J. DIFFERENTIAL GEOMETRY
48 (1998) 165-203

THE TRACE CLASS CONJECTURE FOR ARITHMETIC GROUPS

LIZHEN JI

Abstract

In this paper we show that the counting function of the discrete spectrum of an arithmetic subgroup of a semisimple Lie group satisfies a polynomial upper bound, and we use it to prove the trace class conjecture in complete generality. To get this upper bound on the discrete spectrum, we introduce new spaces which are principal bundles over locally symmetric spaces and prove that the counting function of the eigenvalues of their pseudo-Laplacian satisfies the Weyl law.

1. Introduction

1.1.

Let $\mathbf{G}$ be a semisimple algebraic group defined over $\mathbb{Q}$, and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ an arithmetic subgroup. Fix a Haar measure on $G = \mathbf{G}(\mathbb{R})$, the real locus of $\mathbf{G}$. Then G has a right regular representation in $L^2(\Gamma\backslash G)$, denoted by R: for $f \in L^2(\Gamma\backslash G)$ and $g \in G$, $R(g)f(x) = f(xg)$. In the following, we assume that $\Gamma\backslash G$ is noncompact. Then the representation R has both a discrete spectrum and a continuous spectrum. Denote by $L^2_d(\Gamma\backslash G)$ the direct sum of irreducible subrepresentations of R and by $L^2_c(\Gamma\backslash G)$ the orthogonal complement of $L^2_d(\Gamma\backslash G)$ in $L^2(\Gamma\backslash G)$. The restriction of R to the discrete subspace $L^2_d(\Gamma\backslash G)$ is denoted by R_d.

For any $\alpha \in L^1(G)$, define an operator $R_d(\alpha)$ on $L^2_d(\Gamma\backslash G)$ by

$$R_d(\alpha) = \int_G \alpha(g) R_d(g) dg.$$

Received November 18, 1996, and, in revised form, December 8, 1997. Partially Supported by NSF postdoctoral fellowship 9407427 and NSF grant DMS 9704434.

In the theory of the Selberg trace formula and automorphic forms, a longstanding important problem is the following trace class conjecture [33, Open problem 4] [25, §4] [5, §4.7] [22] [32, §2.3] [13, p.14] [34].

Conjecture 1.1.1. *For any $\alpha \in C_0^\infty(G)$, the operator $R_d(\alpha)$ is of the trace class.*

Let $\mathcal{C}^1(G)$ be Harish-Chandra's Schwartz space of integrable rapidly decreasing functions on G (see [33, p.34] for definition), which clearly contains $C_0^\infty(G)$. The main result in this paper is the following.

Theorem 1.1.2 (§7). *For any $\alpha \in \mathcal{C}^1(G)$, the operator $R_d(\alpha)$ is of the trace class; in particular, Conjecture 1.1.1 holds.*

A different proof of this theorem is given by Müller in [23] and announced in [24].

If the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to one, the trace class conjecture was proved by Langlands [18]. If the convolution function α is assumed to be K-finite for a maximal compact subgroup K of G, then the trace class conjecture was proved by Müller [22]. Earlier in [10], Donnelly proved the conjecture for K-finite convolution functions α when the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to one.

One of the motivations for the trace class conjecture is as follows: Decompose $L_d^2(\Gamma\backslash G)$ into the irreducible subspaces:

$$L_d^2(\Gamma\backslash G) = \sum_{\pi\in\hat{G}} \oplus m(\pi)\pi,$$

where $\hat{G}$ is the set of unitary irreducible representations of G, and $m(\pi)$ is the multiplicity of π. An important problem in automorphic forms is to understand the set of π with nonzero multiplicity $m(\pi)$ and the values of $m(\pi)$. A powerful method for this problem is the Selberg trace formula, which is an equality between the spectral expression and the geometric expression of $tr(R_d(\alpha))$. The first step in establishing such a trace formula is the solution of the above trace class conjecture. Osborne and Warner have tried to establish such a trace formula under the assumption that the trace class conjecture holds in a series of papers (see [26] and the references there). On the other hand, by considering the restriction of R to the cuspidal subspace $L_{cus}^2(\Gamma\backslash G)$ and the decomposition of $L^2(\Gamma\backslash G)$ according to association classes of cuspidal data (see §3.4), Arthur [1] has developed a trace formula when Γ is a congruence subgroup, which is very useful for the purpose of comparing spectra of different groups, i.e., the functorial transfer in Langlands' program.

The proof of Theorem 1.1.2 follows easily from a polynomial upper bound on the discrete spectrum of the Laplace operator on $L^2(\Gamma\backslash G)$ in Theorem 1.1.3 below.

Fix a maximal compact subgroup K of G. Let $\mathfrak{g}$ and $\mathfrak{k}$ be the Lie algebras of G and K respectively. Then K induces the Cartan decomposition: $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$. Define an inner product on $\mathfrak{g}$ which is equal to the Killing form on $\mathfrak{p}$ and the negative of the Killing form on $\mathfrak{k}$. Then this inner product on $\mathfrak{g}$ extends to a left G-invariant Riemannian metric on G, which descends to a Riemannian metric on $\Gamma\backslash G$. Denote by Δ the Beltrami–Laplace operator of this metric on $\Gamma\backslash G$. Since this metric commutes with the right K-action, the Laplace operator Δ also commutes with the right K-action. This K-invariance of Δ will play an important role in this paper.

The Laplace operator Δ preserves the decomposition

$$L^2(\Gamma\backslash G) = L^2_d(\Gamma\backslash G) + L^2_c(\Gamma\backslash G),$$

and has a discrete spectrum $\mathrm{Spec}_d(\Delta)$ on $L^2_d(\Gamma\backslash G)$ and a continuous spectrum $\mathrm{Spec}_c(\Delta)$ on $L^2_c(\Gamma\backslash G)$. Denote the counting function of the discrete spectrum by

$$N_d(\lambda) = |\{\lambda_i \in \mathrm{Spec}_d(\Delta) \mid \lambda_i \leq \lambda\}|.$$

To state the the polynomial upper bound on $N_d(\lambda)$ mentioned earlier, we need some notation. For any proper rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, its real locus P admits the (rational) Langlands decomposition $P = N_P M_P A_P$. Define $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P}) = \dim A_P$, and $\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})$ to be the maximum of $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P})$ for all $\mathbf{P} \subset \mathbf{G}$. The symmetric space $X_P = M_P/(M_P \cap K)$ is called the boundary symmetric space associated with $\mathbf{P}$.

Theorem 1.1.3 (§6). *The counting function $N_d(\lambda)$ of the discrete spectrum satisfies the following upper bound:*

$$N_d(\lambda) \leq (1 + o(1))(4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash G)}{\Gamma(\frac{n}{2}+1)}\lambda^{n/2} + O(1)\lambda^{m/2},$$

where $n = \dim G$, m is the maximum of $(\mathrm{rank}_{\mathbb{Q}}(\mathbf{P})+1)(\dim X_P + \dim K)$ for all proper rational parabolic subgroups $\mathbf{P}$ of $\mathbf{G}$ such that

$$\mathrm{rank}_{\mathbb{Q}}(\mathbf{P}) \leq \mathrm{rank}_{\mathbb{Q}}(\mathbf{G}) - 1,$$

and $o(1)$ is a quantity going to zero as $\lambda \to +\infty$ while $O(1)$ is bounded. In particular, there exists a positive constant c such that

$$N_d(\lambda) \leq c(1 + \lambda^{\frac{n}{2}\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})}).$$

This improves a result of Borel and Garland [5, Theorem 3] that $N_d(\lambda) < +\infty$ for all $\lambda > 0$. Theorem 1.1.2 follows from Theorem 1.1.3 by a standard argument. In fact, any polynomial upper bound on $N_d(\lambda)$ implies Theorem 1.1.2 (see §7).

Theorem 1.1.4 (5.2.3). *If the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 1, then the counting function $N_d(\lambda)$ of the discrete spectrum satisfies the Weyl upper bound, i.e.,*

$$\lim_{\lambda \to +\infty} \sup \frac{N_d(\lambda)}{\lambda^{n/2}} \leq (4\pi)^{-n/2} \frac{\mathrm{vol}(\Gamma \backslash G)}{\Gamma(\frac{n}{2}+1)}, \quad n = \dim G.$$

Theorem 1.1.4 is a corollary of Theorem 1.1.3. In fact, it follows from the first part of the proof of Theorem 1.1.3 (see §5).

Besides this application to the solution of the trace class conjecture, Theorems 1.1.3 and 1.1.4 are interesting from the point of view of spectral geometry. It gives a quantitative description of the distribution of the irreducible subrepresentations of $L^2_d(\Gamma \backslash G)$.

The upper bound in Theorem 1.1.4 is sharp. In fact, if $\Gamma \backslash G$ is compact, then Δ has only a discrete spectrum on $L^2(\Gamma \backslash G)$, whose counting function $N_d(\lambda)$ satisfies the Weyl law, i.e., the equality holds in Theorem 1.1.4. A natural question is under what conditions on noncompact quotients $\Gamma \backslash G$, the Weyl law holds for the discrete spectrum, i.e., the inequality in Theorem 1.1.4 becomes an equality. It is certainly conceivable that this is the case whenever Γ is a congruence subgroup, but it is not known even for $G = \mathrm{SL}(2, \mathbb{R})$. More is known about the Weyl law for locally symmetric spaces; see [15, §1.2] for discussions on this problem for locally symmetric spaces. Since $L^2(\Gamma \backslash G)$ involves all representations of K, this problem will be more complicated for $L^2(\Gamma \backslash G)$.

1.2.

In the rest of this introduction, we outline the basic idea of the proof of Theorem 1.1.3 and the difficulties involved.

The upper bound in Theorem 1.1.3 is similar to the bound in [15, Theorem 1.1.3] for the Laplace operator on $\Gamma \backslash X$, where $X = G/K$ is the Riemannian symmetric space associated with G. In [15], the pseudo-Laplace operator Δ_T of $\Gamma\backslash X$ plays an important role. Inspired by this, we introduce the pseudo-Laplace operator Δ_T on $\Gamma\backslash G$ whose domain consists of those functions with vanishing constant terms above the height T along all proper parabolic subgroups of $\mathbf{G}$ (see §4 for a precise definition). An important property of Δ_T is that it has only a discrete spectrum. Denote the counting function of the spectrum of Δ_T by $N_T(\lambda)$. Using the precise reduction theory in [29], the bounds on the heat kernel as in [15] and Arthur's truncation operator Λ^T [2], we prove the following result.

Theorem 1.2.1 (4.2.2). *The counting function of the eigenvalues of the pseudo-Laplace operator Δ_T of $\Gamma\backslash G$ satisfies the Weyl law, i.e.,*

$$\lim_{\lambda\to+\infty}\frac{N_T(\lambda)}{\lambda^{n/2}}=(4\pi)^{-n/2}\frac{\operatorname{vol}(\Gamma\backslash G)}{\Gamma(\frac{n}{2}+1)},\quad n=\dim\Gamma\backslash G.$$

To derive Theorem 1.1.3 from Theorem 1.2.1, the problem is to relate the discrete spectrum of Δ to the eigenvalues of Δ_T. This is the same problem that we met in [15]. As in [15], our guiding philosophy is that Δ_T is a good approximation to Δ and the discrete spectrum of Δ can be uniformly approximated by a part of the eigenvalues of Δ_T. We will show that such a uniform approximation holds for a majority of the rank-one residual discrete spectrum of Δ and then derive Theorem 1.1.3 from Theorem 1.2.1. Though only the upper bound for $N_T(\lambda)$ is used in the proofs of Theorems 1.1.3 and 1.1.2, the lower bound is of independent interests. As pointed out in [15, §1.2], Theorem 1.2.1 can be interpreted as an analogue of the Weyl–Selberg law.

Let K be a maximal compact subgroup of G, and $X=G/K$ be the associated symmetric space as above. Then $\Gamma\backslash G$ is a principal K-bundle over $\Gamma\backslash X$. The basic idea is that since the Laplace operator Δ of $\Gamma\backslash G$ commutes with the right K-action, Δ can be decomposed as a sum of vector-valued Laplace operators on $\Gamma\backslash X$ which are studied in [15]. Similarly, the pseudo-Laplacian Δ_T can also be decomposed according to the action of K. Then we can use the results for these operators in [15] to compare the discrete spectrum of Δ and the eigenvalues of Δ_T.

To point out new difficulties occurring in this case, we recall the proof of the upper bound on the counting function of the discrete spectrum in [15]. Since the truncation of the constant terms has no effect on cuspidal functions, the cuspidal discrete spectrum is contained in the spectrum of the pseudo-Laplace operator. The problem is to bound the residual discrete spectrum. The residual eigenfunctions are given by iterated residues of Eisenstein series associated with various rational parabolic subgroups of $\mathbf{G}$. Based on the rank of the parabolic subgroups, the residual discrete spectrum can be decomposed into two parts: the rank-one and higher rank residual discrete spectra. The bound on the counting function of the residual discrete spectrum is obtained in two steps: (1) The majority of the rank-one residual discrete spectrum can be approximated uniformly by a part of the spectrum of the pseudo-Laplacian. (2) The counting function of the higher rank residual discrete spectrum can be bounded in terms of the counting function of the rank-one residual discrete spectrum and the pseudo-Laplacian of spaces of smaller dimension. We will use the same strategy to prove Theorem 1.1.3, but there is a new problem of bounding the multiplicities which occur in the decomposition of Δ according to the right action of K and restrictions of representations of K to its subgroups.

More precisely, any unitary irreducible representation σ of K defines a homogeneous bundle E_σ over $\Gamma\backslash X$. Denote the space of L^2-sections of E_σ by $L^2(\Gamma\backslash X, \sigma)$. Then

$$L^2(\Gamma\backslash G) = \sum_{\sigma\in\hat{K}} \oplus(\dim\sigma) L^2(\Gamma\backslash X, \sigma).$$

The restriction of the Laplace operator Δ of G to each subspace $L^2(\Gamma\backslash X, \sigma)$ is denoted by Δ_σ and is a shift of the Bochner–Laplace operator on $L^2(\Gamma\backslash X, \sigma)$ defined in [15] by a constant depending on σ. Similarly, the pseudo-Laplace operator Δ_T admits a decomposition with respect to the right K-action, and the restriction of Δ_T to the subspace corresponding to $L^2(\Gamma\backslash X, \sigma)$ is denoted by $\Delta_{T,\sigma}$ and is a shift of the pseudo-Laplace operator defined in [15] by the same constant determined by σ as above.

For every association class $\mathcal{C}$ of rational parabolic subgroups of $\mathbf{G}$, there is a subspace $L^2_{\mathcal{C}}(\Gamma\backslash G)$ of $L^2(\Gamma\backslash G)$ (§3.4). The subspace $L^2_{\mathcal{C}}(\Gamma\backslash G)$ also admits a decomposition according to the action of K:

$$L^2_{\mathcal{C}}(\Gamma\backslash G) = \sum_{\sigma\in\hat{K}} \oplus(\dim\sigma) L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma).$$

Denote the counting function of the discrete spectrum of Δ in $L^2_{\mathcal{C}}(\Gamma\backslash G)$ by $N_{\mathcal{C}}(\lambda)$, and that of Δ_σ in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ by $N_{\mathcal{C},\sigma}(\lambda)$. Similarly, the corresponding counting functions of the pseudo-Laplacians Δ_T and $\Delta_{T,\sigma}$ are denoted respectively by $N_{T,\mathcal{C}}(\lambda)$ and $N_{T,\mathcal{C},\sigma}(\lambda)$.

Suppose $\mathcal{C}$ is of rank-one, i.e., the parabolic subgroups in $\mathcal{C}$ are of rank-one. Let $\mathbf{P}_1, \cdots, \mathbf{P}_r$ be a set of representatives of Γ-conjugacy classes in $\mathcal{C}$. Let $P_i = N_{P_i} A_{P_i} M_{P_i}$ be the Langlands decomposition of P_i. For every $\sigma \in \hat{K}$ and $\mathbf{P}_i$, denote the counting function of the cuspidal discrete spectrum in $L^2(\Gamma_{M_{P_i}} \backslash X_{P_i}, \sigma_{M_{P_i}})$ by $N_{i,cus,\sigma}(\lambda)$, where $\sigma_{M_{P_i}}$ is the restriction of σ to the maximal compact subgroup $K_i = K \cap M_{P_i}$ of M_{P_i}. Denote by $|\rho|$ the common norm of the half sum of the roots in $\Sigma(P_i, A_{P_i})$ with multiplicity, $i = 1, \cdots, r$. Since Δ_σ and $\Delta_{T,\sigma}$ are shifts of the corresponding Laplace and pseudo-Laplace operators used in [15] by the same constant, by [15, Proposition 5.2.8], $N_{\mathcal{C},\sigma}(\lambda)$ is bounded from above as follows:

$$N_{\mathcal{C},\sigma}(\lambda) \leq N_{T,\mathcal{C},\sigma}(\lambda + |\rho|^2) + \sum_{i=1}^{r} N_{i,cus,\sigma}(\lambda).$$

Summing over representations $\sigma \in \hat{K}$, we get the following bound on the rank-one residual spectrum:

$$\begin{aligned} N_{\mathcal{C}}(\lambda) = \sum_{\sigma \in \hat{K}} (\dim \sigma) N_{\mathcal{C},\sigma}(\lambda) &\leq N_{T,\mathcal{C}}(\lambda + |\rho|^2) \\ &+ \sum_{i=1}^{r} \sum_{\sigma \in \hat{K}} (\dim \sigma) N_{i,cus,\sigma}(\lambda). \end{aligned}$$

Naturally, one expects to bound the sum $\sum_{\sigma \in \hat{K}} (\dim \sigma) N_{i,cus,\sigma}(\lambda)$ by the counting function $N_{i,cus}(\lambda)$ of the cuspidal spectrum of $L^2(\Gamma_{M_{P_i}} \backslash M_{P_i})$ since $\Gamma_{M_{P_i}} \backslash M_{P_i}$ is the natural analogue of $\Gamma \backslash G$ associated with $\mathbf{P}_i$. But

$$N_{i,cus}(\lambda) = \sum_{\delta \in \hat{K}_i} (\dim \delta) N_{i,cus,\delta}(\lambda),$$

where $K_i = K \cap M_{P_i}$ is a maximal compact subgroup of M_{P_i}. Since $K_i \neq K$, the restriction $\sigma_{M_{P_i}}$ to K_i of an irreducible representation σ of K is not irreducible in general, i.e.,

$$\sigma_{M_{P_i}} = \sum_{\delta \in \hat{K}_i} m(\delta) \delta,$$

where the multiplicity $m(\delta)$ will be nonzero for more than one δ in general. Then

$$L^2(\Gamma_{M_{P_i}} \backslash X_{P_i}, \sigma_{M_{P_i}}) = \sum_{\delta \in \hat{K}_i} m(\delta) L^2(\Gamma_{M_{P_i}} \backslash X_{P_i}, \delta),$$

and hence

$$N_{i,cus,\sigma}(\lambda) = \sum_{\delta \in \hat{K}_i} m(\delta) N_{i,cus,\delta}(\lambda),$$

$$\sum_{\sigma \in \hat{K}} (\dim \sigma) N_{i,cus,\sigma}(\lambda) = \sum_{\sigma \in \hat{K}} \dim \sigma \Big(\sum_{\delta \in \hat{K}_i} m(\delta) N_{i,cus,\delta}(\lambda)\Big).$$

Therefore, we can not use the counting function $N_{i,cus}(\lambda)$ of the cuspidal spectrum of $L^2(\Gamma_{M_{P_i}} \backslash M_{P_i})$ to bound the sum above $\sum_{\sigma \in \hat{K}} (\dim \sigma) N_{i,cus,\sigma}(\lambda)$. The basic reason is that $\Gamma_{M_{P_i}} \backslash M_{P_i}$ is a K_i-principal bundle over $\Gamma_{M_{P_i}} \backslash X_{P_i}$, while $\Gamma \backslash G$ is a K-principal bundle over $\Gamma \backslash X$ and K is strictly greater than K_i. Similar problems occur in the second step above in bounding the higher rank residual discrete spectrum. To overcome these problems, we need to study certain K-principal bundles over $\Gamma_{M_{P_i}} \backslash X_{P_i}$ (§2).

The study of these K-principal bundles is one of the new features of this paper and plays an important role in proving the upper bound on the discrete spectrum of $L^2(\Gamma \backslash G)$ in Theorem 1.1.3. From the discussions in §2, these spaces are natural and interesting in themselves.

1.3.

The rest of the paper is organized as follows. In §2, we review the Langlands decomposition of parabolic subgroups and their boundary spaces, and define the K-principal bundles over them which are needed to bound the residual discrete spectrum of $L^2(\Gamma \backslash G)$ as mentioned above. In §3, we generalize a few standard concepts in the theory of automorphic forms to these bundles and review the spectral decomposition of them, in particular $\Gamma \backslash G$, using the spectral decomposition of the locally symmetric spaces associated with them. In §4, we introduce the pseudo-Laplace operator on these K-principal bundles and show that the counting function of its eigenvalues satisfies the Weyl law. In §5, we show that the majority of the rank-one residual discrete spectrum of Δ can be approximated uniformly by the eigenvalues of Δ_T. In §6, we bound the counting function of the higher rank residual discrete spectrum and prove Theorem 1.1.3. In §7, we prove Theorem 1.1.2 by using Theorem 1.1.3.

1.4.

This paper is a continuation of [15], and several results and their proofs from [15] are used here, but we have tried to make this paper as selfcontained as possible. Both papers are motivated by and depend essentially on [22], and the results have been announced in [14]. Because of a gap in the last step in [15] bounding the counting function of the higher rank residual discrete spectrum (see [15, Remark 7.2.3]), the Weyl upper on $N_d(\lambda)$ announced in [14, Theorem 1.4.1] has to be replaced by the weaker one in Theorem 1.1.3 here. But this does not affect the solution of the trace class conjecture announced there ([14, Theorem 1.3.3]).

I would like to thank the referees for their careful reading and very helpful comments. I would also like to thank Professor A. Borel for his comments on the neatness assumption on the arithmetic subgroup Γ to avoid V-manifolds.

2. Boundary spaces $\Gamma_{M_P}\backslash X_P$ and K-principal bundles $\Gamma_{M_P}\backslash B_P$ over them

2.1.

In this section we define boundary spaces and K-principal bundles over them for rational parabolic subgroups of $\mathbf{G}$ (§2.2, §2.3). Then we extend the reduction theory of arithmetic subgroups to such spaces (§2.4).

2.2.

First we recall the construction of the left invariant Riemannian metric on G. Fix a maximal compact subgroup K of G as above. Let $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ be the Cartan decomposition of the Lie algebra $\mathfrak{g}$ of G. Then the Killing form is positive definite on $\mathfrak{p}$ and negative definite on $\mathfrak{k}$. By identifying the tangent space of $X = G/K$ at K with $\mathfrak{p}$, we define an invariant Riemannian metric on X. To get a Riemannian metric on G, we define an inner product on $\mathfrak{g}$ as follows: It is equal to the Killing form on $\mathfrak{p}$ and the negative of the Killing form on $\mathfrak{k}$. Under the left translation, this inner product defines a left G-invariant Riemannian metric on G. Since the adjoint action of K preserves the Killing form and the Cartan decomposition, this left invariant metric is also right K-invariant.

We recall the Langlands decomposition of rational parabolic subgroups of $\mathbf{G}$. For any proper rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, its nilpotent radical $\mathbf{N_P}$ is an algebraic group defined over $\mathbb{Q}$, and the Levi quotient $\mathbf{L_P} = \mathbf{N_P}\backslash\mathbf{P}$ is also an algebraic group defined over $\mathbb{Q}$. Denote by $X(\mathbf{L_P})_{\mathbb{Q}}$ the group of characters of $\mathbf{L_P}$ defined over $\mathbb{Q}$. Define

$\mathbf{M_P} = \cap_{\chi \in X(\mathbf{L_P})_{\mathbb{Q}}} \mathrm{Ker}(\chi^2)$, the anisotropic part of $\mathbf{L_P}$. Then $\mathbf{M_P}$ is an algebraic group defined over $\mathbb{Q}$. Let $\mathbf{S_P}$ denote the maximal $\mathbb{Q}$-split torus in the center of $\mathbf{L_P}$, and $A_P = \mathbf{S_P}(\mathbb{R})^0$, the connected component of the real locus. Then $L_P = M_P A_P$, where $L_P = \mathbf{L_P}(\mathbb{R})$, $M_P = \mathbf{M_P}(\mathbb{R})$.

Denote the Cartan involution of G associated with K by θ, and the real locus $\mathbf{P}(\mathbb{R})$ by P. Then there exists a unique lift $i_0 : L_P \to P$ of the Levi quotient L_P such that the image $i_0(L_P)$ is invariant under the Cartan involution θ. For simplicity, we identify L_P with its lift in P. Similarly, we also identify M_P and A_P with their lifts in P.

Denote the real locus $\mathbf{N_P}(\mathbb{R})$ by N_P. Then we have the following (rational) Langlands decomposition of P:

$$P = N_P M_P A_P, \tag{1}$$

i.e., the map $(n, m, a) \in N_P \times M_P \times A_P \to nma \in P$ is a diffeomorphism. Since A_P commutes with M_P, the Langlands decomposition can also be written as $P = N_P A_P M_P$.

Denote the intersection $M_P \cap K$ by K_{M_P}. Then K_{M_P} is a maximal compact subgroup of M_P. Define

$$X_P = M_P / K_{M_P}.$$

Then X_P is a product of a Riemannian symmetric space of noncompact type with a possible Euclidean factor. The Langlands decomposition of P induces the following horospherical decomposition of X:

$$X = N_P \times X_P \times A_P. \tag{2}$$

This horospherical decomposition of X plays an important role in compactifications of X and $\Gamma\backslash X$ (see [12] [16] and the references there), and X_P and its quotient often appear as a part of the boundary components of the compactifications. Inspired by this, we call X_P the boundary symmetric space associated with P.

The boundary symmetric space X_P can be canonically identified with a totally geodesic submanifold of X as follows: $mK_{M_P} \in X_P \to mK \in X$.

The map $\pi : G \to X = G/K$ gives the Lie group G a K-principal bundle structure over X. Identify X_P with the totally geodesic submanifold of X as above, and denote the inverse image $\pi^{-1}(X_P)$ by B_P.

Lemma 2.2.1. *The space B_P is invariant under the left multiplication by M_P and the right multiplication by K and hence is a K-principal bundle over the boundary symmetric space X_P.*

Proof. By definition $B_P = \{mk \in G \mid m \in M_P, k \in K\}$. Then it is clear that B_P is invariant under the left multiplication by M_P and the right multiplication by K. Since $X_P = M_P/M_P \cap K$, the right K-multiplication defines a K-principal structure on B_P over X_P.

Lemma 2.2.2. *The left invariant metric on G induces a Riemannian metric on B_P which is invariant under the left multiplication by M_P and the right multiplication by K.*

Proof. This lemma follows from the invariance of B_P under the left multiplication by M_P and the fact that the left invariant metric on G commutes with the right K-multiplication.

Remark 2.2.3. The space B_P may seem strange at first. The reason is that X_P is usually considered as a boundary space of X, but we need to embed X_P inside X in order to get the space B_P. We note that when $\mathbf{P} = \mathbf{G}$, $B_P = G$, which is the basic space in this paper. However, we need other spaces B_P when $\mathbf{P}$ are proper rational parabolic subgroups in order to bound the residual discrete spectrum of $L^2(\Gamma\backslash G)$ as explained in the introduction (see the proofs of Propositions 5.2.1 and 6.2.1 below). For other interpretations of B_P, see Remarks 2.3.3 and 2.3.4.

2.3.

We use the spaces B_P above to define K-bundles over the boundary spaces of $\Gamma\backslash X$ which will play an important role below.

Let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup. To avoid finite quotient singularities and V-manifolds, from now on, we assume for simplicity that Γ is a neat arithmetic subgroup (see [4, §17] for definition). Since every arithmetic subgroup admits a neat subgroup of finite index, this restriction is not severe.

For any rational parabolic subgroup $\mathbf{P}$, define $\Gamma_P = \Gamma \cap P$. Then $\Gamma_{N_P} = \Gamma_P \cap N_P$ is a cocompact subgroup in N_P. The quotient $\Gamma_{N_P}\backslash\Gamma_P$ is an arithmetic subgroup in $\mathbf{M_P}$ which can be identified with the image of Γ_P in M_P under the projection $P = N_P A_P M_P \to M_P$. Denote the image of Γ_P in M_P by Γ_{M_P}. Then Γ_{M_P} is an arithmetic subgroup of $\mathbf{M_P}$ and acts on the boundary symmetric space X_P. The quotient $\Gamma_{M_P}\backslash X_P$ is a locally symmetric space of finite volume and called the boundary locally symmetric space of $\Gamma\backslash X$ associated with the rational parabolic subgroup $\mathbf{P}$. In fact, $\Gamma_M\backslash X_P$ is the boundary component associated with $\mathbf{P}$ in the reductive Borel–Serre compactification of $\Gamma\backslash X$ (see [16]).

Lemma 2.3.1. *The group Γ_{M_P} acts isometrically and discretely on B_P, and the quotient $\Gamma_{M_P}\backslash B_P$ is a K-principal bundle over $\Gamma_{M_P}\backslash X_P$.*

Proof. By Lemma 2.2.1, B_P is a K-principal bundle over X_P. Since Γ_{M_P} acts discretely over X_P from the left, Γ_{M_P} acts discretely on B_P. Note that since Γ is neat, Γ_{M_P} is torsion free, and $\Gamma_{M_P}\cap K$ contains only the identity element. This implies that the quotient $\Gamma_{M_P}\backslash B_P$ inherits the K-principal bundle structure from B_P, since the K-principal bundle structure is given by the right K-multiplication.

Definition 2.3.2. The K-principal bundle $\Gamma_{M_P}\backslash B_P$ over $\Gamma_{M_P}\backslash X_P$ is called the boundary bundle associated with $\mathbf{P}$.

Remark 2.3.3. Using the reductive Borel-Serre compactification $\overline{\Gamma\backslash X}^{RBS}$ of $\Gamma\backslash X$ and the K-principal structure of $\Gamma\backslash G$ over $\Gamma\backslash X$, we can get a compactification of $\Gamma\backslash G$ which is also a K-principal bundle over $\overline{\Gamma\backslash X}^{RBS}$. Then the boundary bundle $\Gamma_{M_P}\backslash B_P$ is the boundary component of this compactification of $\Gamma\backslash G$ associated with $\mathbf{P}$. From this point of view, these boundary bundles are natural objects to consider.

Remark 2.3.4. After this paper was submitted, we realize that by writing $B_P = N_P A_P\backslash G$, B_P is similar to the space $\mathbf{N_P}(\mathbb{A})A_P\backslash\mathbf{G}(\mathbb{A})$, and $\Gamma_{M_P}\backslash B_P$ is similar to $\mathbf{N_P}(\mathbb{A})\mathbf{M_P}(\mathbb{Q})A_P\backslash\mathbf{G}(\mathbb{A})$, where $\mathbb{A}$ is the ring of adeles. The latter two spaces occur naturally in the spectral theory of $L^2(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A}))$ and automorphic forms (see [3] for example). On the other hand, our construction of B_P here is more geometric and its K-principal structure is evident, and the embedding into $\mathbf{G}$ naturally induces an invariant Riemannian metric on B_P.

Remark 2.3.5. Using the equality $B_P = N_P A_P\backslash G$, we can give another more intrinsic description of the compactification of $\Gamma\backslash G$ mentioned in Remark 2.3.3. Specifically, for every rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, define its boundary component $e(\mathbf{P})$ in the partial compactification $G\cup\coprod_{\mathbf{P}} e(\mathbf{P})$ of G by $e(\mathbf{P}) = B_P$, where B_P is glued on via the decomposition $G = N_P\times A_P\times(M_P K)$. Then the quotient $\Gamma\backslash G\cup\coprod_{\mathbf{P}} e(\mathbf{P})$ is the above compactification of $\Gamma\backslash G$. This compactification admits a right K-action. In fact, it admits a natural right G-action. For more details about this compactification and the G-action, see [6].

2.4.

For any rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, we generalize the reduction theory of arithmetic subgroups to the action of Γ_{M_P} on B_P.

For any rational parabolic subgroup $\mathbf{Q}$ of $\mathbf{M_P}$, let $Q = N_Q M_Q A_Q$ be the Langlands decomposition of $Q = \mathbf{Q}(\mathbb{R})$ (see Equation 2.2.1). Denote the Lie algebra of A_Q and N_Q by $\mathfrak{a}_Q$ and $\mathfrak{n}_Q$ respectively. Then $\mathfrak{a}_Q$ acts on $\mathfrak{n}_Q$ by the Lie bracket, and the set of roots is denoted by $\Sigma(Q, A_Q)$.

Let $\mathbf{Q}_1, \cdots, \mathbf{Q}_m$ be a set of representatives of Γ_{M_P}-conjugacy classes of maximal rational parabolic subgroups $\mathbf{Q}$ of $\mathbf{M_P}$, i.e., those $\mathbf{Q}$ with $\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) = 1$. Let $\mathfrak{a}_j = \mathfrak{a}_{Q_j}$, and define

$$\mathfrak{a}_0 = \oplus_{j=1}^m \mathfrak{a}_j.$$

Then for any rational parabolic subgroup $\mathbf{Q}$ of $\mathbf{M_P}$, there is a well-defined map

$$I_Q : \mathfrak{a}_0 \to \mathfrak{a}_Q$$

such that if $\mathbf{Q} = \mathbf{Q}_j$, $j = 1, \cdots, m$, I_{Q_j} is the projection from $\mathfrak{a}_0$ to the summand $\mathfrak{a}_j$ [27, p.330] [29].

For $j = 1, \cdots, m$, let $\rho_j \in \mathfrak{a}_j^*$ be the half sum of the roots in $\Sigma(Q_j, A_{Q_j})$ with multiplicity. Then ρ_j defines a vector H_{ρ_j} in $\mathfrak{a}_j$ under the duality between $\mathfrak{a}_j$ and $\mathfrak{a}_j^*$ defined by the Killing form. These vectors H_{ρ_j} define a vector H_ρ in $\mathfrak{a}_0$ such that $I_{Q_j}(H_\rho) = H_{\rho_j}$ for $j = 1, \cdots, m$.

Fix a large positive number t and a vector $T = tH_\rho \in \mathfrak{a}_0$. For any rational parabolic subgroup $\mathbf{Q}$, define subsets of A_Q:

$$A_{Q,T} = \{e^H \in A_Q \mid \alpha(H) > \alpha(I_Q(T)), \alpha \in \Sigma(Q, A_Q)\},$$

$$A_Q^T = \{e^H \in A_Q \mid \langle I_Q(T) - H, V\rangle \geq 0, \text{ for all } \ V \in \mathfrak{a}_Q^+\},$$

where $\mathfrak{a}_Q^+ = \{a \in \mathfrak{a}_Q \mid \alpha(H) > 0, \alpha \in \Sigma(Q, A_Q)\}$ is the positive (or dominant) cone. In other words, $A_{Q,T}$ is a shift of the positive cone $A_{Q,0} = \exp \mathfrak{a}_Q^+$, while A_Q^T is a shift of the negative of the obtuse cone dual to the positive cone $A_{Q,0}$.

For any compact subset $\omega \subset N_Q M_Q$, the set $\omega A_{Q,T}$ is called a Siegel set in M_P associated with the rational parabolic subgroup $\mathbf{Q}$. Inspired by this, we call the set $\omega A_{Q,T} K$ a Siegel set in $B_P = M_P K$.

Denote the projection $B_P \to \Gamma_{M_P}\backslash B_P$ by Proj. Then $\mathrm{Proj}(N_Q A_Q^T M_Q K)$ is a subset in $\Gamma_{M_P}\backslash B_P$ which is obtained by removing a neighborhood of the cusp of $\Gamma_{M_P}\backslash B_P$ associated with $\mathbf{Q}$.

Lemma 2.4.1. *For $t \gg 0$ and $T = tH_\rho$ as above, the intersection $\cap_{\mathbf{Q}}\mathrm{Proj}(N_Q A_Q^T M_Q K)$ over all proper rational parabolic subgroups $\mathbf{Q}$ of $\mathbf{M_P}$ is a compact submanifold with corner. This compact submanifold is called the compact core of $\Gamma_{M_P}\backslash B_P$ and denoted by $(\Gamma_{M_P}\backslash B_P)_T$.*

Proof. Denote the projection from X_P to $\Gamma_{M_P}\backslash X_P$ also by Proj. Identify X_P with $N_Q \times X_Q \times A_Q$ using the horospherical coordinates in Equation (2.2.2). By [29, §0 and §2.2], the intersection

$$\cap_{\mathbf{Q}}\mathrm{Proj}(N_Q \times X_Q \times A_Q^T)$$

over all proper rational parabolic subgroups $\mathbf{Q}$ of $\mathbf{M_P}$ is a compact submanifold with corner in $\Gamma_{M_P}\backslash X_P$. In fact, it is the image in $\Gamma_{M_P}\backslash X_P$ of the central tile in X_P. Since

$$\cap_{\mathbf{Q}}\mathrm{Proj}(N_Q A_Q^T M_Q K) \subset \Gamma_{M_P}\backslash B_P$$

is a K-principal bundle over

$$\cap_{\mathbf{Q}}\mathrm{Proj}(N_Q \times X_Q \times A_Q^T) \subset \Gamma_{M_P}\backslash X_P,$$

the former is also a compact submanifold with corner.

Lemma 2.4.2. *Let $\mathbf{Q}_1, \cdots, \mathbf{Q}_p$ be a set of representatives of Γ_{M_P}-conjugacy classes of proper rational parabolic subgroups of $\mathbf{M_P}$. Then for any $T \in \mathfrak{a}_0$ as above and $j = 1, \cdots, p$, there exists a compact submanifold with corner $\omega_j \subset \Gamma_{Q_j}\backslash N_{Q_j} M_{Q_j}$ such that $\omega_i A_{Q_i,T} K$ is a compact submanifold with corner in $\Gamma_{Q_j}\backslash B_P$ and is mapped homeomorphically into $\Gamma_{M_P}\backslash B_P$, and $\Gamma_{M_P}\backslash B_P$ has the following disjoint decomposition:*

$$\Gamma_{M_P}\backslash B_P = (\Gamma_{M_P}\backslash B_P)_T \cup \coprod_{j=1}^{p} \omega_j A_{Q_j,T} K.$$

Proof. Denote by $(\Gamma_{M_P}\backslash X_P)_T$ the intersection $\cap_{\mathbf{Q}}\mathrm{Proj}(N_Q \times X_Q \times A_Q^T)$ over all rational parabolic subgroups $\mathbf{Q}$ of $\mathbf{M_P}$, which is a compact submanifold with corner of $\Gamma_{M_P}\backslash X_P$ as mentioned in the proof of Lemma 2.4.1. By [29, §2.2], there exist compact submanifolds with corner $\omega'_j \subset \Gamma_{Q_j}\backslash N_{Q_j} \times X_{Q_j}$ such that $\Gamma_{M_P}\backslash X_P$ admits the following disjoint decomposition:

$$\Gamma_{M_P}\backslash X_P = (\Gamma_{M_P}\backslash X_P)_T \cup \coprod \omega'_j \times A_{Q_j,T}.$$

Since $\Gamma_{M_P}\backslash B_P$ is a principal K-bundle over $\Gamma_{M_P}\backslash X_P$, we get the disjoint decomposition of $\Gamma_{M_P}\backslash B_P$ in the lemma by lifting the above disjoint decomposition of $\Gamma_{M_P}\backslash X_P$.

3. The spectral decomposition of $L^2(\Gamma_{M_P}\backslash B_P)$

3.1.

For any rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, let $\Gamma_{M_P}\backslash B_P$ be the boundary bundle associated with $\mathbf{P}$ in Definition 2.3.2. In this section, we describe the spectral decomposition of $L^2(\Gamma_{M_P}\backslash B_P)$ (§3.4). The idea is to reduce the problem to the spectral decomposition of the locally symmetric space $\Gamma_{M_P}\backslash X_P$.

3.2.

We first generalize a few concepts of automorphic forms to the boundary bundle $\Gamma_{M_P}\backslash B_P$. Any function on $\Gamma_{M_P}\backslash B_P$ can be identified with a left Γ_{M_P}-invariant function on B_P. We recall that M_P acts on B_P from the left by multiplication.

For any function f on $\Gamma_{M_P}\backslash B_P$ and any rational parabolic subgroup $\mathbf{Q}$ of $\mathbf{M_P}$, the constant term f_Q of f along Q is defined by

$$f_Q(x) = \int_{\Gamma_{N_Q}\backslash N_Q} f(nx)dn,$$

where $\Gamma_{N_Q} = N_Q \cap \Gamma_{M_P}$.

A function f on $\Gamma_{M_P}\backslash B_P$ is called cuspidal if its constant terms over all proper rational parabolic subgroups of $\mathbf{M_P}$ vanish. The subspace of cuspidal functions of $L^2(\Gamma_{M_P}\backslash B_P)$ is denoted by $L^2_{cus}(\Gamma_{M_P}\backslash B_P)$, called the cuspidal subspace.

3.3.

The boundary bundle $\Gamma_{M_P}\backslash B_P$ has an induced invariant Riemannian metric induced from the left G-invariant metric on G (see Lemma 2.2.2). Denote the Beltrami–Laplace operator of $\Gamma_{M_P}\backslash B_P$ by Δ. Since the metric on $\Gamma_{M_P}\backslash B_P$ is invariant under the right K-multiplication, Δ commutes with the right K-multiplication. To get the spectral decomposition of Δ, we need to decompose $L^2(\Gamma_{M_P}\backslash B_P)$ according to the right K–action.

Remark 3.3.1. Denote the Casimir element of G by Ω_G and the Casimir element of K by Ω_K. Then $-\Omega_G+2\Omega_K$ defines the same left invariant differential operator as the Laplace operator Δ on G (see [5, §3.2]

[33, p.39]), and hence the Beltrami–Laplace operator of $\Gamma\backslash B_G = \Gamma\backslash G$. Since $-\Omega_G + 2\Omega_K$ commutes with the K action, the differential operator is also easily seen to be right K-invariant. The Laplace operator Δ on $\Gamma_{M_P}\backslash B_P$ is a natural generalization.

Lemma 3.3.2. *The right regular representation of K on $L^2(K)$ can be decomposed as follows:*

$$L^2(K) = \oplus_{\sigma\in\hat{K}} (\dim\sigma)\sigma.$$

Proof. The decomposition of the regular representation into a sum of the irreducible ones follows from the famous Peter–Weyl theorem, and the formula for the multiplicity follows from the Frobenius reciprocity law. See [17, pp. 16–20] for details. q.e.d.

For any irreducible unitary representation σ of K, denote the restriction of σ to K_{M_P} by σ_{M_P}, where $K_{M_P} = K \cap M_P$ is a maximal compact subgroup of M_P (see §2.2). This representation σ_{M_P} of K_{M_P} defines a homogeneous bundle $\tilde{E}_{\sigma_{M_P}}$ over $X_P = M_P/K_{M_P}$. The quotient of $\tilde{E}_\sigma$ by Γ_{M_P} is a locally homogeneous bundle $E_{\sigma_{M_P}}$ over $\Gamma_{M_P}\backslash X_P$. Denote by $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ the space of L^2-sections of $E_{\sigma_{M_P}}$.

Lemma 3.3.3. *For any $\sigma \in \hat{K}$, K acts diagonally on $L^2(\Gamma_{M_P}\backslash B_P)\otimes \sigma$. Denote the subspace of K-invariant vectors in $L^2(\Gamma_{M_P}\backslash B_P)\otimes\sigma$ by $(L^2(\Gamma_{M_P}\backslash B_P)\otimes\sigma)^K$. Then $(L^2(\Gamma_{M_P}\backslash B_P)\otimes\sigma)^K \cong L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$.*

Proof. It follows from the fact that $E_{\sigma_{M_P}} = \Gamma_{M_P}\backslash B_P \otimes_K \sigma$.

Proposition 3.3.4. *The space $L^2(\Gamma_{M_P}\backslash B_P)$ admits the decomposition:*

$$L^2(\Gamma_{M_P}\backslash B_P) = \oplus_{\sigma\in\hat{K}} (\dim\sigma) L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}).$$

The Laplace operator Δ preserves this decomposition, and the restriction of Δ to the subspace $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ is denoted by Δ_σ.

Proof. Since $(L^2(\Gamma_{M_P}\backslash B_P)\otimes L^2(K))^K = L^2(\Gamma_{M_P}\backslash B_P)$, the above decomposition of $L^2(\Gamma_{M_P}\backslash B_P)$ follows from the previous two lemmas. Since the Laplace operator Δ commutes with the right K-multiplication, it preserves this decomposition of $L^2(\Gamma_{M_P}\backslash B_P)$.

3.4.

We get the spectral decomposition of $L^2(\Gamma_{M_P}\backslash B_P)$ using the spectral decomposition of the subspaces $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$.

Each space $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ can be decomposed according to association classes of rational parabolic subgroups of $\mathbf{M_P}$ and the cuspidal spectra of the boundary locally symmetric spaces, and these subspaces can further be decomposed using discrete spectra and Eisenstein series. We recall briefly several concepts which will be used below. For a summary of the spectral decomposition of $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$, see [15, §2]. Besides the original book by Langlands [19], there are two other monographs [28] [21] treating the spectral decomposition of $L^2(\Gamma\backslash G)$ in great details.

Recall that two rational parabolic subgroups $\mathbf{Q_1}, \mathbf{Q}_2$ of $\mathbf{M_P}$ are defined to be associated if their split components A_{Q_1}, A_{Q_2} are conjugate under some element of M_P. For each association class $\mathcal{C}$ of rational parabolic subgroups of $\mathbf{M_P}$, define

$$\mathrm{Spec}_{cus,\sigma}(\mathcal{C}) = \cup_{\mathbf{Q}\in\mathcal{C}}\mathrm{Spec}_{cus}(\Gamma_{M_Q}\backslash X_Q, \sigma_{M_Q}),$$

where $\mathrm{Spec}_{cus}(\Gamma_{M_Q}\backslash X_Q, \sigma_{M_Q})$ is the spectrum of Δ_σ on the cuspidal subspace $L^2_{cus}(\Gamma_{M_Q}\backslash X_Q, \sigma_{M_Q})$ and

$$\mathrm{Spec}_{cus}(\mathcal{C}) = \cup_{\sigma\in\hat{K}}\mathrm{Spec}_{cus,\sigma}(\mathcal{C}).$$

For any $\mu \in \mathrm{Spec}_{cus,\sigma}(\mathcal{C})$, denote by $\mathcal{E}_{cus}(\Gamma_{M_Q}\backslash X_Q, \sigma_{M_Q}, \mu)$ the space of cuspidal eigenfunctions of eigenvalue μ in $L^2(\Gamma_{M_Q}\backslash X_Q, \sigma_{M_Q})$. For every $\Phi \in \mathcal{E}_{cus}(\Gamma_{M_Q}\backslash X_Q, \sigma_{M_Q}, \mu)$ and $f \in C_0^\infty(\mathfrak{a}_Q)$, the pseudo-Eisenstein series $E(Q, \Phi, f)$ is defined by

$$E(Q, \Phi, f) = \sum_{\gamma\in\Gamma_Q\backslash\Gamma_{M_P}} f(H_Q(\gamma x))\Phi(z_Q(\gamma x)),$$

where $\Gamma_Q = Q \cap \Gamma_{M_P}$,

$$\gamma x = (u_Q(\gamma x), z_Q(\gamma x), \exp H_Q(\gamma x)) \in N_Q \times X_Q \times A_Q = X_P,$$

the horospherical decomposition of X_P with respect to Q (Equation 2 in §2.2). Denote by $L^2_{\mathcal{C},\mu}(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ the subspace of $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ spanned by the pseudo-Eisenstein series $E(Q, \Phi, f)$, where $\mathbf{Q} \in \mathcal{C}$, $\Phi \in \mathcal{E}_{cus}(\Gamma_{M_Q}\backslash X_Q, \sigma_{M_Q}, \mu)$, $f \in C_0^\infty(\mathfrak{a}_Q)$. If we replace $f(H)$ in the above equation defining $E(Q, \Phi, f)$ by $\exp(\rho_Q + \Lambda)(H)$, $\Lambda \in \mathfrak{a}_Q^* \otimes \mathbb{C}$, which is clearly not of compact support, we get the Eisenstein series $E(Q, \Phi, \Lambda)$ associated with the parabolic subgroup $\mathbb{Q}$ and the eigenfunction Φ. Define

$$L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}) = \sum_{\mu\in\mathrm{Spec}_{cus,\sigma}(\mathcal{C})} \oplus L^2_{\mathcal{C},\mu}(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}).$$

Lemma 3.4.1 ([15, Lemma 2.5.3]). *With the above notation,*

$$
\begin{aligned}
L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}) &= \sum_{\mathcal{C}} \oplus L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}) \\
&= \sum_{\mathcal{C}} \sum_{\mu\in \mathrm{Spec}_{cus,\sigma}(\mathcal{C})} \oplus L^2_{\mathcal{C},\mu}(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}),
\end{aligned}
$$

where $\mathcal{C}$ runs over all the association classes of rational parabolic subgroups in $\mathbf{M_P}$.

Proposition 3.4.2. *For every association class $\mathcal{C}$ of rational parabolic subgroups in $\mathbf{M_P}$, define*

$$
L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash B_P) = \sum_{\sigma\in\hat{K}} (\dim\sigma) L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}).
$$

Then

$$
\begin{aligned}
L^2(\Gamma_{M_P}\backslash B_P) &= \sum_{\mathcal{C}} L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash B_P) \\
&= \sum_{\mathcal{C}} \sum_{\sigma\in\hat{K}} \dim\sigma \sum_{\mu\in \mathrm{Spec}_{cus,\sigma}(\mathcal{C})} \oplus L^2_{\mathcal{C},\mu}(\Gamma_{M_P}\backslash X_P, \sigma_{M_P}),
\end{aligned}
$$

where $\mathcal{C}$ runs over all the association classes of rational parabolic subgroups in $\mathbf{M_P}$.

Proof. It follows from Proposition 3.3.4 and the previous lemma.

Corollary 3.4.3. *The space $L^2(\Gamma\backslash G)$ admits the following orthogonal decomposition:*

$$
L^2(\Gamma\backslash G) = \sum_{\mathcal{C}} L^2_{\mathcal{C}}(\Gamma\backslash G) = \sum_{\mathcal{C}} \sum_{\sigma\in\hat{K}} \dim\sigma \sum_{\mu\in \mathrm{Spec}_{cus,\sigma}(\mathcal{C})} \oplus L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma),
$$

where $\mathcal{C}$ is over all association classes of rational parabolic subgroups in $\mathbf{G}$.

Remark 3.4.4. For a proper rational parabolic subgroup $\mathbf{P}$, $\dim \Gamma_{M_P}\backslash B_P < \dim \Gamma\backslash G$, but every representation σ of K appears with the same multiplicity $\dim\sigma$ in the above decompositions of $L^2(\Gamma_{M_P}\backslash B_P)$ and $L^2(\Gamma\backslash G)$. This is the reason for introducing the auxiliary boundary bundles $\Gamma_{M_P}\backslash B_P$ (see the proofs of Propositions 5.2.1 and 6.2.1 below).

Generalizing the analogous concepts for $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ to $L^2(\Gamma_{M_P}\backslash B_P)$, we define $L^2_d(\Gamma_{M_P}\backslash B_P)$ to be the subspace of $L^2(\Gamma_{M_P}\backslash B_P)$

spanned by L^2-eigenfunctions of Δ, and $L^2_c(\Gamma_{M_P}\backslash B_P)$ the orthogonal complement of $L^2_d(\Gamma_{M_P}\backslash B_P)$ in $L(\Gamma_{M_P}\backslash B_P)$. Then

$$L^2_{cus}(\Gamma_{M_P}\backslash B_P) \subset L^2_d(\Gamma_{M_P}\backslash B_P).$$

The orthogonal complement of $L^2_{cus}(\Gamma_{M_P}\backslash B_P)$ in $L^2_d(\Gamma_{M_P}\backslash B_P)$ is denoted by $L^2_{res}(\Gamma_{M_P}\backslash B_P)$ and called the residual discrete subspace, because it is spanned by iterated residues of Eisenstein series. Using the decomposition in Proposition 3.4.2 and the spectral decomposition of each subspace there, we get the following more refined decompositions.

Lemma 3.4.5. *Define*

$$L^2_{\mathcal{C},res}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}) = L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}) \cap L^2_{res}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P})$$

and

$$L^2_{\mathcal{C},c}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}) = L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}) \cap L^2_{c}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}).$$

Define $L^2_{\mathcal{C},\mu,res}(\Gamma_{M_P}\backslash B_P,\sigma_{M_P})$ *and* $L^2_{\mathcal{C},\mu,c}(\Gamma_{M_P}\backslash B_P,\sigma_{M_P})$ *similarly. Then*

$$\begin{aligned} L^2_{cus}(\Gamma_{M_P}\backslash B_P) &= \sum_{\mathcal{C}=\{\mathbf{M_P}\}} L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash B_P) \\ &= \sum_{\mathcal{C}=\{\mathbf{M_P}\}} \sum_{\sigma\in\hat{K}} \dim\sigma \sum_{\mu\in \mathrm{Spec}_{cus,\sigma}(\mathcal{C})} L^2_{\mathcal{C},\mu}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}), \end{aligned}$$

$$\begin{aligned} L^2_{res}(\Gamma_{M_P}\backslash B_P) &= \sum_{\mathcal{C}\neq\{\mathbf{M_P}\}} L^2_{\mathcal{C},res}(\Gamma_{M_P}\backslash B_P) \\ &= \sum_{\mathcal{C}\neq\{\mathbf{M_P}\}} \sum_{\sigma\in\hat{K}} \oplus(\dim\sigma) L^2_{\mathcal{C},res}(\Gamma_{M_P}\backslash B_P,\sigma_{M_P}) \\ &= \sum_{\mathcal{C}\neq\{\mathbf{M_P}\}} \sum_{\sigma\in\hat{K}} \dim\sigma \sum_{\mu\in \mathrm{Spec}_{cus,\sigma}(\mathcal{C})} \oplus L^2_{\mathcal{C},\mu,res}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}), \end{aligned}$$

$$\begin{aligned} L^2_{c}(\Gamma_{M_P}\backslash B_P) &= \sum_{\mathcal{C}\neq\{\mathbf{M_P}\}} L^2_{\mathcal{C},c}(\Gamma_{M_P}\backslash B_P) \\ &= \sum_{\mathcal{C}\neq\{\mathbf{M_P}\}} \sum_{\sigma\in\hat{K}} \oplus(\dim\sigma) L^2_{\mathcal{C},c}(\Gamma_{M_P}\backslash B_P,\sigma_{M_P}) \\ &= \sum_{\mathcal{C}\neq\{\mathbf{M_P}\}} \sum_{\sigma\in\hat{K}} \dim\sigma \sum_{\mu\in \mathrm{Spec}_{cus,\sigma}(\mathcal{C})} \oplus L^2_{\mathcal{C},\mu,c}(\Gamma_{M_P}\backslash X_P,\sigma_{M_P}). \end{aligned}$$

In particular, the cuspidal subspace $L^2_{cus}(\Gamma_{M_P}\backslash B_P)$ comes from the association class $\mathcal{C} = \{\mathbf{M_P}\}$, while the residual discrete spectrum comes from the association classes of proper rational parabolic subgroups of $\mathbf{M_P}$.

These decompositions are preserved by Δ and hence induce the corresponding decompositions of Δ. If the rank of an association class $\mathcal{C}$, i.e., the rank of the parabolic subgroups in $\mathcal{C}$, is equal to r, the residual discrete spectrum in $L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash B_P)$ is said of rank r. In particular, the total residual discrete spectrum is divided into two parts: the rank-one residual discrete spectrum and the higher rank residual discrete spectrum. It will be seen below that it is easier to bound the counting function of the former.

Remark 3.4.6. There are finitely many association classes $\mathcal{C}$ of rational parabolic subgroups of $\mathbf{G}$, and each subspace $L^2_{\mathcal{C}}(\Gamma\backslash G)$ is invariant under the right action of G. Another more important reason for grouping the rational parabolic subgroups into association classes is that the subspaces for different association classes are orthogonal to each other, and finer divisions of the rational parabolic subgroups will not, in general, produce orthogonal (or direct sum) decompositions of $L^2(\Gamma\backslash G)$. The further decomposition above according to the cuspidal eigenvalues μ and the representations $\sigma \in \hat{K}$ basically separates out different layers of the spectrum and makes it easier to understand the structure of the spectrum.

Remark 3.4.7. For each association class $\mathcal{C}$, $L^2_{\mathcal{C},res}(\Gamma\backslash G)$ is spanned by residues of Eisenstein series $E(P, \Phi, \Lambda)$, where $\mathbf{P} \in \mathcal{C}$ and Φ is a cuspidal eigenfunction of eigenvalue in $\mathrm{Spec}_{cus}(\mathcal{C})$; and $L^2_{\mathcal{C},c}(\Gamma\backslash G)$ is spanned by superpositions (i.e., wave packets) of the Eisenstein series $E(P, \Phi, \Lambda)$ above when $\mathrm{Re}(\Lambda) = 0$.

4. The Weyl law for the Pseudo-Laplacian Δ_T

4.1.

In this section, we introduce the pseudo-Laplacian Δ_T associated with the Laplacian Δ on $\Gamma_{M_P}\backslash B_P$ and show that the counting function of its spectrum satisfies the Weyl law (4.2.2). The upper bound for $N_T(\lambda)$ plays an important role in the upper bound on the discrete spectrum of

Δ in Theorem 1.1.3. As a corollary, we show that the cuspidal discrete spectrum of Δ satisfies the Weyl upper bound (4.2.3). Basically, Δ_T is obtained from Δ by restricting it to the subspace of functions whose constant terms along every rational parabolic subgroup of $\mathbf{M_P}$ vanish above the height T.

4.2.

Before defining the pseudo-Laplacian Δ_T, we notice that for every rational parabolic subgroup $\mathbf{Q}$ of $\mathbf{M_P}$, the Langlangs decomposition of Q (see Equation 1 in §2.1) induces the following (generalized) horospherical decomposition of B_P:

$$B_P = N_Q A_Q M_Q K = N_Q \times A_Q \times (M_Q K),$$

i.e., every element $x \in B_P$ can be written uniquely in the form $x = u_Q(x) \exp H_Q(x) m_Q(x)$, where $u_Q(x) \in N_Q$, $H_Q(x) \in \mathfrak{a}_Q$, $m_Q(x) \in M_Q K$.

Denote by $H^1(\Gamma_{M_P} \backslash B_P)$ the Sobolev space of functions f on $\Gamma_{M_P} \backslash B_P$ satisfying

$$\int_{\Gamma_{M_P} \backslash B_P} |f|^2 + |\bigtriangledown f|^2 < +\infty.$$

For a large truncation parameter $T \in \mathfrak{a}_0$ as in §2.4, define a subspace $H^1_T(\Gamma_{M_P} \backslash B_P)$ of $H^1(\Gamma_{M_P} \backslash B_P)$ as follows:

$$\begin{aligned} H^1_T(\Gamma_{M_P} \backslash B_P) =& \{ f \in H^1(\Gamma_{M_P} \backslash B_P) \mid f_Q(am) = 0 \\ & \text{for } a \in A_{Q,T}, m \in M_Q K \} \end{aligned}$$

for all the proper rational parabolic subgroups $\mathbf{Q}$ of $\mathbf{M_P}$. In other words, the subspace $H^1_T(\Gamma_{M_P} \backslash B_P)$ consists of functions whose constant terms vanish outside the compact core $(\Gamma_{M_P} \backslash B_P)_T$.

Since $H^1_T(\Gamma_{M_P} \backslash B_P)$ is a closed subspace of $H^1(\Gamma_{M_P} \backslash B_P)$, the Dirichlet quadratic form

$$D(f) = \int_{\Gamma_{M_P} \backslash B_P} |\bigtriangledown f|^2$$

restricts to $H^1_T(\Gamma_{M_P} \backslash B_P)$ and defines a self-adjoint operator Δ_T on the closure of $H^1_T(\Gamma_{M_P} \backslash B_P)$ in $L^2(\Gamma_{M_P} \backslash B_P)$. This operator is called the pseudo-Laplace operator associated with the Laplace operator Δ at the height T. The various decompositions of $L^2(\Gamma_{M_P} \backslash B_P)$ in §3.3 and §3.4

also induce the corresponding decompositions of Δ_T. This can be proved by similar arguments as in [15, §3.8]. For example,

$$\Delta_T = \sum_{\sigma\in\hat{K}} \oplus(\dim\sigma)\Delta_{T,\sigma} = \sum_{\mathcal{C}}\sum_{\sigma\in\hat{K}} \oplus(\dim\sigma)\Delta_{T,\mathcal{C},\sigma},$$

where $\Delta_{T,\sigma}$ is the pseudo-Laplacian for $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ and $\Delta_{T,\mathcal{C},\sigma}$ is the pseudo-Laplacian for $L^2_{\mathcal{C}}(\Gamma_{M_P}\backslash B_P, \sigma_{M_P})$. (As mentioned in §1.2, $\Delta_{T,\sigma}$, $\Delta_{T,\mathcal{C},\sigma}$ are shifts of the corresponding operators defined in [15] by a constant depending on σ.)

Remark 4.2.1. For locally symmetric spaces of $G = \mathrm{SL}(2, R)$, the notion of the pseudo-Laplacian was first defined by Lax and Phillips [20] and used by Colin de Verdiere [8] to study the meromorphic continuation of Eisenstein series and the discrete spectrum. For general locally symmetric spaces, the pseudo-Laplacian was defined by Müller [22]. Our definition of the pseudo-Laplacian here is a direct generalization to the boundary bundles.

The main result of this section is the following.

Theorem 4.2.2. *For any rational parabolic subgroup* $\mathbf{P}$ *of* $\mathbf{G}$ *and its associated boundary bundle* $\Gamma_{M_P}\backslash B_P$*, the spectrum* $\mathrm{Spec}(\Delta_T)$ *of the pseudo-Laplacian* Δ_T *is discrete, and its counting function*

$$N_{T,P}(\lambda) = |\{\lambda_i \in \mathrm{Spec}(\Delta_T) \mid \lambda_i \leq \lambda\}|$$

satisfies the Weyl law:

$$\lim_{\lambda\to+\infty} \frac{N_{T,P}(\lambda)}{\lambda^{d/2}} = (4\pi)^{-d/2}\frac{\mathrm{vol}(\Gamma_{M_P}\backslash B_P)}{\Gamma(\frac{d}{2}+1)},$$

where $d = \dim \Gamma_{M_P}\backslash B_P$.

When $\mathbf{P} = \mathbf{G}$, Theorem 4.2.2 reduces to Theorem 1.2.1 in the introduction. The proof of Theorem 4.2.2 is given in the next two subsections.

Corollary 4.2.3. *Denote the counting function of the cuspidal discrete spectrum of* Δ *on* $L^2_{cus}(\Gamma_{M_P}\backslash B_P)$ *by* $N_{cus,P}(\lambda)$*. Then*

$$\lim_{\lambda\to+\infty} \sup \frac{N_{cus,P}(\lambda)}{\lambda^{d/2}} \leq (4\pi)^{-d/2}\frac{\mathrm{vol}(\Gamma_{M_P}\backslash B_P)}{\Gamma(\frac{d}{2}+1)},$$

where $d = \dim \Gamma_{M_P}\backslash B_P$.

Proof. Since the truncation of the constant terms above the height T does not change the cuspidal functions, every cuspidal eigenfunction is also an eigenfunction of Δ_T and hence $N_{cus,P}(\lambda) \leq N_{T,P}(\lambda)$. Then this corollary follows from Theorem 4.2.2. q.e.d.

When $\mathbf{P} = \mathbf{G}$, this corollary improves the well-known result of Gelfand and Piatetski–Shapiro [11] [13, Theorem 3] that the spectrum of Δ on $L^2_{cus}(\Gamma\backslash G)$ is discrete, i.e., $N_{cus}(\lambda) < +\infty$ for every $\lambda > 0$.

4.3.

First we prove the upper bound for $N_{T,P}(\lambda)$ in Theorem 4.2.2.

Proposition 4.3.1. *With the above notation, we have*

$$\lim_{\lambda\to+\infty} \sup \frac{N_{T,P}(\lambda)}{\lambda^{d/2}} \leq (4\pi)^{-d/2}\frac{\mathrm{vol}(\Gamma_{M_P}\backslash B_P)}{\Gamma(\frac{d}{2}+1)}.$$

This upper bound follows from similar arguments as in the proof of [9, Theorem 1.1]. We recall the basic steps in the proof of [9, Theorem 1.1] and indicate necessary modifications.

Let $\mathbf{Q}_1, \cdots, \mathbf{Q}_p$ be a set of representatives of Γ_{M_P}-conjugacy classes of proper rational parabolic subgroups of $\mathbf{M_P}$ as in §2.4. For $k \geq 1$, define $T_k = kH_\rho \in \mathfrak{a}_0$, where H_ρ is defined in §2.4. By Lemma 2.4.2, for any $j = 1, \cdots, p$ and $k \geq 1$, there exists a compact submanifold with corner $\omega_{j,k} \subset \Gamma_{Q_j}\backslash N_{Q_j}M_{Q_j}$ such that $\Gamma_{M_P}\backslash B_P$ admits the following disjoint decomposition:

$$\Gamma_{M_P}\backslash B_P = (\Gamma_{M_P}\backslash B_P)_{T_k} \cup \coprod_{j=1}^{p} \omega_{j,k}A_{Q_j,T_k}K.$$

By slightly enlarging the submanifolds $\omega_{j,k}A_{Q_j,T_k}K$ and smoothing out their corners, we get submanifolds $Y_{j,k}$ in $\Gamma_{Q_j}\backslash B_P$ satisfying the following conditions:

1. The inverse image of $Y_{j,k}$ in B_P is left N_{Q_j}-invariant.

2. Every $Y_{j,k}$ is mapped homeomorphically into $\Gamma_{M_P}\backslash B_P$.

3. $Y_{1,k}, \cdots, Y_{p,k}$ and $(\Gamma_{M_P}\backslash B_P)_{T_k}$ cover $\Gamma_{M_P}\backslash B_P$.

4. $\cup_{j=1}^{p} Y_{j,k} \subset \Gamma_{M_P}\backslash B_P - (\Gamma_{M_P}\backslash B_P)_{T_{k-1}}$ and hence

$$\sum_{j=1}^{p} \mathrm{vol}(Y_{j,k}) \to 0, \qquad \text{as } k \to +\infty.$$

Since the inverse image of $Y_{j,k}$ in B_P is invariant under N_{Q_j}, the cuspidal subspace $L^2_{cus}(Y_{j,k})$ can be defined by

$$L^2_{cus}(Y_{j,k}) = \{f \in L^2(Y_{j,k}) \mid f_Q = 0 \text{ for all } \mathbf{Q} \supset \mathbf{Q}_j\}.$$

Denote the Laplace operator acting on $L^2_{cus}(Y_{j,k})$ and satisfying the Neumann boundary condition by $\Delta_{j,k,N}$. Let $N_{j,k,N}(\lambda)$ be the counting function of the eigenvalues of $\Delta_{j,k,N}$ on $L^2_{cus}(Y_{j,k})$, $j = 1, \cdots, p$, $k \geq 1$.

Proposition 4.3.2. *For $j = 1, \cdots, p$ and $k \geq 1$, the spectrum of $\Delta_{j,k,N}$ on $L^2_{cus}(Y_{j,k})$ is discrete, and there exist a sequence of constants $C_{j,k} \to 0$ as $k \to +\infty$ such that*

$$\lim_{\lambda \to +\infty} \sup \frac{N_{j,k,N}(\lambda)}{\lambda^{d/2}} \leq C_{j,k},$$

where $d = \dim \Gamma_{M_P} \backslash B_P$.

The proof of this proposition will be sketched below. Assume this proposition first. Denote by $N_{k,N}(\lambda)$ the counting function of the Neumann eigenvalues of the compact submanifold $(\Gamma_{M_P} \backslash B_P)_{T_k}$.

Proposition 4.3.3. *For any truncation parameter $T = tH_\rho \in \mathfrak{a}_0$ as above, when $k \geq t+1$, $(\Gamma_{M_P} \backslash B_P)_T$ is contained in $(\Gamma_{M_P} \backslash B_P)_{T_k}$, and for $j = 1, \cdots, p$, the submanifolds $Y_{j,k}$ are contained in the compelment of $(\Gamma_{M_P} \backslash B_P)_T$ in $\Gamma_{M_P} \backslash B_P$, and the counting function $N_{T,P}(\lambda)$ of the pseudo-Laplacian Δ_T of $\Gamma_{M_P} \backslash B_P$ is bounded as follows:*

$$N_{T,P}(\lambda) \leq N_{k,N}(\lambda) + \sum_{j=1}^{p} N_{j,k,N}(\lambda).$$

Proof. Since $Y_{j,k}$ is contained in the complement of $(\Gamma_{M_P} \backslash B_P)_T$, for every $f \in H^1_T(\Gamma_{M_P} \backslash B_P)$, $f_Q = 0$ on $Y_{j,k}$ for every $\mathbf{Q} \supset \mathbf{Q}_j$. This implies that the closure of $\{f|_{Y_{j,k}} \mid f \in H^1_T(\Gamma_{M_P} \backslash B_P)\}$ in $L^2(Y_{j,k})$ belongs to $L^2_{cus}(Y_{j,k})$. Then the proposition follows from the principle of Neumann bracketing. q.e.d.

Proof of Proposition 4.3.1

Since $(\Gamma_{M_P}\backslash B_P)_{T_k}$ is a compact manifold with corners, $N_{k,N}(\lambda)$ satisfies the Weyl law (see [31, Corollary 2.5]):

$$\lim_{\lambda\to+\infty}\frac{N_{k,N}(\lambda)}{\lambda^{d/2}}=(4\pi)^{-d/2}\frac{\mathrm{vol}((\Gamma_{M_P}\backslash B_P)_{T_k})}{\Gamma(\frac{d}{2}+1)}.$$

By Propositions 4.3.2 and 4.3.3, for any $\varepsilon>0$, when $k\gg 1$,

$$\begin{aligned}\lim_{\lambda\to+\infty}\sup\frac{N_{T,P}(\lambda)}{\lambda^{d/2}}&\le(4\pi)^{-d/2}\frac{\mathrm{vol}((\Gamma_{M_P}\backslash B_P)_{T_k})}{\Gamma(\frac{d}{2}+1)}+\varepsilon\\&\le(4\pi)^{-d/2}\frac{\mathrm{vol}(\Gamma_{M_P}\backslash B_P)}{\Gamma(\frac{d}{2}+1)}+\varepsilon.\end{aligned}$$

Since ε is arbitrary, we get

$$\lim_{\lambda\to+\infty}\sup\frac{N_{T,P}(\lambda)}{\lambda^{d/2}}\le(4\pi)^{-d/2}\frac{\mathrm{vol}(\Gamma_{M_P}\backslash B_P)}{\Gamma(\frac{d}{2}+1)}.$$

q.e.d.

Proof of Proposition 4.3.2

Proposition 4.3.2 is the analogue of [9, Corollary 7.6], and the same arguments as in the proof of [9, Theorem 7.2, Corollary 7.6] will prove Proposition 4.3.2.

In [9], Corollary 7.6 follows from Theorem 6.2. Two key estimates in the proof of Theorem 6.2 there are Lemma 3.3 and the upper bounds on the heat kernel and its derivatives in Inequality (5.2).

Since B_P is a K-principal bundle over X_P, in particular admits a cocompact quotient, the heat kernel of B_P satisfies the analogous upper bounds to those in [9, Inequality (5.2)].

Lemma 3.3 in [9] follows from [9, Lemma 3.1]. The analogue of [9, Lemma 3.1] is as follows: Let $\mathfrak{n}_{Q_j}$ be the Lie algebra of N_{Q_j}. Then the Langlands decomposition shows that B_P is diffeomorphic to $\mathfrak{n}_{Q_j}\times(A_{Q_j}M_{Q_j}K)$. Identify $\mathfrak{n}_{Q_j}\times(A_{Q_j}M_{Q_j}K)$ with B_P. For any $a\in A_{Q_j}M_{Q_j}K$, denote by g_a the Riemannian metric on $\mathfrak{n}_{Q_j}\times\{a\}$ induced from B_P. The Lie algebra $\mathfrak{n}_{Q_j}$ can be identified with a subspace of the tangent space of $\mathfrak{n}_{Q_j}\times(A_{Q_j}M_{Q_j}K)$ at $(0,a)$, and hence the metric g_a defines an inner product on $\mathfrak{n}_{Q_j}$ which depends on a. This inner product defines a flat metric on $\mathfrak{n}_{Q_j}$, denoted by h_a. Identifying $\mathfrak{n}_{Q_j}\times\{a\}$ with $\mathfrak{n}_{Q_j}$, we get two metrics g_a and h_a on $\mathfrak{n}_{Q_j}$.

Since K is compact, the same proof of [9, Lemma 3.1] gives the following result.

Lemma 4.3.4. *For a sufficiently small $\varepsilon > 0$, there exists a positive constant C_1 such that for any $a \in A_{Q_j} M_{Q_j} K$, $g_a \geq C_1 h_a$ in a ball of radius ε about the origin in $\mathfrak{n}_{Q_j}$ with respect to the metric g_a.*

Using the above lemma and the same argument as in [9], we can prove the analogue of [9, Lemma 3.3]. Once we have the analogue of [9, Lemma 3.3], the same arguments in [9] prove Proposition 4.3.2 as mentioned above.

4.4.

We now prove the lower bound for $N_{T,P}(\lambda)$ in Theorem 4.2.2.

Proposition 4.4.1. *With the above notation,*

$$\lim_{\lambda \to +\infty} \inf \frac{N_{T,P}(\lambda)}{\lambda^{d/2}} \geq (4\pi)^{-d/2} \frac{\mathrm{vol}(\Gamma_{M_P} \backslash B_P)}{\Gamma(\frac{d}{2}+1)}.$$

The proof is similar to the proof of the lower bound in [15, §3.5]. We outline the main steps involved.

First we need to generalize Arthur's truncation operator Λ^T [2] to $\Gamma_{M_P} \backslash B_P$. Recall from §2.4 that $\mathbf{Q}_1, \cdots, \mathbf{Q}_p$ are a set of representatives of Γ_{M_P}-conjugacy classes of proper rational parabolic subgroups of $\mathbf{M_P}$. If f belongs to $L^2(\Gamma_{M_P} \backslash B_P)$ and is continuous, for a sufficiently large truncation parameter T, define

$$\begin{aligned}\Lambda^T f(x) = & f(x) + \sum_{j=1}^{p} (-1)^{\dim A_{Q_j}} \sum_{\gamma \in \Gamma_{Q_j} \backslash \Gamma_{M_P}} \chi_j(H_{Q_j}(\gamma x) \\ & - I_{Q_j}(T)) f_{Q_j}(\gamma x),\end{aligned}$$

where χ_j is the characteristic function of the closed obtuse cone ${}^{+}\mathfrak{a}_{Q_j}$ dual to the positive cone $\mathfrak{a}_{Q_j}^{+} = \{H \in \mathfrak{a}_{Q_j} \mid \alpha(H) > 0, \alpha \in \Sigma(Q_j, A_{Q_j})\}$, and $H_{Q_j}(x)$ is the $\mathfrak{a}_{Q_j}$ component of x in the horospherical decomposition (§4.2), and $I_{Q_j}(T)$ is the image of T in $\mathfrak{a}_{Q_j}$. A basic property of Λ^T is that for every $\mathbf{Q} \subset \mathbf{M_P}$, $x \in N_Q M_Q A_{Q,T} K$, $(\Lambda^T f)_Q(x) = 0$, i.e., Λ^T satisfies precisely the vanishing conditions in the definition of $H^1_T(\Gamma_{M_P} \backslash B_P)$, and $\Lambda^T f$ belongs to the closure of $H^1_T(\Gamma_{M_P} \backslash B_P)$ in $L^2(\Gamma_{M_P} \backslash B_P)$. In fact, Λ^T extends to an orthogonal projection on $L^2(\Gamma_{M_P} \backslash B_P)$ (see [2, p. 92]) and its image is equal to the closure of

$H^1_T(\Gamma_{M_P}\backslash B_P)$ in $L^2(\Gamma_{M_P}\backslash B_P)$. As pointed out in [15, Remark 3.5.2], it might be better to define

$$H^1_T(\Gamma_{M_P}\backslash B_P) = \Lambda^T L^2(\Gamma_{M_P}\backslash B_P) \cap H^1(\Gamma_{M_P}\backslash B_P)$$

and hence Δ_T by using Λ^T. Then the relation between the pseudo-Laplace operator Δ_T and Arthur's truncation operator Λ^T becomes more direct.

On the other hand, if f is smooth, $\Lambda^T f$ is not smooth in general. The discontinuities of $\Lambda^T f$ come from the discontinuities of the characteristic function χ_i in the formula for $\Lambda^T f$. As pointed out in [15, §3.5] (see [1, p. 19]), since there are only finitely many non-zero terms in $\Lambda^T f$ for x in every compact subset in B_P, these discontinuities of $\Lambda^T f$ lie along a family of locally finite hypersurfaces in $\Gamma_{M_P}\backslash B_P$.

We are now ready to use the precise reduction theory to prove the lower bound for $N_T(\lambda)$. In the notation of Lemma 2.4.2, $\Gamma_{M_P}\backslash B_P$ admits the following disjoint decomposition:

$$\Gamma_{M_P}\backslash B_P = (\Gamma_{M_P}\backslash B_P)_T \cup \coprod_{j=1}^{p} \omega_j A_{Q_j,T} K.$$

By taking an exhausting family of compact submanifolds with boundaries of the complement of the discontinuity hyperplanes in every $\omega_i A_{Q_i,T} K$, we can show that for any $\varepsilon > 0$, there exist disjoint compact submanifolds with boundaries $W_j \subset \Gamma_{Q_j}\backslash B_P$, $j = 1, \cdots, p$, and $W_0 \subset \Gamma_{M_P}\backslash B_P$ satisfying the following conditions:

1. $W_1, \cdots, W_p$ are mapped homeomorphically into
$$\omega_1 A_{Q_1,T} K, \cdots, \omega_p A_{Q_p,T} K$$
respectively and are disjoint from the discontinuity hypersurfaces.
2. $|\mathrm{vol}(\Gamma_{M_P}\backslash B_P) - \mathrm{vol}(W_0) - \sum_{j=1}^{p} \mathrm{vol}(W_j)| < \varepsilon$.

In particular, if $f \in C^\infty(\Gamma_{M_P}\backslash B_P)$, $\Lambda^T f$ restricts to a smooth function on W_j for $j = 1, \cdots, p$.

For $j = 1, \cdots, p$, define the cuspidal subspace $H^1_{cus,D}(W_j)$ with the Dirichlet boundary condition by

$$H^1_{cus,D}(W_j) = \{f|_{W_j} \mid f \in H^1_T(\Gamma_{M_P}\backslash B_P), f|_{\partial W_j} = 0\}.$$

The Dirichlet quadratic form $D(f)$ induces a self-adjoint operator, denoted by $\Delta_{j,D}$, on the closure of $H^1_{cus,D}(W_j)$ in $L^2(W_j)$.

Lemma 4.4.2. *For $j = 1, \cdots, p$, the spectrum of $\Delta_{j,D}$ is discrete and its counting function $N_{j,D}(\lambda)$ satisfies the Weyl law:*

$$\lim_{\lambda \to +\infty} \frac{N_{j,D}(\lambda)}{\lambda^{d/2}} = (4\pi)^{-d/2} \frac{\mathrm{vol}(W_j)}{\Gamma(\frac{d}{2}+1)}.$$

Proof. By the same proof of Proposition 4.3.2, we can show that the spectrum of $\Delta_{j,D}$ is discrete and $N_{j,D}(\lambda)$ is bounded by $O(1)\lambda^{d/2}$. We need to determine the precise behavior of $N_{j,D}(\lambda)$. For this purpose, let $h(x,x,t)$ be the heat kernel of $\Delta_{j,D}$. The point is to show that $h(x,x,t)$ satisfies the following small time asymptotics: for $x \in \mathrm{Int}(W_j)$, as $t \to 0$, $h(x,x,t) \sim (4\pi t)^{-\frac{d}{2}}$.

Let $E(x,y,t)$ be the heat kernel of B_P. Since W_j does not meet the discontinuity hypersurfaces of the truncation operator Λ^T above, the heat kernel $h(x,y,t)$ of W_j can be constructed from $E(x,y,t)$ in three steps:

1. Average over Γ_{M_P} to get the heat kernel of
$$\Gamma_{M_P} \backslash B_P : F(x,y,t) = \sum_{\gamma \in \Gamma_{M_P}} E(x, \gamma y, t).$$
2. Apply the truncation operator Λ^T to get $\Lambda^T F(x,y,t)$.
3. Restrict $\Lambda^T F(x,y,t)$ to W_j and modify it to satisfy the Dirichlet boundary condition on ∂W_j.

Clearly, $E(x,x,t)$ satisfies the small time asymptotics: $E(x,x,t) \sim (4\pi t)^{-d/2}$, and we need to show that the three steps above do not change the small time asymptotics. Because of the Gauss factor $\exp -\frac{d^2(x,y)}{(4+\varepsilon)t}$ in the upper bound for $E(x,y,t)$, Step 1 does not change the asymptotics. For $x, y \in W_j$, there are only finitely many nonzero terms appearing in the infinite sum for $\Lambda^T F(x,y,t)$. By the proof of [15, Proposition 3.5.5], the constant term of the heat kernel along a proper parabolic subgroup and hence every such nonzero term $\chi_j(H_{Q_j}(\gamma x) - I_{Q_j}(T)) F_{Q_j}(x, \gamma x, t)$ is of smaller order than $t^{-d/2}$, and hence Step 2 does not change the asymptotics either. Step 3 is achieved by the method of double layer potential (see [7, pp. 161–164]) and does not change the small time asymptotics either. Therefore, the heat kernel $h(x,x,t)$ of $\Delta_{j,D}$ has the desired small time asymptotics. (See [15, §3.5] for similar discussions.)

Then the Weyl law for $\Delta_{j,D}$ follows from this small time asymptotics of the heat kernel via the standard argument using the Karamata's version of the Tauberian theorem (see [15, Proposition 3.5.4] for example).
q.e.d.

Proof of Proposition 4.4.1

Let $N_{0,D}(\lambda)$ be the counting function of the Dirichlet eigenvalues of W_0. The principle of Neumann bracketing shows that

$$N_{T,P}(\lambda) \geq N_{0,D}(\lambda) + \sum_{j=1}^{p} N_{j,D}(\lambda).$$

Since $N_{0,D}(\lambda)$ satisfies the Weyl law, Lemma 4.4.2 and the choice of the submanifolds W_j imply that

$$\lim_{\lambda \to +\infty} \inf \frac{N_{T,P}(\lambda)}{\lambda^{d/2}} \geq (4\pi)^{-d/2} \frac{\mathrm{vol}(\Gamma_{M_P} \backslash B_P) - \varepsilon}{\Gamma(\frac{d}{2}+1)}.$$

Since ε is arbitrary, we get the lower bound for $N_{T,P}(\lambda)$ in Proposition 4.4.1.

5. Bound on the rank 1 residual discrete spectrum

5.1.

In this section, we show that up to a negligible term, the counting function of the rank-one residual discrete spectrum is bounded by the corresponding counting function of the pseudo-Laplacian. This follows from the fact that up to a negligible subset, the rank-one residual discrete spectrum of Δ can be approximated uniformly by the corresponding part of the pseudo-Laplacian Δ_T. The reason why such an approximation exists is that we can construct eigenfunctions of the pseudo-Laplacian by truncating off the constant terms of certain Eisenstein series of rank-one parabolic subgroups. This bound on the rank-one residual discrete spectrum proves Theorem 1.1.4 and is the first half of the proof of Theorem 1.1.3. In the next section, we will show that the counting function of the higher rank residual discrete spectrum can be bounded in terms of rank 1-data and finish the proof of Theorem 1.1.3.

5.2.

Let $\mathcal{C}$ be an association class of rational parabolic subgroups of $\mathbf{G}$ of rank-one. Let $N_{\mathcal{C}}(\lambda)$ be the counting function of the discrete spectrum in $L^2_{\mathcal{C}}(\Gamma\backslash G)$, and $N_{T,\mathcal{C}}(\lambda)$ be the counting function of Δ_T restricted to the closure of $H^1_T(\Gamma\backslash G)\cap L^2_{\mathcal{C}}(\Gamma\backslash G)$ in $L^2_{\mathcal{C}}(\Gamma\backslash G)$ as in the introduction (see also §4.2). Let $\mathbf{P}_1, \cdots, \mathbf{P}_m$ be a set of representatives of Γ-conjugacy classes in $\mathcal{C}$. For every $j = 1, \cdots, m$, let $N_{cus,P_j}(\lambda)$ be the counting function of the cuspidal discrete spectrum in $L^2_{cus}(\Gamma_{P_j}\backslash B_{P_j})$. Denote by $|\rho_{\mathcal{C}}|$ the common norm of the half sum of roots in $\Sigma(P_j, A_{P_j})$ with multiplicity, $j = 1, \cdots, m$. Then we have the following.

Proposition 5.2.1. *For an association class $\mathcal{C}$ of rank 1,*

$$N_{\mathcal{C}}(\lambda) \leq N_{T,\mathcal{C}}(\lambda + |\rho_{\mathcal{C}}|^2) + \sum_{j=1}^{m} N_{cus,P_j}(\lambda).$$

Proof. By Proposition 3.3.4, for any $j = 1, \cdots, m$,

$$L^2(\Gamma_{M_{P_j}}\backslash B_{P_j}) = \sum_{\sigma\in\hat{K}} (\dim \sigma_{M_{P_j}}) L^2(\Gamma_{M_{P_j}}\backslash X_{P_j}, \sigma).$$

Similarly, by Lemma 3.4.5 and Corollary 3.4.3, we get that

$$L^2_{cus}(\Gamma_{M_{P_j}}\backslash B_{P_j}) = \sum_{\sigma\in\hat{K}} (\dim \sigma) L^2_{cus}(\Gamma_{M_{P_j}}\backslash X_{P_j}, \sigma_{M_{P_j}}),$$

$$L^2_{\mathcal{C}}(\Gamma\backslash G) = \sum_{\sigma\in\hat{K}} (\dim \sigma) L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma).$$

For every $\sigma \in \hat{K}$, denote the counting function of the discrete spectrum of Δ_σ in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ by $N_{\mathcal{C},\sigma}(\lambda)$, the corresponding counting function of $\Delta_{T,\sigma}$ by $N_{T,\mathcal{C},\sigma}(\lambda)$, and the counting function of the cuspidal discrete spectrum of Δ_σ in $L^2(\Gamma_{P_j}\backslash X_{P_j}, \sigma_{M_{P_j}})$ by $N_{cus,P_j,\sigma}(\lambda)$. Then

$$N_{\mathcal{C}}(\lambda) = \sum_{\sigma\in\hat{K}} (\dim \sigma) N_{\mathcal{C},\sigma}(\lambda),$$

$$N_{T,\mathcal{C}}(\lambda) = \sum_{\sigma\in\hat{K}} (\dim \sigma) N_{T,\mathcal{C},\sigma}(\lambda),$$

$$N_{cus,P_j}(\lambda) = \sum_{\sigma\in\hat{K}} (\dim \sigma) N_{cus,P_j,\sigma}(\lambda).$$

For every representation σ of K and the space $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$, the operators Δ_σ and $\Delta_{T,\sigma}$ differ from the corresponding operators in [15] by the same constant, and hence by [15, Proposition 5.2.8], we get that

$$N_{\mathcal{C},\sigma}(\lambda) \leq N_{T,\mathcal{C},\sigma}(\lambda + |\rho_{\mathcal{C}}|^2) + \sum_{j=1}^{m} N_{cus,P_j,\sigma}(\lambda).$$

The reason is that except for at most $\sum_{j=1}^m N_{cus,P_j,\sigma}(\lambda)$ of them, every rank-one residual eigenvalue in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ below λ lies in a $|\rho_{\mathcal{C}}|^2$-neighborhood of an eigenvalue of the pseudo-Laplacian. This is achieved as follows: for suitable cuspidal eigenfunctions Φ, between every pair of poles of the Eisenstein series $E(P_j, \Phi, \Lambda)$ (note $\dim \mathfrak{a}_{P_j} = 1$), there exists a point Λ_0 such that $\Lambda^T E(P_j, \Phi, \Lambda_0)$ is an eigenfunction of the pseudo-Laplacian Δ_T (see [15, §3.2] for details). Summing over all representations σ of K, we get the bound for $N_{\mathcal{C}}(\lambda)$ in the proposition.

q.e.d.

Let $\mathcal{C}_1, \cdots, \mathcal{C}_q$ be all the association classes of rational parabolic subgroups of rank 1, and $\mathbf{P}_1, \cdots, \mathbf{P}_p$ be a set of representatives of Γ-conjugacy classes in $\cup_{j=1}^q \mathcal{C}_j$. Denote the counting function of the discrete spectrum in $\oplus_{j=1}^q L^2_{\mathcal{C}_j}(\Gamma\backslash G)$ by $N^1_{res}(\lambda)$, which is the counting function of the total rank-one residual discrete spectrum (see Lemma 3.4.5). Denote the corresponding counting function of Δ_T by $N^1_T(\lambda)$. Denote the maximum of the norms of half sums $|\rho_{\mathcal{C}_1}|, \cdots, |\rho_{\mathcal{C}_q}|$ by $|\rho^1|$. Recall that $N_{cus}(\lambda)$ denotes the counting function of the cuspidal discrete spectrum of $L^2(\Gamma\backslash G)$.

Then we have the following.

Proposition 5.2.2. *With the above notation,*

$$N^1_{res}(\lambda) \leq N^1_T(\lambda + |\rho^1|^2) + \sum_{j=1}^{p} N_{cus,P_j}(\lambda),$$

$$N_{cus}(\lambda) + N^1_{res}(\lambda) \leq N_T(\lambda + |\rho^1|^2) + \sum_{j=1}^{p} N_{cus,P_j}(\lambda),$$

and hence

$$\lim_{\lambda\to+\infty} \sup \frac{N_{cus}(\lambda) + N^1_{res}(\lambda)}{\lambda^{n/2}} \leq (4\pi)^{-n/2} \frac{\mathrm{vol}(\Gamma\backslash G)}{\Gamma(\frac{n}{2}+1)},$$

where $n = \dim \Gamma\backslash G$.

Proof. The first inequality follows from Proposition 5.2.1. The second inequality follows from the first inequality and the fact that every cuspidal cuspidal eigenfunction of Δ is also an eigenfunction of Δ_T which is orthogonal to eigenfunctions of Δ_T belonging to $\cup_{j=1}^{q}\mathcal{C}_j$. Since $\dim B_{P_j} < \dim G$, the third inequality follows from Theorem 4.2.2 and Corollary 4.2.3. q.e.d.

As a corollary, we get the following result.

Corollary 5.2.3. *Theorem 1.1.4 holds, i.e., if the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to one, the counting function $N_d(\lambda)$ of the discrete spectrum of Δ in $L^2(\Gamma\backslash G)$ satisfies the Weyl upper bound:*

$$\lim_{\lambda\to+\infty}\sup\frac{N_d(\lambda)}{\lambda^{n/2}} \leq (4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash G)}{\Gamma(\frac{n}{2}+1)},$$

where $n = \dim\Gamma\backslash G$.

Proof. If the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 1, $N_d(\lambda) = N_{cus}(\lambda) + N_{res}^{1}(\lambda)$, and the corollary follows from the previous proposition.

Remark 5.2.4. The above arguments show that except for at most $\sum_{j=1}^{p} N_{cus,P_j}(\lambda)$ of them, every rank-one residual eigenvalue in $L^2(\Gamma\backslash G)$ below λ lies in a $|\rho^1|^2$-neighborhood of an eigenvalue of the pseudo-Laplacian. It is conceivable that such a uniform approximation should hold for every residual eigenvalue. If so, the Weyl upper bound on $N_d(\lambda)$ would follow immediately from the Weyl law of the pseudo-Laplacian, i.e., Theorem 1.2.1.

6. Bound on the higher rank residual discrete spectrum

6.1.

In this section, we bound the counting function of the higher rank residual discrete spectrum and hence complete the proof of Theorem 1.1.3. The basic idea is as follows. The residual eigenfunctions are given as iterated residues of Eisenstein series. Since the singularities of the Eisenstein series are contained in the singularities of the scattering matrices, which appear in their constant terms, the number of residual eigenvalues can bounded in terms of the number of singular hyperplanes of the scattering matrices. For scattering matrices of rank-one parabolic subgroups, the number of the singular hyperplanes can be bounded in terms of the counting function of the pseudo-Laplacian, while for higher rank parabolic subgroups, their scattering matrices can be written as products of rank-one scattering matrices (see [15, §5–7]). Therefore, the counting function of the higher rank residual discrete spectrum can be bounded.

6.2.

Let $\mathcal{C}$ be an association class of rational parabolic subgroups of $\mathbf{G}$ of rank $r \geq 2$. Let $\mathbf{P}_1, \cdots, \mathbf{P}_l$ be a set of representatives of Γ-conjugacy classes in $\mathcal{C}$. Denote the counting function of the discrete spectrum in $L^2_{\mathcal{C}}(\Gamma\backslash G)$ by $N_{\mathcal{C}}(\lambda)$ as above.

Proposition 6.2.1. *For an association $\mathcal{C}$ of rank $r \geq 2$, the counting function $N_{\mathcal{C}}(\lambda)$ is bounded by*

$$c\left(\sum_{\mathbf{Q}} N_{T,Q}(\lambda + |\rho_P|^2 + 1) + \sum_{\mathbf{Q}}\sum_{\mathbf{P}'} N_{cus,P'}(\lambda) + 1\right)^r,$$

where (1) $\mathbf{Q}$ runs over a set of representatives of Γ-equivalence classes of all the rational parabolic subgroups containing a group $\mathbf{P}$ in $\mathcal{C}$ such that $\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) = \mathrm{rank}_{\mathbb{Q}}(\mathbf{P}) - 1$, and (2) for every $\mathbf{Q}$ in the first sum, $\mathbf{P}'$ runs over the set of rational parabolic subgroups of $\mathbf{G}$ contained in $\mathbf{Q}$ that correspond to a set of representatives $'\mathbf{P}$ of Γ_{M_Q}-conjugacy classes of rank-1 rational parabolic subgroups of $\mathbf{M_Q}$ in the sense $M_{P'} = M_{'P}$, $N_{P'} = N_Q N_{'P}$, (3) and c is a constant depending only on $\mathbf{G}$ and T. In particular, $N_{\mathcal{C}}(\lambda)$ satisfies the following bound:

$$N_{\mathcal{C}}(\lambda) = O(1)\lambda^{\frac{m}{2}}, \quad \textit{as } \lambda \to +\infty,$$

where m is the maximum of

$$(\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) + 1)\dim B_Q = (\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) + 1)(\dim X_Q + \dim K)$$

for all rational parabolic $\mathbf{Q}$ of $\mathbf{G}$ with $\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) \leq \mathrm{rank}_{\mathbb{Q}}(\mathbf{G}) - 1$ as defined in Theorem 1.1.3.

Proof. Consider the set $\mathcal{C}^+$ of rational parabolic subgroups of $\mathbf{G}$ which contain a conjugate of some $\mathbf{P}_1, \cdots, \mathbf{P}_l$ and whose rank is equal to $r-1$. Let $\mathbf{Q}_1, \cdots, \mathbf{Q}_s$ be a set of representatives of Γ-conjugacy classes in the set $\mathcal{C}^+$.

For every $\mathbf{Q}_i$, $i = 1, \cdots, s$, consider the set $\mathcal{C}_i^-$ of maximal (i.e., rank-one) rational parabolic subgroups of $\mathbf{M_{Q_i}}$. Let $'\mathbf{P}_{i1}, \cdots, {'\mathbf{P}}_{it_i}$ be a set of representatives of $\Gamma_{M_{Q_i}}$-conjugacy classes in $\mathcal{C}_i^-$. Any rational parabolic subgroup $'\mathbf{P}$ of $\mathbf{M_{Q_i}}$ uniquely determines a rational parabolic subgroup $\mathbf{P}'$ of $\mathbf{G}$ contained in $\mathbf{Q_i}$ such that $M_{P'} = M_{'P}$, $A_{P'} = A_{Q_j} A_{'P}$, $N_{P'} = N_{Q_j} N_{'P}$. Denote the rational subgroups of $\mathbf{G}$ corresponding to $'\mathbf{P}_{i1}, \cdots, '\mathbf{P}_{it_i}$ by $\mathbf{P}'_{i1}, \cdots, \mathbf{P}'_{it_i}$.

As in §5.2, let $N_{\mathcal{C}}(\lambda)$ be the counting function of the discrete spectrum of Δ in $L^2_{\mathcal{C}}(\Gamma\backslash G)$, and $N_{\mathcal{C},\sigma}(\lambda)$ the counting function of the discrete spectrum of Δ_σ in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$. For any $i = 1, \cdots, s$, $j = 1, \cdots, t_i$, let $N_{cus,P'_{ij}}(\lambda)$ be the counting function of the cuspidal discrete spectrum of Δ in $L^2_{cus}(\Gamma_{M_{P'_{ij}}}\backslash B_{P'_{ij}})$, and for $\sigma \in \hat{K}$, $N_{cus,P'_{ij},\sigma}(\lambda)$ be the counting function of the cuspidal discrete spectrum of Δ_σ in $L^2_{cus}(\Gamma_{M_{P'_{ij}}}\backslash X_{P'_{ij}}, \sigma)$. Denote the counting function of the spectrum of the pseudo-Laplacian Δ_T on $\Gamma_{M_{Q_i}}\backslash B_{Q_i}$ by $N_{T,Q_i}(\lambda)$, and its σ-component $\Delta_{T,\sigma}$ by $N_{T,Q_i,\sigma}(\lambda)$. Then by the decompositions in §3.3 and §3.4, we get that

$$N_{\mathcal{C}}(\lambda) = \sum_{\sigma\in\hat{K}} (\dim\sigma) N_{\mathcal{C},\sigma}(\lambda),$$

$$N_{cus,P'_{ij}}(\lambda) = \sum_{\sigma\in\hat{K}} (\dim\sigma) N_{cus,P'_{ij},\sigma}(\lambda),$$

$$N_{T,Q_i}(\lambda) = \sum_{\sigma\in\hat{K}} (\dim\sigma) N_{T,Q_i,\sigma}(\lambda).$$

Since the operators Δ_σ and $\Delta_{T,\sigma}$ are shifts of the corresponding operators in [15] by the same constant, by [15, Proposition 7.2.2], there exists a positive constant c depending only on $\mathbf{G}$ and the truncation parameter T such that

$$N_{\mathcal{C},\sigma}(\lambda) \leq c(\sum_{i=1}^{s} \quad (N_{T,Q_i,\sigma}(\lambda + |\rho_P|^2 + 1) + \sum_{j=1}^{t_i} N_{cus,P'_{ij},\sigma}(\lambda)) + 1)^r.$$

As mentioned earlier, this bound on $N_{\mathcal{C},\sigma}(\lambda)$ is obtained as follows. Since the residual discrete spectrum is generated by iterated residues of Eisenstein series and the singularities of the Eisenstein series are contained in the singularities of the scattering matrices, the number of residual discrete eigenvalues is bounded by the number of complete flags of singular hyperplanes of the scattering matrices. For the rank-one scattering matrices, there is a close connection between their poles and eigenfunctions of the pseudo-Laplacian (see [15, §6]), and hence the number of the poles of the rank-one scattering matrices can be bounded by the counting functions of the pseudo-Laplacian and of the cuspidal discrete spectrum of lower dimensional spaces. Using the fact that the higher rank scattering matrices can be written as products of rank-one scattering matrices [15, §2.7], we get a bound on the number of singular hyperplanes of the higher rank scattering matrices, and hence a bound on the number of complete flags of such singular hyperplanes by raising the former bound to the power of the rank.

Combining the above equalities and the inequality, we get

$$N_{\mathcal{C}}(\lambda) \leq c(\sum_{i=1}^{s} (N_{T,Q_i}(\lambda + |\rho_P|^2 + 1) + \sum_{j=1}^{t_i} N_{cus,P'_{ij}}(\lambda)) + 1)^r.$$

This proves the first upper bound on $N_{\mathcal{C}}(\lambda)$ in the proposition.

By Theorem 4.2.2, as $\lambda \to +\infty$,

$$N_{T,Q_i}(\lambda + |\rho_P|^2 + 1) = O(1)\lambda^{\frac{1}{2}\dim B_{Q_i}},$$

and by Corollary 4.2.3,

$$N_{cus,P'_{ij}}(\lambda) = O(1)\lambda^{\frac{1}{2}\dim B_{P'_{ij}}}.$$

Since $\dim B_{P'_{ij}} \leq \dim B_{Q_i}$ and $\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}_i) = r - 1 \leq \mathrm{rank}_{\mathbb{Q}}(\mathbf{G}) - 1$, the above bound on $N_{\mathcal{C}}(\lambda)$ implies that as $\lambda \to +\infty$,

$$N_{\mathcal{C}}(\lambda) = O(1)\lambda^{\frac{m}{2}}.$$

This completes the proof of Proposition 6.2.1. q.e.d.

Proof of Theorem 1.1.3

We combine the above results to prove Theorem 1.1.3. When $\mathcal{C} = \{\mathbf{G}\}$, $N_{\mathcal{C}}(\lambda) = N_{cus}(\lambda)$ (see Lemma 3.4.5). Denote by $N^1_{res}(\lambda)$ the counting function of the rank-one residual discrete spectrum of $L^2(\Gamma\backslash G)$ as in Proposition 5.2.2, and by $N^2_{res}(\lambda)$ the counting function of the higher rank residual discrete spectrum of $L^2(\Gamma\backslash G)$. Then Lemma 3.4.5 shows that

$$N_d(\lambda) = N_{cus}(\lambda) + N^1_{res}(\lambda) + N^2_{res}(\lambda).$$

Since there are only finitely many association classes $\mathcal{C}$ of rank greater than or equal to 2, Proposition 6.2.1 implies that

$$N^2_{res}(\lambda) = O(1)\lambda^{\frac{m}{2}}.$$

Combined with the bound on $N_{cus}(\lambda) + N^1_{res}(\lambda)$ in Proposition 5.2.2, this gives the upper bound for $N_d(\lambda)$ in Theorem 1.1.3.

7. Proof of the trace class conjecture

7.1.

In this section we use the polynomial upper bound on the discrete spectrum in Theorem 1.1.3 to prove the trace class conjecture, i.e., Theorem 1.1.2.

First we prove the following.

Proposition 7.1.1. *If $\alpha \in C_0^{2k}(G)$ with $k > \frac{1}{2}\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})n$, k being an integer, or $\alpha \in \mathcal{C}^1(G)$, then the operator $R_d(\alpha)$ on $L^2_d(\Gamma\backslash G)$ is a Hilbert-Schmidt operator.*

Proof. By Lemma 3.4.5,

$$L^2_d(\Gamma\backslash G) = \sum_{\sigma\in\hat{K}} \oplus(\dim\sigma) L^2_d(\Gamma\backslash X, \sigma).$$

For each subspace $L^2_d(\Gamma\backslash X,\sigma)$, choose a basis $e_{\sigma,i}$ of $L^2_d(\Gamma\backslash X,\sigma)$ consisting of orthonormal eigenfunctions of the Laplacian Δ, where i belongs to an index set $I(\sigma)$. Denote the eigenvalue of $e_{\sigma,i}$ by $\lambda_{\sigma,i}$. Then $\lambda_{\sigma,i}\geq 0$, and Theorem 1.1.3 implies that

$$\sum_{\sigma\in\hat{K}} \sum_{i\in I(\sigma)} (\dim\sigma)(\lambda_{\sigma,i}+1)^{-k} < +\infty.$$

Since $e_{\sigma,i}$, where $\sigma\in\hat{K}$ and $i\in I(\sigma)$, forms an orthonormal basis of $L^2_d(\Gamma\backslash G)$, the Hilbert-Schmidt norm $\|R_d(\alpha)\|$ of $R_d(\alpha)$ is given as follows:

$$\|R_d(\alpha)\|^2 = \sum_{\sigma\in\hat{K}} \sum_{i\in I(\sigma)} (\dim\sigma)(R(\alpha)e_{\sigma,i}, R(\alpha)e_{\sigma,i})$$

$$= \sum_{\sigma\in\hat{K}} \sum_{i\in I(\sigma)} (\dim\sigma)(\lambda_{\sigma,i}+1)^{-k}(R(\alpha)(\Delta+1)^k e_{\sigma,i}, R(\alpha)e_{\sigma,i})$$

$$= \sum_{\sigma\in\hat{K}} \sum_{i\in I(\sigma)} (\dim\sigma)(\lambda_{\sigma,i}+1)^{-k}\int_G (\Delta+1)^k\alpha(g)(R(g)e_{\sigma,i}, R(\alpha)e_{\sigma,i})dg$$

$$\leq \|(\Delta+1)^k\alpha\|_{L^1(G)}\|\alpha\|_{L^1(G)} \sum_{\sigma\in\hat{K}} \sum_{i\in I(\sigma)} (\dim\sigma)(\lambda_{\sigma,i}+1)^{-k} < +\infty,$$

where we have used the fact that $\Delta+1$ is self-adjoint to do integration by part in the third equality. Therefore, $R_d(\alpha)$ is a Hilbert-Schmidt operator. q.e.d.

Proof of Theorem 1.1.2

By [33, Lemma 4.5], for any $\alpha \in \mathcal{C}^1(G)$, there exist $\beta \in \mathcal{C}^1(G)$, $\mu, \nu \in C_0^{2\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})n}(G)$ such that

$$\alpha = \beta * \mu + \alpha * \nu.$$

Then we have

$$R_d(\alpha) = R_d(\beta)R_d(\mu) + R_d(\alpha)R_d(\nu).$$

By Proposition 7.1.1, $R_d(\beta), R_d(\mu), R_d(\alpha)$, and $R_d(\nu)$ are all Hilbert-Schmidt operators. Therefore, $R_d(\alpha)$ is of the trace class.

Remark 7.1.2. The arguments in this section are the same as in [22, pp. 525–526] and [33, pp. 39–40]. The above proof shows that the trace class conjecture also holds for more general functions α on G whose derivatives up to $\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})n+2$ are all integrable. From the above proof, it is also clear that any polynomial upper bound on $N_d(\lambda)$ implies Theorem 1.1.2.

Remark 7.1.3. By the same argument as in [33, pp. 39–40], we can show that the linear map $\mathcal{C}^1(G) \to \mathbb{C} : \alpha \to tr(R_d(\alpha))$ is continuous.

References

[1] J. Arthur, *The trace formula and Hecke operators*, Number theory, trace formulas and discrete groups, (eds. K.Aubert et al.), Academic Press, New York, 1989, 11–27.

[2] ———, *A trace formula for reductive groups. II*, Compositio Math. **40** (1980) 87–121.

[3] ———, *Eisenstein series and the trace formula*, Automorphic forms, representations, and L-functions, (eds. A.Borel and W. Casselman), Proc. of Symp. in Pure Math. Vol 33, Part 1, Amer. Math. Soc., 1977, 253–274.

[4] A. Borel, *Introduction aux groupes arithmetiques*, Hermann, Paris, 1969.

[5] A. Borel & H. Garland, *Laplacian and the discrete spectrum of an arithmetic group*, Amer. J. Math. **105** (1983) 309–335.

[6] A. Borel & L. Ji, *Compactifications of locally symmetric spaces*, in preparation.

[7] I. Chavel, *Eigenvalues in Riemannian geometry*, Academic Press, New York, 1984.

[8] Y.Colin de Verdiére, *Pseudo-Laplaciens II*, Ann. Inst. Fourier (Grenoble) **33** (1983) 87-113.

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

[10] ———, *Eigenvalue estimates for certain noncompact manifolds*, Michigan Math. J. **31** (1984) 349–357.

[11] I. M. Gelfand & I. I. Piatetski–Shapiro, *Unitary representations in homogeneous spaces with discrete stationary groups*, Sov. Math., Dokl. **3** (1962) 1528–1531.

[12] Y. Guivarc'h, L. Ji & J. C. Taylor, *Compactifications of symmetric spaces*, to appear in **Progr. Math.**, Vol. 156, Birkhäuser, Boston, 1998.

[13] Harish-Chandra, *Automorphic forms on semisimple Lie groups*, Lecture Notes in Math. Vol. 62, Springer, Berlin, 1968.

[14] L. Ji, *The trace class conjecture and the Weyl upper bound on the discrete spectrum of arithmetic groups*, Math. Res. Lett. **4** (1997) 501–511.

[15] ———, *The Weyl upper bound on the discrete spectrum of locally symmetric space*, revised version 1997.

[16] ———, *GCQ of Borel–Serre compactification and the toroidal compactifications, and Baily–Borel compactification*, Preprint 1996.

[17] A. Knapp, *Representation Theory of semisimple Groups: An overview based on examples*, Princeton Univ. Press, 1986.

[18] R. Langlands, *Rank–one residues of Eisenstein series*, Festschrift in honor of I. I. Piatetski–Shapiro on the occasion of his sixtieth birthday, Part II, Israel Math. Conf. Proc. Vol.3, Weizmann, Jerusalem, 1990, 111–125.

[19] ———, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Math. Vol. 544, Springer, Berlin, 1976.

[20] P. Lax & R. Phillips, *Scattering theory for automorphic functions*, Princeton Univ. Press, 1976.

[21] C. Moeglin & J. Waldspurger, *Spectral decomposition and Eisenstein series*, Cambridge Univ. Press, 1995.

[22] W. Müller, *The trace class conjecture in the theory of automorphic forms*, Ann. of Math. **130** (1989) 473–529.

[23] ———, *The trace class conjecture in the theory of automorphic forms. II*, Preprint 1996.

[24] ———, *The trace class conjecture without the K-finiteness assumption*, C.R. Acad. Sci. Paris, **324** (1997) 1333–1338.

[25] M. Osborne & G. Warner, *The Selberg trace formula. V*, Trans. Amer. Math. Soc. **286** (1984) 351–376.

[26] ———, *The Selberg trace formula. VIII*, Trans. Amer. Math. Soc. **324** (1991) 623–653.

[27] ———, *The Selberg trace formula.II*, Pacific J. Math. **106** (1983) 307–496.

[28] ———, *The theory of Eisenstein system*, Academic Press, New York, 1981.

[29] L. Saper, *Tilings and finite energy retractions of locally symmetric spaces*, Comment. Math. Helv. **72** (1997) 167–202.

[30] A. Selberg, *Harmonic analysis*, Collected papers of A. Selberg, Vol. 1, Springer, Berlin, 1989, 626–674.

[31] H. Urakawa, *Geometry of the Laplace-Beltrami operator on a complete Riemannian manifold*, Progr. in differential geometry, (eds. K. Shiohama), Adv. Stud. Pure Math. Vol. 22, 1993, 347–406.

[32] N. Wallach, *On the distribution of eigenvalues of the Laplacian of a locally symmetric space*, Differential Geometry, (ed. B.Lawson), Pitman Monographs and Surveys in Pure and Appl. Math. Vol. 52, Longman Scientific & Technical, 1991, 337–350.

[33] G. Warner, *Selberg's trace formula for nonuniform lattices: The $\mathbb{R}$–rank 1 case*, Adv. in Math. Suppl. Stud., Vol.6, Academic Press, New York, 1979, 1–142.

[34] ———, *Traceability on the discrete spectrum*, The Selberg trace formula and related topics, (eds. D.Hejhal et al.), Contemp. Math. Vol 53, Amer. Math. Soc., 1984, 529–533.

University of Michigan