilon} q^{d_0(1-1/p+\epsilon)}.$$

Applying Theorem 3.6 carefully allows us to deduce Theorem 1.6.

To prove Theorem 1.10, Theorem 1.9 and Theorem 1.8, which deduce multiplicity bounds from the weak injective radius property, we use equation (3.2) of Proposition 3.3. To deduce from it upper bounds on multiplicity, one needs *lower bounds* on traces on representations of functions  $h \in C_c^\infty(G)$ . The following definition and proposition capture the situation:

**Definition 3.7.** Let  $A \subset \Pi(G)$  be a precompact subset. We say that a family of functions  $\{f_{d_0}\} \subset C_c^\infty(G)$ ,  $d_0 \in \mathbb{R}$ ,  $d_0 \geq D$  is *good* for  $A$  if it holds that:

1. For every  $\pi \in A$ , and  $\epsilon > 0$ ,

$$q^{d_0(1-1/p(\pi)-\epsilon)} \ll_{A, \epsilon} \text{tr}(\pi(f_{d_0})).$$

2. It holds that for every  $g \in G$ ,  $\epsilon > 0$ ,  $f_{d_0}(g) \ll_{\epsilon} q^{d_0\epsilon} \psi_{d_0}(g)$ , where  $\psi_{d_0}(g)$  is from Definition 3.1.
3. For every representation  $\pi' \in \Pi(G)$ , it holds that  $0 \leq \text{tr}(\pi'(f_{d_0}))$ .

**Remark 3.8.** If we replace (2) by the slightly stronger condition  $|f_{d_0}(g)| \ll_{\epsilon} q^{d_0\epsilon} \psi_{d_0}(g)$  and further assumes that  $f_{d_0}$  is left and right  $K$ -finite, then one actually has by Theorem 3.5

$$\text{tr}(\pi(f_{d_0})) \ll_{\pi, \epsilon} q^{d_0(1-1/p(\pi)+\epsilon)},$$

so the lower bound is rather tight.

**Proposition 3.9.** *Let  $A \subset \Pi(G)$  be a precompact subset, and assume that it has a good family of functions. Under this condition, if the sequence  $(\Gamma_N)$  of cocompact lattices satisfies the weak injective radius property with parameter  $\alpha$  then for every  $N \geq 1$ ,  $p > 2$ ,  $\epsilon > 0$ ,*

$$M(A, \Gamma_N, p) \ll_{A, \epsilon} [\Gamma_1 : \Gamma_N]^{1-\alpha(1-2/p)+\epsilon}.$$

Finding general lower bounds on traces (uniformly for a family of representations) is not well studied. Two special cases, which appear (somewhat implicitly) in the work of Sarnak and Xue, correspond to Theorem 1.9 and Theorem 1.10:

1. In rank 1, one has a simple classification of spherical irreducible unitary representations. In the Archimedean case, for each  $2 < p \leq \infty$  there is at most a single spherical irreducible unitary representation  $(\pi, V)$  with  $p(\pi) = p$  (with a corresponding spherical function  $\Xi_p(g)$ ), and one can easily deduce lower bounds on the trace of  $f_{d_0} = \chi_{d_0/2} * \chi_{d_0/2}$ , and deduce Theorem 1.9. In the non-Archimedean case, Theorem 1.9 reduces to some statement on graphs; see Subsection 4.4.
2. If one is interested in a single representation, one has the following, from which we deduce Theorem 1.10. In the Archimedean case, it follows from the asymptotic behavior of leading exponents ([12, 43, Chapter VIII]). The non-Archimedean case is easier, and in any case follows from Theorem 3.11 below.

**Theorem 3.10** (See Section 9). *Let  $(\pi, V) \in \Pi(G)$ . Then the set  $A = \{\pi\}$  has a good family of functions.*

Finally, we provide the following theorem, which implies Theorem 1.8 together with Proposition 3.9.

**Theorem 3.11** (See Theorem 8.5). *Let  $G$  be non-Archimedean. Then there exists a set  $\{K'\}$  of arbitrarily small open-compact subgroups of  $G$ , such that for every  $K'$ ,  $\Pi(G)_{K'\text{-sph}}$  has a good family.*

The proof of Theorem 3.11 is based on two sources. The first is the connection between the Ihara graph zeta function and expansion (see [33]), and the second is Bernstein’s description of the Hecke algebra  $C_c(K' \backslash G / K')$  (see [3]). A precise connection for  $(q + 1)$ -regular graphs between  $p(\pi)$  for spherical function and poles of the Ihara zeta function may be found in [39]. In recent years, there were various generalizations of the graph zeta function to higher-dimensional buildings (see, e.g., [41] and the references within). In [38], the second named author generalized the connection between  $p(\pi)$  for representations  $\pi \in \Pi(G)$  with Iwahori-fixed vector and the poles of some generalized zeta function. By a slight variant of those ideas, one may prove the special case of Theorem 3.11 when  $K'$  is the Iwahori subgroup. For more general  $K'$ , we follow the same ideas, by using Bernstein’s description of the Hecke algebra  $C_c(K' \backslash G / K')$ . See Section 8 for details.

If we consider only the spherical case, it would be useful if the functions  $f_{d_0}$  in the definition of a good family will be left and right  $K$ -invariant, that is,  $f_{d_0} \in C_c^\infty(K \backslash G / K)$ . Recently, Matz and Templier proved a similar theorem for  $G = \text{PGL}_n$  using the Satake isomorphism ([55]). However, their results are less precise than we desire – they find a spherical function  $f_{d_0} \in C_c^\infty(K \backslash G / K)$  which satisfies  $f_{d_0}(g) \ll_\epsilon q^{d_0 \epsilon} \psi_{d_0}(g)$ , with a lower bound  $q^{\beta d_0(1-1/p(\pi))} \ll \text{tr}(\pi(f_{d_0}))$  for some  $\beta < 1$ , instead of the optimal bound  $q^{d_0(1-1/p(\pi))} \ll \text{tr}(\pi(f_{d_0}))$ .

Let us finish this discussion with the following conjecture, which concerns only spherical functions. For  $g \in G$ , let  $S(g)$  be the  $K$ -bi-invariant function such that  $\int_G f(g)dg = \int_K \int_K \int_{A_+} f(kak')S(a)dadkdk'$ . It holds that for  $g$  ‘far from the walls’  $S(g) \approx q^{l(g)}$  (see Subsection 4.1).

**Conjecture 3.12.** *There exist  $D > 0, L > 0$  such that for every  $\epsilon > 0$  and every  $(\pi, V) \in \Pi(G)_{\text{sph}}$  (i.e., a unitary irreducible spherical representation) with  $p(\pi) > 2$ , if  $v \in V, \|v\| = 1$  is  $K$ -fixed, then:*

1. *In the non-Archimedean case, for  $d_0 > D$*

$$\int_{l(g) \leq d_0} S(g)|\langle v, \pi(g)v \rangle|^2 dg \gg_\epsilon q^{2d_0(1-1/p(\pi)-\epsilon)}.$$

2. *In the Archimedean case, for  $d_0 > D$ ,*

$$\int_{l(g) \leq d_0} S(g)|\langle v, \pi(g)v \rangle|^2 dg \gg_\epsilon (\lambda(\pi) + 1)^{-L} q^{2d_0(1-1/p(\pi)-\epsilon)}.$$

The exponents in the conjecture are tight, as the corresponding upper bounds can be deduced from Theorem 3.5.

## 4. Preliminaries

### 4.1. Distances and length of elements

Besides our definition of length, the following is standard; see, for example, [26, Section 3]. We mainly follow [43] when  $G$  is Archimedean and [10] when  $G$  is non-Archimedean.

Let  $k$  be  $\mathbb{R}$  or a  $p$ -adic field, and  $|\cdot|_k : k \rightarrow \mathbb{R}_+$  its standard nontrivial valuation. Let  $G$  be a semisimple noncompact algebraic group over  $k$ , of  $k$ -rank  $r$ . Let  $T \cong G_m^r \subset G$  be a maximal  $k$ -split torus. The choice of  $T$  determines the set of weights  $X^*(T)$ , that is, of rational characters of  $T$ . Let  $\Phi(T, G) \subset X^*(T)$  be the set of roots of  $G$  with respect to  $T$ .

In the Archimedean case, if  $T_0 \cong \{\pm 1\}^r$  is the maximal compact subgroup of  $T$ , the connected component  $A \cong T/T_0$  of the identity of  $T$  is the Lie group of a real Cartan subalgebra  $\mathfrak{a}$  of  $\mathfrak{g}$ , and we define  $\nu: T \rightarrow A \rightarrow \mathfrak{a} \cong \mathbb{R}^r$  by the logarithm map.

In the non-Archimedean case, let  $T_0$  be a maximal compact subgroup of  $T$ . Then  $T/T_0 \cong \mathbb{Z}^r$ , this identification defines  $\nu: T \rightarrow \mathbb{Z}^r \subset \mathbb{R}^r$ , and we identify  $\mathbb{R}^r$  with  $\mathfrak{a}$ .

Let  $X^*(T)_{\mathbb{R}} \cong X^*(T) \otimes \mathbb{R}$  be the weight space. For an element  $\alpha \in X^*(T)$ , we let  $\chi_{\alpha} \in \mathfrak{a}^*$  be the linear functional defined such that for  $t \in T|\alpha(t)|_k = q^{\chi_{\alpha}(\nu(t))}$ , where  $q = e$  in the Archimedean case and otherwise the size of the quotient field of  $k$ . For  $\alpha \in X^*(T)_{\mathbb{R}}$ , we define  $\chi_{\alpha} \in \mathfrak{a}^*$  by extension of the action above. This isomorphism (as linear spaces) between  $X^*(T)_{\mathbb{R}}$  and  $\mathfrak{a}^*$  defines an isomorphism between the coweight space  $(X^*(T)_{\mathbb{R}})^*$  and  $\mathfrak{a}$ .

Choose an ordering on the root system which defines the positive roots  $\Phi_+ \subset \Phi(T, G)$  and let  $\Delta = \{\alpha_1, \dots, \alpha_r\} \subset X^*(T)_{\mathbb{R}} \cong \mathfrak{a}^*$  be the simple roots of  $\Phi$  with respect to this ordering. Let  $\{\omega_1, \dots, \omega_r\} \subset \mathfrak{a}$  be the set of simple coweights, that is, the dual basis to  $\Delta$ . The set  $\{\sum_{i=1}^r x_i \omega_i : x_i \geq 0\} \subset \mathfrak{a}$  is called the dominant sector, or the positive Weyl chamber. It is isomorphic to  $\mathfrak{a}/W$ , where  $W$  is the Weyl group of the root system.

Let  $\Phi_+^{\vee} \subset \mathfrak{a}$  be the corresponding coroot system, let  $\Delta^{\vee} = \{\alpha_1^{\vee}, \dots, \alpha_r^{\vee}\}$  be the dual basis of simple coroots and let  $\{\omega_1^{\vee}, \dots, \omega_r^{\vee}\} \subset \mathfrak{a}^*$  be the set of simple weights. We define a partial ordering on  $\mathfrak{a}$  by  $\alpha \geq_{\mathfrak{a}} \alpha'$  if and only if  $\omega_i^{\vee}(\alpha) \geq \omega_i^{\vee}(\alpha')$  for every simple weight, or alternatively  $\alpha - \alpha'$  is a nonnegative sum of elements of  $\Delta^{\vee}$ .

Let  $P$  be the Borel subgroup with respect to the set of positive roots. It holds that  $P = MN$ , where  $M$  is the centralizer of  $T$  in  $G$  and  $N$  is the unipotent radical of  $P$  ([10, p. 134]). Let  $K$  be a maximal special compact open subgroup (i.e., in the  $p$ -adic case we choose it to be ‘good’ in the sense of Bruhat and Tits [8]). The Iwasawa decomposition  $G = KP$  holds ([43, Proposition 1.2],[10, p. 140]). It holds that  $M = (M \cap K) \cdot T$ , and we extend  $\nu: M \rightarrow \mathbb{R}^r$  by  $\nu(k) = 1$  for  $k \in M \cap K$ .

Let us recall the Cartan decomposition  $G = KA_+K$ :

- In the real case, following [43, Theorem 5.20],  $A \cong T/T_0$  is the Lie group of the Cartan subalgebra  $\mathfrak{a}$  of  $\mathfrak{g}$ .  $A_+$  is the exponent of the closure of the dominant sector in  $\mathfrak{a}$ , that is, the set of elements  $A_+ = \{t \in A : \forall \alpha \in \Phi_+, \alpha(t) \geq 1\}$ . It is isomorphic (as a set, using the exponential map) to the dominant sector in  $\mathfrak{a}$ .
- In the  $p$ -adic case, following [10, p. 140], let  $\Lambda \cong M/M^0$ , where  $M^0$  is the maximal compact subgroup of  $M$ . We identify elements of  $\Lambda$  with elements of  $M \subset P$ . Elements of the weight space are unramified characters of  $M$ , and therefore are characters of  $\Lambda$  as well. We let  $A_+ = \{\lambda \in \Lambda : \forall \alpha \in \Phi_+, \alpha(\lambda) \geq 1\}$ . The action  $\nu: T/T_0 \rightarrow \mathfrak{a}$  extends to  $\nu: M/M^0 \rightarrow \mathfrak{a}$ . Then  $A_+$  is isomorphic as a set with the intersection of  $\nu(\Lambda)$  with the dominant sector in  $\mathfrak{a}$ . It is also isomorphic to a subset of the special vertices in the dominant sector in an apartment in the Bruhat–Tits building of  $G$ .

In both cases, there exists a map  $\nu: A_+ \rightarrow \mathfrak{a}$ . It is also useful to extend it to a map  $H: G \rightarrow \mathfrak{a}$ , using the Iwasawa decomposition  $G = KMN$ , by  $H(kmn) = \nu(m)$ .

We will need the following fundamental technical lemma:

**Lemma 4.1.** *For  $a \in A_+$  and  $k \in K$*

$$H(ak) \leq_{\mathfrak{a}} H(a).$$

*Proof.* For the non-Archimedean case, see [8, Proposition 4.4.4(i)]. For the Archimedean case, see [24, Corollary 3.5.3]. □

**Corollary 4.2.** *Let  $a, a', a'' \in A_+$ . If  $KaKa'K \cap Ka''K \neq \emptyset$ , then  $\nu(a'') \leq_{\mathfrak{a}} \nu(a) + \nu(a')$ .*

*Proof.* We first notice the following property of the  $H$ -function: for  $k \in K, g \in G, m \in M, n \in N$

$$H(kgmn) = H(g) + H(m)$$



Now, if  $KaKa'K \cap Ka''K \neq \emptyset$ , then  $a'' = k_0ak_1a'k_2$ . Applying the Iwasawa decomposition to  $a'k_2$  we have  $a'k_2 = k'_2mn$  with  $H(m) \leq_a H(a')$  by Lemma 4.1. Applying Lemma 4.1 again, we have

$$\begin{aligned} \nu(a'') &= H(k_0ak_1k'_2mn) = H(k_0ak_1k'_2) + H(m) \\ &\leq_a H(a) + H(a') = \nu(a) + \nu(a'). \end{aligned} \quad \square$$

Corollary 4.2 in the non-Archimedean case is [8, Proposition 4.4.4(iii)], and is deduced from Lemma 4.1 in the same way.

Let  $\delta(\tilde{p})$  be the left modular character of  $P$ , that is, if  $d\tilde{p}$  is a left Haar measure on  $P$ , then  $\delta(\tilde{p})d\tilde{p}$  is a right Haar measure. Normalize the measures so that for  $f \in C_c(G)$ ,  $\int_G f(g)dg = \int_K \int_P f(k\tilde{p})\delta(\tilde{p})d\tilde{p}dk = \int_K \int_P f(\tilde{p}k)d\tilde{p}dk, \int_K dk = 1$ .

It can also be defined as follows:  $M$  acts by conjugation on the Lie algebra  $\mathfrak{n}$  of  $N$ . Then for  $m \in M$ ,  $\delta(m) = |\text{Det Ad}_{\mathfrak{n}}(m)|_k$  ([10, p. 135], [43, Proposition 5.25]). Unwinding the definitions,  $\delta(m) = q^{2\rho(\nu(m))}$ , where  $\rho = \frac{1}{2} \sum_{\alpha \in \Phi_+} (\dim \mathfrak{g}_{\alpha})\chi_{\alpha}$ . Here,  $\mathfrak{g}_{\alpha}$  is the root space of  $\alpha$  in the Lie algebra  $\mathfrak{g}$  of  $G$ . Also, recall that  $q = e$  for  $k = \mathbb{R}$  and otherwise is the size of the quotient field of  $k$ . As an example, for  $G = \text{SL}_n(\mathbb{R})$ , and the matrix  $a = \text{diag}(a_0, \dots, a_{n-1})$ ,  $\delta(a) = \prod a_i^{n-1-2i}$ .

We associate with each element  $a \in A_+ \subset G$  a length  $l: A_+ \rightarrow \mathbb{R}_{\geq 0}$  by  $l(a) = \log_q \delta(a) = 2\rho(a)$ . We extend  $l: G \rightarrow \mathbb{R}_{\geq 0}$  by  $l(kak') = l(a)$ . By definition,  $l$  is left and right  $K$ -invariant.

For  $a \in T$ , we may identify  $l(a)$  by the entropy (taken with logarithm in base  $q$ ) of the dynamical system of translation of  $\Gamma \backslash G$  by  $a$ , with respect to the Haar measure ( $\Gamma$  here is an arbitrary lattice; see [20, Theorem 7.9]). Using this fact, we have for  $a \in A$ ,  $l(a) = l(a^{-1})$  and therefore for every  $g \in G$ ,  $l(g) = l(g^{-1})$ . The same fact can be proven directly.

**Proposition 4.3.** *For  $g_1, g_2 \in G$ , it holds that  $l(g_1g_2) \leq l(g_1) + l(g_2)$ .*

*Proof.* The proposition actually states that if  $KaKa'K \cap Ka''K \neq \emptyset$  for  $a, a', a'' \in A_+$ , then  $l(a'') \leq l(a) + l(a') = l(aa')$ . Since  $l(a) = 2\rho(a)$ , it follows from Corollary 4.2. □

For  $a \in A_+$ , define  $S(a) = \int \delta(ak)dk$ . For  $f \in C_c(G)$ , we have

$$\int f(g)dg = \int_K \int_K \int_{A_+} f(kgk')S(a)dkdk'da.$$

We interpret  $S(a)$  as the measure of the ‘circle’  $KaK$ . In the non-Archimedean case,  $S(a) \asymp \delta(a)$  ([10, p. 141]). In the Archimedean case, by [43, Proposition 5.28],

$$S(a) = \prod_{\alpha \in \Phi_+} (\sinh(\chi_{\alpha}(a)))^{\dim \mathfrak{g}_{\alpha}}.$$

Since  $\sinh(x) \asymp_{\beta} e^x$  for  $x > \beta$ , for  $a \in A_+$  ‘far from the walls’, that is, with  $\chi_{\alpha}(a) > \beta$  for every  $\alpha \in \Delta$  (and therefore  $\chi_{\alpha}(a) > \beta$  for every  $\alpha \in \Phi_+$ ), we have  $S(a) \asymp_{\beta} \delta(a) = e^{l(a)}$ . Near the walls where  $\sinh x \approx x$  this approximation fails, but we still have  $S(a) \ll \delta(a)$ . In any case, if we choose some norm  $\|\cdot\|_a$  on  $A_+$ , then for every  $\tau > 0$ ,  $a \in A_+$ ,  $\mu(\{g : \|a_g - a\|_a \leq \tau\}) \asymp_{\tau} q^{l(a)}$ , since the set  $\{a' \in A_+ : \|a' - a\|_a \leq \tau\}$  contains elements that are far enough (with respect to  $\tau$ ) from the walls.

We deduce that the size of balls for  $l \gg 1$ ,

$$q^l \ll \mu(\cup_{a': l(a') \leq l} Ka'K) = \int_{a \in A_+ : l(a) \leq l} S(a)da \ll p(l)q^l \ll_{\epsilon} q^{l(1+\epsilon)}, \tag{4.1}$$

for some polynomial  $p$ .

**Remark 4.4.** The literature has two popular choices of ‘distance’ or ‘length’ on  $G/K$  or  $G$ .

1. For  $G \subset \mathrm{SL}_n$ , where  $K = G \cap K'$  for  $K'$  maximal compact in  $\mathrm{SL}_n$ , one defines  $\tilde{l}(g) = \log \|g\|$ , where  $\|v\|$  is some matrix norm on  $\mathrm{GL}_n$ . Recall that we are mainly interested in distances as  $g \rightarrow \infty$ , so the specific choice of matrix norm does not matter. Such choice (without calling it a distance) is studied in [58, 19].
2. For  $G$  Archimedean, let  $\mathfrak{g}, \mathfrak{k}$  be the Lie algebras of  $G$  and  $K$ , and let  $B: \mathfrak{g} \times \mathfrak{g} \rightarrow \mathbb{C}$  the Killing form of  $G$ . Let  $\mathfrak{p} = \{X \in \mathfrak{g} : B(X, Y) = 0 \forall Y \in \mathfrak{k}\}$ . Then  $B|_{\mathfrak{p} \times \mathfrak{p}}$  is positive definite.  $\mathfrak{p}$  can be identified with the tangent space of  $G/K$  at the identity, and it defines a natural Riemannian structure on  $G/K$ , with length  $\hat{l}(g) = \hat{d}(g, 1)$ . See, for example, [16, Section 2], [58]. A similar distance is used in the  $p$ -adic case in [62, 2.3].

Since we mostly care about far distances in the group, and since  $K$  is compact, by the Cartan decomposition it suffices to compare  $l$  to other distances on  $A_+$ . In general,  $l$  and  $\tilde{l}$  look like an  $L^1$ -norm on  $A_+$ , and  $\hat{l}$  looks like an  $L^2$ -norm.

Let us concentrate on  $G = \mathrm{SL}_n(\mathbb{R})$  and  $\tilde{l}(g) = \ln(\|g\|_2)$ ,  $\|g\|_2^2 = \mathrm{tr}(gg^t)$ . For  $G = \mathrm{SL}_2(\mathbb{R})$ , its symmetric space is the hyperbolic plane with the standard metric of curvature  $-1$ , and  $l$  we defined above coincides with the hyperbolic metric. For example, consider the matrix  $g = \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}$ . Then  $l(g) = t$ , and  $\tilde{l}(g) = \frac{1}{2} \ln(e^t + e^{-t}) \approx \frac{t}{2}$ . In fact, for every  $g \in \mathrm{SL}_2(\mathbb{R})$ ,  $l(g) - 2\tilde{l}(g) = O(1)$ , and  $l$  and  $\tilde{l}$  are equal up to an additive constant.

For  $G = \mathrm{SL}_3(\mathbb{R})$ , it is no longer true. For

$$g_1 = \begin{pmatrix} e^{t/3} & & \\ & e^{t/3} & \\ & & e^{-2t/3} \end{pmatrix}, g_2 = \begin{pmatrix} e^{2t/3} & & \\ & e^{-t/3} & \\ & & e^{-t/3} \end{pmatrix},$$

it holds that  $l(g_1) = l(g_2) = 2t$ , while  $\tilde{l}(g_1) \approx t/3$ ,  $\tilde{l}(g_2) \approx 2t/3$ . So the two distances  $l, \tilde{l}$  are not equal up to an additive constant, but are only Lipschitz-equivalent, with  $\frac{3}{2}\tilde{l} \leq l + O(1) \leq 3\tilde{l}$ . However, if we chose  $\tilde{\tilde{l}}(g) = 2(\tilde{l}(g) + \tilde{l}(g^{-1}))$ , then  $l(g) - 2\tilde{\tilde{l}}(g) = O(1)$ . This solution no longer works for  $\mathrm{SL}_4(\mathbb{R})$ .

### 4.2. Growth of matrix coefficients and Harish-Chandra’s bounds

For  $2 \leq p \leq \infty$ , let  $\Xi_p(g) = \int \delta(gk)^{-1/p} dk$  be the  $p$ -th version of Harish-Chandra’s function. Let  $\Xi(g) = \Xi_2(g)$  be the standard Harish-Chandra’s function. Note that since  $\Xi(g)$  is left and right  $K$ -invariant it only depends on  $a \in A_+$  from the Cartan decomposition of  $g$ .

An explicit upper bound on Harish-Chandra’s function is given by the following theorem:

**Theorem 4.5.** *There is a constant  $C$  such that for every  $g \in G$  and  $\epsilon > 0$ ,*

$$\Xi_p(g) \leq \Xi^{2/p}(g) \leq q^{(1/2-1/p)l(g)} \Xi(g) \ll (l(g) + 1)^C q^{-l(g)/p} \ll_{\epsilon} q^{-l(g)(1/p-\epsilon)}.$$

*Proof.* By the Cartan decomposition, it suffices to verify the theorem for  $a \in A_+$ , where  $q^{l(a)} = \delta(a)$ . The first inequality follows from convexity, the second inequality follows from  $\Xi(g) \geq q^{-l(g)/2}$  (see equation (4.4)), and the fourth inequality is trivial. We are left with the third inequality.

For the Archimedean case, see [32, Theorem 3] or [43, Proposition 7.15]. For the non-Archimedean case, the standard reference is [61, 4.2.1], where there is an assumption that  $\mathrm{char} k = 0$ , but the same proof works in the general case with minor modifications. In any case, the general non-Archimedean result follows from the slightly more general results of [38, Theorem 28.3] for arbitrary affine buildings.  $\square$

A representation  $(\pi, V)$  of  $G$  is called tempered if  $p(\pi) \leq 2$ . The following theorem is the standard reference for upper bounds on matrix coefficients:

**Theorem 4.6** [14]. *Let  $(\pi, V)$  be a unitary irreducible tempered representation of  $G$ , and let  $v_1, v_2 \in V$  be two  $K$ -finite vectors such that  $\dim \text{span } Kv_1 = d_1$ ,  $\dim \text{span } Kv_2 = d_2$ . Then  $|\langle v_1, gv_2 \rangle| \leq \sqrt{d_1 d_2} \|v_1\| \|v_2\| \Xi(g)$ .*

The work [14] also provides a bound when  $p(\pi) > 2$ :

**Theorem 4.7** [14]. *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq 2k$ ,  $k \in \mathbb{N}$  and let  $v_1, v_2 \in V$  be two  $K$ -finite vectors such that  $\dim \text{span } Kv_1 = d_1$ ,  $\dim \text{span } Kv_2 = d_2$ . Then  $|\langle v_1, gv_2 \rangle| \leq \sqrt{d_1 d_2} \|v_1\| \|v_2\| \Xi^{1/k}(g)$ .*

This theorem is not sufficient for this work since we need more precise bounds when  $p(\pi) \notin 2\mathbb{N}$ . The following theorem contains a general upper bound that is good enough for all the applications of this paper.

**Theorem 4.8.** *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq p$ ,  $p \geq 2$  and let  $v_1, v_2 \in V$  be two  $K$ -finite vectors such that  $\dim \text{span } Kv_1 = d_1$ ,  $\dim \text{span } Kv_2 = d_2$ . Then*

$$|\langle v_1, gv_2 \rangle| \leq \sqrt{d_1 d_2} \|v_1\| \|v_2\| \Xi_p(g).$$

Let us discuss older similar results, slightly weaker than Theorem 4.8.

We first consider bounds on matrix coefficients of a single representation  $\pi$ . In the Archimedean case, we may consider the theory of leading exponents ([43, Chapter VIII]), which we describe in Subsection 9.1. In the non-Archimedean case, analogous results hold ([11, Section 4]).

**Theorem 4.9** (See Theorems 9.1,9.2). *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq p$ ,  $p \geq 2$ , and let  $v_1, v_2 \in V$  be two  $K$ -finite vectors. Then for every  $\epsilon > 0$ ,*

$$|\langle v_1, gv_2 \rangle| \ll_{v_1, v_2, \pi, \epsilon} q^{-l(g)^{(1/p-\epsilon)}}.$$

Next, we consider bounds on matrix coefficients when  $v_1, v_2$  are  $K$ -fixed. In such case (if  $v_1, v_2 \neq 0$ ), the representation is spherical and well understood, at least in the Archimedean case (see Subsection 4.4 below). From those bounds, we have:

**Theorem 4.10** (See [26, Section 3]). *Let  $G$  be Archimedean, and let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq p$ , and let  $v_1, v_2 \in V$  be two  $K$ -fixed vectors. Then for every  $\epsilon > 0$ ,*

$$|\langle v_1, \pi(g)v_2 \rangle| \leq \|v_1\| \|v_2\| \Xi(g) \delta^{1/2-1/p} \ll_{\epsilon} \|v_1\| \|v_2\| q^{-l(g)^{(1/p-\epsilon)}}.$$

Let us provide a proof of Theorem 4.8 when  $v_1, v_2$  are  $K$ -fixed. A bit more work can give bounds for  $K$ -finite vectors as well, using the ideas of [14]. Let us first prove the following lemma, which is based on the proof of [14, Theorem 2]. For  $f \in L^p(G)$  and  $g \in G$ , we let  $gf \in L^p(G)$  be  $gf(g') = f(g^{-1}g')$ .

**Lemma 4.11.** *If  $f_1 \in L^{p/(p-1)}(K \backslash G)$ ,  $f_2 \in L^p(K \backslash G)$  are two  $K$ -fixed functions, then  $|\langle f_1, gf_2 \rangle| \leq \|f_1\|_{p/(p-1)} \|f_2\|_p \Xi_p(g)$ .*

*Proof.* To avoid integrability questions, we assume that  $f_1, f_2 \in C_c(G)$  and deduce the theorem by density. Denote  $p' = \frac{p}{p-1}$ . For  $f \in C_c(G)$ , it holds that

$$\int_G f(x) dx = \int_K \int_P f(k\tilde{p}) \delta(\tilde{p}) d\tilde{p} dk,$$

so

$$\begin{aligned} |\langle f_1, g f_2 \rangle| &\leq \int_K \int_P |f_1(k\tilde{p})| |f_2(g^{-1}k\tilde{p})| \delta(\tilde{p}) d\tilde{p} dk = \\ &= \int_K \int_P |f_1(k\tilde{p})| \delta(\tilde{p})^{1/p'} |f_2(g^{-1}k\tilde{p})| \delta(\tilde{p})^{1/p} d\tilde{p} dk \\ &\leq \int_K \left( \int_P |f_1(k\tilde{p})|^{p'} \delta(\tilde{p}) d\tilde{p} \right)^{1/p'} \left( \int_P |f_2(g^{-1}k\tilde{p})|^p \delta(\tilde{p}) d\tilde{p} \right)^{1/p} dk. \end{aligned}$$

Since  $f_1$  is  $K$ -fixed, we have for every  $k \in K$ ,

$$\left( \int_P |f_1(k\tilde{p})|^{p'} \delta(\tilde{p}) d\tilde{p} \right)^{1/p'} = \|f_1\|_{p'}.$$

write  $g^{-1}k = k_0\tilde{p}_0$ . Then

$$\begin{aligned} \left( \int_P |f_2(g^{-1}k\tilde{p})|^p \delta(\tilde{p}) d\tilde{p} \right)^{1/p} &= \left( \int_P |f_2(k_0\tilde{p}_0\tilde{p})|^p \delta(\tilde{p}) d\tilde{p} \right)^{1/p} \\ &= \left( \int_P |f_2(\tilde{p}_0\tilde{p})|^p \delta(\tilde{p}_0\tilde{p}) d\tilde{p} \right)^{1/p} \delta(\tilde{p}_0)^{-1/p} \\ &= \left( \int_P |f_2(\tilde{p})|^p \delta(\tilde{p}) d\tilde{p} \right)^{1/p} \delta(\tilde{p}_0)^{-1/p} \\ &= \|f_2\|_p \delta(\tilde{p}_0)^{-1/p}. \end{aligned}$$

Since  $\delta$  is left  $K$ -fixed and  $\tilde{p}_0 = k_0^{-1}g^{-1}k$

$$\begin{aligned} \int_{K_0} \left( \int_P |f_2(g^{-1}k\tilde{p})|^p \delta(\tilde{p}) d\tilde{p} \right)^{1/p} dk &= \|f_2\|_p \int \delta(g^{-1}k)^{-1/p} dk = \\ &= \|f_2\|_p \Xi_p(g). \end{aligned} \quad \square$$

The lemma has a nice corollary: For  $a \in A_+$  let  $A_a$  be the operator  $A_a: C(K \backslash G) \rightarrow C(K \backslash G)$  defined by  $A_a f(g) = \int_K f(a^{-1}k^{-1}g) dk = \int_K \int_K f(k'^{-1}a^{-1}k^{-1}g) dk dk'$ . We may define  $A_g$  for  $g \in G$ , but it only depends on the  $A_+$  component of  $g$  from the Cartan decomposition. Since  $A_a$  is a translation followed by an average, its  $L^p$ -norm is bounded by 1 on  $L^p(G) \cap C(K \backslash G)$ , and therefore it defines an operator  $A_a: L^p(K \backslash G) \rightarrow L^p(K \backslash G)$ .

**Corollary 4.12.** *The norm of  $A_a: L^p(K \backslash G) \rightarrow L^p(K \backslash G)$  is bounded by  $\Xi_p(a)$ .*

*Proof.* Let  $p' = p/(p-1)$ . Let  $f \in L^p(K \backslash G)$ . Let  $f_1 \in L^{p'}(K \backslash G)$  with  $\|f_1\|_{p'} = 1$  be such that  $\langle f_1, A_a f \rangle = \|A_a f\|_p$ . Then since  $f_1$  is left  $K$ -invariant  $\langle f_1, A_a f \rangle = \langle f_1, a f \rangle$ . Applying Lemma 4.11, we have

$$\|A_a f\|_p \leq \Xi_p(a) \|f\|_p,$$

as needed. □

Note that if  $(\pi, V)$  is a unitary representation of  $G$ , we may define  $\pi(A_a): V^K \rightarrow V^K$  by the same arguments as above, as

$$\pi(A_a)v = \int_K \pi(ka)v \, dk = \int_K \int_K \pi(kak')v \, dkdk', \tag{4.2}$$

and  $\|A_a\|_V \leq 1$ . By a standard argument, this operation commutes with taking matrix coefficients – denote for  $v_1, v_2 \in V^K$ ,  $\varphi_{v_1, v_2}(g) = \langle v_1, \pi(g)v_2 \rangle$ ,  $\varphi_{v_1, v_2}(g) \in L^\infty(G)$  and then  $A_a\varphi_{v_1, v_2}(g) = \varphi_{A_a v_1, v_2}(g)$ .

**Corollary 4.13.** *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq p$ , and let  $v_1, v_2 \in V$  be two  $K$ -fixed vectors. Then*

$$|\langle \pi(g)v_1, v_2 \rangle| \leq \|v_1\| \|v_2\| \Xi_p(g).$$

*Proof.* We may assume that  $v_1, v_2$  are of norm 1. Since both of them are left  $K$ -invariant  $\langle \pi(g)v_1, v_2 \rangle = \langle \pi(A_{a_g})v_1, v_2 \rangle$ , where  $a_g \in A_+$  is the  $A_+$  component of the Cartan decomposition of  $g$ . If  $\pi$  is irreducible, it is well known that the subspace  $V^K$  of  $K$ -fixed vectors is one-dimensional ([43, Theorem 8.1], [10, Theorem 4.3]). Therefore,  $v_1$  is an eigenvector of  $\pi(A_{a_g})$  on  $V^K$ . Therefore,  $c_{v_1, v_2}(g)$  is an eigenvector of  $A_{a_g}$  on  $L^{p+\epsilon}(g)$  for every  $\epsilon > 0$ , with eigenvalue  $\langle \pi(g)v_1, v_2 \rangle$ . As the norm of  $A_{a_g}$  on  $L^{p+\epsilon}(K \backslash G)$  is bounded by  $\Xi_{p+\epsilon}(g)$ , each of its eigenvalues is bounded by  $\Xi_{p+\epsilon}(g)$  as well. Therefore,  $|\langle \pi(g)v_1, v_2 \rangle| \leq \Xi_{p+\epsilon}(g)$  for every  $\epsilon > 0$ . By taking  $\epsilon \rightarrow 0$ , we deduce  $|\langle v_1, \pi(A_g)v_2 \rangle| \leq \Xi_p(g)$ , as required.  $\square$

### 4.3. Upper bounds on operators

Let  $(\pi, V)$  be a unitary representation of  $G$ . For  $a \in A_+$ , let  $\pi(A_a): V^K \rightarrow V^K$  be from equation (4.2). Similarly, for  $d_0 \in \mathbb{R}_{\geq 0}$ , let  $\chi_{d_0}, \psi_{d_0} \in C_c^\infty(G)$  be as in Definition 3.1. Recall that  $\chi_{d_0}$  is a smooth approximation for the characteristic function of  $\{g : l(g) \leq d_0\}$  and  $\psi_{d_0}$  is a smooth approximation for  $q^{(d_0-l(g))/2} \chi_{d_0}$ .

For  $h \in C_c(G)$  we define as usual  $\pi(h)v = \int_G h(g)\pi(g)v \, dg$ .

**Corollary 4.14.** *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$  with  $p(\pi) \leq p$ , and let  $v \in V$  be a  $K$ -fixed vector. Then*

$$\begin{aligned} \|\pi(A_a)v\| &\ll_\epsilon q^{-l(a)(1/p-\epsilon)} \|v\| \\ \|\pi(\chi_{d_0})v\| &\ll_\epsilon q^{d_0(1-1/p+\epsilon)} \|v\| \\ \|\pi(\psi_{d_0})v\| &\ll_\epsilon q^{d_0(1-1/p+\epsilon)} \|v\|. \end{aligned}$$

*Proof.* The bound on  $\pi(A_a)$  follows from Corollary 4.13, using the explicit bounds of Theorem 4.5.

We will only prove the estimate for  $\psi_{d_0}$  since the proof for  $\chi_{d_0}$  is similar and a little easier.

$$\begin{aligned} \|\pi(\psi_{d_0})v\| &\leq \int_{l(g) \leq d_0+1} q^{(d_0-l(g))/2} \|\pi(g)v\| \, dg \\ &\ll_\epsilon \int_{l(g) \leq d_0+1} q^{d_0/2-l(g)(1/2+1/p-\epsilon)} \|v\| \, dg \\ &= \int_K \int_K \int_{l(a) \leq d_0+1} S(a) q^{d_0/2-l(k_1 a k_2)(1/2+1/p-\epsilon)} \|v\| \, dk_1 \, da \, dk_2 \end{aligned}$$

$$\begin{aligned} &\ll \int_{l(a) \leq d_0+1} q^{l(a)} q^{d_0/2-l(a)(1/2+1/p-\epsilon)} \|v\| da \\ &\ll \epsilon \int_{l(a) \leq d_0+1} q^{d_0/2+l(a)(1/2-1/p+\epsilon)} \|v\| da \\ &\leq \int_{l(a) \leq d_0+1} q^{d_0/2+d_0(1/2-1/p+\epsilon)} \|v\| da \\ &\ll \epsilon q^{d_0(1-1/p+\epsilon)} \|v\|. \end{aligned}$$

In the last line, we used the fact that  $\int_{l(a) \leq d_0+1} da \ll \epsilon q^{d_0 \epsilon}$ . □

#### 4.4. Spherical Functions

Let us recall the definition of spherical functions. For  $\lambda \in \mathfrak{a}_{\mathbb{C}}^* = \mathfrak{a}^* \otimes \mathbb{C}$  dominant, we let  $\lambda_P : P \rightarrow \mathbb{C}^\times$  be  $\lambda_P(mn) = q^{\lambda(\nu(m))} \delta^{-1/2}(m) = q^{(\lambda-\rho)(\nu(m))}$ . Extend it to  $\tilde{\lambda} : G \rightarrow \mathbb{C}$  by  $\tilde{\lambda}(kp) = \lambda_P(p)$ . Finally, define the spherical function

$$\varphi_\lambda(g) = \int_K \tilde{\lambda}(gk) dk.$$

Note that  $\varphi_0 = \Xi$  and  $\varphi_{(1-2/p)\rho} = \Xi_p$  for  $2 \leq p \leq \infty$ .

The theory of spherical functions was developed by Harish-Chandra in the Archimedean case and by Satake in the non-Archimedean case (see [26, Subsection 3.2] and the reference therein). We will need some very basic properties of spherical functions.

Let  $(\pi, V) \in \Pi(G)_{\text{sph}}$  be a unitary spherical representation, that is, a unitary irreducible representation with a nontrivial  $K$ -fixed vector. If  $v_1, v_2$  are  $K$ -fixed, then there exists a dominant  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$  such that

$$\langle v_1, \pi(g)v_2 \rangle = \langle v_1, v_2 \rangle \varphi_\lambda(g).$$

Upper bounds on spherical functions are given as follows. By [26, Lemma 3.3], it holds that for  $a \in A_+, k, k' \in K$  and  $\epsilon > 0$ ,

$$\varphi_\lambda(kak') = \varphi_\lambda(a) \ll q^{\text{Re } \lambda(\nu(a))} \Xi(a) \ll \epsilon q^{\text{Re } \lambda(\nu(a))} q^{-l(a)(1/2-\epsilon)}.$$

Let  $\omega_1, \dots, \omega_r \in \mathfrak{a}$  be the fundamental coweights. Since  $\lambda$  is dominant  $\text{Re } \lambda(\omega_i) \geq 0$  for  $1 \leq i \leq r$ . Using the above bounds, we deduce that

$$\text{Re } \lambda(\omega_i) \leq (1 - 2/p)\rho(\omega_i) \quad \forall 1 \leq i \leq r \implies p(\pi) \leq p. \tag{4.3}$$

See also [26, Lemma 3.2], which has a small misprint.

In the Archimedean case, the other direction of the implication of equation (4.3) is also true ([43, Theorem 8.48]), which gives a proof of Theorem 4.10 in this case.

However, the other direction is no longer true for the general non-Archimedean case. The simplest example is when the Bruhat–Tits building of  $G$  is a bipartite  $(d_0, d_1)$ -regular tree,  $d_0 \neq d_1$ , where for one of the choices of a maximal compact subgroup, there is a spherical function that is square integrable but  $\text{Re } \lambda(\omega) > 0$  (this result was first observed by [54]). In general, the other direction is true in the standard case according to MacDonalD ([52, Chapter V]), which excludes the case above. In particular, every group has a standard maximal compact subgroup for which equation (4.3) is an equivalence.

Lower bounds on spherical functions can be given in general if  $\text{Re } \lambda = \lambda$ . By Lemma 4.1, if  $\text{Re } \lambda = \lambda$  and  $\text{Re } \lambda(\omega_i) \leq \rho(\omega_i)$ , then for  $a \in A_+, k, k' \in K$

$$\varphi_\lambda(kak') = \varphi_\lambda(a) = \int_K \tilde{\lambda}(ak)dk \geq \int_K \tilde{\lambda}(a)dk = \lambda_P(a). \tag{4.4}$$

In the case when  $G$  is Archimedean, it is well known that the dominant  $\lambda \in \mathfrak{a}_\mathbb{C}^*$  which occur this way for unitary spherical representations must satisfy  $-\tilde{\lambda} = w\lambda$  for some  $w \in W$  ([44]).

If we moreover assume that  $G$  is of rank 1, then  $\mathfrak{a}_\mathbb{C}^* \cong \mathbb{C}$  and  $W = \{1, s\}$  acts by  $s\lambda = -\lambda$ . Therefore, the only dominant  $\lambda \in \mathfrak{a}_\mathbb{C}^*$  which may occur satisfy either  $\text{Re } \lambda = 0$  or  $\text{Re } \lambda = \lambda$ . In the case  $\text{Re } \lambda = 0$ , the corresponding unitary representation satisfies  $p(\pi) = 2$ . If  $\text{Re } \lambda = \lambda$ , then  $\lambda = \alpha\rho$  with  $\alpha \geq 0$ . By equation (4.3),  $\alpha \leq 1$ . Write  $\alpha = 1 - 1/p, 2 \leq p \leq \infty$ . Using equation (4.3) and the definition of  $\Xi_p(g)$ , we conclude that for  $p > 2$ , there is at most a single unitary irreducible representation  $(\pi, V)$  with a nontrivial  $K$ -fixed vector and  $p(\pi) = p$ . The corresponding spherical function is  $\Xi_p(g)$ . We conclude:

**Proposition 4.15.** *Let  $G$  be Archimedean of rank 1,  $p > 2$  and  $(\pi, V)$  a unitary irreducible representation of  $G$  with  $p(\pi) = p$ , having a non-trivial  $K$ -fixed vector  $v$ . Then  $\pi(A_a)v = \lambda_1 v, \pi(\chi_{d_0})v = \lambda_2 v$ , with*

$$e^{-l(a)/p} \leq \lambda_1$$

$$e^{d_0(1-1/p)} \ll \lambda_2 \text{ for } d_0 \geq 1.$$

Moreover, if  $h \in C_c(K \backslash G / K)$  satisfies  $h(g) \geq \chi_{d_0}(g)$  for every  $g \in G$ , then  $\pi(h)v = \lambda_h v$ , with  $\lambda_h \geq \lambda_2$ .

*Proof.* It is well known that in this case the set of all  $K$ -fixed vectors in  $V$  is one-dimensional and equals  $\text{span}\{v\}$  ([10, Theorem 4.3]). Since  $\pi(A_a)v, \pi(\chi_{d_0})v, \pi(h)v$  are  $K$ -fixed, we get that  $\pi(A_a)v = \lambda_1 v, \pi(\chi_{d_0})v = \lambda_2 v$  and  $\pi(h)v = \lambda_h v$ . Applying matrix coefficients, we see that

$$\lambda_1 = \Xi_p(a)$$

$$\lambda_2 = \int_G \chi_{d_0}(g)\Xi_p(g)dg$$

$$\lambda_h = \int_G h(g)\Xi_p(g)dg \geq \int_G \chi_{d_0}(g)\Xi_p(g)dg = \lambda_2.$$

Using equation (4.4), we get

$$\Xi_p(g) \geq e^{-\frac{2}{p}\rho(v(a))} = e^{-l(a)/p}.$$

And therefore  $\lambda_1 \geq e^{-l(a)/p}$  and for  $d_0 \geq 1$ ,

$$\lambda_2 = \int_G \chi_{d_0}(g)\Xi_p(g)dg \geq$$

$$\geq \int_0^{d_0} \sinh(t)e^{-t/p} dt \gg e^{(1-1/p)d_0}. \quad \square$$

A similar analysis can be done for the non-Archimedean rank 1 case, whenever the Bruhat–Tits tree of  $G$  is regular. However, when the Bruhat–Tits tree of  $G$  is not regular, the description of the spherical unitary dual is more complicated, and in particular, it is different for the two nonconjugate maximal compact subgroups. The analysis of this can be done by analyzing the slightly more general case of regular and biregular trees. See [33] or [39] for an analysis of this case.



Since the calculations are a bit long and are not the main focus of this work, we skip them and state the final result. Recall that we have a map  $\nu: M/M^0 \rightarrow \mathfrak{a}$  with a discrete image. We identify its image with  $\mathbb{Z}$ .

**Proposition 4.16.** *Let  $G$  be non-Archimedean of rank 1,  $p > 2$  and  $(\pi, V)$  a unitary irreducible representation of  $G$  with  $p(\pi) = p$ , having a nontrivial  $K$ -fixed vector  $v$ . Then for  $a \in A_+$  with  $\nu(a) \in \mathbb{Z}$  even, we have  $\pi(A_a)v = \lambda_1 v$ , with*

$$q^{-l(a)/p} \ll \lambda_1.$$

We can finally conclude:

**Theorem 4.17.** *Let  $G$  be of rank 1. Then the set  $\Pi(G)_{\text{sph,nt}}$  of spherical nontempered unitary representations has a good family of functions.*

*Proof.* For  $G$  Archimedean, we choose  $f_{d_0} = \chi_{d_0/2} * \chi_{d_0/2}$ . For  $G$  non-Archimedean, we choose  $a \in A_+$  with  $\nu(a) \in \mathbb{Z}$  even, and with  $l(a) = d_0/2 \pm O(1)$ , and choose  $f_{d_0} = A_a * A_a^*$ .

The first property of a good family follows from Proposition 4.15, Proposition 4.16 and the Convolution Lemma 4.18 below.

The second and third properties follow from simple properties of convolutions. □

### 4.5. Convolution of Operators

In this section, we analyze the function  $\chi_{d_0} * \chi_{d_0}(g)$ . Similar analysis can be found for rank 1 in [58, Lemma 3.1]. Our analysis is less accurate but is more abstract and works for every rank.

**Lemma 4.18** (Convolution Lemma). *It holds that  $c_{d_0} = \chi_{d_0} * \chi_{d_0} \in C_c^\infty(K \backslash G / K)$  and satisfies the inequality*

$$c_{d_0}(g) \ll_\epsilon q^{d_0 \epsilon} \psi_{2d_0+2}(g).$$

*The same bound holds for the convolution of every  $f_1, f_2 \in C_c(K \backslash G / K)$  such that  $f_1(g), f_2(g) \ll_\epsilon q^{d_0 \epsilon} \chi_{d_0}(g)$ .*

*Proof.* The second statement is a consequence of the first, so we only need to prove the first statement.

The idea is to look at the action (by right convolution) of  $\chi_{d_0}$  on  $L^2(G)$ . By Lemma 4.11 and the same arguments as in Corollary 4.12, the norm of  $\chi_{d_0}$  on  $L^2(G)$  is bounded by  $\ll_\epsilon q^{d_0(1/2+\epsilon)}$ . Therefore, the norm of  $c_{d_0}$  on  $L^2(G)$  is bounded by  $\ll_\epsilon q^{d_0(1+\epsilon)}$ . Now, we need to use some continuity arguments to deduce pointwise bounds.

Notice that since  $\chi_{d_0} \in C_c^\infty(K \backslash G / K)$ , the same is true for  $c_{d_0}$ . In the non-Archimedean case, the arguments are simpler: We look at the action of  $c_{d_0}$  on the characteristic function  $\mathbf{1}_K$  of  $K$ . Then

$$\|\mathbf{1}_K * c_{d_0}\|_{L^2(G)}^2 \ll_\epsilon q^{2d_0(1+\epsilon)} \|\mathbf{1}_K\|_{L^2(G)} = q^{2d_0(1+\epsilon)}.$$

But if  $c_{d_0}(g) = R$  then  $\mathbf{1}_K * c_{d_0}(g) = R$ , so

$$\|\mathbf{1}_K * c_{d_0}\|_{L^2(G)}^2 \geq \mu(KgK)R^2 \gg q^{l(g)}R^2.$$

Therefore,  $c_{d_0}(g) = R \ll_\epsilon q^{d_0(1+\epsilon)-l(g)/2}$  as needed.

In the Archimedean case, assume that  $c_{d_0}(g) = R$ . Then  $c_{d_0+2}(g') \geq R$  for every  $g' \in G$  with  $|l(g) - l(g')| \leq 1$ . We consider  $\mathbf{1}_{B_1}$ , where  $B_1$  is the ball of radius 1 around the identity. It holds that

$$\|\mathbf{1}_{B_1} * c_{d_0+2}\|_{L^2(G)}^2 \ll_\epsilon q^{2d_0(1+\epsilon)} \|\mathbf{1}_{B_1}\|_{L^2(G)} \ll q^{2d_0(1+\epsilon)}.$$

It also holds that  $\mathbf{1}_{B_1} * c_{d_0+2}(g) \gg R$ , for  $g'$  satisfying  $|l(g) - l(g')| \leq 1$ , so

$$\|\mathbf{1}_K * c_{d_0+2}\|_{L^2(G)}^2 \gg \mu(KB_1gK)R^2 \gg q^{l(g)}R^2,$$

and  $c_{d_0}(g) = R \ll_\epsilon q^{d_0(1+\epsilon)-l(g)/2}$  as needed. □

#### 4.6. Traces of operators on irreducible unitary representations

Our goal in this section is to relate the lower and upper bounds of the previous sections to lower and upper bounds on traces.

Let us recall how to define traces of an operator  $h \in C_c(G)$  on a unitary irreducible representation  $(\pi, V)$ . We will only consider the case when  $h$  is left and right  $K$ -finite, which simplifies the theory. In such case, there is an orthogonal projection  $e_h : C(K) \rightarrow C(K)$ , with a finite-dimensional image, such that  $h = e_h * h * e_h$  (see, e.g., [14, proof of Theorem 2]).

Since  $\pi$  is admissible, the image of  $\pi(h)$  is finite-dimensional and therefore is trace class ([43, Chapter X]), with trace given by

$$\text{tr } \pi(h) = \sum_i \langle u_i, \pi(h)u_i \rangle,$$

where  $\{u_i\} \subset V$  is an orthonormal basis.

By uniform admissibility ([43, Theorem 10.2], [3]) and the fact the image of  $\pi(h)$  is supported on a finite number of  $K$ -types, the image of  $\pi(h)$  is of bounded dimension, depending only on the projection  $e_h$ .

Finally, recall that the norm of a finite-dimensional operator is larger than the largest absolute value of an eigenvalue. We conclude:

**Proposition 4.19.** *Assume that  $h \in C_c(G)$  is left and right  $K$ -finite and  $(\pi, V) \in \Pi(G)$ . Then*

$$|\text{tr } \pi(h)| \ll_{e_h} \|\pi(h)\|,$$

*the bound depending only on the projection  $e_h$  such that  $e_h * h * e_h$ .*

As an example, if  $h \in C_c(K \backslash G / K)$  is left and right  $K$ -invariant, the image of  $\pi(h)$  is of dimension 1 or 0. If the dimension is 0, obviously  $\text{tr } \pi(h) = \pi(h) = 0$ . If the dimension is 1,  $V$  has a  $K$ -invariant vector  $v \in V$ ,  $\|v\| = 1$  and

$$|\text{tr } \pi(h)| = \|\pi(h)\| = |\langle v, \pi(h)v \rangle|.$$

One may also deduce lower bounds on traces of nonnegative self-adjoint operators. It follows from the same considerations as in the finite-dimensional case.

**Proposition 4.20.** *Assume that  $h \in C_c(G)$  is left and right  $K$ -finite and  $(\pi, V) \in \Pi(G)$ . Moreover, assume that  $\pi(h)$  is self-adjoint and nonnegative. Then*

$$\|\pi(h)\| = \sup_{v: \|v\|=1} \langle v, \pi(h)v \rangle \leq \text{tr } \pi(h).$$

#### 4.7. The pretrace formula

Let us recall the pretrace formula ([25, Chapter 1]). Let  $h \in C_c(G)$ , and let  $\Gamma \subset G$  be a cocompact lattice. Denote by  $\hat{h} \in C_c(G)$  the function  $\hat{h}(g) = h(g^{-1})$ . Then we have an operator  $h : L^2(\Gamma \backslash G) \rightarrow L^2(\Gamma \backslash G)$ , acting on  $f \in L^2(\Gamma \backslash G)$  by

$$\begin{aligned} (hf)(x) &= f * \hat{h}(x) = \int_G f(xg)h(g)dg = \int_G f(y)h(x^{-1}y)dy \\ &= \int_{\Gamma \backslash G} f(y) \left( \sum_{\gamma \in \Gamma} h(x^{-1}\gamma y) \right) dy = \int_{\Gamma \backslash G} K(x, y) f(y) dy, \end{aligned}$$

for  $K(x, y) = \sum_{\gamma \in \Gamma} h(x^{-1}\gamma y)$ . By [25, Chapter 1], if  $h$  is also self-adjoint, then it is trace-class on  $L^2(\Gamma \backslash G)$  and

$$\text{tr } h|_{L^2(\Gamma \backslash G)} = \int_{\Gamma \backslash G} \sum_{\gamma \in \Gamma} h(x^{-1}\gamma x) dx.$$

Moreover, if  $h \in C_c^\infty(G)$  and  $L^2(\Gamma \backslash G) \cong \bigoplus_{\pi \in \Pi(G)} m(\pi, \Gamma)$  is the decomposition into irreducible representations, then we have the pretrace formula:

$$\begin{aligned} \text{tr } h|_{L^2(\Gamma \backslash G)} &= \int_{\Gamma \backslash G} \sum_{\gamma \in \Gamma} h(x^{-1}\gamma x) dx \\ &= \sum_{\pi \in \Pi(G)} m(\pi, \Gamma) \text{tr } \pi(h), \end{aligned}$$

where  $\text{tr}(\pi(h))$  is the usual trace on the representation space  $(\pi, V)$ .

The following lemma is immediate from the pretrace formula but essential for our work.

**Lemma 4.21.** *If  $h_1, h_2 \in C_c(G)$  satisfy that  $h_1(g) \geq h_2(g)$  for every  $g \in G$  then  $\text{tr } h_1|_{L^2(\Gamma \backslash G)} \geq \text{tr } h_2|_{L^2(\Gamma \backslash G)}$ . In particular, if  $h_1(g) \geq 0$  then  $\text{tr } h_1|_{L^2(\Gamma \backslash G)} \geq 0$ .*

#### 4.8. Spectral decomposition of a characteristic function of a small ball

For  $x \in \Gamma_N \backslash \Gamma_1 \subset X_N = \Gamma_N \backslash G/K$ , let  $b_{x, \delta} \in L^2(X_N)$  be defined as follows:

- o In the non-Archimedean case, choose  $h_\delta \in C_c(K \backslash G/K)$  to be the characteristic function of  $K$ .
- o In the Archimedean case, choose  $0 < \delta < 1/4$  such that  $l(\gamma) > \delta$  for every  $\gamma \in \Gamma_1$  with  $l(\gamma) > 0$ . Choose a function  $h_\delta \in C_c^\infty(K \backslash G/K)$  such that:
  - $0 \leq h_\delta(g) \leq \frac{2}{\mu(B_\delta(e))}$  for all  $g \in G$ , where  $B_\delta(e)$  is the ball of radius  $\delta$  around the identity  $e \in G$ .
  - $h_\delta(g)$  is supported on  $\{g \in G : l(g) \leq \delta\}$ .
  - $\int_G h_\delta(g) dg = 1$
  - $h(g) = h(g^{-1})$ , that is,  $h = \hat{h}$ .

Finally, let  $b_{x, \delta} \in L^2(X_N)$  be

$$b_{x, \delta}(y) = \sum_{\gamma \in \Gamma_N} h_\delta(x^{-1}\gamma y).$$

We fix  $\delta > 0$  once and for all (depending on  $\Gamma_1$ ) and suppress the dependence on it from now on. We notice that

$$\int_{X_N} b_{x, \delta}(y) dy = 1.$$

By the properties of  $\delta$ , the sum defining it is over at most  $|\Gamma_1 \cap K|$  elements, and therefore (recall that we suppress the dependence on  $\Gamma_1$  from our notations),

$$\|b_{x, \delta}\|_\infty \ll 1, \quad \|b_{x, \delta}\|_2 \ll 1.$$

Let us remark that for our uses in the Archimedean rank 1 case one can simply choose instead

$$h_\delta(g) = \begin{cases} \frac{1}{\mu(B_\delta(e))} & l(g) \leq \delta \\ 0 & \text{else} \end{cases},$$

and for higher rank, we make this choice so that one can apply the Paley–Wiener theorem for spherical functions due to Harish-Chandra, which is used as follows:

**Lemma 4.22.** *Let  $f \in L^2(X_N)$ . Then,*

$$\sum_{x \in \Gamma_N \backslash \Gamma_1} |\langle b_{x,\delta}, f \rangle|^2 \ll \|f\|_2^2.$$

*Moreover, in the Archimedean case, if  $f \in L^2(X_N)$  is the  $K$ -fixed function of some irreducible representation  $\pi \subset L^2(\Gamma_N \backslash G)$  with  $\lambda(\pi) = \lambda$ , then for every  $L' > 0$ ,*

$$\sum_{x \in \Gamma_N \backslash \Gamma_1} |\langle b_{x,\delta}, f \rangle|^2 \ll_{L'} (1 + \lambda)^{-L'} \|f\|_2^2.$$

*(The last result will only be used in the Archimedean rank  $\geq 2$  case.)*

*Proof.* By our assumption on  $\delta$ , the balls  $B_\delta(x)$  for  $x \in \Gamma_N \backslash \Gamma_1$  are all either equal (with multiplicity at most  $|\Gamma_1 \cap K|$ ) or distinct, so by Cauchy–Schwartz

$$\begin{aligned} \sum_{x \in \Gamma_N \backslash \Gamma_1} |\langle f, b_{x,\delta} \rangle|^2 &\leq \sum_{x \in \Gamma_N \backslash \Gamma_1} \|f|_{B_\delta(x)}\|_2^2 \|b_{x,\delta}\|_2^2 \leq |\Gamma_1 \cap K| \|f\|_2^2 \max_{x \in \Gamma_N \backslash \Gamma_1} \|b_x\|_2^2 \\ &\ll_\delta \|f\|_2^2. \end{aligned}$$

For the moreover part, note that

$$\langle f, b_{x,\delta} \rangle = f(x) \operatorname{tr} \pi(h_\delta).$$

It is well known (see [60] for an exact statement) that there is a constant  $M > 0$  such that

$$|f(x)| \ll (1 + \lambda)^M \|f|_{B_\delta(x)}\|_2^2.$$

By the Paley–Wiener theorem for spherical functions ([18, Subsection 3.4]),

$$\operatorname{tr} \pi(h_\delta) \ll_{L'} (1 + \lambda)^{-L'-M}.$$

Combining both estimates we get the required inequality. □

### 5. The weak injective radius property

In this section, we will study the weak injective radius property and deduce spectral results from it when the lattices are cocompact.

Consider  $\chi_{d_0} \in C_c^\infty(G)$  from Section 3. It is self-adjoint since  $l(g) = l(g^{-1})$ . Since  $\chi_{d_0}$  is left and right  $K$ -invariant, it acts on  $L^2(\Gamma \backslash G/K)$ .

Note that, by the definition of  $N(\Gamma, d_0, y)$ , it holds that for every  $x \in \Gamma_n \backslash \Gamma_1$ ,

$$N(\Gamma, d_0, x) \leq \sum_{\gamma \in \Gamma} \chi_{d_0}(x^{-1}\gamma x) \leq N(\Gamma, d_0 + 1, x). \tag{5.1}$$

If  $\chi_d$  had been the characteristic function of  $\{g \in G : l(g) \leq d\}$ , then the first inequality in equation (5.1) would simply be an equality.

Our initial observation is:

**Lemma 5.1.** For every  $x_0 \in \Gamma_N \setminus \Gamma$

$$\langle b_{x_0, \delta}, \chi_{d_0-2} b_{x_0, \delta} \rangle \ll N(\Gamma_N, d_0, x_0) \ll \langle b_{x, \delta}, \chi_{d_0+2} b_{x_0, \delta} \rangle.$$

*Proof.* By unfolding, we get that

$$\begin{aligned} \langle b_{x, \delta}, \chi_{d_0} b_{x_0, \delta} \rangle &= \int_{\Gamma_N \setminus G} \int_{\Gamma_N \setminus G} \sum_{\gamma \in \Gamma_N} b_{x_0, \delta}(x) \chi_d(x^{-1} \gamma y) b_{x_0, \delta}(y) dx dy \\ &= \int_{\Gamma_N \setminus G} \int_{\Gamma_N \setminus G} \sum_{\gamma_1 \in \Gamma_N} h_\delta(x_0^{-1} \gamma_1 x) \sum_{\gamma \in \Gamma_N} \chi_d(x^{-1} \gamma y) \sum_{\gamma_2 \in \Gamma_N} h_\delta(x_0^{-1} \gamma_2 y) dx dy \\ &= \int_{\Gamma_N \setminus G} \int_{\Gamma_N \setminus G} \sum_{\gamma_1 \in \Gamma_N} h_\delta(x_0^{-1} \gamma_1 x) \sum_{\gamma \in \Gamma_N} \chi_d(x^{-1} \gamma y) \sum_{\gamma_2 \in \Gamma_N} h_\delta(y^{-1} \gamma_2 x_0) dx dy \\ &= \int_G \int_G h_\delta(x_0^{-1} x) \sum_{\gamma \in \Gamma_N} \chi_d(x^{-1} \gamma y) h_\delta(y^{-1} x_0) dx dy \\ &= \sum_{\gamma \in \Gamma_N} (h_\delta * \chi_d * h_\delta)(x_0^{-1} \gamma x_0). \end{aligned}$$

The lemma follows from equation (5.1), and the simple pointwise estimates

$$\chi_{d-2}(g) \ll h_\delta * \chi_d * h_\delta(g) \ll \chi_{d+2}(g). \quad \square$$

We can now prove that some bounds on  $N(\Gamma_N, d_0, x)$  imply bounds on  $N(\Gamma_N, d_0 + d, x)$ .

**Lemma 5.2.** Assume that for some  $0 \leq d_0, x \in \Gamma_N \setminus \Gamma_1$  it holds for some  $M$  that for every  $d \leq d_0$

$$N(\Gamma_N, d, x) \ll M q^{d/2}.$$

Then for every  $d_1 \geq 0, \epsilon > 0$  it holds that

$$N(\Gamma_N, d_0 + d_1, x) \ll_\epsilon M q^{(d_0+d_1)\epsilon} q^{d_0/2+d_1}.$$

*Proof.* We choose  $\tilde{d} = (d_0 - 4)/2$  and calculate  $\|\chi_{\tilde{d}} b_{x, \delta}\|_2^2$  using Lemma 4.18 and Lemma 5.1:

$$\begin{aligned} \|\chi_{\tilde{d}} b_{x, \delta}\|_2^2 &= \langle \chi_{\tilde{d}} b_{x, \delta}, \chi_{\tilde{d}} b_{x, \delta} \rangle \\ &= \langle \chi_{\tilde{d}} * \chi_{\tilde{d}} b_{x, \delta}, b_{x, \delta} \rangle \\ &\ll_\epsilon q^{d_0 \epsilon} \langle \psi_{2\tilde{d}+2} b_{x, \delta}, b_{x, \delta} \rangle \\ &\ll q^{d_0 \epsilon} \int_0^{d_0-4} q^{(d_0-d)/2} \langle \chi_{d+2} b_{x, \delta}, b_{x, \delta} \rangle dd \end{aligned}$$

$$\begin{aligned} &\ll q^{d_0\epsilon} \int_0^{d_0-4} q^{(d_0-d)/2} N(\Gamma_N, d+4, x) dd \\ &\ll q^{d_0\epsilon} M \int_0^{d_0-4} q^{(d_0-d)/2} q^{d/2} dd \\ &\ll Mq^{d_0\epsilon} q^{\bar{d}} \leq Mq^{d_0\epsilon} q^{d_0/2}. \end{aligned}$$

Now, for  $d_1 \geq 0$ , we use the inequality

$$\chi_{d_0+d_1+2}(g) \ll (\chi_{\bar{d}} * \chi_{d_1+4} * \chi_{\bar{d}})(g),$$

to deduce

$$\begin{aligned} N(\Gamma, d_0 + d_1, x) &\ll \langle \chi_{d_0+d_1+2} b_{x, \delta}, b_{x, \delta} \rangle \\ &\ll \langle \chi_{\bar{d}} * \chi_{d_1+4} * \chi_{\bar{d}} b_{x, \delta}, b_{x, \delta} \rangle \\ &= \langle \chi_{d_1+4} * \chi_{\bar{d}} b_{x, \delta}, \chi_{\bar{d}} b_{x, \delta} \rangle \\ &\leq \|\pi(\chi_{d_1+4})\| \|\chi_{\bar{d}} b_{x, \delta}\|_2^2 \\ &\ll_\epsilon q^{d_1(1+\epsilon)} \|\chi_{\bar{d}} b_{x, \delta}\|_2^2 \\ &\ll q^{(d_0+d_1)\epsilon} Mq^{d_0/2+d_1}. \end{aligned} \quad \square$$

Lemma 5.2 allows us to slightly modify the definition of the weak injective radius property. In particular, the following two claims are equivalent to each other and the weak injective radius property with parameter  $\alpha$ , for  $C$  some fixed constant (say  $C = 100$ ):

- o For every  $d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N]) - C, \epsilon > 0$ ,

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{x \in \Gamma_N \setminus \Gamma} N(\Gamma_N, d_0, x) dx \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon q^{d_0/2}.$$

- o For every  $d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N]) + C, \epsilon > 0$ ,

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{x \in \Gamma_N \setminus \Gamma} N(\Gamma_N, d_0, x) dx \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon q^{d_0/2}.$$

It also implies the following proposition, which should be compared with [58, Conjecture 2]:

**Proposition 5.3.** *Let  $(\Gamma_N)$  be a sequence of lattices. Assume that the weak injective radius property holds with parameter  $\alpha = 1$ , that is, for every  $0 \leq d_0 \leq 2 \log_q([\Gamma_1 : \Gamma_N]), \epsilon > 0$ ,*

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{y \in \Gamma_N \setminus \Gamma_1} N(\Gamma_N, d_0, y) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon q^{d_0(1/2+\epsilon)}.$$

Then for every  $d_0 \geq 0, \epsilon > 0$  it holds that

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{y \in \Gamma_N \setminus \Gamma_1} N(\Gamma_N, d_0, y) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon q^{d_0\epsilon} \left( \frac{q^{d_0}}{[\Gamma_1 : \Gamma_N]} + q^{d_0/2} \right).$$

*Proof.* It is sufficient to prove that for  $d_1 \geq 0, d_0 = 2 \log_q([\Gamma_1 : \Gamma_N])$ , it holds that

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{y \in \Gamma_N \setminus \Gamma_1} N(\Gamma_N, d_0 + d_1, y) \ll_\epsilon q^{(d_0+d)\epsilon} q^{d_0/2+d_1}.$$

This follows from Lemma 5.2. □

We can now present the proof of Proposition 3.3, in the following slightly more general claim.

**Corollary 5.4.** *Let  $(\Gamma_N)$  be a sequence of cocompact lattices, and  $0 < \alpha \leq 1$ . The following are equivalent:*

1. For every  $d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N]), \epsilon > 0$ ,

$$\text{tr } \chi_{d_0}|_{L^2(X_N)} = \int_{\Gamma_N \setminus G} \sum_{\gamma \in \Gamma_N} \chi_{d_0}(x^{-1}\gamma x) dx \ll_\epsilon [\Gamma_1 : \Gamma_N]^{1+\epsilon} q^{d_0(1/2+\epsilon)}.$$

2. The weak injective radius property with parameter  $\alpha$  – for every  $d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N]), \epsilon > 0$ ,

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{x \in \Gamma_N \setminus \Gamma} N(\Gamma_N, d_0, x) dx \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon q^{d_0(1/2+\epsilon)}.$$

3. For every  $h \in C_c(G)$  self-adjoint and satisfying  $h(g) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon \psi_{\tilde{d}}(g)$  for  $\tilde{d} = 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ , it holds that

$$\text{tr } h|_{L^2(X_N)} \ll_\epsilon [\Gamma_1 : \Gamma_N]^{1+\epsilon} q^{\tilde{d}(1/2+\epsilon)} \asymp [\Gamma_1 : \Gamma_N]^{1+\alpha+\epsilon}. \tag{5.2}$$

4. For  $\tilde{d} = 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ , it holds that

$$\text{tr } \psi_{\tilde{d}}|_{L^2(X_N)} \ll_\epsilon [\Gamma_1 : \Gamma_N]^{1+\epsilon} q^{\tilde{d}(1/2+\epsilon)} \asymp [\Gamma_1 : \Gamma_N]^{1+\alpha+\epsilon}.$$

*Proof.* (3) obviously implies (4), and the fact that (4) implies (3) is a result of Lemma 4.21.

Since  $\Gamma_1$  is cocompact, it has a finite diameter  $D$ . It implies that the balls of radius  $D$  around points in  $\Gamma_N \setminus \Gamma_1$  cover the entire space  $\Gamma_N \setminus G$ . Moreover, if two points  $x, y$  are of distance  $d$  apart, then

$$\sum_{\gamma \in \Gamma_N} \chi_{d_0-2d-1}(x^{-1}\gamma x) \leq \sum_{\gamma \in \Gamma_N} \chi_{d_0}(y^{-1}\gamma y) \leq \sum_{\gamma \in \Gamma_N} \chi_{d_0+2d+1}(x^{-1}\gamma x).$$

Using equation (5.1), we deduce that if  $B_{x_0,D}$  is the ball of radius  $D$  around  $x_0 \in \Gamma_N \setminus \Gamma_1$ , then for every  $d_0$ ,

$$\int_{B_{x_0,D}} \sum_{\gamma \in \Gamma_N} \chi_{d_0-2D-2}(x^{-1}\gamma x) dx \leq N(\Gamma_N, d_0, x) \leq \int_{B_{x_0,D}} \sum_{\gamma \in \Gamma_N} \chi_{d_0+2D+2}(x^{-1}\gamma x) dx.$$

Summing over  $x_0 \in \Gamma_N \setminus \Gamma_1$  and using the discussion after Lemma 5.2 to change  $d_0$  by a constant allows us to deduce the equivalence between (1) and (2).

To show that (4) implies (1), note that for  $d_0 \leq \tilde{d} = 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ , it holds that  $\chi_{d_0} \ll q^{(d_0-\tilde{d})/2} \psi_{\tilde{d}}(g)$ . Then if (4) holds, then



$$\begin{aligned} \text{tr } \chi_{d_0}|_{L^2(X_N)} &\ll q^{(d_0-\bar{d})/2} \text{tr } \psi_{\bar{d}}|_{L^2(X_N)} \\ &\ll q^{(d_0-\bar{d})/2} [\Gamma_1 : \Gamma_N]^{1+\epsilon} q^{\bar{d}(1/2+\epsilon)} \\ &\ll [\Gamma_1 : \Gamma_N]^{1+\epsilon} q^{d_0(1/2+\epsilon)}. \end{aligned}$$

Finally, we prove that (1) implies (4). Note that for every  $g \in G$ ,

$$\psi_{\bar{d}}(g) \ll \int_0^{\bar{d}} q^{(\bar{d}-d_0)/2} \chi_{d_0}(g) dd_0.$$

Then if (1) holds for every  $d_0 \leq \bar{d}$ , we have

$$\begin{aligned} \text{tr } \psi_{\bar{d}} &\ll \int_0^{\bar{d}} q^{(\bar{d}-d_0)/2} \text{tr } \chi_{d_0}|_{L^2(X_N)} dd_0 \\ &\ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{1+\epsilon} \int_0^{\bar{d}} q^{(\bar{d}-d_0)/2} q^{d_0(1/2+\epsilon)} dd_0 \\ &= [\Gamma_1 : \Gamma_N]^{1+\epsilon} \bar{d} q^{\bar{d}(1/2+\epsilon)} \\ &\ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{1+\epsilon} q^{\bar{d}(1/2+\epsilon)}. \end{aligned} \quad \square$$

Another simple property of the weak injective radius is:

**Proposition 5.5.** *Let  $(\Gamma_N)$  be a sequence of lattices. If  $\alpha$  is the weak injective radius parameter of the sequence, then  $\alpha \leq 1$ .*

*Proof.* For simplicity, we assume that the lattices  $(\Gamma_N)$  are cocompact, so we can use equation (5.2). The arguments can change to deal with the nonuniform case as well.

Note that Corollary 5.4 did not assume that  $\alpha \leq 1$ . It is therefore enough to prove that for every  $\alpha > 1$  equation (5.2) from Corollary 5.4 does not hold for  $\alpha$ .

Let  $d_0 = 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ ,  $d_1 = d_0/2 - 1$  and  $c_{d_1} = \chi_{d_1} * \chi_{d_1}$ . By Lemma 4.18,  $c_{d_1} \ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{\epsilon} \psi_{d_0}(g)$ , so the condition before equation (5.2) holds, and

$$\text{tr } c_{d_1}|_{L^2(X_N)} \ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{1+\alpha+\epsilon}.$$

Since  $\chi_{d_1}$  is self-adjoint,  $\text{tr } \pi(c_{d_1}) \geq 0$  for every  $\pi \in \Pi(G)$ , so

$$\text{tr } c_{d_1}|_{L^2(X_N)} \geq \text{tr}(\pi_{\text{triv}}(c_{d_0})),$$

where  $\pi_{\text{triv}}$  is the trivial representation. On the other hand, it holds that

$$\text{tr}(\pi_{\text{triv}}(c_{d_0})) = \int_G c_{d_1}(g) dg = \left( \int_G \chi_{d_1} dg \right)^2 \gg q^{2d_1} \gg [\Gamma_1 : \Gamma_N]^{2\alpha}.$$

As  $\alpha > 1$ , we get a contradiction for  $[\Gamma_1 : \Gamma_N]$  big enough. □

We can now prove Proposition 3.9:

*Proof.* Recall that we assume that the sequence  $(\Gamma_N)$  of cocompact lattices satisfies the weak injective radius property with parameter  $\alpha$ , let  $d_0 = \alpha \log_q([\Gamma_1 : \Gamma_N])$  and let  $f_{d_0}$  from the definition of a good family.

By Corollary 5.4, the trace of  $f_{d_0}$  on  $L^2(\Gamma_N \backslash G)$  satisfies

$$\text{tr } f_{d_0}|_{L^2(\Gamma_N \backslash G)} \ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{1+\alpha+\epsilon}.$$

Let us calculate the spectral side of the trace. From the second and first properties of a good family,

$$\begin{aligned} \text{tr } f_{d_0}|_{L^2(\Gamma_N \backslash G)} &\geq \sum_{\pi \in A} m(\pi, \Gamma_N) \text{tr } \pi(f_{d_0}) \\ &\gg_{\epsilon, A} \sum_{\pi \in A} m(\pi, \Gamma_N) q^{d_0(1-1/p(\pi)-\epsilon)} \\ &= \sum_{\pi \in A} m(\pi, \Gamma_N) [\Gamma_1 : \Gamma_N]^{2\alpha(1-1/p(\pi)-\epsilon)} \\ &\geq M(A, \Gamma_N, p) [\Gamma_1 : \Gamma_N]^{2\alpha(1-1/p-\epsilon)}. \end{aligned}$$

We deduce that for every  $N, p > 2, \epsilon > 0$ ,

$$M(A, \Gamma_N, p) \ll_{A, \epsilon} [\Gamma_1 : \Gamma_N]^{1-\alpha(1-2/p)+\epsilon},$$

as needed. □

### 6. The weak injective radius property implies the optimal lifting property

In this section, we prove Theorem 1.5.

#### 6.1. Reduction to a spectral argument

Recall that assuming the weak injective radius property and spectral gap, we should prove that for every  $\epsilon > 0$ , for every  $a \in A_+$  with  $l(a) \geq (1 + \epsilon) \log_q(\mu(X_N))$ ,

$$\#\{(x, y) \in (\Gamma_1/\Gamma_N)^2 : \exists \gamma \in \Gamma_1 \text{ s.t. } \pi_N(\gamma)x = y, \|a_\gamma - a\|_a < \epsilon \|a\|_a\} = (1 - o_\epsilon(1)) [\Gamma_1 : \Gamma_N]^2. \tag{6.1}$$

For  $(x, y) \in (\Gamma_1/\Gamma_N)^2, a \in A_+$  and  $\epsilon > 0$ , we say that  $\gamma \in \Gamma_1$  is *good* for  $(x, y, a, \epsilon)$  if  $\pi_N(\gamma)x = y$  and  $\|a_\gamma - a\|_a < \epsilon \|a\|_a$ .

**Lemma 6.1.** *Let  $(x, y) \in (\Gamma_1/\Gamma_N)^2$ , and assume that there is no good  $\gamma \in \Gamma_1$  for  $(x, y, a, \epsilon)$ . Identify  $x, y$  with elements  $x, y \in X_N = \Gamma_N \backslash G/K$ . Let  $f_a \in L^1(G)$  be a function supported on the set  $\{g \in G : \|a_g - a\|_a < \epsilon/2 \|a\|_a\}$ , and for  $\delta$  small enough with respect to  $\epsilon \|a\|_a$  let  $b_{x, \delta}$  as in Subsection 4.8. Then*

$$f_a b_{x, \delta}(y) = 0.$$

Moreover, for every  $y' \in B_\delta(y)$  it holds that

$$f_a b_{x, \delta}(y') = 0.$$

*Proof.* We think of  $f_a b_{x, \delta}$  as a left  $\Gamma_N$ -invariant function on  $G$ , and identify  $x, y$  with some lifts of them in  $\Gamma_1 \subset G$ . The support of  $f_a b_{x, \delta} = b_{x, \delta} * \hat{f}_a$  is contained in the set

$$\{y' \in G : \exists \gamma \in \Gamma_N, x' \in G, f_a(y'^{-1}\gamma x') > 0, d(x, x') \leq \delta\}.$$

Assume by contradiction that  $y' \in B_\delta(y)$  is in the support of  $f_a b_{x, \delta}$ , and let  $\gamma \in \Gamma_N, x' \in G$  be such that  $f_a(y'^{-1}\gamma x') > 0, d(x, x') \leq \delta$ . By the assumption on the support of  $f_a$ ,

$$\|a_{y'^{-1}\gamma x'} - a\|_a < \epsilon/2 \|a\|_a.$$

Look at

$$\gamma' = y^{-1}\gamma x \in \Gamma_1.$$

Then  $a_{\gamma'} = a_{y^{-1}\gamma x}$ , and if  $\delta$  is small enough with respect to  $\epsilon \|a\|_a$ , then  $\|a_{y^{-1}\gamma x} - a_{y'^{-1}\gamma x'}\|_a < \epsilon/2 \|a\|_a$ . Therefore,

$$\|a_{\gamma'} - a\|_a < \epsilon \|a\|_a,$$

and

$$\pi_N(\gamma')x = x\gamma'^{-1} = \gamma y.$$

Since  $\gamma \in \Gamma_N$ , it says that  $\gamma'$  sends  $x \in \Gamma_N \setminus \Gamma_1$  to  $y \in \Gamma_N \setminus \Gamma_1$ , as needed. □

Let  $\pi \in L^2(X_N)$  be the uniform probability distribution, that is,  $\pi(x) = \frac{1}{\mu(X_N)}$ .

**Lemma 6.2.** *Equation (6.1) holds if for every  $\epsilon > 0$ , for every  $a \in A_+$  with  $l(a) > (1 + \epsilon) \log_q([\Gamma_1 : \Gamma_N])$ , there exists a probability function  $f_a \in C_c^\infty(G)$  supported on  $\{g \in G : \|a_g - a\|_a < \epsilon \|a\|_a\}$  such that*

$$\sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta} - \pi\|_2^2 = o_\epsilon(1). \tag{6.2}$$

*Proof.* We assume that equation (6.2) holds and want to prove that equation (6.1) holds.

Let  $\epsilon > 0$ . For  $x, y \in (\Gamma_1/\Gamma_N)^2$ , by Lemma 6.1, if there is no good  $\gamma$  for  $(x, y, a, \epsilon/2)$ , then for  $y'$  in the  $\delta$ -neighborhood of  $y$  it holds that  $|(f_a b_{x,\delta} - \pi)(y')| = \pi(y') = \frac{1}{\mu(X_N)} \asymp [\Gamma_1 : \Gamma_N]^{-1}$ . Therefore, for a fixed  $x \in \Gamma_1/\Gamma_N$  each  $y$  without good  $\gamma$  contributes  $\gg_\delta [\Gamma_1 : \Gamma_N]^{-2}$  to  $\|f_a b_{x,\delta} - \pi\|_2^2$ . Moreover, the contributions are distinct for  $y, y'$  whose image in  $X_N = \Gamma_N \setminus G/K$  is different.

Therefore,

$$\begin{aligned} & \#\{(x, y) \in (\Gamma_1/\Gamma_N)^2 : \text{There is no good } \gamma \text{ for } (x, y, a, \epsilon/2)\} \\ & \ll \sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta} - \pi\|_2^2 [\Gamma_1 : \Gamma_N]^2 \\ & = [\Gamma_1 : \Gamma_N]^2 \sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta} - \pi\|_2^2 \\ & = o([\Gamma_1 : \Gamma_N]^2), \end{aligned}$$

where we used equation (6.2) in the last step. This implies equation (6.1) for  $\epsilon/2$ . □

The following lemma explains where spectral gap is used.

**Lemma 6.3.** *Equation (6.1) follows from the following two conditions:*

1. *Spectral gap holds for  $(\Gamma_N)$ .*
2. *For every  $\epsilon > 0$ , for some  $\delta > 0$ , for every  $a \in A_+$  with  $l(a) \geq (1 + \epsilon) \log_q([\Gamma_1 : \Gamma_N])$ , there exists a probability function  $f_a \in C_c^\infty(G)$  supported on  $\{g \in G : \|a_g - a\|_a < \epsilon \|a\|_a\}$ , such that for every  $\epsilon_1 > 0$ ,*

$$\sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta}\|_2^2 \ll_{\epsilon, \epsilon_1} q^{\epsilon_1 l(a)}. \tag{6.3}$$

*Proof.* First, we note that  $\|\pi\|_2^2 = \int_{X_N} \mu(X_N)^{-2} dx = \mu(X_N)^{-1} \asymp [\Gamma_1 : \Gamma_N]^{-1}$ . Therefore, if equation (6.3) holds, then also

$$\sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta} - \pi\|_2^2 \leq \sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta}\|_2^2 + \sum_{x \in \Gamma_1/\Gamma_N} \|\pi\|_2^2 \ll_{\epsilon, \epsilon_1} q^{\epsilon l(a)},$$

which is similar to equation (6.2), but  $o(1)$  is replaced with  $O_{\epsilon, \epsilon_1}(q^{\epsilon l(a)})$ .

Let  $\epsilon' > 0$ , and let  $a \in A_+$  be such that  $l(a') = \epsilon' l(a)$ . Let  $f'_a = A_{a'} * f_a$ . Assuming  $\epsilon'$  is small enough,  $f'_a$  is supported on

$$\{g \in G : g \in G : \|a_g - a\|_a < 2\epsilon \|a\|_a\}.$$

We will show that equation (6.2) holds for  $f'_a$ .

Notice that  $A_{a'}\pi = \pi$  and  $f_a b_{x,\delta} - \pi \perp \pi$ . By the spectral gap assumption and Corollary 4.14, for some  $p' < \infty$ ,

$$\begin{aligned} \|f'_a b_{x,\delta} - \pi\|_2 &= \|A_{a'}(f_a b_{x,\delta} - \pi)\|_2 \\ &\ll q^{-\epsilon' l(a)/p'} \|f_a b_{x,\delta} - \pi\|_2. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{x \in \Gamma_1/\Gamma_N} \|f'_a b_{x,\delta} - \pi\|_2^2 &\ll q^{-\epsilon' l(a)/p'} \sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta} - \pi\|_2^2 \\ &\ll_{\epsilon, \epsilon_1} q^{-\epsilon' l(a)/p'} q^{\epsilon l(a)}. \end{aligned}$$

If we choose  $\epsilon_1$  small enough, this is  $o(1)$  and equation (6.2) holds. Applying Lemma 6.2, we get that equation (6.1) holds as well. □

### 6.2. Completing the proof of Theorem 1.5

Recall that Theorem 1.5 states that spectral gap and the weak injective radius property imply the optimal lifting property. In Lemma 6.3, we reduced it to some spectral statement, equation (6.3). We now claim:

**Lemma 6.4.** *Assume that the weak injective radius property holds for a sequence  $(\Gamma_n)$ . Then for every  $\epsilon > 0$  sufficiently small, for some  $\delta > 0$ , for every  $a \in A_+$  with  $l(a) \geq \log_q([\Gamma_1 : \Gamma_N])$ , there is a probability function  $f_a \in C_c^\infty(G)$  supported on  $\{g \in G : \|a_g - a\|_a < \epsilon\}$  such that*

$$\sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta}\|_2^2 \ll_{\epsilon, \epsilon_1} q^{\epsilon l(a)}.$$

Notice the function  $f_a$  in Lemma 6.4 satisfies slightly stronger conditions than required by Lemma 6.3, so together they imply Theorem 1.5.

*Proof.* Let  $\epsilon > 0$  be sufficiently small. By a standard argument, there is a smooth probability function  $f_a \in C_c^\infty(G)$  supported on  $\{g \in G : \|a_g - a\|_a < \epsilon\}$  and satisfies  $f_a(x) \ll q^{-l(a)} \chi_{l(a)+1}(x)$ .

Therefore, by the Convolution Lemma 4.18 and Lemma 5.1,

$$\begin{aligned} \|f_a b_{x,\delta}\|_2^2 &\ll_\epsilon q^{-2l(a)} \|\chi_{l(a)+1} b_{x,\delta}\|_2^2 = q^{-2l(a)} \langle \chi_{l(a)+1} b_{x,\delta}, \chi_{l(a)+1} b_{x,\delta} \rangle \\ &= q^{-2l(a)} \langle \chi_{l(a)+1} * \chi_{l(a)+1} b_{x,\delta}, b_{x,\delta} \rangle \\ &\ll_{\epsilon_1} q^{-2(1-\epsilon_1)l(a)} \langle \psi_{2(l(a)+2)} b_{x,\delta}, b_{x,\delta} \rangle \\ &\ll q^{-2(1-\epsilon_1)l(a)} \int_0^{2(l(a)+2)} q^{(2l(a)-d_0)/2} \langle \chi_{d_0} b_{x,\delta}, b_{x,\delta} \rangle dd_0. \\ &\ll q^{-2(1-\epsilon_1)l(a)} \int_0^{2(l(a)+2)} q^{(2l(a)-d_0)/2} N(x, \Gamma_N, d_0 + 2) dd_0. \end{aligned}$$

Therefore, applying Proposition 5.3, we get

$$\begin{aligned} \sum_{x \in \Gamma_1/\Gamma_N} \|f_a b_{x,\delta}\|_2^2 &\ll_{\epsilon_1} q^{-2(1-\epsilon_1)l(a)} \int_0^{2(l(a)+2)} q^{(2l(a)-d_0)/2} (q^{d_0} + q^{d_0/2} [\Gamma_1 : \Gamma_N]) dd_0 \\ &\ll_{\epsilon_1} q^{l(a)\epsilon_1} (1 + q^{-l(a)} [\Gamma_1 : \Gamma_N]). \end{aligned}$$

Since  $l(a) \geq \log_q([\Gamma_1 : \Gamma_N])$ , this is  $O(q^{l(a)\epsilon_1})$  as needed. □

## 7. The spectral to geometric direction

### 7.1. Some technical calculations

For ease of reference, we give here a couple of technical bounds.

**Lemma 7.1.** *Let  $G$  be non-Archimedean or rank 1. The following are equivalent for a sequence  $(\Gamma_N)$ :*

1. *The spherical density hypothesis with parameter  $\alpha$ : For every  $\epsilon > 0$  and  $p > 2$*

$$M(\Pi(G)_{\text{sph}}, \Gamma_N, p) \ll_\epsilon [\Gamma_1 : \Gamma_N]^{1-\alpha(1-2/p)+\epsilon}.$$

2. *For every  $\epsilon > 0$*

$$\sum_{\pi \in \Pi(G)_{\text{sph}}, p(\pi) > 2} [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p(\pi))} m(\pi, \Gamma_N) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon.$$

*Proof.* The fact that (2) implies (1) is simple and is left to the reader.

The fact that (1) implies (2) follows from a standard trick of integration by parts ([31, Theorem 421]):

$$\begin{aligned} &\sum_{\pi \in \Pi(G)_{\text{sph}}, p(\pi) > 2} [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p(\pi))} m(\pi, \Gamma_N) \\ &= \lim_{p \rightarrow 2, p > 2} M(\Pi(G)_{\text{sph}}, \Gamma_N, p) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} \\ &\quad + \int_2^\infty M(\Pi(G)_{\text{sph}}, \Gamma_N, p) \frac{\partial}{\partial p} ([\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)}) dp \\ &= \lim_{p \rightarrow 2, p > 2} M(\Pi(G)_{\text{sph}}, \Gamma_N, p) [\Gamma_1 : \Gamma_N]^{-1} \end{aligned}$$

$$\begin{aligned}
 &+ \int_2^\infty M(\Pi(G)_{\text{sph}}, \Gamma_N, p) 2p^{-2} \alpha \ln([\Gamma_1 : \Gamma_N]) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} dp \\
 &\ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon \left( 1 + \int_2^\infty p^{-2} dp \right) \asymp [\Gamma_1 : \Gamma_N]^\epsilon. \quad \square
 \end{aligned}$$

For the higher-rank Archimedean case we have:

**Lemma 7.2.** *Let  $G$  be Archimedean. The following are equivalent for a sequence  $(\Gamma_N)$ :*

1. *The spherical density hypothesis with parameter  $\alpha$ : for some  $L > 0$  large enough and for every  $\lambda \geq 0, N \geq 1, p > 2, \epsilon > 0,$*

$$M(\Pi(G)_{\text{sph}}, \Gamma_N, p, \lambda) \ll_\epsilon (1 + \lambda)^L [\Gamma_1 : \Gamma_N]^{1-\alpha(1-2/p)+\epsilon}$$

2. *For some  $L' > 0$  large enough and every  $\epsilon > 0,$*

$$\sum_{\pi \in \Pi(G)_{\text{sph}}, p(\pi) > 2} [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p(\pi))} (1 + \lambda(\pi))^{-L'} m(\pi, \Gamma_N) \ll_\epsilon [\Gamma_1 : \Gamma_N]^\epsilon. \quad (7.1)$$

*Proof.* The fact that (2) implies (1) is again simple and is left to the reader.

The fact that (1) implies (2) is again done by integration by parts, with two variables. Let us state it formally. If  $((p_i, \lambda_i))_{i=1}^\infty \subset (2, \infty) \times (0, \infty)$  is a sequence of values without limit points,  $f(p, \lambda)$  is a nonnegative smooth function, and  $M'(p, \lambda) = \#\{i : p_i \geq p, \lambda_i \leq \lambda\}$  then whenever everything absolutely converges,

$$\begin{aligned}
 \sum_i f(p_i, \lambda_i) &= \lim_{p \rightarrow 2} \lim_{\lambda \rightarrow \infty} M'(p, \lambda) f(p, \lambda) - \lim_{p \rightarrow 2} \int_0^\infty M'(p, \lambda) \frac{\partial}{\partial \lambda} f(p, \lambda) d\lambda \\
 &+ \lim_{\lambda \rightarrow \infty} \int_2^\infty M'(p, \lambda) \frac{\partial}{\partial p} f(p, \lambda) dp \\
 &- \int_0^\infty \int_2^\infty M'(p, \lambda) \frac{\partial}{\partial \lambda} \frac{\partial}{\partial p} f(p, \lambda) dp d\lambda.
 \end{aligned}$$

Applying the integration by parts formula to the left-hand side of equation (7.1), with  $M'(p, \lambda) = M(\Pi(G)_{\text{sph}}, \Gamma_N, p, \lambda), f(p, \lambda) = [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} (1 + \lambda)^{-L'}$ , we get

$$\begin{aligned}
 L.H.S &= \lim_{p \rightarrow 2} \lim_{\lambda \rightarrow \infty} M(\Pi(G)_{\text{sph}}, \Gamma_N, p, \lambda) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} (1 + \lambda)^{-L'} \\
 &+ \lim_{p \rightarrow 2} \int_0^\infty M(\Pi(G)_{\text{sph}}, \Gamma_N, p, \lambda) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} L' (1 + \lambda)^{-L'-1} d\lambda \\
 &+ \lim_{\lambda \rightarrow \infty} \int_2^\infty M(\Pi(G)_{\text{sph}}, \Gamma_N, p, \lambda) 2\alpha p^{-2} \ln([\Gamma_1 : \Gamma_N]) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} (1 + \lambda)^{-L'} dp \quad (7.2) \\
 &+ \int_0^\infty \int_2^\infty M(\Pi(G)_{\text{sph}}, \Gamma_N, p, \lambda) 2\alpha p^{-2} \ln([\Gamma_1 : \Gamma_N]) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} L' (1 + \lambda)^{-L'-1} dp d\lambda.
 \end{aligned}$$

Note that by the spherical density hypothesis,

$$\lim_{p \rightarrow 2} M(\Pi(G)_{\text{sph}}, \Gamma_N, p, \lambda) \ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{1+\epsilon} (1 + \lambda)^L.$$

A similar and more precise bound may be derived directly from Weyl’s law ([18]). Note also that we may assume that  $L' > L + 10$ .

Therefore, the first summand in equation (7.2) is bounded by

$$\ll_{\epsilon} \lim_{\lambda \rightarrow \infty} [\Gamma_1 : \Gamma_N]^{1+\epsilon} [\Gamma_1 : \Gamma_N] (1 + \lambda)^L (1 + \lambda)^{-L'} = 0.$$

The second summand is bounded similarly by

$$\ll_{\epsilon} \int_0^{\infty} [\Gamma_1 : \Gamma_N]^{1+\epsilon} [\Gamma_1 : \Gamma_N]^{-1} (1 + \lambda)^L L' (1 + \lambda)^{-L'-1} d\lambda \ll [\Gamma_1 : \Gamma_N]^{\epsilon}.$$

The third summand is bounded by

$$\begin{aligned} &\ll_{\epsilon} \lim_{\lambda \rightarrow \infty} \int_2^{\infty} (1 + \lambda)^L [\Gamma_1 : \Gamma_N]^{1-\alpha(1-2/p)+\epsilon} 2\alpha p^{-2} \ln([\Gamma_1 : \Gamma_N]) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} (1 + \lambda)^{-L'} dp \\ &\ll \lim_{\lambda \rightarrow \infty} (1 + \lambda)^{L-L'} \int_2^{\infty} [\Gamma_1 : \Gamma_N]^{\epsilon} p^{-2} dp = 0. \end{aligned}$$

The final summand is bounded by

$$\begin{aligned} &\ll_{\epsilon} \int_0^{\infty} \int_2^{\infty} (1 + \lambda)^L [\Gamma_1 : \Gamma_N]^{1-\alpha(1-2/p)+\epsilon} 2\alpha p^{-2} \ln([\Gamma_1 : \Gamma_N]) [\Gamma_1 : \Gamma_N]^{-1+\alpha(1-2/p)} L' (1 + \lambda)^{-L'-1} dp d\lambda \\ &\ll [\Gamma_1 : \Gamma_N]^{\epsilon} \int_0^{\infty} L' (1 + \lambda)^{L-L'-1} d\lambda \int_2^{\infty} p^{-2} dp \ll [\Gamma_1 : \Gamma_N]^{\epsilon}. \end{aligned}$$

By combining all the bounds we get equation (7.1). □

### 7.2. Proof of Theorem 1.6

In this subsection, we prove Theorem 1.6, or more explicitly we prove that the spherical density hypothesis with parameter  $\alpha$  implies the weak injective radius property with parameter  $\alpha$ . The most natural proof of the claim is to analyze the spectral side of the pretrace formula for the function  $\chi_{d_0}$ . We will instead discretize and prove directly the weak injective radius property.

*Proof.* Recall that we should prove that for every  $d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ ,  $\epsilon > 0$ ,

$$\frac{1}{[\Gamma_1 : \Gamma_N]} \sum_{x \in \Gamma_1/\Gamma_N} N(\Gamma_N, d_0, x) dx \ll_{\epsilon} [\Gamma_1 : \Gamma_N]^{\epsilon} q^{d_0(1/2+\epsilon)}.$$

For  $x \in \Gamma_N \setminus \Gamma_1$ , let  $b_{x,\delta} \in L^2(X_N)$  as in Subsection 4.8.

Recall from Lemma 5.1 that

$$N(X_N, d_0, x) \ll \langle b_{x,\delta}, \chi_{d_0+2} b_{x,\delta} \rangle.$$

Let  $\{\pi_i\}_{i=1}^T$  be an orthogonal basis of irreducible subrepresentations of  $L^2(\Gamma_N \backslash G)$  with  $K$ -fixed vectors and  $p(\pi_i) > 2$  ( $T$  is finite in the  $p$ -adic or real rank 1 case, otherwise  $T$  may be  $\infty$ ). Recall that the set of  $K$ -invariant vectors of each irreducible representation  $\pi_i$  is a one-dimensional vector space. Let  $u_i \in L^2(X_N)$  be a  $K$ -invariant vector in  $\pi_i$  with  $\|u_i\| = 1$ . Let  $p_0 = 2$ , and let  $V_0$  the orthogonal complement of  $\text{span}\{\pi\} \oplus (\oplus_i \text{span}\{u_i\})$  in  $L^2(X_N)$  ( $\pi$  here is the uniform probability function on  $X_N$ ). Note that the  $G$ -representation generated by  $V_0$  is (2-tempered).

Decompose  $b_{x,\delta} = \pi + v_{0,x} + v_{1,x} + \dots$ , according to the decomposition  $L^2(X_N) = \text{span}\{\pi\} \oplus V_0 \oplus \text{span } u_1 \oplus \dots$ , that is, for  $i = 1, 2, \dots, T$ ,  $v_{i,x} = \langle u_i, b_{x,\delta} \rangle u_i$ . Then,

$$\begin{aligned} \sum_{x \in \Gamma_1 / \Gamma_N} \langle b_{x,\delta}, \chi_{d_0+1} b_{x,\delta} \rangle &= \sum_{x \in \Gamma_1 / \Gamma_N} \langle \pi, \chi_{d_0+1} \pi \rangle + \\ &\sum_{x \in \Gamma_1 / \Gamma_N} \langle v_{0,x}, \chi_{d_0+1} v_{0,x} \rangle + \sum_{x \in \Gamma_1 / \Gamma_N} \sum_{i=1}^T \langle v_{i,x}, \chi_{d_0+1} v_{i,x} \rangle. \end{aligned} \tag{7.3}$$

The first summand in equation (7.3) equals

$$\begin{aligned} \lambda_{\text{triv}}(\chi_{d_0+C})[\Gamma_1 : \Gamma_N] \|\pi\|_2^2 &\ll_\epsilon q^{d_0(1+\epsilon)} [\Gamma_1 : \Gamma_N] \mu^{-1}(X_N) \ll q^{d_0(1+\epsilon)} \\ &\ll q^{d_0(1/2+\epsilon)} [\Gamma_1 : \Gamma_N], \end{aligned}$$

where  $\lambda_{\text{triv}}(\chi_{d_0+C})$  is the trivial eigenvalue of  $\chi_{d_0+C}$ , and we used the fact that  $d_0 \leq 2\alpha \log_q([\Gamma_1 : \Gamma_N])$ .

Since  $V_0$  spans a tempered representation, by Corollary 4.14 the second summand in equation (7.3) is bounded by

$$\ll_\epsilon q^{d_0(1/2+\epsilon)} \sum_{x \in \Gamma_1 / \Gamma_N} \|v_{0,x}\|_2^2 \ll q^{d_0(1/2+\epsilon)} [\Gamma_1 : \Gamma_N] \|b_{x,\delta}\|_2^2 \ll q^{d_0(1/2+\epsilon)} [\Gamma_1 : \Gamma_N].$$

To analyze the final summand in equation (7.3), we first assume that  $G$  is  $p$ -adic or rank 1. By Lemma 4.22,  $\sum_{x \in \Gamma_1 / \Gamma_N} \|v_{i,x}\|_2^2 \ll_\delta 1$ . Write  $d_0 = 2\alpha' \log_q([\Gamma_1 : \Gamma_N])$ , for  $\alpha' \leq \alpha$ . Then using Corollary 4.14,

$$\begin{aligned} \sum_{x \in \Gamma_1 / \Gamma_N} \sum_{i=1}^T \langle v_{i,x}, \chi_{d_0+1} v_{i,x} \rangle &\ll_\epsilon \sum_{x \in \Gamma_1 / \Gamma_N} \sum_{i=1}^T q^{d_0(1-1/p(\pi_i)+\epsilon)} \|v_{i,x}\|_2^2 \\ &\ll \sum_{i=1}^T q^{d_0(1-1/p(\pi_i)+\epsilon)} \\ &\ll \sum_{\pi \in \Pi(G)_{\text{sph}, p(\pi) > 2}} m(\pi, \Gamma_N) q^{d_0(1-1/p(\pi)+\epsilon)} \\ &= \sum_{\pi \in \Pi(G)_{\text{sph}, p(\pi) > 2}} m(\pi, \Gamma_N) [\Gamma_1 : \Gamma_N]^{2\alpha'(1-1/p(\pi)+\epsilon)}. \end{aligned}$$

Applying Lemma 7.1 (for  $\alpha' \leq \alpha$ ) and arranging, we get

$$\ll_\epsilon [\Gamma_1 : \Gamma_N]^{1+\alpha'+\epsilon} = q^{d_0(1/2+\epsilon)} [\Gamma_1 : \Gamma_N].$$

This finishes the proof of the non-Archimedean and the rank 1 case.



For the Archimedean high-rank case, by Lemma 4.22, for  $L$  large enough

$$\sum_{x \in \Gamma_1/\Gamma_N} \|v_{i,x}\|_2^2 \ll_L (1 + \lambda(\pi_i))^{-L}.$$

The rest of the argument is as above, but using Lemma 7.2 instead of Lemma 7.1. □

**Remark 7.3.** The proof of Theorem 1.6 for hyperbolic spaces actually works for noncompact quotients as well. The reason is that the entire continuous spectrum is tempered, that is, contained in  $V_0$  ([46]). For more general results about hyperbolic surfaces, see [28].

### 7.3. A strong version of Theorem 1.5, assuming the spherical density hypothesis

The goal of this subsection is to prove the following theorem:

**Theorem 7.4.** *Let  $(\Gamma_N)$  be a sequence satisfying the spherical density hypothesis (with parameter  $\alpha = 1$ ) and spectral gap. Then for every  $\epsilon > 0$ , for every  $a \in A_+$  with  $l(a) \geq (1 + \epsilon) \log_q(\mu(X_N))$ ,*

$$\#\{(x, y) \in (\Gamma_1/\Gamma_N)^2 : \exists \gamma \in \Gamma_1 \text{ s.t. } \pi_N(\gamma)x = y, \|a_\gamma - a\|_a < 1\} = (1 - o_\epsilon(1))[\Gamma_1 : \Gamma_N]^2.$$

The result of Theorem 7.4 is stronger than in Theorem 1.5 – here, we determine the  $A_+$ -component of the Cartan decomposition of  $\gamma$  in far greater precision. In the non-Archimedean case, it says that we may choose  $a \in A_+$  precisely.

*Proof.* As before, let  $\pi \in L^2(X_N)$ , be the uniform probability distribution, that is,  $\pi(x) = \frac{1}{\mu(X_N)}$ .

Using the same arguments as in Lemma 6.2, to prove Theorem 7.4 it suffices to prove under its assumptions that for  $a \in A_+$  with  $l(a) > (1 + \epsilon) \log_q([\Gamma_1 : \Gamma_N])$

$$\sum_{x \in \Gamma_1/\Gamma_N} \|A_a b_{x,\delta} - \pi\|_2^2 = o_{\epsilon,\delta}(1). \tag{7.4}$$

Let us first show that equation (7.4) is immediate if  $M(\Pi(G)_{\text{sph}}, \Gamma_N, p) = 0$  for  $p > 2$  (the Ramanujan case). If  $l(a) > (1 + \epsilon) \log_q([\Gamma_1 : \Gamma_N])$  and then  $q^{-l(a)} \ll [\Gamma_1 : \Gamma_N]^{-(1+\epsilon)}$ . Then if we apply Corollary 4.14, for  $\epsilon'$  sufficiently small,

$$\begin{aligned} \|A_a b_{x_0,\delta} - \pi\|_2^2 &= \|A_a(b_{x_0,\delta} - \pi)\|_2^2 \ll_{\epsilon'} q^{-l(a)(1+\epsilon')} \|b_{x_0,\delta}\|_2^2 \\ &\ll [\Gamma_1 : \Gamma_N]^{-(1+\epsilon)(1-\epsilon')} = o([\Gamma_1 : \Gamma_N]^{-1}). \end{aligned} \tag{7.5}$$

Summing over  $x \in \rho_N^{-1}(x_0)$ , we get the required bound. As a matter of fact, we proved the stronger result that if  $M(X_N, p) = 0$  for  $p > 2$  then for every  $x \in \Gamma_1/\Gamma_N$

$$\#\{y \in \Gamma_1/\Gamma_N : \exists \gamma \in \Gamma_1 \text{ s.t. } \pi_N(\gamma)x = y, \|a_\gamma - a\|_a < \delta\} = (1 - o_{\epsilon,\delta}(1))[\Gamma_1 : \Gamma_N].$$

The proof in the general case is basically the same as the proof of Theorem 1.6 in the previous subsection. Let us quickly give the differences in the proofs, while using the same notations.

The decomposition of  $b_{x,\delta}$  is the same, but instead of bounding

$$\sum_{x \in \Gamma_1/\Gamma_N} \langle b_{x,\delta}, \chi_{d_0+1} b_{x,\delta} \rangle,$$

we bound for  $l(a) > (1 + \epsilon) \log_q([\Gamma_1 : \Gamma_N])$

$$\sum_{x \in \Gamma_1/\Gamma_N} \|A_a(b_{x_0,\delta} - \pi)\|_2^2.$$

We get in the same way

$$\sum_{x \in \Gamma_1/\Gamma_N} \|A_a(b_{x_0, \delta} - \pi)\|_2^2 \ll \sum_{x \in \Gamma_1/\Gamma_N} \|A_a v_{0,x}\|_2^2 + \sum_{x \in \Gamma_1/\Gamma_N} \sum_{i=1}^T \|A_a v_{i,x}\|_2^2. \tag{7.6}$$

Instead of using the bound on  $\chi_{d_0}$ , we use the bound on  $A_a$ , which is very similar. For  $\epsilon'$  small enough, the first summand of equation (7.6) is bounded by

$$\begin{aligned} \sum_{x \in \Gamma_1/\Gamma_N} \|A_a v_{0,x}\|_2^2 &\ll_{\epsilon'} \sum_{x \in \Gamma_1/\Gamma_N} q^{-l(a)(1-\epsilon')} \|v_{0,x}\|_2^2 \\ &\ll [\Gamma_1 : \Gamma_N]^{-(1+\epsilon)(1-\epsilon')} [\Gamma_1 : \Gamma_N] \\ &= o(1). \end{aligned}$$

The second summand of equation (7.6) is bounded in the  $p$ -adic or rank 1 case by

$$\begin{aligned} \sum_{x \in \Gamma_1/\Gamma_N} \sum_{i=1}^T \|A_a v_{i,x}\|_2^2 &\ll_{\epsilon'} \sum_{x \in \Gamma_1/\Gamma_N} \sum_{i=1}^T q^{-l(a)(2/p(\pi_i)-\epsilon')} \|v_{i,x}\|_2^2 \\ &\ll \sum_{i=1}^T [\Gamma_1 : \Gamma_N]^{-(1+\epsilon)(2/p(\pi_i)-\epsilon')} [\Gamma_1 : \Gamma_N] \\ &\ll [\Gamma_1 : \Gamma_N]^{-\epsilon(2/p'-\epsilon')} \sum_{\pi \in \Pi(G)_{\text{sph}, p(\pi) > 2}} m(\pi, \Gamma_N) [\Gamma_1 : \Gamma_N]^{(1-2/p(\pi)+\epsilon')}, \end{aligned}$$

where  $p'$  satisfies  $p(\pi_i) \leq p'$  for every  $i$ , by the spectral gap assumption. Using Lemma 7.1 we get for  $\epsilon' > 0$  small enough relative to  $\epsilon$ ,

$$\sum_{x \in \Gamma_1/\Gamma_N} \sum_{i=1}^T \|A_a v_{i,x}\|_2^2 \ll_{\epsilon'} [\Gamma_1 : \Gamma_N]^{-\epsilon(2/p'-\epsilon')} [\Gamma_1 : \Gamma_N]^{\epsilon'} = o(1),$$

as needed.

The proof of the Archimedean high-rank case is also similar. □

### 8. The Bernstein theory of nonbacktracking operators

The results of this section are the main technical contribution of this work. We restrict to the case of a *non-Archimedean* field, and we base our work on the results of [3].

Let  $K' \subset K$  be a compact open subgroup. For  $g \in G$ , it holds that  $\mu(KgK)[K : K']^{-2} \leq \mu(K'gK') \leq \mu(KgK)$ , and therefore  $\mu(K'gK') \asymp_{K'} \mu(KgK) \asymp q^{l(g)}$ .

Consider the Hecke algebra  $H_{K'} = \mathbb{C}_c(K' \backslash G / K')$ . For  $g \in K' \backslash G / K'$ , denote  $h_{K',g} = \frac{1}{\mu(K')} K'gK'$ , and let  $q_{K',g} = \mu(K'gK')\mu^{-1}(K')$  be the number of right (or left)  $K'$  cosets in  $K'gK'$ . It holds that  $q_{K',g} \asymp_{K'} q^{l(g)}$ . By the representation theory of  $p$ -adic groups ([10]), given a smooth representation  $V$  of  $G$ , the Hecke algebra  $H_{K'}$  acts on the  $K'$ -fixed vectors  $V^{K'}$  of  $V$ .

We first discuss the Iwahori–Hecke algebra. Let  $I \subset K$  be the Iwahori–Hecke subgroup, that is, the pointwise stabilizer of a chamber in the Bruhat–Tits building of  $G$ . Let  $W$  be the affine Weyl group of the root system of  $G$  (relative to the maximal  $k$ -torus  $T$ ) and  $\hat{W}$  the extended affine Weyl group. By the Iwahori decomposition, we have  $G = I\hat{W}I$ , where  $W \subset \hat{W} \subset \hat{W}$  is some intermediate subgroup. For  $w \in \hat{W}$ , denote  $q_w = \mu(IwI)/\mu(I)$  which is a natural number. Let  $H = \mathbb{C}_c(I \backslash G / I)$  be the Iwahori–Hecke algebra of  $G$  and  $h_w \in H$  be the element  $\frac{1}{\mu(I)} IwI$ .

Let  $\beta_1, \dots, \beta_r \in \tilde{W}$  ( $r = \text{rank } G$ ) be some fixed multiples of the simple coweights of the root system of  $W$  (the simple coweight themselves belong to  $\hat{W}$ , so we cannot use them directly). Then  $h_{\beta_i}$  satisfies that  $h_{\beta_i}^m = h_{\beta_i^m}$ , that is,  $h_{\beta_i}$  is a *nonbacktracking operator* (when acting on the building  $\mathcal{B}$  of  $G$ , it is indeed a nonbacktracking operator, or ‘collision free’ in the notions of [47]). Then it holds that:

**Theorem 8.1** (See [38, Theorem 22.1]). *There exist two finite sets  $A, B \subset \tilde{W}$  such that each  $w \in \tilde{W}$  can be written uniquely as  $w = a\beta_1^{m_1} \cdot \dots \cdot \beta_r^{m_r} b$  with  $a \in A, b \in B, m_i \geq 0$ , and moreover  $l_{\tilde{W}}(w) = l_{\tilde{W}}(a) + \sum_{i=1}^r m_i l_{\tilde{W}}(\beta_i) + l_{\tilde{W}}(b)$ , where  $l_{\tilde{W}} : \tilde{W} \rightarrow \mathbb{N}$  is the length function of the group  $\tilde{W}$  as (an extended) Coxeter group.*

As a corollary, it holds that in the Iwahori–Hecke algebra,

$$h_w = h_a h_{\beta_1}^{m_1} \cdot \dots \cdot h_{\beta_r}^{m_r} h_b.$$

Let us now generalize this theorem to arbitrarily small compact open subgroups  $K' \subset G$ . The following theorem is based on the results of [3] (see also [4, Chapter II, Section 2]).

**Theorem 8.2** (Bernstein’s decomposition). *There exist arbitrarily small compact open subgroups  $K'$  such that for each  $K'$  there exist two finite sets  $A, B \subset K' \backslash G / K'$  and  $\beta_1, \dots, \beta_r \in K' \backslash G / K', r = \text{rank}_k G$ , such that:*

1. For each  $1 \leq i \leq r$  and  $m \geq 0, h_{K', \beta_i}^m = h_{K', \beta_i^m}$ .
2. The operators  $h_{K', \beta_i}, 1 \leq i \leq r$ , commute.
3. For each  $g \in G$ , there exist  $b \in B, a \in A$  and  $m_i \geq 0$  such that

$$h_{K', g} = h_{K', a} h_{K', \beta_1}^{m_1} \cdot \dots \cdot h_{K', \beta_r}^{m_r} h_{K', b}. \tag{8.1}$$

**Remark 8.3.** Equation (8.1) is equivalent to the double coset decomposition

$$K'gK' = K'aK'\beta_1^{m_1}K' \cdot \dots \cdot K'\beta_r^{m_r}K'bK' = K'a\beta_1^{m_1} \cdot \dots \cdot \beta_r^{m_r}bK'.$$

Unlike in Theorem 8.2, there is no uniqueness in the claim.

*Proof.* We follow [4, Chapter II, Section 2]. Start from the Cartan decomposition  $KA_+K$ , where  $A_+ \subset \tilde{P}$  are the dominant elements in the lattice  $M / (M \cap K)$  ( $A_+$  is denoted  $\Lambda_+$  in [4]). Let  $\beta_1, \dots, \beta_r \in A_+$  be generators of a free semigroup  $\tilde{A}_+$  in  $A_+$  (they may be chosen so that their lift to  $M$  commute, not just as elements in  $M / (M \cap K)$ ). Let  $\mu_1, \dots, \mu_l \in A_+$  be elements such that  $A_+ = \cup_{i=1}^l \tilde{A}_+ \mu_i$ , the union being disjoint.

By Bruhat’s theorem in [4], there exist arbitrary small compact open subgroups  $K' \subset K$  such that:

- o  $K'$  is normal in  $K$ .
- o For every  $a, b \in A_+$ , it holds that  $h_{K', a} h_{K', b} = h_{K', ab}$ , that is,

$$K'aK'bK' = K'abK'. \tag{8.2}$$

Let  $x_1, \dots, x_m \in K$  be representatives of right cosets of  $K'$  in  $K$ . Since  $K'$  is normal in  $K$  they are also representatives of left cosets. Let  $A = \{x_1, \dots, x_r\}$ , and let  $B = \{\mu_i x_j : i = 1, \dots, l, j = 1, \dots, r\}$ . By the Cartan decomposition for each  $g \in G$  there exist  $x \in A, \mu x' \in B$  and  $m_1, \dots, m_r$  such that  $K'gK' = K'x\beta_1^{m_1} \cdot \dots \cdot \beta_r^{m_r}\mu x'K'$ . It remains to show that

$$K'xK'\beta_1^{m_1}K' \cdot \dots \cdot \beta_r^{m_r}K'\mu x'K' = K'x\beta_1^{m_1} \cdot \dots \cdot \beta_r^{m_r}\mu x'K'. \tag{8.3}$$

Since  $\beta_1, \dots, \beta_m, \mu \in A_+, K'\beta_1^{m_1}K' \cdot \dots \cdot \beta_r^{m_r}K'\mu K' = K'\beta_1^{m_1} \cdot \dots \cdot \beta_r^{m_r}\mu K'$  by equation (8.2). Finally, since  $K'$  is normal in  $K$  and  $x, x' \in K, K'x = xK'$  and  $K'x' = x'K'$ . Applying those equalities, we get equation (8.3). □

8.1. Nonbacktracking operators and temperedness

We may now deduce:

**Theorem 8.4.** *Let  $(\pi, V)$  be a unitary irreducible representation of  $G$ , let  $V^{K'}$  the  $K'$ -fixed vectors and assume that  $V^{K'} \neq \{0\}$ . Consider the action of  $H_{K'} = C_c(K' \backslash G / K')$  on  $V^{K'}$ . Then  $\pi$  is  $p$ -tempered if and only if, for every  $1 \leq i \leq r$ , for every eigenvalue  $\lambda$  of  $\pi(h_{K', \beta_i})$  on  $V^{K'}$  it holds that  $|\lambda| \leq q_{K', \beta_i}^{1-1/p}$ .*

*Proof.* First, note that  $V$  is  $p$ -tempered if and only if for every  $0 \neq v_0 \in V^{K'}$  and  $p' > p$ ,

$$\begin{aligned} \int_G |\langle v_0, \pi(g) \cdot v_0 \rangle|^{p'} dg &= \sum_{g \in [K' \backslash G / K']} \mu(K'gK') \left( \frac{|\langle v_0, \pi(K'gK')v_0 \rangle|}{\mu(K'gK')} \right)^{p'} \\ &= \sum_{g \in [K' \backslash G / K']} (\mu(K'gK'))^{1-p'} (\mu(K') |\langle v_0, \pi(h_{K',g})v_0 \rangle|)^{p'} \\ &\asymp \sum_{g \in [K' \backslash G / K']} (q_{K',g})^{1-p'} (|\langle v_0, \pi(h_{K',g})v_0 \rangle|)^{p'} < \infty. \end{aligned}$$

The ‘only if’ part is easier and does not require Bernstein’s decomposition. For  $p = 2$ , it is essentially the main result of [47].

Assume that some eigenvalue  $\lambda$  of  $h_{K', \beta_i}$  satisfies  $|\lambda| > q_{K', \beta_i}^{1-1/p}$ . Let  $v_0 \in V^{K'}$  be an eigenvector of  $h_{K', \beta_i}$  with eigenvalue  $\lambda$ . Then, for  $p' > p$  such that  $|\lambda| \geq q_{K', \beta_i}^{1-1/p'}$ ,

$$\begin{aligned} \int_G |\langle v_0, \pi(g) \cdot v_0 \rangle|^{p'} dg &\gg \sum_{g \in [K' \backslash G / K']} (q_{K',g})^{1-p'} (|\langle v_0, \pi(h_{K',g})v_0 \rangle|)^{p'} \\ &\geq \sum_{m=0}^{\infty} (q_{K', \beta_i^m})^{1-p'} (|\langle v_0, \pi(h_{K', \beta_i^m})v_0 \rangle|)^{p'} \\ &= \sum_{m=0}^{\infty} (q_{K', \beta_i})^{m(1-p)} (|\langle v_0, \pi(h_{K', \beta_i})^m v_0 \rangle|)^{p'} \\ &= \sum_{m=0}^{\infty} (q_{K', \beta_i})^{m(1-p)} (|\lambda|^m \|v_0\|^2)^{p'} \\ &\geq \|v_0\|^{2p} \sum_{m=0}^{\infty} (q_{K', \beta_i})^{m((1-p)+p(1-1/p))} = \|v_0\|^{2p} \sum_{m=0}^{\infty} 1 = \infty, \end{aligned}$$

and  $V$  is not  $p$ -tempered.

We now prove the ‘if’ part. One should prove that for  $p' > p$ , the matrix coefficients are in  $L^{p'}(G)$ . By Bernstein decomposition,

$$\begin{aligned} \int_G |\langle v_0, \pi(g) \cdot v_0 \rangle|^{p'} dg &\asymp \sum_{g \in [K' \backslash G / K']} (q_{K',g})^{1-p'} (|\langle v_0, \pi(h_{K',g})v_0 \rangle|)^{p'} \\ &\leq \sum_{a \in A, b \in B} \sum_{m_1 \geq 0, \dots, m_r \geq 0} (q_{K',a} q_{K', \beta_1}^{m_1} \cdots q_{K', \beta_r}^{m_r} q_{K',b})^{1-p'} \\ &\quad \cdot (|\langle \pi(h_{K',a})^* v_0, \pi(h_{K', \beta_1}^{m_1} \cdots h_{K', \beta_r}^{m_r}) \pi(h_{K',b}) v_0 \rangle|)^{p'} \\ &\ll_{K'} \sum_{a \in A, b \in B} (q_{K',a} q_{K',b})^{1-p'} \|\pi(h_{K',a})\|^{p'} \|\pi(h_{K',b})\|^{p'} \end{aligned}$$

$$\begin{aligned} & \cdot \sum_{m_1 \geq 0, \dots, m_r \geq 0} (q_{K', \beta_1}^{m_1} \cdot \dots \cdot q_{K', \beta_r}^{m_r})^{1-p'} \cdot \left\| \pi(h_{K', \beta_1}^{m_1} \cdot \dots \cdot h_{K', \beta_r}^{m_r}) \right\|^{p'} \\ & \ll_{K'} \left( \sum_{m_1 \geq 0} q_{K', \beta_1}^{m_1(1-p')} \left\| \pi(h_{K', \beta_1}^{m_1}) \right\|^{p'} \right) \left( \sum_{m_2 \geq 0} q_{K', \beta_2}^{m_2(1-p')} \left\| \pi(h_{K', \beta_2}^{m_2}) \right\|^{p'} \right) \\ & \cdot \dots \cdot \left( \sum_{m_r \geq 0} q_{K', \beta_r}^{m_r(1-p')} \left\| \pi(h_{K', \beta_r}^{m_r}) \right\|^{p'} \right). \end{aligned}$$

Since all the eigenvalues of  $\pi(h_{K', \beta_i})$  are bounded by  $q_{K', \beta_i}^{1-1/p}$ , it follows from the theory of matrix norms that the sum converges (note that  $\pi(h_{K', \beta_i})$  is usually not unitary or self-adjoint, so one should be a bit careful here). □

For  $d_0 \in \mathbb{R}_{\geq 0}$ , choose  $m_i$ , so that  $q_{\beta_i}^{m_i-1} \leq q^{d_0/2} \leq q_{\beta_i}^{m_i}$  and denote  $f_{d_0} = \sum_{i=1}^r (h_{\beta_i}^{m_i})^* h_{\beta_i}^{m_i}$ , where  $f^*(g) = \overline{f(g^{-1})}$  for a function  $f: G \rightarrow \mathbb{C}$ . The following theorem proves Theorem 3.11

**Theorem 8.5.** *The functions  $f_{d_0} \in H_{K'}$  are a good family for the set  $\Pi(G)_{K'-\text{sph}}$ :*

1. *For every  $(\pi, V) \in \Pi(G)$  with a nontrivial  $K'$ -invariant vector, it holds that*

$$q^{d_0(1-1/p(\pi))} \ll_{K'} \text{tr}(\pi(f_{d_0})).$$

2. *For every  $(\pi, V) \in \Pi(G)$  without a nontrivial  $K'$ -invariant vector, it holds that*

$$\text{tr}(\pi(f_{d_0})) = 0.$$

3. *It holds that  $f_{d_0}(g) \ll_{\epsilon} q^{d_0 \epsilon} \psi_{d_0}(g)$ .*

*Proof.* The second condition is obvious since if there is no nontrivial  $K'$ -invariant vector, then  $\pi(f_{d_0}) = 0$ .

The fact that  $f_{d_0}(g) \ll_{\epsilon} q^{d_0 \epsilon} \psi_{d_0}(g)$  follows from Lemma 4.18, as  $h_{\beta_i}^{m_i}(g) \ll \chi_{d_0/2}$ .

It remains to prove the first condition. Since  $\pi(f_{d_0})$  is nonnegative and self-adjoint it is diagonalizable on  $V^{K'}$  and all its eigenvalues are nonnegative.

On the other hand,  $\text{tr} \pi(f_{d_0}) \geq \left\| \pi(f_{d_0}) \right\|$  (as a matter of fact, by uniform admissibility ([3]),  $\dim V^{K'} \ll_{K'} 1$ , so  $\text{tr} \pi(f_{d_0}) \asymp \left\| \pi(f_{d_0}) \right\|$ ). It holds that

$$\max_{i=1}^r \left\| h_{\beta_i}^{m_i} \right\|^2 \leq \left\| \pi(f_{d_0}) \right\|.$$

Since  $V$  is not  $p'$ -tempered for  $p' < p$ , by Theorem 8.4, for some  $1 \leq i \leq r$  we have an eigenvalue  $\lambda$  of  $h_{\beta_i}$  with  $|\lambda| \geq q_{K', \beta_i}^{1-1/p}$ . Therefore,  $h_{\beta_i}^{m_i}$  has an eigenvalue  $\lambda^{m_i}$  with  $|\lambda^{m_i}| \geq q_{K', \beta_i}^{m_i(1-1/p)} \asymp_{K'} q^{l(\beta_i)m_i(1-1/p)} \asymp q^{d_0/2(1-1/p)}$ . Therefore,  $\left\| \pi(f_{d_0}) \right\| \geq \left\| h_{\beta_i}^{m_i} \right\|^2 \gg q^{d_0(1-1/p)}$ , as needed. □

### 9. Lower Bounds on Matrix Coefficients for a Specific Representation

In this section, we prove Theorem 3.10. We assume that  $G$  is Archimedean and  $(\pi, V)$  is an irreducible unitary representation of  $G$ .

We will prove that  $\{\pi\}$  has a good family of functions. Let  $v \in V^{\tau}$ ,  $\|v\| = 1$ , belong to a fixed  $\tau$ -type of  $K$ , that is,  $K$  acts on  $U_0 = \text{span}\{Kv\}$  as a  $K$ -irreducible representation  $\tau$ . In particular,  $v$  is  $K$ -finite. Let

$$g_{d_0/2}(g) = \chi_{d_0/2}(g) q^{l(g)p(\pi)^{-1} \overline{\langle v, \pi(g)v \rangle}},$$

and let

$$f_{d_0} = g_{d_0/2}^* * g_{d_0/2}.$$

We claim that  $f_{d_0}$  satisfies the condition of Definition 3.7 for  $\{\pi\}$ . We first verify conditions (2) and (3). Since  $f_{d_0}$  is self-adjoint and nonnegative, (2) follows.

By the theory of leading exponents (which is described below),  $|\langle v, \pi(g)v \rangle| \ll_{\pi, v, \epsilon} q^{-l(g)(1/\rho(\pi)+\epsilon)}$ , so  $|g_{d_0/2}(g)| \ll q^{d_0\epsilon} \chi_{d_0/2}(g)$ . It follows that also  $|g_{d_0/2}^*(g)| \ll q^{d_0\epsilon} \chi_{d_0/2}(g)$ . By Lemma 4.18,  $|f_{d_0}(g)| \ll_{\epsilon} q^{d_0\epsilon} \psi_{d_0+2}(g)$ .

It remains to prove (1), which will concern the rest of this section. We note that since  $f_{d_0}$  is self-adjoint and nonnegative

$$\text{tr}(\pi(f_{d_0})) \gg \|\pi(f_{d_0})\| = \|\pi(g_{d_0/2})\|^2.$$

Now,

$$\begin{aligned} \langle v, \pi(g_{d_0/2})v \rangle &= \int_G g_{d_0/2}(g) \langle v, \pi(g)v \rangle \\ &= \int_G \chi_{d_0/2}(g) q^{l(g)\rho(\pi)^{-1}} \overline{\langle v, \pi(g)v \rangle} \langle v, \pi(g)v \rangle dg \\ &\geq \int_{l(g) \leq d_0/2} q^{l(g)\rho(\pi)^{-1}} |\langle v, \pi(g)v \rangle|^2 dg, \end{aligned}$$

and conclude that

$$\sqrt{\text{tr}(\pi(f_{d_0}))} \gg \int_{l(g) \leq d_0/2} q^{l(g)\rho(\pi)^{-1}} |\langle v, \pi(g)v \rangle|^2 dg.$$

Therefore, to prove that  $\text{tr}(\pi(f_{d_0})) \gg_{\epsilon} q^{d_0(1-1/\rho(\pi)-\epsilon)}$ , one should prove (after changing  $d_0$  and  $d_0/2$ )

$$\int_{l(g) \leq d_0} q^{l(g)\rho(\pi)^{-1}} |\langle v, \pi(g)v \rangle|^2 dg \stackrel{!}{\gg}_{\pi, \epsilon} q^{d_0(1-\rho(\pi)^{-1}-\epsilon)}. \tag{9.1}$$

So far, our proof is essentially the same as (part of) the proof of [58, Theorem 3], which only concerns rank 1.

We start by simplifying the left-hand side of equation (9.1).

Applying the Cartan decomposition we get

$$\int_{l(g) \leq d_0} q^{l(g)\rho(\pi)^{-1}} |\langle v, \pi(g)v \rangle|^2 dg = \int_K \int_K \int_{a \in A_+, l(a) \leq d_0} q^{l(a)\rho(\pi)^{-1}} S(a) |\langle v, \pi(ka k')v \rangle|^2 dk dk' da.$$

Using the logarithm map, we identify  $A_+ \subset \mathfrak{a}$ , and give  $a \in A_+$  coordinates  $(x_1, \dots, x_r)$ ,  $x_i \geq 0$  by  $(x_1, \dots, x_r) \rightarrow \sum_{i=1}^r x_i \omega_i$ , where  $\omega_1, \dots, \omega_r$  are the fundamental coweights. Recall that for  $a \in A_+$ ,  $l(a) = 2\rho(a)$ . For  $\kappa > 0$ , denote by  $A_+^{\kappa} \subset A_+$  the set of  $a \in A_+$  with  $x_i > \kappa$ . Then for  $a \in A_+^{\kappa}$  it holds that  $S(a) \asymp_{\kappa} q^{l(a)} = q^{2\rho(a)}$ . Then:

$$\begin{aligned} & \int_K \int_K \int_{a \in A_+, l(a) \leq d_0} q^{l(a)p(\pi)^{-1}} S(a) |\langle v, \pi(kak')v \rangle|^2 dk dk' da \\ & \gg_\kappa \int_K \int_K \int_{a \in A_+, l(a) \leq d_0} q^{2\rho(a)(1+p(\pi)^{-1})} |\langle v, \pi(kak')v \rangle|^2 dk dk' da, \end{aligned}$$

so we should prove that for some  $\kappa > 0$ ,

$$\int_K \int_K \int_{a \in A_+, l(a) \leq d_0} q^{2\rho(a)(1+p(\pi)^{-1})} |\langle v, \pi(kak')v \rangle|^2 dk dk' da \stackrel{!}{\gg}_\epsilon q^{d_0(1-p(\pi)^{-1}-\epsilon)}. \tag{9.2}$$

### 9.1. Leading exponents

We recall the Casselman–Harish Chandra–Milicic theory of *Leading Exponents*. We follow [43, Chapter VIII].

Equation (9.2) is very similar to [43, Theorem 8.48, (b) implies (a)], which is one of the more technical parts of the theory. We will follow the same proof closely while deriving an explicit expression. A side benefit is that the proofs below somewhat simplify the proof of [43, Theorem 8.48, (b) implies (a)].

By the theory of leading exponents, we may associate with an irreducible unitary representation (actually, to any irreducible admissible representation)  $(\pi, V)$  a finite subset called leading exponent  $\mathcal{F} \subset \mathfrak{a}_\mathbb{C}^*$  such that the following two theorems hold:

**Theorem 9.1** ([43, Theorem 8.47]). *The following are equivalent:*

- For  $\nu_0: a \rightarrow \mathbb{R}$  a real character, every  $K$ -finite matrix coefficient  $\langle v, av \rangle$  for  $a \in A_+$  is bounded in absolute value by  $\ll_{\pi, v} e^{(\nu_0 - \rho)(a)} l(a)^N$ , where  $N$  is some constant.
- For every  $\nu \in \mathcal{F}$ , and every fundamental weight  $\omega_i$ ,  $1 \leq i \leq r$ ,  $\text{Re } \nu(\omega_i) \leq \nu_0(\omega_i)$ .

**Theorem 9.2** ([43, Theorem 8.48]). *The following are equivalent:*

- Every  $K$ -finite matrix coefficient  $\phi(g) = \langle v, \pi(g)v \rangle$  is in  $L^p(G)$ .
- For every  $\nu \in \mathcal{F}$  and every fundamental weight  $\omega_i$ ,  $1 \leq i \leq r$ ,  $\text{Re } \nu(\omega_i) < (1 - \frac{2}{p})\rho(\omega_i)$ .

Note that the second theorem implies that

$$p(\pi) = \min \left\{ p : \forall v \in \mathcal{F}, 1 \leq i \leq r, \text{Re } \nu(\omega_i) \leq (1 - \frac{2}{p})\rho(\omega_i) \right\}$$

and that some matrix coefficient is not in  $L^{p(\pi)}(g)$ .

To state the main theorem about leading exponents, let us set some notations. Assume that  $0 \neq v \in V$  is  $K$ -finite, and let  $E_0: V \rightarrow U_0$  be a projection onto a finite-dimensional  $K$ -invariant subspace  $U_0 \subset V$  such that  $v \in U_0$ . We define  $F: A_+ \rightarrow \text{End}_\mathbb{C}(U_0)$  by  $F(a) = E_0\pi(a)E_0$ .

We denote by  $\mathcal{H}_{\text{End } U_0}$  the set of holomorphic functions  $f: D^r \rightarrow \text{End}(U_0)$ , where  $D = \{z \in \mathbb{C} : |z| < 1\}$  is the open unit ball. Each such function has a convergent multiple power series, which is absolutely and uniformly convergent on compact subsets of  $D^r$ .

As before, we identify  $a \in A_+ \subset \mathfrak{a}$  with coordinates  $(x_1, \dots, x_r)$ ,  $x_i \geq 0$  by  $(x_1, \dots, x_r) \rightarrow \sum_{i=1}^r x_i \omega_i$ .

We say that  $\nu, \nu' \in \mathfrak{a}_\mathbb{C}^*$  are integrally equivalent if their difference  $\nu - \nu'$  is an integral combination of simple roots. If the difference is a *nonnegative* integral combination of simple roots, we write  $\nu' \leq \nu$ .

**Theorem 9.3** [43, Theorem 8.32]. *There exist  $n_0 \in \mathbb{N}$  and a finite set  $\mathcal{F}'$  satisfying:*

1.  $\mathcal{F} \subset \mathcal{F}'$ .
2. Each  $\nu' \in \mathcal{F}'$  satisfies  $\nu' \leq \nu$  for some  $\nu \in \mathcal{F}$ .

3. It holds that for  $x_1 > 0, \dots, x_r > 0$ ,

$$F(a) = F(x_1, \dots, x_r) = \sum_{\nu \in \mathcal{F}'} \sum_{1 \leq n_1 \leq n_0 \dots 1 \leq n_r \leq n_0} G_{\nu, n_1, \dots, n_r}(x_1, \dots, x_r) e^{(\nu-\rho)(x_1, \dots, x_r)} x_1^{n_1} \dots x_r^{n_r} \tag{9.3}$$

such that for  $\nu \in \mathcal{F}'$ ,  $1 \leq n_i \leq n_0$ ,  $G_{\nu, n_1, \dots, n_r} : (0, \infty)^r \rightarrow \text{End}_{\mathbb{C}}(U_0)$  are functions given by  $G_{\nu, n_1, \dots, n_r}(x_1, \dots, x_r) = f_{\nu, n_1, \dots, n_r}(e^{-x_1}, \dots, e^{-x_r})$ , where  $f_{\nu, n_1, \dots, n_r} \in \mathcal{H}_{\text{End} U_0}$ .

Moreover, if  $f_{\nu, n_1, \dots, n_r} \neq 0$ , then  $f_{\nu, n_1, \dots, n_r}(0, \dots, 0) \neq 0$  and for each  $\nu \in \mathcal{F}'$  there exist  $n_1, \dots, n_r$  with  $f_{\nu, n_1, \dots, n_r} \neq 0$ .

*Proof.* The theorem follows from [43, Theorem 8.32] and the discussion following it. Let us explain: By [43, Theorem 8.32],  $F$  has a decomposition like equation (9.3) for a certain set  $\mathcal{F}'$ . If  $f_{\nu, n_1, \dots, n_r}(0, \dots, 0) = 0$ , we may use its power series expansion to replace  $\nu$  by other elements in  $\mathfrak{a}_{\mathbb{C}}^*$ , which are integrally equivalent to it.

It remains to prove that  $\mathcal{F} \subset \mathcal{F}'$  and that each  $\nu' \in \mathcal{F}'$  has  $\nu \in \mathcal{F}$  with  $\nu' \leq \nu$ . By power series expansion, we have a *unique* decomposition (see [43, Equation 8.52])

$$F(x_1, \dots, x_r) = \sum_{\nu \in \mathfrak{a}_{\mathbb{C}}^*} F_{\nu-\rho}(x_1, \dots, x_r)$$

$$F_{\nu-\rho}(x_1, \dots, x_r) = \sum_{1 \leq n_1 \leq n_0, \dots, 1 \leq n_r \leq n_0} c_{\nu, n_1, \dots, n_r} e^{(\nu-\rho)(x_1, \dots, x_r)} x_1^{n_1} \dots x_r^{n_r},$$

for some  $c_{\nu, n_1, \dots, n_r} \in \text{End} U_0$ . Each term  $F_{\nu-\rho}$  can be calculated from  $G_{\nu', n_1, \dots, n_r}$ , for  $\nu \leq \nu'$ . The set of leading exponents is the set of maximal elements relative to  $\leq$  for  $\nu \in \mathcal{F}$  with  $F_{\nu-\rho} \neq 0$ . This immediately implies that  $\mathcal{F} \subset \mathcal{F}'$ . Moreover, each  $\nu'' \in \mathfrak{a}_{\mathbb{C}}^*$  with  $F_{\nu''-\rho} \neq 0$  satisfies  $\nu'' \leq \nu$  for some  $\nu \in \mathcal{F}$ , which says that each  $\nu' \in \mathcal{F}'$  satisfies  $\nu' \leq \nu$  for some  $\nu \in \mathcal{F}$ . □

We remark that Theorem 9.3 does not directly imply the upper bound given in Theorem 9.1 since it does not give bounds for  $x_i \rightarrow 0$ . Such bounds are available using asymptotic expansion near the walls ([43, Chapter VIII, Section 12]).

### 9.2. Some technical lemmas

**Lemma 9.4** (Compare [43, Lemma B.24]). *Let  $f : \mathbb{R} \rightarrow \mathbb{C}$  be a function defined as  $f(x) = e^{\beta x} \sum_{i=1}^k c_i e^{-\alpha_i x} x^{n_i}$  with  $\alpha_i, c_i \in \mathbb{C}$ ,  $\text{Re}(\alpha_i) \geq 0$ ,  $\beta \in \mathbb{R}$ ,  $\beta \geq 0$ ,  $n_i \in \mathbb{N}$ .*

*Assume that there is  $0 \leq i \leq k$  such that  $\text{Re}(\alpha_i) = 0$ ,  $c_i \neq 0$ , and let*

$$n_0 = \max_{1 \leq i \leq k} \{n_i : \text{Re}(\alpha_i) = 0 \text{ and } c_i \neq 0\}.$$

*Then for  $T$  large enough,*

$$\int_0^T |f(x)| dx \gg_f e^{\beta T} T^{n_0}.$$

*Moreover, if we assume that  $n_0 = \max_{1 \leq i \leq k} \{n_i : \text{Re}(\alpha_i) = 0\}$ , then the underlying lower bound on  $T$  and the constants are continuous for small perturbations of the  $c_i$ .*

**Remark 9.5.** The condition on  $n_0$  in the ‘moreover part’ comes to deal with the case that after a small perturbation, some  $c_i = 0$  becomes nonzero.



*Proof.* During the proof,  $\gg$  may depend on  $f$ . Fix  $M$  large enough, depending on  $f$ , to be chosen later. Then

$$\int_0^T |f(x)|dx \geq \int_{T-M}^T |f(x)|dx.$$

Let  $\alpha_0 = \min_{1 \leq i \leq k} \{\text{Re}(\alpha_i) : \text{Re}(\alpha_i) \neq 0\}$ ,  $N = \max_{1 \leq i \leq k} \{n_i\}$ .

After rearranging the summands, write  $f = f_0 + f_1$ , with  $f_0(x) = e^{\beta x} \sum_{i=1}^l c_i e^{-\alpha_i x} x^{n_0}$ ,  $f_1 = e^{\beta x} \sum_{i=l+1}^k c_i e^{-\alpha_i x} x^{n_i}$ , where the summands  $1 \leq i \leq l$  contain all factors with  $\text{Re}(\alpha_i) = 0$  and  $n_i = n_0$ . Moreover, we may assume that all the  $\alpha_i$ ,  $1 \leq i \leq l$  are different.

Then for  $\epsilon > 0$  small enough and  $T$  large enough,

$$\int_{T-M}^T |f_1(x)|dx \leq M \max_{T-M \leq x \leq T} \{f_1(x)\} \ll ((e^{(\beta-\alpha_0)(T-M)} + e^{(\beta-\alpha_0)T})T^N + e^{\beta T} T^{n_0-1}) = o(e^{\beta T} T^{n_0}).$$

Now,

$$\begin{aligned} \int_{T-M}^T |f_0(x)|dx &= \int_{T-M}^T x^{n_0} e^{\beta x} \left| \sum_{i=1}^l c_i e^{-\alpha_i x} \right| dx \\ &\geq (T-M)^{n_0} e^{\beta(T-M)} \int_{T-M}^T \left| \sum_{i=1}^l c_i e^{-\alpha_i x} \right| dx. \end{aligned}$$

Note that  $|\sum_{i=1}^l c_i e^{\alpha_i x}| \gg |\sum_{i=1}^l c_i e^{-\alpha_i x}|^2 = \sum_{i=1}^l |c_i|^2 + \sum_{1 \leq i \neq j \leq l} c_i \bar{c}_j e^{(\alpha_i - \alpha_j)x}$  since this value is bounded.

$$\begin{aligned} \int_{T-M}^T \left| \sum_{i=1}^l c_i e^{\alpha_i x} \right| dx &\gg \int_{T-M}^T \left( \sum_{i=1}^l |c_i|^2 + \sum_{1 \leq i \neq j \leq l} c_i \bar{c}_j e^{(\alpha_i - \alpha_j)x} \right) dx \\ &\geq M \left( \sum_{i=1}^l |c_i|^2 \right) - \sum_{1 \leq i \neq j \leq l} \left| \frac{c_i \bar{c}_j}{\alpha_i - \alpha_j} \right| \\ &\gg M - O(1). \end{aligned}$$

For  $M$  large enough, the last value is  $\gg 1$ , so

$$\int_{T-M}^T |f_0(x)|dx \gg e^{\beta T} T^{n_0}$$

and

$$\int_0^T |f(x)|dx \geq \int_{T-M}^T |f(x)|dx \geq \int_{T-M}^T |f_0(x)|dx - \int_{T-M}^T |f_1(x)|dx \gg e^{\beta T} T^{n_0}.$$

For the ‘moreover part’, one follows the proof carefully and notices that it remains true for small perturbations in the  $c_i$ . □

**Remark 9.6.** For  $\beta > 0$ , our lower bound agrees with a similar upper bound. For  $\beta = 0$  it is no longer true, but a similar proof will give the right lower bound  $T^{n_0+1}$ .

**Lemma 9.7.** Let  $M$  be an open subset of a smooth Riemannian manifold (e.g., a Lie group),  $F : M \times [R, \infty)$  a function defined by

$$F(m, x) = \sum_{i=1}^k e^{s_i x} x^{n_i} F_i(m, x), \tag{9.4}$$

such that:  $s_i \in \mathbb{C}, n_i \in \mathbb{N}$  and  $F_i(m, x) = f_i(m, e^{-x})$  for some function  $f_i(m, z)$  real analytic on  $M \times \overline{D}_r$ , where  $\overline{D}_r \subset \mathbb{C}$  is the closed ball of radius  $r$ , and holomorphic in the second variable. Assume also that for each  $1 \leq i \leq k$  there is  $m \in M$  with  $f_i(m, 0) \neq 0$ . Let  $s_0 = \max_{1 \leq i \leq k} \operatorname{Re} s_i$ , and assume that  $s_0 \geq 0$ . Then for  $T$  large enough

$$\int_M \int_R^T |F(s, r)| dr \gg e^{s_0 T}.$$

*Proof.* By restriction to a compact subset, we may assume that the closure of  $M$  is compact.

Decompose  $f_i(m, z) = c_i(m) + g_i(m, z)z$ , where  $g_i(m, z)$  is also holomorphic in  $|z| \leq e^{-R}$ . Then

$$F_i(m, r) = c_i(m) + e^{-r} G_i(m, r),$$

where  $G_i(m, r)$  is bounded for  $m, r \in M \times [R, \infty)$ . Therefore,

$$F(m, r) = \sum_{i=1}^k e^{s_i r} r^{n_i} c_i(m) + \sum_{i=1}^k e^{(s_i-1)r} r^{n_i} G_i(m, r). \tag{9.5}$$

Without loss of generality,  $\operatorname{Re}(s_1) = s_0$  and  $n_1 = \max\{n_i : \operatorname{Re}(s_i) = s_0\}$ . Let  $m_0 \in M$  be a point with  $f_1(m_0, 0) \neq 0$ . Choose a small enough neighborhood  $M_0 \subset M$  of  $m_0$ . We have

$$\int_M \int_R^T |F(m, r)| dr dm \gg \int_{M_0} \int_R^T |F(m, r)| dr dm.$$

Since  $G_i(m, r)$  is bounded on  $M_0$ , for  $T$  large enough the second summand of equation (9.5) satisfies

$$\int_{M_0} \int_R^T \left| \sum_{i=1}^k e^{(s_i-1)r} r^{n_i} G_i(m, r) \right| dr = o(e^{s_0 T}).$$

As for the first summand of equation (9.5), by Lemma 9.4 and the fact that  $c_1(m_0) \neq 0$ , it holds that

$$\int_{M_0} \int_R^T \left| \sum_{i=1}^k e^{s_i r} r^{n_i} c_i(m) \right| dr \gg T^{n_1} e^{s_0 T}$$

and we are done. □

We can finally prove equation (9.2).

*Proof.* Recall that we chose  $v \in V, \|v\| = 1$  to span a representation  $\tau$  of  $K$ . We choose in Theorem 9.3  $U_0 = \operatorname{span} K v$ .

Using Theorem 9.2, choose  $v_0 \in \mathcal{F}'$  and  $1 \leq i \leq r$  such that  $\operatorname{Re} v_0(\omega_i) = (1 - \frac{2}{p(\pi)})\rho(\omega_i)$ . Without loss of generality  $i = 1$ . Moreover, we assume that among all  $v \in \mathcal{F}'$  satisfying this condition,  $v_0$  has maximal  $0 \leq N_1 \leq n_0$  such that for some constants  $z_2, \dots, z_l \neq 0, 0 \leq n_2, \dots, n_r \leq n_0, \lim_{z_1 \rightarrow 0} f_{v, N_1, \dots, n_k}(z_1, z_2, \dots, z_r) \neq 0$ , where  $f_{v, N_1, \dots, n_k}(z_1, z_2, \dots, z_r)$  is taken from Theorem 9.3 (notice that  $v_0$  may belong to  $\mathcal{F}' - \mathcal{F}$ , so it may not be a *leading exponent*).

By Theorem 9.3, if we let  $M = K \times (0, \infty)^{r-1} \times K$ , we identify  $m \in M$  with  $m = (k_1, x_2, \dots, x_r, k_2)$ , and let  $G : M \times (0, \infty) \rightarrow \infty$  be

$$G(m, x_1) = \langle v, \pi(k_1 a(x_1, \dots, x_r) k_2) v \rangle,$$

then  $G(m, x_1)$  has the form of equation (9.4), with  $s_1 = (v_0 - \rho)(\omega_1) = -\frac{2}{p(\pi)}\rho(\omega_1), n_1 = N_1$ . Note that  $G^2(m, x_1)$  also has this form, with  $s_1 = -\frac{4}{p(\pi)}\rho(\omega_1)$ . We Let

$$F(m, x_1) = e^{2\rho(\omega_1)(1+p(\pi)^{-1})x_1} G^2(m, x_1),$$

and  $F$  also has a similar form, with  $s_1 = 2\rho(\omega_1)(1 - p(\pi)^{-1})$ . Let  $m_0 = (k_1, x_2, \dots, x_r, k_2)$  be a point where the condition of Lemma 9.7 holds. By Lemma 9.7, for a small neighborhood  $M_0$  of  $m_0$ , it holds for  $d_0$  large enough and some constant  $C$

$$\int_{M_0} \int_1^{(d_0-C)/2\rho(\omega_1)} |F(m, x_1)| dx_1 dm \gg e^{d_0(1-p(\pi)^{-1})}$$

Finally, for  $M_0, \kappa$  small enough, for each  $m = (k_1, x_2, \dots, x_r, k_2) \in M_0$  and  $1 \leq x_1 \leq (d_0 - C)/2\rho(\omega_1)$ , it holds that  $a = (x_1, \dots, x_r) \in A_+^K$  and  $l(a) \leq d_0$ . Therefore,

$$\begin{aligned} & \int_K \int_K \int_{a \in A_+^K, l(a) \leq d_0} q^{2\rho(a)(1-p(\pi)^{-1})} |\langle v, \pi(kak') v \rangle|^2 dk dk' da \\ & \gg \int_{M_0} \int_0^{(d_0-1)/2\rho(\omega_1)} |F(m, x_1)| dx_1 dm \\ & \gg e^{d_0(1-p(\pi)^{-1})} \end{aligned}$$

as needed in equation (9.2). □

### Index

The following notations appear throughout the paper.

- $k$  – a local field.
- $q$  – if  $k$  is Archimedean  $q = e$ . Otherwise,  $q$  is the size of the quotient field of  $k$ ,
- $G$  – the  $k$ -rational points of a semisimple algebraic group over  $k$ .
- $\Gamma$  – a lattice in  $G$ . If there is a sequence  $(\Gamma_N)$  of lattices, then  $\Gamma_N$  is a finite index subgroup of  $\Gamma_1$ , with  $[\Gamma_1 : \Gamma_N] \rightarrow \infty$ .
- $K$  – a good maximal compact subgroup of  $G$ .
- $X$  – the locally symmetric space  $\Gamma \backslash G/K$ .
- $\Pi(G)$  – the set of equivalence classes of irreducible unitary representations of  $G$ . A representation is usually denoted by  $(\pi, V)$ .
- $p(\pi)$  – the minimal  $p$  such that all  $K$ -finite matrix coefficients of  $(\pi, V)$  are in  $L^{p+\epsilon}(G)$ .
- $\lambda$  – an eigenvalue of the Casimir operator. Appears only in the Archimedean case.

- $l: G \rightarrow \mathbb{R}_{\geq 0}$  – a length function on  $G$ . A length is usually denoted  $d_0$ .
- $\chi_{d_0}$  – a smooth approximation for the characteristic set  $\{g \in G : l(g) \leq d_0\}$ .
- $\psi_{d_0}$  – a smooth approximation for  $q^{(d_0-l(g))/2} \chi_{d_0}$ .
- $b_{x_0, \delta}$  – for  $x \in \Gamma \backslash G/K$ ,  $\delta \in \mathbb{R}_{>0}$ . In the non-Archimedean case, it is the characteristic function of  $\{x\}$ . In the Archimedean case, it is a smooth approximation for the characteristic function of a ball of radius  $\delta$  around  $x$ .

**Acknowledgements.** The authors are thankful to Amos Nevo and Peter Sarnak for their support of this project.

**Funding statement.** During the work on this project, the first author was supported by the SNF grant 200020-169106 at ETH Zurich. This work was part of the Ph.D. thesis of the second author at the Hebrew University of Jerusalem, under the guidance of Prof. Alexander Lubotzky, and was supported by the ERC grant 692854.

**Competing interests.** The authors have no conflicts of interest to declare.

## References

- [1] M. Abert, N. Bergeron, L. Biringer, T. Gelander, N. Nikolov, J. Raimbault and I. Samet, ‘On the growth of  $L^2$ -invariants for sequences of lattices in Lie groups’, *Annals of Mathematics* **185**(3) (2017), 711–790.
- [2] E. Assing and V. Blomer, ‘The density conjecture for principal congruence subgroups’, Preprint, 2022, [arXiv:2204.08868](https://arxiv.org/abs/2204.08868).
- [3] J. Bernstein, ‘All reductive p-adic groups are tame’, *Functional Analysis and its applications* **8**(2) (1974), 91–93.
- [4] J. Bernstein, ‘Notes of lectures on Representations of p-adic groups,’ 1992, [http://www.math.tau.ac.il/~bernstei/Unpublished\\_texts/unpublished\\_texts/Bernstein93new-harv.lect.from-chic.pdf](http://www.math.tau.ac.il/~bernstei/Unpublished_texts/unpublished_texts/Bernstein93new-harv.lect.from-chic.pdf).
- [5] V. Blomer, ‘Density theorems for  $GL(n)$ ’, *Invent. Math.* **232**(2) (2023), 683–711.
- [6] V. Blomer, J. Buttcane and P. Maga, ‘Applications of the Kuznetsov formula on  $GL(3)\text{II}$ : The level aspect’, *Mathematische Annalen* **369**(1–2) (2017), 723–759.
- [7] C. Bordenave and H. Lacoïn, ‘Cutoff at the entropic time for random walks on covered expander graphs’, *Journal of the Institute of Mathematics of Jussieu* (2021), 1–46.
- [8] F. Bruhat and J. Tits, ‘Groupes réductifs sur un corps local’, *Publications Mathématiques de l’Institut des Hautes Études Scientifiques* **41**(1) (1972), 5–251.
- [9] M. Burger, J.-S. Li and P. Sarnak, ‘Ramanujan duals and automorphic spectrum’, *Bulletin of the American Mathematical Society* **26**(2) (1992), 253–257.
- [10] P. Cartier, ‘Representations of p-adic groups: A survey’, in *Automorphic Forms, Representations and L-functions* Proc. Sympos. Pure Math., vol. 1 (Oregon State Univ., Corvallis, OR, 1977), 111–155. Part (1979).
- [11] B. Casselman, ‘Introduction to admissible representations of p-adic groups’, *unpublished notes* (1974).
- [12] W. Casselman and D. Milicic, ‘Asymptotic behavior of matrix coefficients of admissible representations’, *Duke Math. J* **49**(4) (1982), 869–930.
- [13] M. Chapman and O. Parzanchevski, ‘Cutoff on Ramanujan complexes and classical groups’, *Comment. Math. Helv.* **97**(3) (2022), 431–456.
- [14] M. Cowling, U. Haagerup and R. Howe, ‘Almost  $L^2$  matrix coefficients’, *Journal für die Reine und Angewandte Mathematik* **387** (1988), 97–110.
- [15] G. Davidoff, P. Sarnak and A. Valette, *Elementary Number Theory, Group Theory and Ramanujan Graphs*, vol. 55 (Cambridge University Press, Cambridge, 2003).
- [16] D. DeGeorge and N. Wallach, ‘Limit formulas for multiplicities in  $L^2(G/\Gamma)$ ’, *Annals of Mathematics* (1978), 133–150.
- [17] P. Deligne, ‘La conjecture de weil. I’, *Publications Mathématiques de l’Institut des Hautes Études Scientifiques* **43**(1) (1974), 273–307.
- [18] J. Duistermaat, J. Kolk and V. Varadarajan, ‘Spectra of compact locally symmetric manifolds of negative curvature’, *Inventiones mathematicae* **52**(1) (1979), 27–93.
- [19] W. Duke, Z. Rudnick and P. Sarnak, ‘Density of integer points on affine homogeneous varieties’, *Duke Mathematical Journal* **71**(1) (1993), 143–179.
- [20] M. Einsiedler and E. Lindenstrauss, ‘Diagonal actions on locally homogeneous spaces’, *Homogeneous Flows, Moduli Spaces and Arithmetic* **155241** (2010).
- [21] S. Evra and O. Parzanchevski, ‘Ramanujan complexes and golden gates in  $PU(3)$ ’, *Geom. Funct. Anal.* **32**(2) (2022), 193–235.
- [22] T. Finis and E. Lapid, ‘An approximation principle for congruence subgroups’, *Journal of the European Mathematical Society* **20**(5) (2018), 1075–1138.
- [23] A. Gamburd, ‘On the spectral gap for infinite index “congruence” subgroups of  $SL_2(\mathbb{Z})$ ’, *Israel Journal of Mathematics* **127**(1) (2002), 157–200.

- [24] R. Gangolli and V. S. Varadarajan, *Harmonic Analysis of Spherical Functions on Real Reductive Groups*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 101 (Springer-Verlag, Berlin, 1988).
- [25] I. M. Gelfand, M. I. Graev and I. Piatetski-Shapiro, *Representation Theory and Automorphic Functions* (Saunders, Philadelphia, Pa.-London-Toronto, Ont. 1968).
- [26] A. Ghosh, A. Gorodnik and A. Nevo, 'Diophantine approximation and automorphic spectrum', *International Mathematics Research Notices* **21** (2013), 5002–5058.
- [27] A. Ghosh, A. Gorodnik and A. Nevo, 'Best possible rates of distribution of dense lattice orbits in homogeneous spaces', *Journal für die reine und angewandte Mathematik (Crelles Journal)* (2014).
- [28] K. Golubev and A. Kamber, 'Cutoff on hyperbolic surfaces', *Geometriae Dedicata* **3** (2019), 1–31.
- [29] K. Golubev and A. Kamber, 'Cutoff on graphs and the Sarnak–Xue density of eigenvalues', *European J. Combin.* **104** (2022), Paper No. 103530, 23.
- [30] L. Guth and A. Lubotzky, 'Quantum error correcting codes and 4-dimensional arithmetic hyperbolic manifolds', *Journal of Mathematical Physics* **55**(8) (2014), 082202.
- [31] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford University Press, New York, 1979).
- [32] Harish-Chandra, 'Spherical functions on a semisimple Lie group, I', *American Journal of Mathematics* (1958), 241–310.
- [33] K. I. Hashimoto, 'Zeta functions of finite graphs and representations of p-adic groups', *Automorphic Forms and Geometry of Arithmetic Varieties* (1989), 211–280.
- [34] R. E. Howe and I. I. Piatetski-Shapiro, 'A counterexample to the generalized Ramanujan conjecture for (quasi-) split groups', in *Proceedings of Symposia in Pure Mathematics, Volume 33, Part 1: Automorphic Forms, Representations and L-functions*, vol. 1 (American Mathematical Soc., Providence, R.I., 1979), 315–322.
- [35] M. N. Huxley, 'Exceptional eigenvalues and congruence subgroups', *The Selberg Trace Formula and Related Topics, Contemp. Math* **53** (1986), 341–349.
- [36] S. Jana and A. Kamber, 'On the local  $L^2$ -bound of the Eisenstein series', Preprint, 2022, [arXiv:2210.16291](https://arxiv.org/abs/2210.16291).
- [37] S. Jana and A. Kamber, 'Optimal diophantine exponents for  $SL(n)$ ', Preprint, 2022, [arXiv:2211.05106](https://arxiv.org/abs/2211.05106).
- [38] A. Kamber, ' $L^p$ -expander complexes', Preprint, 2016, [arXiv:1701.00154](https://arxiv.org/abs/1701.00154).
- [39] A. Kamber, ' $L^p$ -expander graphs', *Israel Journal of Mathematics* **234**(2) (2019), 863–905.
- [40] A. Kamber and H. Lavner, 'Optimal lifting for the projective action of  $SL_3(\mathbb{Z})$ ', *Algebra Number Theory* **17**(3) (2023), 749–774.
- [41] M.-H. Kang, W.-C. W. Li and C.-J. Wang, 'The zeta functions of complexes from  $PGL(3)$ : A representation-theoretic approach', *Israel Journal of Mathematics* **177**(1) (2010), 335–348.
- [42] Y. Katznelson, 'Singular matrices and a uniform bound for congruence groups of  $SL_n(\mathbb{Z})$ ', *Duke Mathematical Journal* **69**(1) (1993), 121–136.
- [43] A. W. Knap, *Representation Theory of Semisimple Groups: An Overview Based on Examples (PMS-36)* (Princeton University Press, Princeton, NJ, 2016).
- [44] B. Kostant, 'On the existence and irreducibility of certain series of representations', *Bulletin of the American Mathematical Society* **75**(4) (1969), 627–642.
- [45] L. Lafforgue, 'Chtoucas de drinfeld et correspondance de langlands', *Inventiones mathematicae* **147**(1) (2002), 1–241.
- [46] P. D. Lax and R. S. Phillips, 'The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces', *Journal of Functional Analysis* **46**(3) (1982), 280–350.
- [47] E. Lubetzky, A. Lubotzky and O. Parzanchevski, 'Random walks on Ramanujan complexes and digraphs', *Journal of the European Mathematical Society* **22**(11) (2020), 3441–3466.
- [48] E. Lubetzky and Y. Peres, 'Cutoff on all Ramanujan graphs', *Geometric and Functional Analysis* **26**(4) (2016), 1190–1216.
- [49] A. Lubotzky, 'High dimensional expanders', in *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018* (World Scientific, Hackensack, NJ, 2018), 705–730.
- [50] A. Lubotzky, B. Samuels and U. Vishne, 'Explicit constructions of Ramanujan complexes of type  $A_d$ ', *European Journal of Combinatorics* **26**(6) (2005), 965–993.
- [51] A. Lubotzky, B. Samuels and U. Vishne, 'Ramanujan complexes of type  $A_d$ ', *Israel Journal of Mathematics* **149**(1) (2005), 267–299.
- [52] I. G. Macdonald, *Spherical functions on a group of p-adic type*, Publications of the Ramanujan Institute, No. 2 (University of Madras, Centre for Advanced Study in Mathematics, Ramanujan Institute, Madras, 1971), p. vii+79.
- [53] M. Magee, 'Quantitative spectral gap for thin groups of hyperbolic isometries', *Journal of the European Mathematical Society* **17**(1) (2015), 151–187.
- [54] H. Matsumoto, 'Fonctions spheriques sur un groupe semi-simple p-adique', *Comptes Rendus de l'Académie des Sciences, Paris, Ser A* **269** (1969), 829–832.
- [55] J. Matz and N. Templier, 'Sato–Tate equidistribution for families of Hecke–Maass forms on  $SL(n, \mathbb{R})/SO(n)$ ', *Algebra & Number Theory* **15**(6) (2021), 1343–1428.
- [56] N. T. Sardari, 'Diameter of Ramanujan graphs and random Cayley graphs', *Combinatorica* **39**(2) (2019), 427–446.
- [57] P. Sarnak, 'Letter to Stephen D. Miller and Naser Talebizadeh Sardari on optimal strong approximation by integral points on quadrics', 2015.

- [58] P. Sarnak and X. X. Xue, 'Bounds for multiplicities of automorphic representations', *Duke Math. J* **64**(1) (1991), 207–227.
- [59] F. Sauvageot, 'Principe de densité pour les groupes réductifs', *Compositio Mathematica* **108**(2) (1997), 151–184.
- [60] A. Seeger and C. D. Sogge, 'Bounds for eigenfunctions of differential operators', *Indiana University Mathematics Journal* **38**(3) (1989), 669–682.
- [61] A. G. Silberger, *Introduction to Harmonic Analysis on Reductive P-adic Groups.(MN-23): Based on Lectures by Harish-Chandra at The Institute for Advanced Study, 1971–73.* (Princeton University Press, Princeton, NJ, 2015).
- [62] J. Tits, 'Reductive groups over local fields', in *Automorphic Forms, Representations and L-functions* Proc. Sympos. Pure Math., vol. 1 (Oregon State Univ., Corvallis, OR, 1977), 29–69. Part (1979).