J. DIFFERENTIAL GEOMETRY
**51** (1999) 97-147

# THE WEYL UPPER BOUND ON THE DISCRETE SPECTRUM OF LOCALLY SYMMETRIC SPACES

LIZHEN JI

## 1. Introduction

**1.1.** Let $G$ be a reductive Lie group with finitely many connected components, and $\Gamma$ a cofinite volume discrete subgroup of $G$. Let $K \subset G$ be a maximal compact subgroup, and $X = G/K$ be the associated symmetric space, which is the product of a symmetric space of noncompact type and a possible Euclidean space. Then $\Gamma\backslash X$ is a locally symmetric space of finite volume. For simplicity, we assume, unless otherwise specified, that there exists a reductive algebraic group $\mathbf{G}$ defined over $\mathbb{Q}$ satisfying the conditions in [18, p. 1] such that $G = \mathbf{G}(\mathbb{R})$, and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ is an arithmetic subgroup.

Any finite dimensional unitary representation $\sigma$ of $K$ defines a homogeneous bundle $\tilde{E}_\sigma$ on $X$ and hence a locally homogeneous bundle $E_\sigma$ on $\Gamma\backslash X$. The bundle $E_\sigma$ admits a locally invariant connection $\bigtriangledown$ which is the push forward of the invariant connection on the homogeneous bundle $\tilde{E}_\sigma$. The connection $\bigtriangledown$ defines a quadratic form $D$ on sections of $E_\sigma$: For any $f \in C_0^\infty(\Gamma\backslash X, \sigma)$,

$$D(f) = \int_{\Gamma\backslash X} | \bigtriangledown f(x)|^2 dx.$$

This quadratic form $D$ defines an elliptic operator $\Delta$ on $L^2(\Gamma\backslash X, \sigma)$, called the Laplace operator, where $L^2(\Gamma\backslash X, \sigma)$ denotes the space of $L^2$-sections of $E_\sigma$. If $\sigma$ is irreducible, $\Delta$ is equal to a shift of the restriction of

Received October 8, 1996, and, in revised form, August 7, 1998. Author was partially supported by NSF postdoctoral fellowship 9407427 and NSF grant DMS-9704434.

the negative of the Casimir element of the Lie algebra of $G$ by a constant determined by $\sigma$ (cf. [5, Theorem 2.5, p. 49] and [32, Equation 6.10, p. 386]).

If $\Gamma\backslash X$ is compact, $\Delta$ has a discrete spectrum $\mathrm{Spec}(\Delta) : \lambda_1 \leq \lambda_2 \leq \cdots$ (repeated with multiplicity) on $L^2(\Gamma\backslash X, \sigma)$, and the spectral counting function $N(\lambda) = |\{\lambda_i \in \mathrm{Spec}(\Delta) \mid \lambda_i \leq \lambda\}|$ satisfies the famous Weyl law:

$$\lim_{\lambda\to+\infty} \frac{N(\lambda)}{\lambda^{n/2}} = (4\pi)^{-n/2} \frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)} \dim\sigma,$$

where $n = \dim\Gamma\backslash X$, $\dim\sigma$ is the dimension of the representation space of $\sigma$, i.e., the rank of the bundle $E_\sigma$, and $\Gamma(\cdot)$ is the Gamma function.

From now on we assume that $\Gamma\backslash X$ is noncompact. Then $\Delta$ has both a continuous spectrum $\mathrm{Spec}_{con}(\Delta)$ and a discrete spectrum $\mathrm{Spec}_{dis}(\Delta)$ : $\lambda_1 \leq \lambda_2 \leq \cdots$ (repeated with multiplicity), which could be finite.

Denote the counting function of the discrete spectrum of $\Delta$ by $N_d(\lambda) = |\{\lambda_i \in \mathrm{Spec}_{dis}(\Delta) \mid \lambda_i \leq \lambda\}|$. In [4, Theorem 1], Borel and Garland showed that $N_d(\lambda)$ is finite for all $\lambda > 0$. Inspired by the above Weyl law for compact quotients and a result of Donnelly [11, Theorem 1.1] (see Lemma 2.3.2 below) on the cuspidal discrete spectrum, they raised the following question [4, §4.7]:

**Question 1.1.1.** Decide whether $N_d(\lambda)$ satisfies the following Weyl upper bound

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

One of the main results of this paper is to answer this question affirmatively in several cases. The results in this paper strongly suggest that the answer to the above question is always positive.

**Theorem 1.1.2** (§5). *If the $\mathbb{Q}$-rank of* $\mathbf{G}$ *is equal to 1 or the $\mathbb{R}$-rank of* $\mathbf{G}$*, i.e., the rank of $X$, is less than or equal to 2, then the counting function $N_d(\lambda)$ of the discrete spectrum of $\Delta$ satisfies the Weyl upper bound, i.e.,*

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

*where $n = \dim\Gamma\backslash X$, and $\dim\sigma$ is the dimension of the representation space of $\sigma$.*

Siegel modular varieties $\Gamma\backslash\mathrm{Sp}(2,\mathbb{R})/\mathrm{U}(2)$ and the Hilbert modular varieties satisfy the condition in the above theorem, and hence the Weyl upper bound holds for them, but the Siegel modular varieties $\Gamma\backslash\mathrm{Sp}(n,\mathbb{R})/\mathrm{U}(n)$ for $n \geq 3$ do not satisfy this condition. On the other hand, we show below that for Siegel modular varieties $\Gamma\backslash\mathrm{Sp}(3,\mathbb{R})/\mathrm{U}(3)$, $N_d(\lambda)$ satisfies a bound of the sharp order, though not the sharp constant.

The sharp upper bound on $N_d(\lambda)$ was known only in the following cases. If $G = \mathrm{SL}(2,\mathbb{R})$, the upper bound in Theorem 1.1.2 is due to Selberg and follows from the Selberg trace formula [46, p. 668]. For the general rank-one $X$, this upper bound is due to Donnelly [9, Theorem 1.1]. The only higher rank case where this upper bound was known to hold is when $G = \mathrm{SL}(n,\mathbb{R})$ and $\Gamma \subset \mathrm{SL}(n,\mathbb{Z})$ is a congruence subgroup, $n \geq 3$; in fact, the upper bound then follows from a result of Moeglin and Waldspurger [33] on an explicit description of the residual discrete spectrum that implies that every Eisenstein series induce at most one residual eigenfunction, and the result of Donnelly mentioned earlier on the Weyl upper bound on the cuspidal discrete spectrum in [11, Theorem 1.1] (see Lemma 2.3.2 below). Donnelly also proved in [10] a (nonsharp) polynomial upper bound on $N_d(\lambda)$ in the $\Gamma$-rank-1, in particular the $\mathbb{Q}$-rank-1, case (see below for the precise bound).

To state general result, we need more notation. For any proper rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, there is a boundary locally symmetric space $\Gamma_{M_P}\backslash X_P$, which is a boundary component of the reductive Borel–Serre compactification $\overline{\Gamma\backslash X}^{RBS}$. The dimension of the split center of the Levi quotient of $\mathbf{P}$ is called the $\mathbb{Q}$-rank (or the split rank, or the parabolic rank) of $\mathbf{P}$, denoted by $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P})$. Then the $\mathbb{Q}$-rank of $\mathbf{G}$, denoted by $\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})$, is equal to the maximum of $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P})$ for all proper rational parabolic subgroups $\mathbf{P}$. Since $\Gamma\backslash X$ is noncompact, $\mathrm{rank}_{\mathbb{Q}}(\mathbf{G}) \geq 1$. (See §2.2 for more details.)

**Theorem 1.1.3** (§7). *The counting function $N_d(\lambda)$ of the discrete spectrum of $\Delta$ satisfies the following upper bound:*

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

*where $n = \dim\Gamma\backslash X$, $m$ is equal to the maximum of*

$$(\mathrm{rank}_{\mathbb{Q}}(\mathbf{P})+1)\dim\Gamma_{M_P}\backslash X_P$$

*for all proper rational parabolic subgroups* $\mathbf{P}$ *of* $\mathbf{G}$ *such that*

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

*where* $o(1)$ *goes to zero, and* $O(1)$ *is a bounded quantity as* $\lambda \to +\infty$. *In particular, there exists a positive constant* $C$ *such that for all* $\lambda > 1$,

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

When $m < n$, this theorem implies that $N_d(\lambda)$ satisfies the Weyl upper bound. The condition $m < n$ is clearly satisfied when $\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})$ is equal to 1 since there is no proper rational parabolic subgroup $\mathbf{P}$ satisfying $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P}) \leq \mathrm{rank}_{\mathbb{Q}}(\mathbf{G}) - 1$. When the $\mathbb{R}$-rank of $\mathbf{G}$ is equal to 2, then $m = 0$ and hence $m < n$. Therefore, this theorem implies Theorem 1.1.2.

For Siegel modular varieties $\Gamma\backslash\mathrm{Sp}(3,\mathbb{R})/\mathrm{U}(3)$, $n = 12$ and $m = 12$. Theorem 1.1.3 implies the following bound of sharp order.

**Corollary 1.1.4.** *When* $\Gamma\backslash X = \Gamma\backslash\mathrm{Sp}(3)/\mathrm{U}(3)$, *there exists a constant* $c$ *such that for all* $\lambda > 0$

$$N_d(\lambda) \leq c(1 + \lambda^{\frac{n}{2}}),$$

*where* $n = \dim \Gamma\backslash X = 12$.

Theorem 1.1.3 improves the following bound on $N_d(\lambda)$ due to Müller [35, Theorem 0.1] [1] that

$$N_d(\lambda) \leq C(1 + \lambda^{\frac{n}{2} + \frac{3n}{2}\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})})$$

for a positive constant $C$. (See also Remark 7.2.3 below for explanations.) When $\mathrm{rank}_{\mathbb{Q}}(\mathbf{G}) = 1$, Donnelly proved earlier in [10, Theorem 4.11] that $N_d(\lambda) \leq C(1 + \lambda^{\frac{1}{2}(n+p)})$ where $p = \max\{\dim \Gamma_{M_P}\backslash X_P \mid \mathbf{P}$ is a proper rational parabolic subgroup of $\mathbf{G}\}$, and Langlands proved independently in [29] that $N_d(\lambda)$ has a polynomial upper bound.

Such a polynomial upper bound on $N_d(\lambda)$ is closely related to the trace class conjecture in the theory of the Selberg trace formula (see [35] and [48]). In fact, the method of proof of Theorem 1.1.3 is used in [21] to prove a polynomial upper bound on the discrete spectrum of $L^2(\Gamma\backslash G)$ and hence the trace class conjecture in full generality, i.e., the $K$–finiteness assumption on the convolution function in [35] will be removed.

[1] It is stated in [35, Theorem 0.1] that $N_d(\lambda) \leq C(1 + \lambda^{2n})$. Actually, the proof only gives the above weaker bound.

**Remark 1.1.5.** The results of this paper along with [21] have been announced in [22]. It is claimed there that the Weyl upper bound is satisfied by the counting function $N_d(\lambda)$ in all higher rank cases also. Unfortunately, in an early version of this paper, there is a gap, as in [35, p. 523], in the last step bounding the residual discrete spectrum associated with rational parabolic subgroups of rank greater than or equal to 2 (see Remark 7.2.3 below for details). Because of this, we can prove the Weyl upper bound only in the cases listed in Theorem 1.1.2 and get the weaker upper bound in Theorem 1.1.3 above for other cases. This weakening of the upper bound on $N_d(\lambda)$, however, does not affect the solution of the trace class conjecture announced in [22].

**1.2.** On the other hand, relatively little is known about lower bounds for $N_d(\lambda)$. Inspired by his upper bound on $N_d(\lambda)$ and results for $G = \mathrm{SL}(2, \mathbb{R})$ (see below), Müller made the following closely related but extremely difficult conjecture [36, Remarks 2, p. 180]. (This question is also raised in [4, §4.7].)

**Conjecture 1.2.1.** *If $\Gamma$ is an arithmetic subgroup of $G$, then the Weyl law holds:*

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

An interesting corollary of Conjecture 1.2.1 is that the discrete subspace $L^2_d(\Gamma\backslash X, \sigma)$ is of infinite dimension, which is not known in general.

In [45, Conjecture 2], Sarnak made an even stronger conjecture that under the same assumption, the counting function for the cuspidal discrete spectrum alone satisfies the Weyl law. It seems that it is implicitly conjectured in [45] that the counting function of the residual discrete spectrum is of smaller order than the Weyl law (see Conjecture 1.3.1 below).

If $\Gamma$ is an irreducible cofinite volume discrete subgroup and the rank of $X$ is greater than or equal to two, then Margulis' superrigidity theorem [31] implies that $\Gamma$ is arithmetic, and hence the assumption that $\Gamma$ is arithmetic is always satisfied in this case. On the other hand, if $G = \mathrm{SL}(2, \mathbb{R})$ and $\Gamma\backslash X$ is a Riemann surface, the theory of disappearance of the cuspidal discrete spectrum under a generic deformation in the Teichmüller space of $\Gamma\backslash X$ developed by Phillips and Sarnak [45], [41] and others shows that the arithmetic assumption on $\Gamma$ is necessary. In fact, in [51], Wolpert showed that under the multiplicity one assumption on new forms, the Weyl law does not hold for a generic Riemann surface which is the sphere with even number of punctures and certain symmetries.

In view of the results of Efrat for function fields (see [15] for a summary), it may be necessary to require $\Gamma$ to be a congruence subgroup in Conjecture 1.2.1.

Conjecture 1.2.1 has been proved only for the following cases:

1. $G = \mathrm{SL}(2)$, $\Gamma$ a congruence subgroup, $\sigma$ the trivial representation, a combination of results of Selberg [46, pp. 668 and 670], Hejhal [19, Chapter 11], and Huxley [20].

2. $G = \mathrm{SO}(n,1)$ or $\mathrm{SU}(1,n)$, $\Gamma$ a congruence subgroup, $\sigma$ the trivial representation, by Reznikov [43].

3. $G = R_{k/\mathbb{Q}}\mathrm{SL}(2)$ [2], $\Gamma$ a congruence subgroup of the Hilbert modular group (any arithmetic subgroup is automatically congruent in this case), $\sigma$ the trivial representation, by Efrat [13, p. 6] (the details are in [14]).

So Conjecture 1.2.1 is still open for a general arithmetic but non-congruence subgroup $\Gamma$ of $\mathrm{SL}(2,\mathbb{Z})$.

In the above cases, the Weyl law is proved in two steps:

1. Obtain the Weyl–Selberg law from the Selberg trace formula:

$$N_d(\lambda) + N_c(\lambda) \sim (4\pi)^{-n/2} \frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)} \lambda^{n/2},$$

   where $N_c(\lambda)$ counts the continuous spectrum and is an integral of the determinant of the scattering (or intertwining) matrix.

2. Study the scattering matrix and show that the term $N_c(\lambda)$ is of smaller order than $\lambda^{n/2}$ as $\lambda \to +\infty$.

In the general case, the trace formula in the original sense of Selberg is not as well developed as in the case of $\mathrm{SL}(2)$, though Arthur has developed a trace formula which expresses the trace of a convolution operator on the cuspidal subspace and is very powerful for applications to number theory and automorphic forms (see [1] and the references there). In particular, the Weyl–Selberg law is not known in general.

[2]In $R_{k/\mathbb{Q}}\mathrm{SL}(2)$, $k$ is a totally real number field, and $R_{k/\mathbb{Q}}$ is the functor of the restriction of the scalar.

Another main result of this paper gives bounds for the counting function of eigenvalues of the pseudo-Laplace operator (Theorem 3.3.2). This result can be considered as an analogue of the Weyl–Selberg law, since the spectrum of the pseudo-Laplacian involves both the discrete and the continuous spectra of the Laplacian. But the second step above is extremely difficult and seems to be out of reach now.

**1.3.** To sketch the proof of Theorem 1.1.2, we need to introduce more notation.

As mentioned above, we assume in the following that $\Gamma\backslash X$ is non-compact. Then the space $L^2(\Gamma\backslash X,\sigma)$ can be decomposed into two non-trivial subspaces:

$$L^2(\Gamma\backslash X,\sigma)=L^2_d(\Gamma\backslash X,\sigma)\oplus L^2_c(\Gamma\backslash X,\sigma),$$

where $\Delta$ has a discrete spectrum on $L^2_d(\Gamma\backslash X,\sigma)$ and a continuous spectrum on $L^2_c(\Gamma\backslash X,\sigma)$.

The discrete subspace $L^2_d(\Gamma\backslash X,\sigma)$ has a further decomposition:

$$L^2_d(\Gamma\backslash X,\sigma)=L^2_{cus}(\Gamma\backslash X,\sigma)\oplus L^2_{res}(\Gamma\backslash X,\sigma),$$

where $L^2_{cus}(\Gamma\backslash X,\sigma)$ is spanned by the cuspidal eigenfunctions of $\Delta$, i.e., those eigenfunctions whose constant terms along all proper rational parabolic subgroups are zero (§2.3), and $L^2_{res}(\Gamma\backslash X,\sigma)$ is the orthogonal complement of $L^2_{cus}(\Gamma\backslash X,\sigma)$ in $L^2_d(\Gamma\backslash X,\sigma)$, called the residual discrete subspace.

In [11, Theorem 1.1] (see Lemma 2.3.2 below), Donnelly proved that the counting function $N_{cus}(\lambda)$ of the cuspidal discrete spectrum, i.e., the spectrum of $\Delta$ on the cuspidal subspace $L^2_{cus}(\Gamma\backslash X,\sigma)$, satisfies the following Weyl upper bound:

$$\lim_{\lambda\to+\infty}\sup\frac{N_{cus}(\lambda)}{\lambda^{n/2}}\le(4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)}\dim\sigma.$$

Let $N_{res}(\lambda)$ be the counting function of the residual discrete spectrum, i.e., the spectrum of $\Delta$ restricted to $L^2_{res}(\Gamma\backslash X,\sigma)$. Then

$$N_d(\lambda)=N_{cus}(\lambda)+N_{res}(\lambda).$$

Therefore, the problem is to bound $N_{res}(\lambda)$ and to get the right constant for $N_d(\lambda)$. Examples, whose residual discrete spectra are understood, and our work below strongly suggest the following conjecture.

**Conjecture 1.3.1.** *The counting function of the residual discrete spectrum $N_{res}(\lambda)$ is of smaller order than $\lambda^{n/2}$ as $\lambda \to +\infty$, where $n = \dim \Gamma\backslash X$.*

**Remark 1.3.2.** From the point of view of representation theory of $G$, the residual discrete subspace is contained in the subspace spanned by the complementary series subrepresentations [3] and hence the residual discrete spectrum is contained in the complementary spectrum defined in [12]. If $\Gamma\backslash X$ is compact, it is proved in [12, Theorem 8.3] that the counting function of the complementary spectrum is indeed of smaller order than the Weyl law, i.e., the analogue of Conjecture 1.3.1 holds for compact spaces $\Gamma\backslash X$.

To bound the residual discrete spectrum, we introduce a pseudo-Laplacian (or cut-off Laplacian) $\Delta_T$, where $T$ is the truncation parameter. The domain of $\Delta_T$ roughly consists of those functions whose constant terms along all proper rational parabolic subgroups above height $T$ vanish (see §3 below for the precise definition). Then the spectrum of $\Delta_T$ is discrete. By modifying the arguments of Donnelly in [11], we prove that the spectral counting function of $\Delta_T$ satisfies the Weyl upper bound and a lower bound of the same order (Theorem 3.3.2), which is an analogue of the Weyl–Selberg law as mentioned earlier.

Then the problem is to understand the relation between the eigenvalues of $\Delta$ and $\Delta_T$. All the cuspidal discrete spectrum of $\Delta$ is contained in the spectrum of $\Delta_T$ since truncating the constant terms above $T$ does not affect the cuspidal functions (3.7.2), but the relation between the residual discrete spectrum of $\Delta$ and the spectrum of $\Delta_T$ is not entirely clear. Our guiding philosophy below is that $\Delta_T$ is a good perturbation of $\Delta$ and hence that the residual discrete spectrum can be approximated uniformly by a part of the spectrum of $\Delta_T$.

To explain our approach to the residual discrete spectrum, we need to introduce further notation. As above, we assume that $G$ is the real locus of a reductive algebraic group, and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ is an arithmetic subgroup.

The space $L^2(\Gamma\backslash X, \sigma)$ can be decomposed according to the association classes $\mathcal{C}$ of rational parabolic subgroups $\mathbf{P}$ and the cuspidal spectra $\{\mu\}$ of the boundary locally symmetric spaces $\Gamma_{M_P}\backslash X_P$ associated with

[3]It is shown in [49] that tempered subrepresentations of $L^2(\Gamma\backslash G)$ is cuspidal, and hence residual discrete subspace is contained in the space spanned by nontempered subrepresentations.

**P** (see Lemma 2.5.3):

$$
\begin{aligned}
L^2(\Gamma\backslash X, \sigma) &= \sum_{\mathcal{C}} \oplus L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma), \\
L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma) &= \sum_{\mu} \oplus L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma).
\end{aligned} \tag{1}
$$

When $\mathcal{C} = \{\mathbf{G}\}$, $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ is the cuspidal subspace $L^2_{cus}(\Gamma\backslash X, \sigma)$. For any $1 \leq r \leq \mathrm{rank}_{\mathbb{Q}}(\mathbf{G})$, the subspace

$$\sum_{\mathcal{C}} L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma) \cap L^2_{dis}(\Gamma\backslash X, \sigma),$$

where the sum over $\mathcal{C}$ is over all the association classes of rational parabolic subgroups of rank $r$, is called the rank $r$ residual discrete spectrum of $\Delta$. The above decomposition of $L^2(\Gamma\backslash X, \sigma)$ also induces a decomposition of the pseudo-Laplacian $\Delta_T$ (§3.8). One significance of this decomposition of $L^2(\Gamma\backslash X, \sigma)$ is that after this decomposition, the discrete spectrum of $\Delta$ can be separated away from the continuous spectrum and the regular perturbation theory can be applied (see the remarks after Lemma 2.5.3 and Remark 4.2.4).

Using these decompositions and some positivity of the scattering matrices, we can show that the majority of the rank-one residual discrete spectrum can be approximated uniformly by the corresponding part of the spectrum of $\Delta_T$ (Corollary 5.2.7); in particular, Theorem 1.1.2 holds,

If such a uniform approximation also holds for the higher rank residual discrete spectrum, then the Weyl upper bound $\Delta_T$ (Theorem 3.3.2) implies immediately that the counting function $N_d(\lambda)$ of the discrete spectrum of $\Delta$ satisfies the Weyl upper bound. We believe that such a uniform approximation holds, but we can not prove it. Instead, we follow Müller's approach in [35] of reducing the bound on the higher rank residual discrete spectrum to bounds on poles of the rank-one scattering matrices. Using the above result on the rank-one residual discrete spectrum, the bound on the eigenvalues of $\Delta_T$, and the decomposition of $\Delta$ and $\Delta_T$ induced from Equation (1) above, we can improve the estimates of Müller and prove Theorem 1.1.3.

**Acknowledgements**

I would like to thank Professor S.T.Yau for suggesting me to study the paper [35] many years ago. This paper depends essentially on [35].

The main difference is the emphasis on the relation between the Laplacian $\Delta$ and the pseudo-Laplacian (or cut-off Laplacian) $\Delta_T$. I would also like to thank Professor W.Müller for his interests in early partial results. Especially, I would like to thank the referees for their extremely careful reading of the manuscript and pointing out several mistakes in an earlier version, and for their kind suggestions. I would also like to thank Dr. W.Hoffmann for meticulous reading of the manuscript, for pointing out several mistakes and helpful suggestions, and for sending me the preprint [8].

## 2. Spectral decomposition of $L^2(\Gamma\backslash X, \sigma)$

**2.1.** In this section, we describe Langlands' theory of Eisenstein series (2.4), various decompositions of $L^2(\Gamma\backslash X, \sigma)$ (2.5), the residual discrete spectrum as iterated residues of the cuspidal Eisenstein series (2.6), and some properties of the scattering matrices needed for the reduction to the rank-one case (2.7).

For convenience, throughout this paper, a function on $\Gamma\backslash X$ is identified with a $\Gamma$-invariant function on $X$, and the same convention applies to other quotient spaces.

**2.2.** Recall from §1.1 and §1.3 that $G$ is the real locus $\mathbf{G}(\mathbb{R})$ of a reductive algebraic group $\mathbf{G}$ defined over $\mathbb{Q}$ and satisfying the conditions in [18, p. 1], and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ an arithmetic subgroup. Fix a maximal compact subgroup $K$ of $G$ and hence a basepoint $x_0$ of $X = G/K$.

For any rational parabolic subgroup $\mathbf{P}$, let $\mathbf{N_P}$ be the unipotent radical of $\mathbf{P}$. Denote the real locus $\mathbf{P}(\mathbb{R})$ of $\mathbf{P}$ by $P$, and the real locus $\mathbf{N_P}(\mathbb{R})$ of $\mathbf{N_P}$ by $N_P$. Then $P$ has a (rational) Langlands decomposition $P = N_P M_P A_P$ such that $M_P A_P$ is stabilized by the Cartan involution determined by $K$, where $A_P$ is a lift of the connected component of the real locus of the split center of the Levi quotient $\mathbf{N_P}\backslash\mathbf{P}$ of $\mathbf{P}$, and $M_P$ is a lift of the real locus of the anisotropic part $\mathbf{M_P}$ of the Levi quotient. Denote the Lie algebras of $A_P$ and $N_P$ by $\mathfrak{a}_P$ and $\mathfrak{n}_P$ respectively. Then $\mathfrak{a}_P$ acts on $\mathfrak{n}_P$, and the set of roots is denoted by $\Sigma(P, A_P)$. The half sum of the roots in $\Sigma(P, A_P)$ with multiplicity is denoted by $\rho_P$.

For any $g \in P$, write

$$g = n_P(g) m_P(g) a_P(g) = n_P(g) m_P(g) \exp(H_P(g)), \tag{1}$$

where $n_P(g) \in N_P, a_P(g) \in A_P$, $H_P(g) \in \mathfrak{a}_P$, $m_P(g) \in M_P$ are uniquely determined by $g$.

Let $K_{M_P} = K \cap M_P$. Then $K_{M_P}$ is a maximal compact subgroup of $M_P$, and $X_P = M_P/K_{M_P}$ is the boundary symmetric space associated with $P$. [4]

Since $P$ acts transitively on $X$, the above decomposition induces the following decomposition of $X$:

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

i.e., any point $x \in X = G/K$ can be uniquely written as

$$x = n_P(x)a_P(x)z_P(x)K,$$

where

$$n_P(x) \in N_P, z_P(x) \in X_P = M_P/K_{M_P},$$

$a_P(x) \in A_P$.

The arithmetic subgroup $\Gamma$ induces several discrete subgroups associated with the Langlands decomposition of $P$. Define $\Gamma_P = \Gamma \cap P$, $\Gamma_{N_P} = \Gamma_P \cap N_P$, and $\Gamma_{M_P}$ to be the image of $\Gamma_P$ under the projection $P = N_P A_P M_P \to M_P$. Then $\Gamma_{N_P}$ is a cocompact subgroup of $N_P$, and $\Gamma_{M_P}$ is an arithmetic subgroup of $M_P$. The quotient $\Gamma_{M_P}\backslash X_P$ is called the boundary locally symmetric space associated with $\mathbf{P}$. As mentioned in §1.1, these boundary locally symmetric spaces form the boundary components of the reductive Borel–Serre compactification $\overline{\Gamma\backslash X}^{RBS}$ (see [24, §7]).

**2.3.** Let $E_\sigma$ be the locally homogeneous bundle over $\Gamma\backslash X$ defined by a finite dimensional unitary representation $\sigma$ of $K$. Let $f$ be an $E_\sigma$-valued, locally bounded measurable function on $\Gamma\backslash X$. The constant term $f_P$ of $f$ along a rational parabolic subgroup $\mathbf{P}$ is defined by

$$f_P(x) = \int_{\Gamma_{N_P}\backslash N_P} f(nx)dn, \tag{1}$$

where $dn$ is the Haar measure on $N_P$ normalized by the condition that $\mathrm{vol}(\Gamma_{N_P}\backslash N_P) = 1$.

The subspace of $L^2(\Gamma\backslash X, \sigma)$ consisting of all functions whose constant terms along all proper rational parabolic subgroups vanish is called the cuspidal subspace and denoted by $L^2_{cus}(\Gamma\backslash X, \sigma)$.

[4]In general, $X_P$ is not of noncompact type; instead, it is the product of a symmetric space of noncompact type and a Euclidean space, i.e., $M_P$ is only reductive, not necessarily semisimple. This is the reason that we start with a reductive, instead of semisimple, algebraic group $\mathbf{G}$.

**Lemma 2.3.1.** *The spectrum of the Laplace operator* $\Delta$ *on* $L^2_{cus}(\Gamma\backslash X, \sigma)$ *is discrete, i.e., is of finite multiplicity and has no finite accumulation point. This spectrum is called the cuspidal spectrum and denoted by* $\mathrm{Spec}_{cus}(\Gamma\backslash X, \sigma)$.

*Proof.* It essentially follows from a result of Gelfand and Piatetski-Shapiro [17] [18, Theorem 3]. See [4] for a complete proof of this result. q.e.d.

Let $N_{cus}(\lambda)$ be the counting function of the cuspidal discrete spectrum in $L^2_{cus}(\Gamma\backslash X, \sigma)$. Then Lemma 2.3.1 implies that $N_{cus}(\lambda)$ is finite for every $\lambda > 0$. As mentioned earlier, Donnelly proved the following Weyl upper bound on $N_{cus}(\lambda)$ [11, Theorem 1.1].

**Lemma 2.3.2.** *The counting function of the cuspidal spectrum* $N_{cus}(\lambda)$ *satisfies the Weyl upper bound:*

$$\lim_{\lambda\to+\infty} \sup \frac{N_{cus}(\lambda)}{\lambda^{n/2}} \leq (4\pi)^{-n/2} \frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)} \dim \sigma.$$

**Remark 2.3.3.** Donnelly proved this result for locally symmetric spaces of noncompact type. But the same method works in the slightly more general situation here. (See also Remark 3.4.3.)

**2.4.** We now introduce cuspidal Eisenstein series associated with a rational parabolic subgroup $\mathbf{P}$. Let $\sigma_M : K_{M_P} \to \mathrm{GL}(V)$ be the restriction of the representation $\sigma : K \to \mathrm{GL}(V)$. Then $\sigma_M$ defines a bundle $E_{\sigma_M}$ on $\Gamma_{M_P}\backslash X_P$ and the space $L^2(\Gamma_{M_P}\backslash X_P, \sigma_M)$ of $E_{\sigma_M}$-valued functions on $\Gamma_{M_P}\backslash X_P$. For any $\mu \in \mathrm{Spec}_{cus}(\Gamma_{M_P}\backslash X_P, \sigma_M)$, and a cuspidal eigenfunction $\Phi$ on $\Gamma_{M_P}\backslash X_P$, $\Delta\Phi = \mu\Phi$, where $\Delta$ is the Laplacian on $L^2(\Gamma_{M_P}\backslash X_P, \sigma_M)$ defined as in §1.1, define an Eisenstein series

$$E(P, \Phi, \Lambda, x) = \sum_{\gamma\in\Gamma_P\backslash\Gamma} e^{(\rho_P+\Lambda)(H_P(\gamma x))}\Phi(z_P(\gamma x)),$$

where $H_P(\gamma x) \in \mathfrak{a}_P$ and $z_P(\gamma x) \in X_P$ are the components of $\gamma x$ in the decomposition of $X$ in Equation 2.2.(2)[5], and $\Lambda \in \mathfrak{a}_P^* \otimes \mathbb{C}$, $\mathrm{Re}(\Lambda) \in \rho_P + \mathfrak{a}_P^{*+}$, $\mathfrak{a}_P^{*+} = \{\Lambda \in \mathfrak{a}_P^* \mid \langle\Lambda, \alpha\rangle > 0 \text{ for all } \alpha \in \Sigma(P, A_P)\}$. This series converges uniformly over compact subsets of $(\rho_P + \mathfrak{a}_P^{*+} + i\mathfrak{a}_P^*) \times X$ and can be continued as a meromorphic function in $\Lambda$ to the whole complex space $\mathfrak{a}_P^* \otimes \mathbb{C}$ (see Proposition 2.4.2 below).

[5] Here Equation 2.2.(2) means equation (2) in §2.2. The same convention applies to the rest of the paper.

The theory of constant terms of the Eisenstein series plays an important role in the spectral theory of $\Gamma\backslash X$. To state it, we need the following definition.

**Definition 2.4.1.** Two rational parabolic subgroups $\mathbf{P}_1, \mathbf{P}_2$ are called associated if there exists $g \in G$ such that $Ad(g)\mathfrak{a}_{P_1} = \mathfrak{a}_{P_2}$. The set of such isomorphisms between $\mathfrak{a}_{P_1}$ and $\mathfrak{a}_{P_2}$ is denoted by $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_2})$. For simplicity, $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1})$ is denoted by $W(\mathfrak{a}_{P_1})$.

Let $\mathbf{P}_1, \mathbf{P}_2$ be two associated rational parabolic subgroups. For any cuspidal eigenvalue $\mu \in \cup_{i=1}^{2} \mathrm{Spec}_{cus}(\Gamma_{M_i}\backslash X_{P_i}, \sigma_{M_i})$ denote the eigenspace of $\mu$ in $L^2_{cus}(\Gamma_{M_i}\backslash X_{P_i}, \sigma_{M_i})$ by $\mathcal{E}_{cus}(\Gamma_{M_i}\backslash X_{P_i}, \sigma_{M_i}, \mu)$. Then for any $s \in W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_2})$ and $\Lambda \in \mathfrak{a}^*_{P_1} \otimes \mathbb{C}$, there is an intertwining operator (or scattering matrix)

$$(1) \qquad C_{P_2,P_1}(s, \Lambda) : \mathcal{E}_{cus}(\Gamma_{M_1}\backslash X_{P_1}, \sigma_{M_1}, \mu) \to \mathcal{E}_{cus}(\Gamma_{M_2}\backslash X_{P_2}, \sigma_{M_2}, \mu)$$

such that for any $\Phi \in \mathcal{E}_{cus}(\Gamma_{M_1}\backslash X_{P_1}, \sigma_{M_1}, \mu)$, the constant term of $E(P_1, \Phi, \Lambda)$ along $\mathbf{P}_2$ is given by

$$(2) \quad E_{P_2}(P_1, \Phi, \Lambda, x) = \sum_{s \in W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_2})} e^{(s\Lambda + \rho_{P_2})(H_{P_2}(x))} C_{P_2,P_1}(s, \Lambda)\Phi(x),$$

where the $s$ action on $\mathfrak{a}^*_{P_1}$ is defined as follows: For any $\Lambda \in \mathfrak{a}^*_{P_1}$, $H \in \mathfrak{a}_{P_2}$, $(s\Lambda)(H) = \Lambda(s^{-1}H)$ (see [18, Theorem 5]).

These intertwining operators satisfy functional equations. Let $\mathcal{C}$ be an association class of proper rational parabolic subgroups, and $\mathcal{C}_1, \cdots, \mathcal{C}_r$ be the $\mathbf{G}(\mathbb{Q})$-conjugacy classes in $\mathcal{C}$. Let $\mathbf{P}_1, \cdots, \mathbf{P}_r$ be representatives of the $\mathbf{G}(\mathbb{Q})$-conjugacy classes. For any $1 \le i \le r$, let $\mathbf{P}_{i1}, \cdots, \mathbf{P}_{ir_i}$ be representatives of the $\Gamma$-conjugacy classes in $\mathcal{C}_i$. Then $\mathbf{P}_{il}$, $1 \le i \le r$, $1 \le l \le r_i$, are representatives of the $\Gamma$-conjugacy classes in $\mathcal{C}$. Denote the split component $\mathfrak{a}_{P_i}$ by $\mathfrak{a}_i$, and $\mathfrak{a}_{P_{il}}$ by $\mathfrak{a}_{il}$. Choose $y_{il} \in K$ such that $P_{il} = y_{il} P_i y_{il}^{-1}$. Then $\mathfrak{a}_{il} = Ad(y_{il})\mathfrak{a}_i$. By duality, $Ad(y_{il})$ defines a map $Ad(y_{il}) : \mathfrak{a}^*_i \to \mathfrak{a}^*_{il}$.

For the association class $\mathcal{C}$, define

$$\mathrm{Spec}_{cus}(\mathcal{C}) = \cup_{i=1}^{r} \cup_{l=1}^{r_i} \mathrm{Spec}_{cus}(\Gamma_{M_{il}}\backslash X_{P_{il}}, \sigma_{M_{il}}),$$

called the cuspidal spectrum of the class $\mathcal{C}$. For every $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$, define

$$(3) \qquad \mathcal{E}_{cus}(\mathcal{C}_i, \mu) = \oplus_{l=1}^{r_i} \mathcal{E}_{cus}(\Gamma_{M_{il}}\backslash X_{P_{il}}, \sigma_{M_{il}}, \mu).$$

For any $\Phi = (\Phi_1, \cdots, \Phi_{r_i}) \in \mathcal{E}_{cus}(\mathcal{C}_i, \mu)$, and $\Lambda \in \mathfrak{a}_i^* \otimes \mathbb{C}$, define

$$E(\mathbf{P}_i, \Phi, \Lambda) = \sum_{l=1}^{r_i} E(P_{il}, \Phi_l, Ad(y_{il})\Lambda).$$

Then for $1 \leq i, j \leq r$, $s \in W(\mathfrak{a}_i, \mathfrak{a}_j)$, $\Lambda \in \mathfrak{a}_i^* \otimes \mathbb{C}$, the scattering matrix

$$C_{ji}(s, \Lambda) : \mathcal{E}_{cus}(C_i, \mu) \to \mathcal{E}_{cus}(C_j, \mu)$$

has entries $C_{P_{jk}, P_{il}}(Ad(y_{jk})s\ Ad(y_{il})^{-1}, Ad(y_{il})\Lambda)$, $1 \leq k \leq r_j$, $1 \leq l \leq r_i$, and the constant term of $E(\mathbf{P}_i, \Phi, \Lambda)$ along $P_{jk}$ is given as follows:

$$E_{P_{jk}}(\mathbf{P}_i, \Phi, \Lambda) = \sum_{s \in W(\mathfrak{a}_i, \mathfrak{a}_j)} e^{(s\Lambda + \rho_{P_{jk}})(H_{P_{jk}}(x))} (C_{ij}(s, \Lambda)\Phi)_{jk}, \tag{4}$$

where $(C_{ij}(s, \Lambda)\Phi)_{jk}$ is the $jk$-component in the decomposition $\mathcal{E}_{cus}(\mathcal{C}_j, \mu)$ in Equation (3) above.

A basic result in the theory of Eisenstein series is the following (see [18, Theorems 7, 8, 9] and [34, Proposition IV.1.11]).

**Proposition 2.4.2.**

1. *For any $1 \leq i, j \leq r$, $s \in W(\mathfrak{a}_i, \mathfrak{a}_j)$, the scattering matrix $C_{ji}(s, \Lambda)$ is meromorphic in $\Lambda \in \mathfrak{a}_i^* \otimes \mathbb{C}$, whose all singularities lie along hyperplanes of the form $\langle \alpha, \Lambda \rangle = c$, where $\alpha \in \Sigma(P_i, A_i)$ and $c$ is a constant. Moreover, the singular hyperplanes in the tube domain over the positive chamber $\mathfrak{a}_i^{*+} + \sqrt{-1}\mathfrak{a}_i^*$ are real and simple.*

2. *For any $\Phi \in \mathcal{E}_{cus}(\mathcal{C}_i, \mu)$, the Eisenstein series $E(\mathbf{P}_i, \Phi, \Lambda)$ satisfies the functional equation*

$$E(\mathbf{P}_i, \Phi, \Lambda) = E(\mathbf{P}_j, C_{ji}(s, \Lambda)\Phi, s\Lambda).$$

3. *The Eisenstein series $E(\mathbf{P}_i, \Phi, \Lambda)$ is meromorphic in $\Lambda \in \mathfrak{a}_i^* \otimes \mathbb{C}$ whose poles (with multiplicity) are contained in the union of the poles of $C_{ij}(s, \Lambda)$, $s \in W(\mathfrak{a}_i, \mathfrak{a}_j)$, $1 \leq j \leq r$.*

**Remark 2.4.3.** To get the spectral decomposition of $L^2(\Gamma\backslash G)$ with respect to all invariant differential operators, one usually requires the cuspidal function $\Phi$ in $E(P, \Phi, \Lambda, x)$ to be an eigenfunction of all invariant differential operators as in [27] [34] [40]. Since we mainly deal with the spectral decomposition of the Laplacian on $L^2(\Gamma\backslash X, \sigma)$, we only require $\Phi$ to be an eigenfunction of the Laplacian on $L^2(\Gamma_{M_P}\backslash X_P, \sigma_M)$ to save notation.

**2.5.** In general,

$$E(P, \Phi, \Lambda, x) \notin L^2(\Gamma\backslash X, \sigma),$$

though

$$E(P, \Phi, \Lambda, x) \in L^2_{loc}(\Gamma\backslash X, \sigma).$$

To get square integrable functions, we need to introduce pseudo-Eisenstein series. For any function $f \in C_0^\infty(\mathfrak{a}_P)$, define

$$E(P, \Phi, f, x) = \sum_{\gamma \in \Gamma_P\backslash\Gamma} f(H_P(\gamma x))\Phi(z_P(\gamma x)).$$

**Lemma 2.5.1** ([18, Corollary to Lemma 26]). *For any such $f$, $E(P, \Phi, f, x) \in L^2(\Gamma\backslash X, \sigma)$.*

The relation between the Eisenstein series and the pseudo-Eisenstein series is as follows. Let $\hat{f}(\Lambda)$ be the Fourier transform of $f$:

$$\hat{f}(\Lambda) = \int_{\mathfrak{a}} f(H)e^{-(\Lambda+\rho_P)(H)}dH.$$

Then

$$E(P, \Phi, f, x) = \int_{\mathrm{Re}(\Lambda)=\lambda} E(P, \hat{f}(\Lambda)\Phi, \Lambda, x)d\Lambda, \tag{1}$$

where $\lambda \in \rho_P + \mathfrak{a}_P^{*+}$ [18, Lemma 28].

**Lemma 2.5.2** ([18, Lemma 39]). *Let $\mathbf{P}_1, \mathbf{P}_2$ be two rational parabolic subgroups, and $\Phi_1, \Phi_2$ are two cuspidal eigenfunctions on $\Gamma_{M_1}\backslash X_{P_1}$ and $\Gamma_{M_2}\backslash X_{P_2}$ with eigenvalues $\mu_1, \mu_2$ respectively. Then*

$$(E(P_1, \Phi_1, f_1), E(P_2, \Phi_2, f_2))_{L^2} = 0$$

*if either $P_1$ and $P_2$ are not associated, or $\mu_1 \neq \mu_2$.*

For any association class $\mathcal{C}$ of rational parabolic subgroups and a cuspidal eigenvalue $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$, which is the union of the cuspidal eigenvalues of $L^2(\Gamma_{M_P}\backslash X_P, \sigma_M)$ for all the parabolic subgroups $\mathbf{P} \in \mathcal{C}$, the pseudo-Eisenstein series $E(P, \Phi, f)$, where $\Phi \in \mathcal{E}_{cus}(\Gamma_M\backslash X_P, \sigma_M, \mu)$, $\mathbf{P} \in \mathcal{C}$, $f \in C_0^\infty(\mathfrak{a}_P)$, span a closed subspace of $L^2(\Gamma\backslash X, \sigma)$, denoted by $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$. For convenience, the pair $(\mathcal{C}, \mu)$ is called a cuspidal pair.

For each association class $\mathcal{C}$, define

$$L_{\mathcal{C}}(\Gamma\backslash X, \sigma) = \sum_{\mu \in \mathrm{Spec}_{cus}(\mathcal{C})} L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma).$$

Then we have the following decomposition of $L^2(\Gamma\backslash X, \sigma)$.

**Lemma 2.5.3** ([18, Lemma 45]). *With the above notation,*

$$L^2(\Gamma\backslash X,\sigma)=\sum_{\mathcal{C}}\oplus L_{\mathcal{C}}(\Gamma\backslash X,\sigma)=\sum_{(\mathcal{C},\mu)}\oplus L^2_{\mathcal{C},\mu}(\Gamma\backslash X,\sigma),$$

*where the second sum is over all the cuspidal pairs $(\mathcal{C},\mu)$. The Laplace operator $\Delta$ preserves this decomposition and hence is the direct sum of the restrictions to the subspaces.*

In general, the discrete spectrum of $\Delta$ on $L^2(\Gamma\backslash X,\sigma)$ is embedded in the continuous spectrum. But after this decomposition of $L^2(\Gamma\backslash X,\sigma)$ in Lemma 2.5.3, the discrete spectrum is separated out. For example, if $\mathcal{C}=\{\mathbf{G}\}$, then $\Delta$ has only a discrete spectrum on $L^2_{\mathcal{C}}(\Gamma\backslash X,\sigma)$, which is equal to the cuspidal subspace $L^2_{cus}(\Gamma\backslash X,\sigma)$. For an association class $\mathcal{C}$ of rank-one parabolic subgroups, the discrete spectrum of $\Delta$ on $L^2_{\mathcal{C},\mu}(\Gamma\backslash X,\sigma)$ is below its continuous spectrum on $L^2_{\mathcal{C},\mu}(\Gamma\backslash X,\sigma)$ (see Lemma 4.2.2 below). For other association classes, it is more complicated (see Remark 4.2.4 below).

**2.6.** Next we recall the description of the residual discrete spectrum as iterated residues of cuspidal Eisenstein series. Besides the cuspidal eigenfunctions and cuspidal Eisenstein series, the residual discrete spectrum and the associated noncuspidal Eisenstein series are also needed to give the spectral decomposition of $L^2(\Gamma\backslash X,\sigma)$.

For any rational parabolic subgroup $\mathbf{P}$, an affine subspace in $\mathfrak{a}_P^*\otimes\mathbb{C}$ is called admissible if it is the intersection of affine hyperplanes of the form $\{\Lambda\in\mathfrak{a}_P^*\otimes\mathbb{C}\mid\langle\alpha,\Lambda\rangle=c\}$, where $\alpha\in\Sigma(P,A_P)$, and $\langle\cdot,\cdot\rangle$ is induced from the Killing form. In particular, the affine hyperplanes $\{\Lambda\in\mathfrak{a}_P^*\otimes\mathbb{C}\mid\langle\alpha,\Lambda\rangle=c\}$ are called admissible hyperplanes.

A complete flag of admissible subspaces of $\mathfrak{a}_P^*$:

$$\mathcal{F}:\mathfrak{a}_P^*\otimes\mathbb{C}=V_r\supset V_{r-1}\supset\cdots\supset V_0,\quad r=\dim\mathfrak{a}_P^*,$$

is called an admissible flag. For any admissible flag, choose real unit vectors $\Lambda_i\in V_i$, $i=1,\cdots,r$ such that $\Lambda_i\perp V_{i-1}$. Then for any meromorphic function $E(\Lambda)$ on $\mathfrak{a}_P^*\otimes\mathbb{C}$ whose singularities lie on admissible hyperplanes, we can define the iterated residue $\mathrm{Res}_{\mathcal{F}}E$ along the admissible flag $\mathcal{F}$ as follows: $E_r(\Lambda)=E(\Lambda)$, and for $i=0,\cdots,r-1$, $\Lambda\in V_i$,

$$E_i(\Lambda)=\delta\int_0^1 E_{i+1}(\Lambda+\delta e^{2\pi\sqrt{-1}\theta}\Lambda_{i+1})d\theta,$$

where $\delta$ is a small positive number so that $E_{i+1}(\Lambda + z\Lambda_{i+1})$ has no singularities for $0 < |z| < \delta$, and

$$\mathrm{Res}_{\mathcal{F}} E(\Lambda) = E_0(\Lambda_0),$$

where $\Lambda_0$ is the unique element in $V_0$.

By Lemma 2.5.3, to describe the residual discrete subspace $L^2_{res}(\Gamma\backslash X, \sigma)$, it suffices to describe the subspaces

$$L^2_{res}(\Gamma\backslash X, \sigma) \cap L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$$

for all cuspidal pairs $(\mathcal{C}, \mu)$. If $\mathcal{C} = \{\mathbf{G}\}$, then

$$L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma) \subset L^2_{cus}(\Gamma\backslash X, \sigma),$$

and this intersection

$$L^2_{res}(\Gamma\backslash X, \sigma) \cap L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$$

is empty. We assume that $\mathcal{C}$ is an association class of proper rational parabolic subgroups. Then

$$L^2_{res}(\Gamma\backslash X, \sigma) \cap L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$$

is given as follows.

For any $\mathbf{P} \in \mathcal{C}$, $\Phi \in \mathcal{E}_{cus}(\Gamma_M\backslash X_P, \sigma_M, \mu)$, by Proposition 2.4.2, the singularities of $E(P, \Phi, \Lambda)$ lie on admissible hyperplanes of $\mathfrak{a}_P^* \otimes \mathbb{C}$. For any $r$ singular hyperplanes $\mathcal{H}_1, \cdots, \mathcal{H}_r$ with $\cap_{i=1}^r \mathcal{H}_i = \{\Lambda_0\}$, where $r = \dim \mathfrak{a}_P^*$, set $V_r = \mathfrak{a}_P^* \otimes \mathbb{C}$, and for $j = 1, \cdots, r$, $V_{r-j} = \mathcal{H}_1 \cap \cdots \cap \mathcal{H}_j$. Then $\mathcal{F} : V_r \supset \cdots V_0 = \{\Lambda_0\}$ is an admissible flag. For every $s \in W(\mathfrak{a}_P)$, $s \cdot \mathcal{F}$ is also an admissible flag.

For any $f \in C_0^\infty(\mathfrak{a}_P)$, let $\overset{\wedge}{f}(\Lambda)$ be its Fourier transformation which is a holomorphic function on $\mathfrak{a}_P^* \otimes \mathbb{C}$. Then $\mathrm{Res}_{\mathcal{F}}(E(P, \Phi, \Lambda)\overset{\wedge}{f}(\Lambda))$ is a function on $\Gamma\backslash X$.

**Proposition 2.6.1** ([27, Theorem 7.1, p. 222] [34, Theorem V.3.13, p. 221]). *Let* $^+\mathfrak{a}_P^*$ *be the closed cone in* $\mathfrak{a}_P^*$ *dual to the positive chamber* $\mathfrak{a}_P^+$, *i.e., the span of the roots in* $\Sigma(P, A_P)$. *If* $\Lambda_0 \in {}^+\mathfrak{a}_P^*$, *then the following sum of iterated residues* $\sum_{s \in W(\mathfrak{a}_P)} \mathrm{Res}_{s\cdot\mathcal{F}}(E(P, \Phi, \Lambda)\overset{\wedge}{f}(\Lambda))$ *is square integrable, i.e., belongs to* $L^2(\Gamma\backslash X, \sigma)$, *and is an eigenfunction of* $\Delta$ *with eigenvalue* $|\rho_P|^2 - |\Lambda_0|^2 + \mu$. *Furthermore, the residual discrete subspace* $L^2_{res}(\Gamma\backslash X, \sigma) \cap L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ *is spanned by these* *iterated residues* $\sum_{s\in W(\mathfrak{a}_P)} \mathrm{Res}_{s\cdot\mathcal{F}}(E(P,\Phi,\Lambda)\overset{\wedge}{f}(\Lambda))$ *at singular points* $\Lambda_0$ *in the bounded, closed domain* $\{\Lambda \in {}^+\mathfrak{a}_P^* \mid |\Lambda| \leq |\rho_P|\}$, *where* $\mathbf{P} \in \mathcal{C}, \Phi \in \mathcal{E}_{cus}(\Gamma_M\backslash X_P, \sigma_M, \mu)$, *and* $f \in C_0^\infty(\mathfrak{a}_P)$.

**Remark 2.6.2.** The sum over $W(\mathfrak{a}_P)$ is necessary for the square integrability of the residues. In fact, each individual iterated residue $\mathrm{Res}_{\mathcal{F}}(E(P,\Phi,\Lambda)\overset{\wedge}{f}(\Lambda))$ may not be square integrable. See [16, Remark 1 at the end of §6] for an explicit example. On the other hand, if $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P}) = 1$ and $\Phi$ is cuspidal, then $\mathrm{Res}_{\mathcal{F}}(E(P,\Phi,\Lambda)\overset{\wedge}{f}(\Lambda))$ is square integrable and the residual discrete subspace of $L^2_{\mathcal{C},\mu}(\Gamma\backslash X,\sigma)$ is spanned by such residues.

**Remark 2.6.3.** The factor $\overset{\wedge}{f}(\Lambda)$ is inserted because $E(P,\Phi,\Lambda)$ could have nonsimple poles outside the positive chamber $\mathfrak{a}_P^{*+}$. This phenomenon can be seen in the example of $G_2$ in [27, Appendix III].

**Remark 2.6.4.** The fact that the singular point $\Lambda_0$ is restricted to the bounded domain $\{\Lambda \in {}^+\mathfrak{a}_P^* \mid |\Lambda| \leq |\rho_P|\}$ could be briefly explained as follows: The Eisenstein series $E(P,\Phi,\Lambda)$ is holomorphic in $\Lambda$ when $\mathrm{Re}(\Lambda) \in \rho_P + \mathfrak{a}_P^{*+}$. The residues are picked up when the contour of integration in Equation 2.5.(1) is moved from $\lambda \in \rho_P + \mathfrak{a}_P^{*+}$ to $\lambda = 0$. In the process, contours of smaller dimension based at one point $\lambda_1$ are moved to another basepoint $\lambda_2$, where $\lambda_1$ belongs to an admissible real affine subspace $V$ which is the intersection of singular hyperplanes of $E(P,\Phi,\Lambda)$, and $\lambda_2$ is the intersection of this admissible subspace $V$ and the subspace spanned by the roots which define $V$. (See [28] and [27, Appendix III] for pictures and illustrations of the deformation of the contours of integration.) So after the whole deformation process, only iterated residues at points in the above bounded domain are picked up.

**Remark 2.6.5.** Once we get the residual discrete spectrum, we can describe the complete spectral decomposition of $L^2_{\mathcal{C},\mu}(\Gamma\backslash X,\sigma)$ for $\mathcal{C} \neq \{\mathbf{G}\}$. The complement of the residual discrete spectrum in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X,\sigma)$ is a continuous spectrum. This continuous subspace is spanned by wave packets of Eisenstein series $E(Q,\Psi,\Lambda)$, $\mathrm{Re}(\Lambda) = 0$, where $\mathbf{Q}$ is a rational parabolic subgroup containing a group $\mathbf{P} \in \mathcal{C}$, and $\Psi$ is a residual eigenfunction on $\Gamma_{M_Q}\backslash X_Q$, which is the residue of an Eisenstein series on $\Gamma_{M_Q}\backslash X_Q$ associated with a cuspidal eigenfunction $\Phi$ on $\Gamma_{M_P}\backslash X_P$. If $\mathbf{Q} = \mathbf{P}$, then $E(Q,\Psi,\Lambda) = E(P,\Phi,\Lambda)$ is a cuspidal Eisenstein series. Otherwise, such Eisenstein series $E(Q,\Psi,\Lambda)$ arise as sums of the iterated residues of the cuspidal Eisenstein series $E(P, \Phi, \Lambda)$ along an affine subspace of positive dimension.

**2.7.** The above description of the residual discrete spectrum as iterated residues and Proposition 2.4.2 show that to bound the counting function of the residual discrete spectrum, it suffices to bound the number of complete flags of singular hyperplanes of the scattering matrices $C(s, \Lambda)$ in the bounded domain in Proposition 2.6.1.

In the rest of this section, we recall from [18, Chap. V] and [35, pp. 523–525] that the scattering matrices of higher rank parabolic subgroups can be expressed as products of the scattering matrices of rank-one parabolic subgroups of some Levi subgroups of $\mathbf{G}$. This reduces the problem to bounding the poles of the scattering matrices of rank-one.

Let $\mathcal{C}$ be an association class of rational parabolic subgroups. Fix a parabolic subgroup $\mathbf{P} \in \mathcal{C}$ with split component $A = A_P$. Then there is a canonical correspondence between the set of rational parabolic subgroups whose split components are equal to $A$ and the set of chambers in $\mathfrak{a}$. Denote the chambers of $\mathfrak{a}$ by $C_1, \cdots, C_q$, and the corresponding parabolic subgroups by $\mathbf{P}_1, \cdots, \mathbf{P}_q$. Since every parabolic subgroup in $\mathcal{C}$ is $\mathbf{G}(\mathbb{Q})$-conjugate to a parabolic subgroup whose split component is equal to $A$, the set of $\mathbf{P}_1, \cdots, \mathbf{P}_q$ contains a set of representatives of the $\mathbf{G}(\mathbb{Q})$-conjugacy classes in $\mathcal{C}$ in §2.4. For any $1 \leq i \leq q$, denote the $\mathbf{G}(\mathbb{Q})$-conjugacy class in $\mathcal{C}$ containing $\mathbf{P}_i$ by $\mathcal{C}_i$. Then for any $s \in W(\mathfrak{a})$ and $1 \leq i, j \leq q$, we can define a scattering matrix $C_{ji}(s, \Lambda)$ as in §2.4.

**Lemma 2.7.1** ([18, Theorem 8]). *For any $1 \leq i, j, k \leq q$, and $s, t \in W(\mathfrak{a})$,*

$$C_{ji}(ts, \Lambda) = C_{jk}(t, s\Lambda) C_{ki}(s, \Lambda).$$

**Lemma 2.7.2** ([18, Lemmas 80 and 115]). *For any $1 \leq i \leq q$, $C_{ii}(1, \Lambda) = Id$. For any $1 \leq i, j \leq q$, $s \in W(\mathfrak{a})$, if $sC_i = C_j$, then $C_{ji}(s, \Lambda)$ is holomorphic in $\Lambda \in \mathfrak{a}^* \otimes \mathbb{C}$.*

**Lemma 2.7.3** ([18, p. 123]). *For any two chambers $C_i$ and $C_j$, there exists a chain of chambers $C_{i_1}, \cdots, C_{i_p}$ with $p \leq |\Sigma(P, A_P)|$ such that $C_{i_1} = C_i$, $C_{i_p} = C_j$, and $C_{i_l}$ and $C_{i_{l+1}}$ are adjacent for $l = 1, \cdots, p-1$.*

From Lemma 2.7.1, we get the following factorization of the scattering matrices.

**Lemma 2.7.4.** *For any $1 \leq i, j \leq q$ and $s \in W(\mathfrak{a})$, let $C_k = sC_i$, and $C_{j_1}, \cdots, C_{j_p}$ a chain connecting $C_j$ and $C_k$ as in Lemma 2.7.3.*

*Then*

$$C_{ji}(s,\Lambda) = C_{jk}(1,s\Lambda)C_{ki}(s,\Lambda) \tag{1}$$
$$= C_{j_1,j_2}(1,s\Lambda)\cdots C_{j_{p-1},j_p}(1,s\Lambda)C_{ki}(s,\Lambda). \tag{2}$$

By Lemma 2.7.2, $C_{ki}(s,\Lambda)$ is holomorphic. Since the chambers $C_{j_l}, C_{j_{l+1}}$ are adjacent, to bound the singularities of $C_{ij}(s,\Lambda)$, it suffices to bound the singularities of $C_{ij}(1,\Lambda)$ when $C_i, C_j$ are adjacent.

Assume that $C_i, C_j$ are adjacent. Let $\mathbf{Q}$ be the unique rational parabolic subgroup containing both $\mathbf{P}_i, \mathbf{P}_j$ with $\overline{\mathfrak{a}_Q^+} = \overline{C_i} \cap \overline{C_j}$, in particular, $\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) = \mathrm{rank}_{\mathbb{Q}}(\mathbf{P}_i) - 1$. Then $\mathbf{P}_i, \mathbf{P}_j$ determine two rank-one parabolic subgroups $'\mathbf{P}_i, '\mathbf{P}_j$ of $\mathbf{M_Q}$. Denote the Lie algebra of the split component of $'P_i$ by $'\mathfrak{a}_i$. Then $\dim {'\mathfrak{a}_i} = 1$. Denote the $\mathbf{M_Q}(\mathbb{Q})$-conjugacy classes of rational parabolic subgroups of $\mathbf{M_Q}$ containing $'\mathbf{P}_i, '\mathbf{P}_j$ by $'\mathcal{C}_i, '\mathcal{C}_j$, and the scattering matrix by $C_{'j,'i}(s, '\Lambda)$, $s \in W('\mathfrak{a}_i)$, $'\Lambda \in ('\mathfrak{a}_i)^* \otimes \mathbb{C}$, where $('\mathfrak{a}_i)^*$ is the dual of $'\mathfrak{a}_i$.

**Lemma 2.7.5** ([18, p. 124-125], [35, p. 524-525]). *An entry of the scattering matrix $C_{ji}(1,\Lambda)$ is either zero or equal to an entry of the rank-one scattering matrix $C_{'j,'i}(s,'\Lambda)$, where $s \in W('\mathfrak{a}_i)$, and $'\Lambda$ is the restriction of $\Lambda$ to $'\mathfrak{a}_i \otimes \mathbb{C}$.*

## 3. The pseudo-Laplacian $\Delta_T$

**3.1.** In this section, we recall the pseudo-Laplacian (or cut-off Laplacian) $\Delta_T$ introduced by Müller [35, §3] and show that its counting function satisfies the Weyl upper bound and a lower bound of the same order (§3.3). This result could be interpreted as an analogue of the Weyl–Selberg law as mentioned in §1.2. Then we characterize the eigenfunctions of $\Delta_T$ in order to establish connections between $\Delta$ and $\Delta_T$ (§3.7).

The pseudo-Laplacian $\Delta_T$ is closely connected with Arthur's truncation operator $\Lambda^T$ in [2], [39]. In fact, $\Delta_T$ can be defined through $\Lambda^T$ (see §3.3).

**3.2.** Before defining the pseudo-Laplacian, we recall the precise reduction theory. Let $\mathbf{P}_1, \cdots, \mathbf{P}_m$ be a set of representatives of $\Gamma$-conjugacy classes of rank-one, i.e., maximal rational parabolic subgroups of $\mathbf{G}$. Let $\mathfrak{a}_j = \mathfrak{a}_{P_j}$, $j = 1, \cdots, m$. Define

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

Then for any rational parabolic subgroup $\mathbf{P}$, there is a well defined map

$$I_P : \mathfrak{a}_0 \to \mathfrak{a}_P \tag{1}$$

such that if $\mathbf{P} = \mathbf{P}_j$, $1 \leq j \leq m$, then $I_{P_j}$ is the projection from $\mathfrak{a}_0$ to the factor $\mathfrak{a}_{P_j}$, [39, p. 330], [44].

Let $\rho_j$ be the half sum of the roots in $\Sigma(P_j, A_j)$ with multiplicity. Then $\rho_j$ defines a vector $H_{\rho_j}$ in $\mathfrak{a}_j$ under the duality defined by the Killing form. These vectors $H_{\rho_j}$, $j = 1, \cdots, m$, define a unique vector $H_\rho$ in $\mathfrak{a}_0$ such that $I_{P_j}(H_\rho) = H_{\rho_j}$.

Fix a large positive number $t$ so that $T = tH_\rho \in \mathfrak{a}_0$. For any rational parabolic subgroup $\mathbf{P}$, define

$$\begin{aligned} A_{P,T} =& \{e^H \in A_P \mid \alpha(H) > \alpha(I_{\mathbf{P}}(T)), \alpha \in \Sigma(P, A_P)\}, \\ A_P^T =& \{e^H \mid \langle I_P(T) - H, V\rangle \geq 0 \text{ for all } V \in \mathfrak{a}_P^+\}, \end{aligned}$$

i.e., $A_P^T$ is a shift of the negative of the obtuse cone dual to the dominant cone $A_{P,0} = \exp \mathfrak{a}^+$. Using the decomposition $X = N_P \times A_P \times X_P$ in Equation 2.2.(2), we get a subset $N_P \times A_P^T \times X_P$ in $X$.

Denote by $X_T$ the intersection $\cap_{\mathbf{P}} N_P \times A_P^T \times X_P$ over all proper rational parabolic subgroups $\mathbf{P}$. Then $X_T$ is a $\Gamma$-invariant noncompact submanifold with corners of $X$ of codimension 0 and is the central tile in [44, Equation (5.1)] associated with the parameter $T$.

**Proposition 3.2.1** ([44, Proposition 2.2, Theorem 5.7]). *For $t \gg 0$ and $T = tH_\rho$, the quotient $\Gamma\backslash X_T$ is a compact submanifold with corners of $\Gamma\backslash X$, denoted by $(\Gamma\backslash X)_T$.*

Intuitively, $(\Gamma\backslash X)_T$ is obtained by chopping off all the cusps at infinity at the height $T$ and hence is a compact core of $\Gamma\backslash X$. In the following, $T$ will be referred to as the truncation parameter.

Let $\mathbf{P}_1, \cdots, \mathbf{P}_p$ be a set of representatives of $\Gamma$-conjugacy classes of proper rational parabolic subgroups. Then the precise reduction theory in [44, Proposition 2.2 and Theorem 5.7] can be stated as follows.

**Proposition 3.2.2.** *For every $1 \leq i \leq p$, there exists a compact submanifold with corners $\omega_i \subset \Gamma_{P_i}\backslash N_{P_i} \times X_{P_i}$ which is left $N_{P_i}$-invariant such that when $t$ is sufficiently large as above and $T = tH_\rho$, the subset $\omega_i \times A_{P_i,T} \subset \Gamma_{P_i}\backslash X$ is mapped injectively into $\Gamma\backslash X$ whose image is still denoted by $\omega_i \times A_{P_i,T}$, and $\Gamma\backslash X$ admits the following disjoint decomposition:*

$$\Gamma\backslash X = (\Gamma\backslash X)_T \cup \coprod_{i=1}^{p} \omega_i \times A_{P_i,T}.$$

**3.3.** Denote by $H^1(\Gamma\backslash X,\sigma)$ the Sobolev space of sections $f$ of the vector bundle $E_\sigma$ such that

$$\int_{\Gamma\backslash X} |f|^2 + |\bigtriangledown f|^2 < +\infty.$$

For the truncation parameter $T \in \mathfrak{a}_0$ as above, define a subspace $H^1_T(\Gamma\backslash X,\sigma)$ of $H^1(\Gamma\backslash X,\sigma)$ as follows:

$$H^1_T(\Gamma\backslash X,\sigma) = \{f \in H^1(\Gamma\backslash X,\sigma) \mid f_P(nza) = 0 \\ \text{for } n \in N_P, z \in X_P, a \in A_{P,T}\},$$

where $\mathbf{P}$ runs over all proper maximal parabolic subgroups of $\mathbf{G}$. For a non-maximal parabolic subgroup $\mathbf{P}$, the constant term $f_P(nza)$ vanishes when $a \in A_P \setminus A_P^T$, $n \in N_P$, $z \in X_P$. Roughly, $H^1_T(\Gamma\backslash X,\sigma)$ is the subspace of functions all whose constant terms vanish outside the compact core $(\Gamma\backslash X)_T$. (See Remark 3.3.3 for another description of $H^1_T(\Gamma\backslash X,\sigma)$.)

Since $H^1_T(\Gamma\backslash X,\sigma)$ is a closed subspace of $H^1(\Gamma\backslash X,\sigma)$, the Dirichlet quadratic form

$$D(f) = \int_{\Gamma\backslash X} |\bigtriangledown f|^2$$

restricts to $H^1_T(\Gamma\backslash X,\sigma)$ and defines a self-adjoint operator $\Delta_T$ on the closure of $H^1_T(\Gamma\backslash X,\sigma)$ in $L^2(\Gamma\backslash X,\sigma)$ [35, p. 489]. This operator $\Delta_T$ is called the pseudo-Laplace (or cut-off Laplace) operator at the height $T$. If $G = \mathrm{SL}(2,\mathbb{R})$, this operator was first defined by Lax and Phillips [30], and used by Colin de Verdiere [7] to study the meromorphic continuation of the Eisenstein series and the discrete spectrum.

**Proposition 3.3.1** ([35, Theorem 3.23]). *The spectrum of $\Delta_T$ is discrete and the counting function $N_T(\lambda) = |\{\lambda_i \in \mathrm{Spec}(\Delta_T) \mid \lambda_i \le \lambda\}|$ satisfies $N_T(\lambda) \le c(1+\lambda^{n/2}), \lambda \ge 0$, where $n = \dim \Gamma\backslash X$ and $c$ is a positive constant.*

The main result of this section is the following bounds for the pseudo-Laplacian $\Delta_T$.

**Theorem 3.3.2.** *Let $N_T(\lambda)$ be the counting function of the eigen-*

*values of* $\Delta_T$. *Then*

$$\begin{aligned}(4\pi)^{-n/2}\frac{\mathrm{vol}((\Gamma\backslash X)_T)}{\Gamma(\frac{n}{2}+1)}\dim\sigma \leq & \lim_{\lambda\to+\infty}\inf\frac{N_T(\lambda)}{\lambda^{n/2}} \\ \leq & \lim_{\lambda\to+\infty}\sup\frac{N_T(\lambda)}{\lambda^{n/2}} \\ \leq & (4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)}\dim\sigma.\end{aligned}$$

Only the upper bound for $N_T(\lambda)$ is used in the proof of Theorems 1.1.2 and 1.1.3, but the lower bound is of independent interest. It will be shown below that $N_T(\lambda)$ involves counting of both the discrete spectrum and the continuous spectrum of $\Delta$, and hence this theorem can be looked upon as an analogue of the Weyl–Selberg law as mentioned in §1.2 (see Remark 3.5.2 for more details).

In the next two subsections §3.4 and §3.5, we prove Theorem 3.3.2. In §3.6, we introduce other pseudo-Laplace operators. In the $\mathbb{Q}$-rank-one case, they are the same as $\Delta_T$ but different in the higher rank cases. These other operators seem to have nicer geometric properties.

**Remark 3.3.3.** Let $L^2_T(\Gamma\backslash X,\sigma)$ be the closure of $H^1_T(\Gamma\backslash X,\sigma)$ in $L^2(\Gamma\backslash X,\sigma)$. This subspace can be described by Arthur's truncation operator $\Lambda^T$, [2, p. 89], [39], [35, pp. 487–489]. Let $\mathbf{P}_1,\cdots,\mathbf{P}_p$ be the set of representatives of $\Gamma$-conjugacy classes of proper rational parabolic subgroups as in §3.2. For each $\mathbf{P}_i$, let $\chi_i$ be the characteristic function of ${}^+\mathfrak{a}_{P_i}=\{H\in\mathfrak{a}_{P_i}\mid\langle H,V\rangle\geq 0 \text{ for all } V\in\mathfrak{a}^+_{P_i}\}$, which is the obtuse cone dual to the dominant cone $\mathfrak{a}^+_{P_i}$. Then for a large truncation parameter $T$ as in §3.2, if $f$ is continuous and belongs to $L^2(\Gamma\backslash X,\sigma)$, define

$$\Lambda^T_{P_i}f(x)=\sum_{\gamma\in\Gamma_{P_i}\backslash\Gamma}\chi_i(H_{P_i}(\gamma x)-I_{P_i}(T))f_{P_i}(\gamma x),$$

where $I_{P_i}(T)$ is the image of $T$ in $\mathfrak{a}_i$ (see Equation 1 in §3.2), and

$$\Lambda^T f(x)=f(x)+\sum_{i=1}^{p}(-1)^{\mathrm{rank}_{\mathbb{Q}}(\mathbf{P}_i)}\Lambda^T_{P_i}f(x).$$

A basic property of the truncation operator $\Lambda^T$ is that for every maximal proper rational parabolic subgroup $\mathbf{P}$, $(\Lambda^T f(x))_P(nza)=0$ for all $n\in N_P$, $z\in X_P$, $a\in A_{P,T}$, i.e., $\Lambda^T f(x)$ satisfies precisely the vanishing conditions in the definition of $H^1_T(\Gamma\backslash X,\sigma)$, hence $\Lambda^T f\in L^2_T(\Gamma\backslash X,\sigma)$;

furthermore, $\Lambda^T$ extends to an orthogonal projection on $L^2(\Gamma\backslash X,\sigma)$ [2, Remark, p. 92], and $L^2_T(\Gamma\backslash X,\sigma) = \Lambda^T L^2(\Gamma\backslash X,\sigma)$. Consequently, we can define $H^1_T(\Gamma\backslash X,\sigma)$ and hence the pseudo-Laplacian $\Delta_T$ by setting $H^1_T(\Gamma\backslash X,\sigma) = H^1(\Gamma\backslash X,\sigma)\cap \Lambda^T L^2(\Gamma\backslash X,\sigma)$. This gives a more direct relation between the pseudo-Laplacian $\Delta_T$ and Arthur's truncation operator $\Lambda^T$.

**3.4.** First we prove the upper bound for $N_T(\lambda)$ in Theorem 3.3.2:

**Proposition 3.4.1.**

$$\lim_{\lambda\to+\infty}\sup \frac{N_T(\lambda)}{\lambda^{n/2}} \le (4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)}\dim\sigma.$$

*Proof.* The proof is a combination of the precise reduction theory and the proof of [11, Theorem 1.1]. Let

$$\Gamma\backslash X = (\Gamma\backslash X)_T \cup \coprod_{i=1}^{p} \omega_i \times A_{P_i,T}$$

be the decomposition in Proposition 3.2.2. Let $Y_{0,k}$, $k\ge 1$, be a family of smooth compact submanifolds with boundary such that (1) $Y_{0,k}\subset Y_{0,k+1}$, and $\cup_{k=1}^{\infty} Y_{0,k} = \Gamma\backslash X$, (2) the interior of $Y_{0,k}$ contains

$$(\Gamma\backslash X)_T \cup \coprod \omega_i \times (A_{P_i,T}\cap A_{P_i}^{T_k}),$$

and $Y_{0,k}$ is contained in the interior of

$$(\Gamma\backslash X)_T \cup \coprod \omega_i \times (A_{P_i,T}\cap A_{P_i}^{T_{k+1}}),$$

where $T_k = kH_\rho + T = (k+t)H_\rho$.

For every pair of $i$ and $k$, let $Y_{i,k}$ be a smooth manifold with boundary in $\Gamma_{P_i}\backslash X$ such that (1) $Y_{i,k}$ is invariant under $N_{P_i}$, (2) the image of $Y_{i,k}$ in $\Gamma\backslash X$ contains $\omega_i\times A_{P_i,T} - Y_{0,k}$, and the images of $Y_{1,k},\cdots,Y_{P,k}$ in $\Gamma\backslash X$ together with $Y_{0,k}$ cover $\Gamma\backslash X$, (3) the image of $Y_{i,k}$ in $\Gamma\backslash X$ is contained in the complement of $Y_{0,k-1}$, and the image in the split component $A_{P_i}$ of points in $Y_{i,k}$ is contained in the complement of $A_{P_i}^{T_{k-1}}$ in $A_{P_i}$ and hence shrinks to infinity, and the image in $\Gamma_{M_{P_i}}\backslash X_{P_i}$ of points in $Y_{i,k}$ is bounded uniformly for $k$ (in fact contained in a fixed neighborhood of $(\Gamma_{M_{P_i}}\backslash X_{P_i})_T$).

For every $i = 1,\cdots,p$, the homogeneous bundle $\tilde{E}_\sigma$ on $X$ induces a bundle $E_\sigma$ on $\Gamma_{P_i}\backslash X$. Denote the space of $L^2$-sections of $E_\sigma$ by

$L^2(\Gamma_{P_i}\backslash X, \sigma)$, and the subspace for $Y_{i,k}$ by $L^2(Y_{i,k}, \sigma)$. Define the cuspidal subspace $H^1_{cus}(Y_{i,k}, \sigma)$ by

$$H^1_{cus}(Y_{i,k}, \sigma) = \{f \in H^1(Y_{i,k}, \sigma) \mid f_P = 0 \text{ for all proper } \mathbf{P} \supset \mathbf{P}_i\}.$$

Since $Y_{i,k}$ is invariant under $N_{P_i}$ and hence all $N_P$ for $\mathbf{P} \supset \mathbf{P}_i$, the cuspidal subspace is well-defined. The Dirichlet quadratic form defines a self-adjoint operator $\Delta_{i,N}$ on the closure of $H^1_{cus}(Y_{i,k}, \sigma)$ in $L^2(Y_{i,k}, \sigma)$, where the subscript $N$ stands for the Neumann boundary condition.

**Lemma 3.4.2.** *The spectrum of $\Delta_{i,N}$ is discrete and its spectral counting function $N_{i,k}(\lambda)$ satisfies the following bound:*

$$N_{i,k}(\lambda) \leq \varepsilon_i(k)\lambda^{\frac{n}{2}}, \quad \lambda \geq 1,$$

*where $\varepsilon_i(k) \to 0$ as $k \to +\infty$.*

*Proof.* This follows from the same proof of [11, Corollary 7.6]. One crucial point here is that the component of $Y_{i,k}$ is $A_{P_i}$ shrinks to infinity, while the component of $Y_{i,k}$ in $\Gamma_{M_{P_i}}\backslash X_{P_i}$ stays bounded as $k \to +\infty$.

**Remark 3.4.3.** As pointed out recently in [8], there is a possible gap in the construction of the cuspidal Neumann heat kernel in [11]. Briefly, the order of the three steps in constructing the cuspidal heat kernel of $X_{i,k}$ with the Neumann boundary condition needs to be changed to the following: (1) Average the heat kernel of $X$ over $\Gamma_{P_i}$ to get the heat kernel of $\Gamma_{P_i}\backslash X$. (2) Use the method of single layer potentials to adjust the heat kernel to satisfy the Neumann boundary condition. (3) Remove the constant term of the heat kernel in Step (2) to make it cuspidal.

In [11], Step (3) was performed before Step (2). This causes a problem since after the constant terms are removed, the heat kernel does not satisfy the bounds needed to carry out the method of single layer potentials. See [8] for details and also [37, pp. 332–334] for some related comments.

*Proof of Proposition 3.4.1.* Denote the counting function of the Neumann eigenvalues of $Y_{0,k}$ by $N_{0,k}(\lambda)$. Since $Y_{0,k}, Y_{1,k}, \cdots, Y_{p,k}$ form a covering of $\Gamma\backslash X$ and the image of $Y_{i,k}$ is contained in the complement of $(\Gamma\backslash X)_T$, we have an inclusion

$$H^1_T(\Gamma\backslash X, \sigma) \subset H^1(Y_{0,k}, \sigma) \oplus H^1_{cus}(Y_{1,k}, \sigma) \oplus \cdots H^1_{cus}(Y_{p,k}, \sigma).$$

The principle of Neumann bracketing implies that

$$N_T(\lambda) \leq N_{0,k}(\lambda) + \sum_{i=1}^{p} N_{i,k}(\lambda).$$

Since $Y_{0,k}$ is a compact submanifold with boundary,

$$\lim_{\lambda\to+\infty} \frac{N_{0,k}(\lambda)}{\lambda^{n/2}} = (4\pi)^{-n/2}\frac{\mathrm{vol}(Y_{0,k})}{\Gamma(\frac{n}{2}+1)} \leq (4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)},$$

and we get from Proposition 3.4.2 that

$$\lim_{\lambda\to+\infty}\sup \frac{N_T(\lambda)}{\lambda^{n/2}} \leq (4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)} + \sum_{i=1}^{p}\varepsilon_i(k).$$

Since $\varepsilon_i(k) \to 0$ as $k \to +\infty$, the upper bound for $N_T(\lambda)$ in Proposition 3.4.1 follows immediately.

**Remark 3.4.4.** In an early vision of this paper, the following slightly different proof of Proposition 3.4.1 was given, which turned out to be problematic, as explained below.

By the precise reduction theory, for each $T' \geq T$, we get a covering of $\Gamma\backslash X$ and hence a bound on $N_T(\lambda)$:

$$\text{(1)}\qquad \lim_{\lambda\to+\infty}\sup \frac{N_T(\lambda)}{\lambda^{n/2}} \leq (4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash X)_{T'}}{\Gamma(\frac{n}{2}+1)}\dim\sigma + c(T'),$$

where $c(T')$ comes from bounds on the counting functions of the Neumann cuspidal eigenvalues of the covering of the infinity: $\omega_i \times A_{P_i,T'}$, $i = 1, \cdots, p$ (or their slight enlargements). We claimed that since $A_{P_i,T'}$ shrinks towards infinity as $T' \to \infty$, $\lim_{T'\to\infty} c(T') = 0$, which would imply the Weyl upper bound for $N_T(\lambda)$. Hoffmann pointed out recently that this is not obvious. Actually, $c(T')$ is a product of two factors, one goes to zero and another to infinity. This can be seen through the example $\Gamma\backslash X = \Gamma_1\backslash\mathbf{H}^2 \times \Gamma_2\backslash\mathbf{H}^2$, where $\mathbf{H}^2$ is the upper half plane, and $\Gamma_1\backslash\mathbf{H}^2$, $\Gamma_2\backslash\mathbf{H}^2$ are non-compact and have finite volume. The reason is that when $P_i$ is not a minimal parabolic subgroup, as $T' \to \infty$, $\omega_i$ becomes unbounded, i.e., not enough constant terms are removed on $\omega_i \times A_{P_i,T'}$ in this upper bound. Briefly, $\omega_i$ is a $\Gamma_{N_{P_i}}\backslash N_{P_i}$ bundle over $(\Gamma_{M_{P_i}}\backslash X_{P_i})_{T'}$, and $(\Gamma_{M_{P_i}}\backslash X_{P_i})_{T'}$ develops cusps as $T' \to \infty$, and hence these new constant terms associated with the cusps of $\Gamma_{M_{P_i}}\backslash X_{P_i}$ (or rational parabolic subgroups of $\mathbf{M_{P_i}}$) need to be removed also. If we remove these new constant terms, then the argument is equivalent to the above proof of Proposition 3.4.1.

Unfortunately, this approach outlined here was also used in [21] to prove the Weyl upper bound for the pseudo-Laplacian of $\Gamma\backslash G$. The above discussions show that the arguments in [21, Proposition 4.3.2] need to be modified as indicated above or replaced by an analogue of the proof of Proposition 3.4.1 here. On the other hand, for a fixed $T' \geq T$, Equation (1) gives a bound $N_T(\lambda) \leq c(1+\lambda^{\frac{n}{2}})$. Such a bound (in fact, any polynomial bound) is sufficient for the proof of the trace class conjecture in [21, Theorem 1.1.2].

The proof of Proposition 3.4.1 combines the covering in [11, p. 243] and the decomposition in Proposition 3.2.2. Since the constant terms of functions in $H^1_T(\Gamma\backslash X, \sigma)$ do not vanish along all horospheres, unlike cuspidal functions, this combination seems necessary.

**3.5.** Next we prove the lower bound for $N_T(\lambda)$ in Theorem 3.3.2:

**Proposition 3.5.1.**

$$\lim_{\lambda\to+\infty}\inf \frac{N_T(\lambda)}{\lambda^{n/2}} \geq (4\pi)^{-n/2}\frac{\mathrm{vol}((\Gamma\backslash X)_T)}{\Gamma(\frac{n}{2}+1)}\dim\sigma.$$

*Proof.* Let $H^1_0((\Gamma\backslash X)_T, \sigma)$ be the Sobolev space with vanishing boundary values. Then it defines through the Dirichlet quadratic form a Laplacian $\Delta_D$ on $L^2((\Gamma\backslash X)_T, \sigma)$ satisfying the Dirichlet boundary condition. Let $N_D(\lambda)$ be the counting function of the eigenvalues of $\Delta_D$. Since $(\Gamma\backslash X)_T$ is a manifold with corners, it clearly satisfies the segment property (see [42, p. 256] for its definition and applications) and hence $N_D(\lambda)$ satisfies the Weyl law (see [47, Corollary 2.5]);

$$\lim_{\lambda\to+\infty}\frac{N_D(\lambda)}{\lambda^{n/2}} = (4\pi)^{-n/2}\frac{\mathrm{vol}((\Gamma\backslash X)_T)}{\Gamma(\frac{n}{2}+1)}\dim\sigma.$$

For every function $f$ in $H^1_0((\Gamma\backslash X)_T, \sigma)$, its extension by zero outside $(\Gamma\backslash X)_T$ gives a function $\tilde{f}$ in $H^1(\Gamma\backslash X, \sigma)$. We claim that $\tilde{f} \in H^1_T(\Gamma\backslash X, \sigma)$. In fact, for any point $x \in X$ which is mapped into the complement of $(\Gamma\backslash X)_T$ in $\Gamma\backslash X$ and any maximal rational parabolic subgroup $\mathbf{P}$, by the definition of $(\Gamma\backslash X)_T$ in §3.2, the orbit $N_P x$ of $x$ under $N_P$ is also projected into the complement of $(\Gamma\backslash X)_T$. This implies that $\tilde{f}$ satisfies the vanishing conditions for the definition of $H^1_T(\Gamma\backslash X, \sigma)$ in §3.3. This proves the claim. Then the mini-max principle implies that

$$N_T(\lambda) \geq N_D(\lambda),$$

and the lower bound for $N_T(\lambda)$ in the proposition follows from the Weyl law for $N_D(\lambda)$ above.

**Remark 3.5.2.** It was announced in [22, Theorem 2.4.1] that $N_T(\lambda)$ satisfies the lower Weyl bound also, i.e., $\mathrm{vol}((\Gamma\backslash X)_T)$ in the above proposition could be replaced by $\mathrm{vol}(\Gamma\backslash X)$. Actually, the arguments only worked in the $\mathbb{Q}$-rank-1 case, and for the higher rank case, they worked for a different pseudo–Laplacian (see §3.6 below for various pseudo-Laplacians in the higher rank case), though we believe that this sharp lower bound is true. One potential application of this lower bound for $N_T(\lambda)$ in Proposition 3.5.1 is to serve as an analogue of the Weyl–Selberg law as mentioned near the end of §1.2. Specifically, suppose we could show that all eigenfunctions of $\Delta_T$ are either cuspidal eigenfunctions of $\Delta$ or truncation by $\Lambda^T$ of some Eisenstein series, and show that the contribution coming from the continuous spectrum of $\Delta$ is of smaller order than $\lambda^{n/2}$, the analogue of Step 2 in §1.2, and that the residual discrete spectrum of $\Delta$ can be approximated well by the corresponding part of the spectrum of $\Delta_T$ (see Corollary 5.2.7 below for the rank-1 case). Then the Weyl upper bound of $N_T(\lambda)$ would imply the Weyl upper bound of $N_d(\lambda)$, and the lower bound for $N_T(\lambda)$ would give a bound

$$\lim_{\lambda\to+\infty}\inf\frac{N_d(\lambda)}{\lambda^{\frac{n}{2}}}\geq(4\pi)^{-\frac{n}{2}}\frac{\mathrm{vol}((\Gamma\backslash X)_T)}{\Gamma(\frac{n}{2}+1)}\dim\sigma.$$

Since $\lim_{T\to\infty}\mathrm{vol}((\Gamma\backslash X)_T)=\mathrm{vol}(\Gamma\backslash X)$, by letting $T\to\infty$, we would get the desired Weyl lower bound and hence the Weyl law for $N_d(\lambda)$.

**Remark 3.5.3.** As mentioned in the previous remark, in the $\mathbb{Q}$-rank-1 case, the pseudo-Laplacian $\Delta_T$ satisfies the Weyl law. In fact, this Weyl law can be used to derive the Weyl–Selberg law. These results, together with discussions of other pseudo-Laplacians in the higher rank, will be treated elsewhere.

**Remark 3.5.4.** In [21, Theorem 1.2.1], it is claimed as in an earlier version of this paper that the pseudo-Laplacian $\Delta_T$ for $\Gamma\backslash G$ satisfies the Weyl lower bound also. As pointed out in Remark 3.5.2, the arguments there actually only work for the $\mathbb{Q}$-rank-1 case, and in the higher rank case, the lower bound in [21, Proposition 4.4.1] should be replaced by a weaker one that is obtained by substituting $\mathrm{vol}((\Gamma\backslash G)_T)$ for $\mathrm{vol}(\Gamma\backslash G)$. Since this lower bound for the pseudo-Laplacian of $\Gamma\backslash G$ is not used in the rest of that paper, in particular, the proof of [21, Theorem 1.1.2], this change does not affect the paper.

**3.6.** In this section, we introduce another pseudo-Laplace operator $\hat{\Delta}_T$. This operator is more geometric and can be generalized to other spaces, and hence is of independent interests, though not used in this paper.

By the precise reduction theory in Proposition 3.2.2, $\Gamma\backslash X$ admits the following disjoint decomposition $\Gamma\backslash X = (\Gamma\backslash X)_T \cup \coprod_{i=1}^{p} \omega_i \times A_{P_i,T}$, where $\omega_i$ is a compact manifold with corners invariant under $N_{P_i}$. Using this decomposition, we define a subspace $\hat{H}_T^1(\Gamma\backslash X, \sigma)$ of $H^1(\Gamma\backslash X, \sigma)$ as follows: $\hat{H}_T^1(\Gamma\backslash X, \sigma) = \{f \in H^1(\Gamma\backslash X, \sigma) \mid$ for every $i$, points $w \in \omega_i$, $a \in A_{P_i,T}$, and all $\mathbf{P}$ containing $\mathbf{P}_i$, $f_P(wa) = 0.\}$ As above, the Dirichlet quadratic form defines a self-adjoint operator $\hat{\Delta}_T$ in the closure of $\tilde{H}_T^1(\Gamma\backslash X, \sigma)$ in $L^2(\Gamma\backslash X, \sigma)$.

It is conceivable that $\hat{\Delta}_T$ has a discrete spectrum, and the counting function $\hat{N}_T(\lambda)$ of its spectrum satisfies the Weyl law. Though each $\omega_i \times A_{P_i,T}$ is not a smooth manifold with boundary, it is a manifold with corners, and it seems that the method of single layers and the procedure of [11] plus the modification in [8] can still be applied to show that the spectrum of $\hat{\Delta}_T$ is discrete and satisfies the Weyl upper bound. On the other hand, since the removal of the constant terms for functions in $\hat{H}_T^1(\Gamma\backslash X, \sigma)$ is simpler, it can be shown that $\hat{\Delta}_T$ also satisfies the Weyl lower bound, and also that $\hat{\Delta}_T$ depends real analytically on $t$, where $T = tH_\rho$.

The pseudo-Laplacian $\hat{\Delta}_T$ can be generalized to other spaces, for example, compact perturbations of $\Gamma\backslash X$ and manifolds with corners whose metrics respect the corner structure. In other words, $\hat{\Delta}_T$ depends only locally on the geometry at infinity. On the other hand, it is not obvious whether the pseudo-Laplacian $\Delta_T$ can be generalized to other spaces.

To avoid the problems with the corners of $\omega_i \times A_{P_i,T}$, there is another way to define a pseudo-Laplacian. By slightly enlarging $\omega_i \times A_{P_i,T}$ and rounding off the corners, we can get a covering of $\Gamma\backslash X$ by smooth manifolds with boundary. Then we can require the suitable constant terms to vanish near infinity to define a cut-off Sobolev space and hence a pseudo-Laplacian. Such a pseudo-Laplacian can also be shown to satisfy the Weyl law.

In the $\mathbb{Q}$-rank-1 case, the three pseudo-Laplacians coincide. But they are different in the higher rank case. One problem with the latter two pseudo-Laplacians is that they are not connected to Arthur's truncation operator $\Lambda^T$. These pseudo-Laplacians and their properties will be discussed elsewhere.

**Remark 3.6.1.** This operator $\hat{\Delta}_T$ was not introduced in the early version of this paper, but $\Delta_T$ was confused with $\hat{\Delta}_T$ in several places.

**3.7.** In order to apply the Weyl upper bound for $\Delta_T$ in Theorem 3.3.2 to study the discrete spectrum of $\Delta$, we need to understand the relation between $\Delta$ and $\Delta_T$.

First we recall the domain of $\Delta_T$ [35, p. 490]. Let $H^{-1}(\Gamma\backslash X, \sigma)$ denote the space of distributions in $\mathcal{D}'(\Gamma\backslash X, \sigma)$ that extend to continuous linear functionals on $H^1(\Gamma\backslash X, \sigma)$.

**Lemma 3.7.1.** *The domain of $\Delta_T$ consists of all $\varphi \in H^1_T(\Gamma\backslash X, \sigma)$ such that there exists a distribution $D \in H^{-1}(\Gamma\backslash X, \sigma)$ which vanishes on $H^1_T(\Gamma\backslash X, \sigma)$ and $\Delta\varphi - D$ belongs to the closure of $H^1_T(\Gamma\backslash X, \sigma)$ in $L^2(\Gamma\backslash X, \sigma)$. Then $\Delta_T\varphi = \Delta\varphi - D$.*

**Lemma 3.7.2.** *Assume that $\lambda \in \mathrm{Spec}_{cus}(\Delta)$ and $\varphi \in L^2_{cus}(\Gamma\backslash X, \sigma)$ with $\Delta\varphi = \lambda\varphi$. Then $\varphi \in \mathrm{Dom}(\Delta_T)$ and $\Delta_T\varphi = \lambda\varphi$. In particular, $\mathrm{Spec}_{cus}(\Delta) \subset \mathrm{Spec}(\Delta_T)$.*

*Proof.* Since $\Delta\varphi = \lambda\varphi \in H^1_T(\Gamma\backslash X, \sigma)$, the distribution $D$ in Lemma 3.7.1 is equal to zero, and hence $\Delta_T\varphi = \lambda\varphi$.

**3.8.** In order to study other eigenfunctions of $\Delta_T$ and their connection with the spectrum of $\Delta$, we decompose $\Delta_T$ according to the decomposition of $L^2(\Gamma\backslash X, \sigma)$ in Lemma 2.5.3.

**Lemma 3.8.1.** *For every cuspidal pair $(\mathcal{C}, \mu)$ as in Lemma 2.5.3, define $H^1_{T,\mathcal{C},\mu}(\Gamma\backslash X, \sigma) = H^1_T(\Gamma\backslash X, \sigma) \cap L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$. Then $H^1_T(\Gamma\backslash X, \sigma)$ is equal to the completed sum in $H^1(\Gamma\backslash X, \sigma)$ of $H^1_{T,\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ over all cuspidal pairs $(\mathcal{C}, \mu)$.*

*Proof.* By Lemma 2.5.3, any $f \in H^1_T(\Gamma\backslash X, \sigma)$ can be decomposed into $f = \sum_{\mathcal{C},\mu} f_{\mathcal{C},\mu}$, where $f_{\mathcal{C},\mu} \in L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$. Since

$$\int_{\Gamma\backslash X} |\bigtriangledown f|^2 = \sum_{\mathcal{C},\mu} \int_{\Gamma\backslash X} |\bigtriangledown f_{\mathcal{C},\mu}|^2,$$

we have $f_{\mathcal{C},\mu} \in H^1(\Gamma\backslash X, \sigma)$.

For any cuspidal pair $(\mathcal{C}, \mu)$ and any rational parabolic subgroup $\mathbf{P}$, either the constant term $(f_{\mathcal{C},\mu})_P$ of $f_{\mathcal{C},\mu}$ along $P$ vanishes, or for any fixed $a \in A_P$, $(f_{\mathcal{C},\mu})_P(za)$ is a cuspidal eigenfunction on $\Gamma_{M_P}\backslash X_P$ of eigenvalue $\mu$. This implies that if $(\mathcal{C}_1, \mu_1) \neq (\mathcal{C}_2, \mu_2)$, $(f_{\mathcal{C}_1,\mu_1})_P \neq 0$, and

$(f_{\mathcal{C}_2,\mu_2})_P \neq 0$, then $(f_{\mathcal{C}_1,\mu_1})_P$ and $(f_{\mathcal{C}_2,\mu_2})_P$ are linearly independent as functions on $\Gamma_{M_P}\backslash X_P$. Therefore, for $a \in A_{P,T}$, $f_P(za) = 0$ implies that $(f_{\mathcal{C},\mu})_P(za) = 0$, and hence $f_{\mathcal{C},\mu} \in H^1_{T,\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ for all $(\mathcal{C}, \mu)$. This completes the proof. q.e.d.

For every cuspidal pair $(\mathcal{C}, \mu)$, the subspace $H^1_{T,\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ defines a self-adjoint operator denoted by $\Delta_{T,\mathcal{C},\mu}$ on the closure of $H^1_{T,\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$. We note that when the parabolic subgroups in $\mathcal{C}$ are of rank-one, this operator $\Delta_{T,\mathcal{C},\mu}$ is the operator $\Delta_t$ defined in [35, p. 500]. The above lemma shows that $\Delta_T$ is the direct sum of these operators $\Delta_{T,\mathcal{C},\mu}$.

**3.9.** When $\mathcal{C} = \{\mathbf{G}\}$, Lemma 3.7.2 shows that $\Delta_{T,\mathcal{C},\mu} = \Delta$ on $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$. In this subsection, we study the eigenfunction of $\Delta_{T,\mathcal{C},\mu}$ when the parabolic subgroups in $\mathcal{C}$ are of rank-one. Fix an association class $\mathcal{C}$ of rank-one parabolic subgroups and a cuspidal eigenvalue $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$.

To characterize the eigenfunctions of $\Delta_T$, we need the truncation operator $\Lambda^T$ [2, p. 89], [39] as recalled in Remark 3.3.3. For functions in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$, $\Lambda^T$ can be simplified as follows. Let $\mathbf{P}_1, \cdots, \mathbf{P}_m$ be a set of representatives of $\Gamma$-conjugacy classes of maximal rational parabolic subgroups. Then

$$\Lambda^T f(x) = f(x) - \sum_{i=1}^{m} \Lambda^T_{P_i} f(x).$$

Recall from [39, p. 370] that a function $f \in L^\infty_{loc}(\Gamma\backslash X, \sigma)$ is of moderate growth if for any invariant differential operator $D$, any rational parabolic subgroup $\mathbf{P}$ and any compact subset $\omega \subset N_P \times X_P$, there exist $\Lambda \in \mathfrak{a}_P^*$ independent of $D$ and a positive constant $c = c(D)$ such that for any $w \in \omega$, $a \in A_{P,T}$,

$$|Df(wa)| \leq c a^\Lambda. \tag{1}$$

A basic property of the truncation operator $\Lambda^T$ is that if a function $f$ on $\Gamma\backslash X$ is of moderate growth, then $\Lambda^T f$ decays rapidly at infinity, and in particular belongs to $L^2(\Gamma\backslash X, \sigma)$ (see [39, Theorem 5.2], [2, Lemma 1.4]).[6]

[6]If $f$ satisfies an equation $\Delta f = \lambda f$ on $\Gamma\backslash X$ and has an upper bound $|f(wa)| \leq c a^\Lambda$ on every Siegel set $\omega A_{P,T}$, then the elliptic theory implies that Inequality (1) above also holds for all other invariant differential operators $D$, and hence $f$ is of moderate growth.

Denote the unique unit vector in $\mathfrak{a}_{P_i}^{*+}$ by $\Lambda_i$. For any

$$\Phi = (\Phi_1, \cdots, \Phi_m),$$

where $\Delta\Phi_i = \mu\Phi_i$, $\Phi_i \in L^2_{cus}(\Gamma_{M_i}\backslash X_{P_i}, \sigma_{M_{P_i}})$, $s \in \mathbb{C}$, define

$$E(\Phi, s) = \sum_{i=1}^{m} E(P_i, \Phi_i, s\Lambda_i), \quad s \in \mathbb{C}.$$

**Lemma 3.9.1** ([35, Lemma 3.14]). *Let $E(\Phi, s)$ be an Eisenstein series defined above for $\Phi \neq 0$. If for a fixed $s$ and all $i = 1, \cdots, m$, $E_{P_i}(\Phi, s, az) = 0$ when $a = \exp I_{P_i}(T)$ and $z \in X_P$, then $\Lambda^T E(\Phi, s)$ is a nonzero eigenfunction of $\Delta_{T,\mathcal{C},\mu}$ with eigenvalue $|\rho_P|^2 - s^2 + \mu$, where $|\rho_P|$ is equal to the norm of the half sum of the roots in $\Sigma(P, A_P)$ with multiplicity for any $\mathbf{P}$ in $\mathcal{C}$.*

**Remark 3.9.2.** For a smooth function $f$ of moderate growth, $\Lambda^T f$ does not belong to $H^1_T(\Gamma\backslash X, \sigma)$ in general. In the above lemma, the vanishing condition on the constant terms implies that

$$\Lambda^T E(\Phi, s) \in H^1_T(\Gamma\backslash X, \sigma).$$

## 4. Perturbation of $\Delta_T$.

**4.1.** In this section, we show that $\Delta_T$ is a continuous family of self-adjoint operators in $T$. This allows us to study branches of eigenvalues of $\Delta_T$. The results in this section are not used in the proofs of Theorem 1.1.2 and 1.1.3, but the behavior of the eigenvalues of $\Delta_T$ as $T$ varies is interesting in itself.

**Lemma 4.1.1.** *For $T = tH_\rho$, $t \gg 0$, $\Delta_T$ depends continuously on $t$, and hence the eigenvalues of $\Delta_T$ depend continuously on $t$.*

*Proof.* Since $\Delta_T$ is defined on the closure of $H^1_T(\Gamma\backslash X, \sigma)$ through the Dirichlet quadratic form, the eigenvalues of $\Delta_T$ are given by the Rayleigh quotient and the mini-max principle. Since the vanishing conditions defining $H^1_T(\Gamma\backslash X, \sigma)$ depend continuously on $T$, the eigenvalues of $\Delta_T$ also change continuously.

**Remark 4.1.2.** In view of the close connection between the pseudo-Laplace operator $\Delta_T$ and Arthur's truncation operator $\Lambda^T$, and the fact that for an invariant integral operator of compact support $K(x, y)$ the trace of the truncated kernel $\Lambda^T\Lambda^T K(x,y)$ is a polynomial in $T$ for *sufficiently large* $t$ [2, p. 88], it is conceivable that $\Delta_T$ depends real analytically on $t$, where $T = tH_\rho$.

**4.2.** We now the study convergence behavior of the continuous branches of eigenvalues of $\Delta_T$ as $t \to +\infty$. Since $H^1_T(\Gamma\backslash X, \sigma)$ converges to $H^1(\Gamma\backslash X, \sigma)$ as $t \to +\infty$ and the eigenvalues of $\Delta_T$ are given by the mini-max principle applied to the Rayleigh quotient associated with the Dirichlet quadratic form $D$ on $H^1_T(\Gamma\backslash X, \sigma)$, it is conceivable that $\mathrm{Spec}(\Delta_T)$ converges to $\mathrm{Spec}(\Delta)$ as $t \to +\infty$.

**Remark 4.2.1.** There are different ways to define the convergence of $\mathrm{Spec}(\Delta_T)$ to $\mathrm{Spec}(\Delta)$. A weak convergence is to treat them as subsets of $\mathbb{R}$. Then $\mathrm{Spec}(\Delta)$ is a union of finitely points and a half line going to the positive infinity (the most interesting eigenvalues are embedded in this half line, the continuous spectrum), while $\mathrm{Spec}(\Delta_T)$ is a union of discrete points. It can be shown that below the continuous spectrum in $\mathrm{Spec}(\Delta)$, the eigenvalues of $\mathrm{Spec}(\Delta_T)$ converge to the corresponding eigenvalues in $\mathrm{Spec}(\Delta)$, and the points in $\mathrm{Spec}(\Delta_T)$ become dense in the half line in $\mathrm{Spec}(\Delta)$. (See [23] for some related discussions about the spectral degeneration.) A more interesting question concerns the behavior of eigenfunctions, in particular those that might converge to the eigenfunctions whose eigenvalues are embedded in the continuous spectrum.

By Lemma 3.7.2, $\mathrm{Spec}_{cus}(\Delta) \subset \mathrm{Spec}(\Delta_T)$. This implies that for any $\lambda_j \in \mathrm{Spec}_{cus}(\Delta)$, there is a constant branch $\lambda_j(T) = \lambda_j \in \mathrm{Spec}(\Delta_T)$. On the other hand, the relation between $\mathrm{Spec}_{res}(\Delta)$ and $\mathrm{Spec}(\Delta_T)$ is not entirely clear.

Let $\mathcal{C}$ be an association class of rational parabolic subgroups of rank-1, and $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$. Consider the restriction $\Delta_{\mathcal{C},\mu}$ of $\Delta$ to $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ and the corresponding component $\Delta_{T,\mathcal{C},\mu}$ of $\Delta_T$ in §3.8.

**Lemma 4.2.2.** *The continuous spectrum of* $\Delta_{\mathcal{C},\mu}$ *on* $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ *is* $[\mu + |\rho_P|^2, +\infty)$*, and the residual discrete spectrum is contained in* $[\mu, \mu + |\rho_P|^2)$*, in particular, is outside the continuous spectrum, where* $|\rho_P|$ *is the norm of the half sum of the roots in* $\Sigma(P, A_P)$ *with multiplicity for any* $\mathbf{P} \in \mathcal{C}$.

*Proof.* This follows from the spectral decomposition of $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ in Remark 2.6.5. More precisely, $E(\Phi, \Lambda)$ satisfies

$$\Delta E(\Phi, \Lambda) = (\mu + |\rho_P|^2 - \Lambda^2) E(\Phi, \Lambda).$$

Since the continuous spectrum is spanned by $E(\Phi, i\Lambda)$, where $\Lambda$ is real, and the residual discrete spectrum by residues of $E(\Phi, \Lambda)$ at points in $(0, |\rho_P|]$, the lemma follows. q.e.d.

Since the residual discrete spectrum of $\Delta_{\mathcal{C},\mu}$ is outside the continuous spectrum and the set of eigenvalues of $\Delta_{T,\mathcal{C},\mu}$ converges to the spectrum of $\Delta_{\mathcal{C},\mu}$ as $t \to +\infty$, we get the following result.

**Proposition 4.2.3.** *Let $\lambda_1 \leq \cdots \leq \lambda_r$ be the residual eigenvalues in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ repeated with multiplicities, and $\lambda_1(t) \leq \cdots \leq \lambda_r(t)$ the first $r$ eigenvalues of $\Delta_{T,\mathcal{C},\mu}$. Then for $i = 1, \cdots, r$, $\lim_{t\to+\infty} \lambda_i(t) = \lambda_i$.*

It will be shown in §5 that except for at most one of them,

$$|\lambda_i - \lambda_i(t)| \leq |\rho_P|^2.$$

**Remark 4.2.4.** If $\mathcal{C}$ is an association class of rational parabolic subgroups of rank greater than one, then the residual discrete spectrum of $\Delta_{\mathcal{C},\mu}$ on $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ could be embedded in the continuous spectrum of $\Delta_{\mathcal{C},\mu}$. On the other hand, if all invariant differential operators are considered, then the residual discrete spectrum is not embedded in the continuous spectrum anymore.

## 5. Bounds on the rank-1 residual discrete spectrum

**5.1.** In this section, we show that the majority of the rank-one residual discrete spectrum can be approximated uniformly by the corresponding eigenvalues of $\Delta_T$ and use it to prove Theorem 1.1.2.

**5.2.** It follows from §3.9 that in studying eigenfunctions of $\Delta_T$, the key point is to understand the constant terms of Eisenstein series.

Let $\mathbf{P}$ be a maximal, i.e., $\mathbb{Q}$-rank-1, rational parabolic subgroup of $\mathbf{G}$, and $\mathcal{C}$ an association class containing $\mathbf{P}$. Denote the opposite group $\mathbf{N}_\mathbf{P}^-\mathbf{A}_\mathbf{P}\mathbf{M}_\mathbf{P}$ of $\mathbf{P}$ by $\mathbf{P}^-$, where the Lie algebra $\mathfrak{n}_\mathbf{P}^-$ of $\mathbf{N}_\mathbf{P}^-$ is spanned by the root spaces of the negative of the roots in $\Sigma(P, A_P)$. In the notation of §2.7, $\mathbf{P}^-$ corresponds to the negative of the positive chamber $\mathfrak{a}_P^+$. Then $\mathbf{P}^- \in \mathcal{C}$, and any group in $\mathcal{C}$ is $\mathbf{G}(\mathbb{Q})$-conjugate to either $\mathbf{P}$ or $\mathbf{P}^-$. And $\mathbf{P}^-$ is $\mathbf{G}(\mathbb{Q})$-conjugate to $\mathbf{P}$ if and only if $-1 \in W(\mathfrak{a}_P)$, i.e., $W(\mathfrak{a}_P)$ contains two elements.

Let $\mathbf{P}_1, \cdots, \mathbf{P}_r$ be a set of representatives of the $\mathbf{G}(\mathbb{Q})$-conjugacy classes in $\mathcal{C}$. Then $r = 1$ or 2. For any $1 \leq i \leq r$, let $\mathbf{P}_{i1}, \cdots, \mathbf{P}_{ir_i}$ be a set of representatives of the $\Gamma$-conjugacy classes in the $\mathbf{G}(\mathbb{Q})$-conjugacy class containing $\mathbf{P}_i$ as in §2.4. Then $\mathbf{P}_{il}$, $1 \le i \le r$, $1 \le l \le r_i$, are representatives of all the $\Gamma$-conjugacy classes in the association class $\mathcal{C}$.

Let $\Lambda_i$ be the unique unit vector in $\mathfrak{a}_i^{*+}$, $1 \le i \le r$. If $r = 1$, then $W(\mathfrak{a}_1) = \{\pm 1\}$. Define

$$C(z) = C_{11}(-1, z\Lambda_1), \quad z \in \mathbb{C}. \tag{1}$$

If $r = 2$, then $W(\mathfrak{a}_i) = \{1\}$, $i = 1, 2$, and $W(\mathfrak{a}_1, \mathfrak{a}_2) = \{s\}$. Define

$$C(z) = \begin{pmatrix} 0 & C_{12}(s^{-1}, z\Lambda_2) \\ C_{21}(s, z\Lambda_1) & 0 \end{pmatrix}. \tag{2}$$

For $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$, define $\mathcal{E}_{cus}(\mathcal{C}, \mu) = \oplus_{i=1}^r \mathcal{E}_{cus}(\mathcal{C}_i, \mu)$. Then

$$C(z) : \mathcal{E}_{cus}(\mathcal{C}, \mu) \to \mathcal{E}_{cus}(\mathcal{C}, \mu), \quad z \in \mathbb{C}. \tag{3}$$

By the functional equation in Lemmas 2.7.1 and 2.7.2, $C(z)$ satisfies the following equations:

$$C(z)C(-z) = Id, \quad C(z)^* = C(\overline{z}), \quad z \in \mathbb{C}. \tag{4}$$

For any $\Phi = (\Phi_i)_{i=1}^r \in \mathcal{E}_{cus}(\mathcal{C}, \mu)$, define

$$E(\Phi, z) = \sum_{i=1}^{r} E(P_i, \Phi_i, z\Lambda_i).$$

By Lemma 2.7.2, $C_{ii}(1, \Lambda) = Id$. Then it follows from Equation 2.4.(4) that for any $\mathbf{P}_{il} \in \mathcal{C}_i$, the constant term of $E(\Phi, z)$ along $\mathbf{P}_{il}$ is given by

$$\begin{aligned} E_{P_{il}}(\Phi, z, x) =& e^{(z\mathrm{Ad}(y_{il})\Lambda_i + \rho_{il})(H_{il}(x))}\Phi_{il} \\ &+ e^{(-z\mathrm{Ad}(y_{il})\Lambda_i + \rho_{il})(H_{il}(x))}(C(z)\Phi)_{il}(x), \end{aligned}$$

where $\Phi_i = (\Phi_{il})_{l=1}^{r_i}$, $C(z)\Phi = ((C(z)\Phi)_{il})$, $i = 1, \cdots, r, l = 1, \cdots, r_i$. For simplicity, define $t_{il}(x) = (\mathrm{Ad}(y_{il})\Lambda)(H_{il}(x))$. Since $\rho_{il}(H_{il}(x)) = |\rho_{il}| t_{il}(x)$ and $|\rho_{il}|$ are the same for different $i$ and $l$, we denote $|\rho_{il}|$ by $|\rho|$ and get

$$E_{P_{il}}(\Phi, z, x) = e^{(z+|\rho|)t_{il}(x)}\Phi_{il} + e^{(-z+|\rho|)t_{il}(x)}(C(z)\Phi)_{il}. \tag{5}$$

**Lemma 5.2.1.**

1. *The poles of $E(\Phi, z)$ are contained in the poles of $C(z)\Phi$.*

2. *The poles of $C(z)$ in the right half plane* $\mathrm{Re}(z) \geq 0$ *are simple, real and contained in the interval* $(0, |\rho|]$.

3. *The residual discrete spectrum in* $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ *is spanned by* $\mathrm{Res}_{z_0}(E(\Phi, z))$, *where* $\Phi \in \mathcal{E}_{cus}(\mathcal{C}, \mu)$, *and* $z_0 \in (0, |\rho|]$ *is a pole of* $C(z)$.

*Proof.* Parts (1) and (2) follow from Proposition 2.4.2, and Part (3) from Proposition 2.6.1 and Remark 2.6.2. q.e.d.

For $u \in (0, |\rho|]$, $C(u)^* = C(u)$, i.e., $C(u)$ is Hermitian symmetric and hence can be diagonalized. Let $u_1, \cdots, u_m \in (0, |\rho|]$ be the poles of $C(z)$ in the half plane $\mathrm{Re}(z) \geq 0$ and $d = \dim \mathcal{E}_{cus}(\mathcal{C}, \mu)$. Then we have the following [35, p. 485 and Proposition 3.6].

**Lemma 5.2.2.** *For* $u \in (0, |\rho|]$, *there exists an analytic family of bases* $\Phi_1(u), \cdots, \Phi_d(u)$ *of* $\mathcal{E}_{cus}(\mathcal{C}, \mu)$ *such that the following hold:*

1. *Each* $\Phi_k(u)$ *is an eigenfunction of* $C(u)$, $C(u)\Phi_k(u) = \lambda_k(u)\Phi_k(u)$.

2. *Every eigenvalue* $\lambda_k(u)$ *is a real valued analytic function in* $u \in (0, |\rho|] \setminus \{u_1, \cdots, u_m\}$.

3. *Near each pole* $u_j$ *of* $C(u)$, $j = 1, \cdots, m$, $\lambda_k(u)$ *has the following expansion*

$$\lambda_k(u) = \frac{\mu_{kj}}{u - u_j} + \sum_{i=0}^{\infty} a_{ki}(u - u_j)^i$$

*with* $\mu_{kj} \geq 0$. *In particular,* $\lambda_k(u)$ *is singular at* $u_j$ *if and only if* $\mu_{kj} > 0$.

*Proof.* The eigenfunctions $\Phi_k(u)$ are not mentioned explicitly in [35, Proposition 3.6]. But the existence of such analytic families of eigenfunctions is a part of Rellich's theorem. (See [26, Chap. 2, §4.5 and 6.2].) q.e.d.

For every $1 \leq k \leq d$, $1 \leq j \leq m$ with $\mu_{kj} > 0$, $E(\Phi_k(u_j), u)$ has a nonzero residue $\mathrm{Res}_{u_j} E(\Phi_k(u_j), u)$ at $u = u_j$ which is an eigenfunction of eigenvalue $\lambda_{k,j} = \mu + |\rho|^2 - u_j^2$, and these eigenfunctions generate the residual discrete subspace in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$.

For each $1 \leq k \leq d$, denote the poles of $\lambda_k(u)$ in $(0, |\rho|]$ by $u_{k,1} < \cdots < u_{k,m_k}$, which are the subsets of the poles $u_1, \cdots, u_m$ with $\mu_{kj} > 0$. Then for any $j = 1, \cdots, m_k$,

$$\lim_{u \to u_{k,j} \pm 0} \lambda_k(u) = \pm\infty.$$

In particular, if $m_k \geq 2$, then in each interval $(u_{k,j-1}, u_{k,j})$, $2 \leq j \leq m_k$, the value of $\lambda_k(u)$ changes from $+\infty$ to $-\infty$. Therefore, for any truncation parameter $T = tH_\rho$, there exists a point $u = u_{k,j}(t) \in (u_{k,j-1}, u_{k,j})$ such that

$$e^{(u+|\rho|)t_{il}(\exp I_{il}(T))} + \lambda_k(u)e^{(-u+|\rho|)t_{il}(\exp I_{il}(T))} = 0. \tag{1}$$

By Lemma 3.9.1, we get the following lemma.

**Lemma 5.2.3.** *For any $1 \leq k \leq d$ with $m_k \geq 2$, $2 \leq j \leq m_k$ and $u_{k,j}(t)$ above, $\Lambda^T E(\Phi_k(u_{k,j}(t)), u_{k,j}(t))$ is a nonzero eigenfunction of the pseudo-Laplacian $\Delta_T$ with eigenvalue $\lambda_{k,j}(t) = \mu + |\rho|^2 - u_{k,j}(t)^2$.*

The above discussions show that between every pair of poles $u_{k,j-1}, u_{k,j}$, there is an eigenfunction of $\Delta_T$. In fact, such an eigenfunction is unique.

**Proposition 5.2.4.** *For $2 \leq j \leq m_k$, there exists a unique solution $u_{k,j}(t)$ in $(u_{k,j-1}, u_{k,j})$ to the above Equation (1) when $t \gg 0$.*

*Proof.* Since $H^1_{T,\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ increases and converges to

$$H^1(\Gamma\backslash X, \sigma) \cap L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$$

as $t \to +\infty$, the mini-max principle implies that the eigenvalues of $\Delta_{T,\mathcal{C},\mu}$ are decreasing functions of $t$ and converge to the spectrum of $\Delta_{\mathcal{C},\mu}$. Let $N_{T,\mathcal{C},\mu}(\lambda)$ be the counting function of $\Delta_{T,\mathcal{C},\mu}$, and $N_{\mathcal{C},\mu}(\lambda)$ the counting function of $\Delta_{\mathcal{C},\mu}$. Then, for any $\lambda \leq \mu + |\rho|^2$, which is the bottom of the continuous spectrum of $\Delta_{\mathcal{C},\mu}$ (see Lemma 4.4.2),

$$N_{T,\mathcal{C},\mu}(\lambda) \leq N_{\mathcal{C},\mu}(\lambda). \tag{2}$$

Since $\lim_{u\to u_{k,1}-0} \lambda_{k,1}(u) = -\infty$, we can find a point $u_{k,1}(t) \in (0, u_{k,1})$ when $t \gg 0$ such that $\Lambda^T E(\Phi_k(u_{k,1}(t)), u_{k,1}(t))$ is an eigenfunction of $\Delta_T$. As shown above, Equation (1) has at least one solution in each interval $(u_{i,j-1}, u_{i,j})$, and each such solution produces an eigenfunction of $\Delta_{T,\mathcal{C},\mu}$. If there are two solutions to Equation (1) in some interval $(u_{i,j-1}, u_{i,j})$, the above inequality (2) will not hold for $\lambda$ slightly smaller than $\mu + |\rho|^2$. This proves the proposition. q.e.d.

As pointed out in the above proof, when $t \gg 0$, we can get $u_{k,1}(t) \in (0, u_{k,1})$ such that $\Lambda^T E(\Phi_k(u_{k,1}(t)), u_{k,1}(t))$ is an eigenfunction of $\Delta_T$. But it is not clear that such a point $u_{k,1}(t)$ should exist for all $t \geq t_0$.

**Lemma 5.2.5.** *For $1 \leq k \leq d$ and $2 \leq j \leq m_k$, the eigenfunction $(u_{k,j}(t) - u_{k,j})\Lambda^T E(\Phi_k(u_{k,j}(t)), u_{k,j}(t))$ of $\Delta_T$ converges to the residual eigenfunction $Res_{u_{k,j}} E(\Phi_k(u_{k,j}), u)$ as $t \to +\infty$.*

*Proof.* Clearly, when $t \to +\infty$, $u_{k,j}(t) \to u_{k,j}$, and hence $\lambda_{k,j}(t) \to \lambda_{k,j}$ and

$$(u_{j,k}(t) - u_{k,j})\Lambda^T E(\Phi_k(u_{k,j}(t)), u_{k,j}(t)) \to \mathrm{Res}_{u_{k,j}} E(\Phi_k(u_{k,j}), u).$$

q.e.d.

**Lemma 5.2.6.** *For every $1 \leq k \leq d$ and $2 \leq j \leq m_k$, the branch $\lambda_{k,j}(t)$ in Lemma 5.2.5 converging to $\lambda_{k,j}$ satisfies the following bound:*

$$\lambda_{k,j} \leq \lambda_{k,j}(t) \leq \lambda_{k,j} + |\rho|^2.$$

*Proof.* It follows from the fact that $u_{k,j}(t) \in (u_{k,j-1}, u_{k,j}) \subset [0, |\rho|)$.

**Corollary 5.2.7.** *For the residual eigenvalues $\{\lambda_k\}$ in $L^2_{\mathcal{C},\lambda}(\Gamma\backslash X, \sigma)$, except for at most $\dim \mathcal{E}_{cus}(\mathcal{C}, \lambda)$ of them, the branches $\lambda_k(t)$ above converging to $\lambda_k$ satisfy the following inequality:*

$$\lambda_k \leq \lambda_k(t) \leq \lambda_k + |\rho|^2.$$

Recall that $\mathcal{C}$ is an association class of rank-one rational parabolic subgroups of $\mathbf{G}$. Let $N_{res,\mathcal{C}}(\lambda)$ be the counting function of the discrete spectrum in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$, $N_{il,cus}(\lambda)$ the counting function of

$$L^2_{cus}(\Gamma_{M_{il}}\backslash X_{P_{il}}, \sigma_{M_{il}}),$$

and $N_{T,\mathcal{C}}(\lambda)$ the counting function of $\Delta_T$ restricted to the closure of

$$H^1_T(\Gamma\backslash X, \sigma) \cap L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$$

in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ (see §3.8). Then we have the following bound on $N_{res,\mathcal{C}}(\lambda)$.

**Proposition 5.2.8.** *For an association class $\mathcal{C}$ of rank-one,*

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

*Proof.* It follows from Corollary 5.2.7 and the fact that for any $\Phi \in \mathcal{E}_{cus}(\mathcal{C}, \mu)$, any residual eigenfunction induced from $E(\Phi, \Lambda)$ has an eigenvalue greater than $\mu$.

**Proposition 5.2.9.** *Let $N'_d(\lambda)$ be the counting function of the discrete spectrum in the subspace $L^2_{cus}(\Gamma\backslash X,\sigma)\oplus\sum_{\mathcal{C}} L^2_{\mathcal{C}}(\Gamma\backslash X,\sigma)$, where $\mathcal{C}$ runs over all association classes of rank-1. Then*

$$\lim_{\lambda\to+\infty}\sup\frac{N'_d(\lambda)}{\lambda^{n/2}}\le(4\pi)^{-n/2}\frac{\mathrm{vol}(\Gamma\backslash X)}{\Gamma(\frac{n}{2}+1)}\dim\sigma.$$

*Proof.* Let $\mathcal{C}_1,\cdots,\mathcal{C}_m$ be all the association classes of rank-one rational parabolic subgroups. Then

$$N'_d(\lambda)=N_{cus}(\lambda)+\sum_{i=1}^{m}N_{res,\mathcal{C}_i}(\lambda),$$

$$N_{cus}(\lambda)+\sum_{i=1}^{m}N_{T,\mathcal{C}_i}(\lambda)\le N_T(\lambda).$$

Here we have used the fact that $L^2_{cus}(\Gamma\backslash X,\sigma)$ is contained in the domain of $\Delta_T$ (Lemma 3.7.2) and orthogonal to $L^2_{\mathcal{C}_i}(\Gamma\backslash X,\sigma)$. For each $\mathcal{C}_i$, let $\mathbf{P}_{i1},\cdots,\mathbf{P}_{ir_i}$ be representatives of the $\Gamma$-conjugacy classes in $\mathcal{C}_i$. Then Proposition 5.2.8 implies that

$$N_{res,\mathcal{C}_i}(\lambda)\le N_{T,\mathcal{C}_i}(\lambda+|\rho|^2)+\sum_{l=1}^{r_i}N_{il,cus}(\lambda),$$

and hence

$$N'_d(\lambda)\le N_T(\lambda+|\rho|^2)+\sum_{i=1}^{m}\sum_{l=1}^{r_i}N_{il,cus}(\lambda).$$

Since $\dim X_{P_{il}}<\dim X=n$, Lemma 2.3.2 gives that for every pair $i,l$ above,

$$N_{il,cus}(\lambda)=O(\lambda^{\dim X_{P_{il}}/2})=o(\lambda^{n/2})$$

as $\lambda\to+\infty$. Then the bound for $N'_d(\lambda)$ follows from the Weyl upper bound in Theorem 3.3.2. q.e.d.

*Proof of Theorem 1.1.2.* If the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 1, then all proper rational parabolic subgroups are of rank-1, and hence all the residual discrete spectrum is of rank-1. Therefore, $N_d(\lambda)=N'_d(\lambda)$, and the Weyl upper bound on $N_d(\lambda)$ is given by Proposition 5.2.9.

If the $\mathbb{R}$-rank of $\mathbf{G}$ is equal to 2, there are two cases to consider. If $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 1, then it is a special case of the above paragraph. If the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 2, then only the minimal rational parabolic subgroups are of rank 2 and hence greater than one. For any minimal rational parabolic subgroup $\mathbf{P}$, $\dim X_P = 0$, since the rank of $X$ is equal to 2 and $\dim A_P = 2$. Let $\mathcal{C}$ be the association class containing $\mathbf{P}$. Then the residual discrete spectrum in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ is of finite dimension. This implies $N_d(\lambda) = N'_d(\lambda) + O(1)$, and the Weyl upper bound on $N_d(\lambda)$ also follows from Proposition 5.2.9.

## 6. Bounds on the poles of the rank-one scattering matrices

**6.1.** In this section, we bound the number of poles of the rank-one scattering matrix on the negative real line, which is needed to bound the poles of the higher rank scattering matrices in the bounded domain $\{\Lambda \in {}^+\mathfrak{a}^*_P \mid |\Lambda| \leq |\rho_P|\}$ in Proposition 2.6.1, i.e., the higher rank residual discrete spectrum.

**6.2.** Let $\mathcal{C}$ be an association class of rank-one rational parabolic subgroups, $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$, and $C(z) : \mathcal{E}_{cus}(\mathcal{C}, \mu) \to \mathcal{E}_{cus}(\mathcal{C}, \mu)$ the scattering matrix defined in §5.2. Then det $C(z)$ is a meromorphic function of $z$.

**Lemma 6.2.1** ([35, Theorem 5.10]). *The determinant* det $C(z)$ *is the quotient of two holomorphic functions of order less than* $n + 2$.

Let $u_1, \cdots, u_l \in (0, |\rho|]$ be the poles of det $C(z)$ in the right half-plane $\mathrm{Re}(z) \geq 0$, and let $\eta_k$, $k \geq 1$, be the poles of det $C(z)$ in the left half-plane $\mathrm{Re}(z) \leq 0$, both repeated with multiplicities. Denote $\dim \mathcal{E}_{cus}(\mathcal{C}, \mu)$ by $d$ as above. Fix a truncation parameter $T = tH_\rho$ with $t \gg 0$ as in §3.2. Using the functional equation for $C(z)$ and Hadamard's factorization theorem, Müller proved the following result for det $C(z)$.

**Proposition 6.2.2** ([35, Theorem 6.9]). *There exists a negative number* $a$ *such that*

$$\det\, C(z) = q^z \prod_{j=1}^{l} \frac{z + u_j}{z - u_j} \prod_{k \geq 1} \frac{z + \overline{\eta_k}}{z - \eta_k}, \quad z \in \mathbb{C},$$

*where* $q = e^{2d(t+1)+1+a}$, *and hence for* $v \in \mathbb{R}$,

$$\frac{d}{dz} \log(\det\, C(iv)) = \log q + \sum_{j=1}^{l} \frac{2u_j}{v^2 + u_j^2} + \sum_{k \geq 1} \frac{2\mathrm{Re}(\eta_k)}{\mathrm{Re}(\eta_k)^2 + (v - \mathrm{Im}(\eta_k))^2}.$$

**6.3.** We use the above factorization of det $C(z)$ to bound the number of its poles on the negative half of the real line.

**Lemma 6.3.1.** *Let $N_{T,\mathcal{C},\mu}(x)$ denote the counting function of the eigenvalues of $\Delta_{T,\mathcal{C},\mu}$ (see §3.8). Then for any $V > 0$,*

$$\begin{aligned}|\int_{-V}^{V} \frac{d}{dz} \log\det\, C(iv)dv| \leq& 2\pi N_{T,\mathcal{C},\lambda}(\mu + |\rho|^2 + V^2)\\ &+ (4t|\rho|V + 4\pi)\dim \mathcal{E}_{cus}(\mathcal{C},\mu).\end{aligned}$$

*Proof.* Recall that $T = tH_\rho$ is the truncation parameter. Define $C_t(z) = e^{-2zt|\rho|}C(z)$. By the functional equation of $C(z)$ in Equation 5.2.(4), $C(iv)$, $v \in \mathbb{R}$, is unitary and hence $C_t(iv)$ is also unitary. So $C_t(iv)$ can be diagonalized. Since $C_t(z)$ is holomorphic in a neighborhood of $\mathrm{Re}(z) = 0$, the regular perturbation theory in [26, Chap. 2, §4.5 and 6.2] implies that there exist real valued analytic functions $\beta_1(v), \cdots, \beta_d(v)$ such that $e^{i\beta_1(v)}, \cdots, e^{i\beta_d(v)}$ are the eigenvalues of $C(iv)$, $v \in \mathbb{R}$. Denote the corresponding analytic branches of eigenfunctions by $\Phi_1(v), \cdots, \Phi_d(v)$. Then

$$\frac{d}{dz}\log\det\, C(iv) = 2t|\rho|d + \frac{d}{dz}\log\det\, C_t(iv) = 2t|\rho|d + \sum_{j=1}^{d}\beta_j'(v),$$

and

$$(1)\qquad |\int_{-V}^{V}\frac{d}{dz}\log\det\, C(iv)dv| \leq 4t|\rho|Vd + \sum_{j=1}^{d}|\int_{-V}^{V}\beta_j'(v)dv|.$$

Each $\beta_j(v)$ is only determined up to $2\pi\mathbb{Z}$. Since $C_t(0)^2 = Id$, we can fix $\beta_j(v)$ by picking either $\beta_j(0) = 0$ or $\pi$. For every $j = 1, \cdots, d$, let $v_{j,1} < \cdots < v_{j,n_j}$ be the all the points in $[-V, V]$ such that $e^{i\beta_j(v)} = -1$. Then

$$\begin{aligned}|\int_{-V}^{v_{j,1}} \beta_j'(v)dv| \leq& 2\pi,\\ |\int_{v_{j,n_j}}^{V} \beta_j'(v)dv| \leq& 2\pi,\\ |\int_{v_{j,1}}^{v_{j,n_j}} \beta_j'(v)dv| \leq& 2\pi n_j,\end{aligned}$$

and hence

$$(2)\qquad |\int_{-V}^{V}\beta_j'(v)dv| \leq 4\pi + 2\pi n_j.$$

For each $v_{j,k}$, $1 \leq k \leq n_j$, $\Lambda^T E(\Phi_k(v_{j,k}), iv_{j,k})$ is a nonzero eigenfunction of $\Delta_{T,\mathcal{C},\mu}$ with eigenvalue $\mu+|\rho|^2+v_{j,k}^2$ as in Lemma 5.2.3. Since different points $v_{j,k}$, $1 \leq j \leq d$, $1 \leq k \leq n_j$, produce linearly independent eigenfunctions $\Lambda^T E(\Phi_k(v_{j,k}), iv_{j,k})$ of $\Delta_{T,\mathcal{C},\mu}$, we obtain

$$\sum_{j=1}^{d} 2\pi n_j \leq 2\pi N_{T,\mathcal{C},\mu}(\mu + |\rho|^2 + V^2).$$

Combined with the above inequalities (1) and (2), this completes the proof of the lemma.

**Proposition 6.3.2.** *For any $V > 0$, there exists a constant $b$ depending on $V$ such that the number of poles, counted with multiplicity, of* $\det C(z)$ *in $[-V, 0]$ is bounded by $b(N_{T,\mathcal{C},\mu}(\mu + |\rho|^2 + 1) + (t + t|\rho|) \dim \mathcal{E}_{cus}(\mathcal{C}, \mu) + 1)$.*

*Proof.* For any positive pole $u_j$ of $\det C(z)$,

$$\int_{-1}^{1} \frac{u_j}{u_j^2 + v^2} dv \leq \int_{-\infty}^{\infty} \frac{1}{1 + v^2} dv = \pi.$$

By Proposition 6.2.2, we get

$$\begin{aligned} &|\int_{-1}^{1} \log q + \sum_{k\geq 1} \frac{2\mathrm{Re}(\eta_k)}{\mathrm{Re}(\eta_k)^2 + (v - \mathrm{Im}(\eta))^2} dv| \\ &\qquad \leq 2l\pi + |\int_{-1}^{1} \frac{d}{dz} \log \det\, C(iv) dv|, \end{aligned}$$

where $l$ is the number of the poles of $\det C(u)$ in $(0, |\rho|]$ counted with multiplicity and is bounded by the dimension of the residual discrete spectrum in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ (see the discussion after Lemma 5.2.2). By Corollary 5.2.7, we get

$$l \leq N_{T,\mathcal{C},\mu}(\mu + |\rho|^2) + d,$$

where $d = \dim \mathcal{E}_{cus}(\mathcal{C}, \mu)$. Combined with Lemma 6.3.1, this implies

$$\begin{aligned} &|\int_{-1}^{1} \log q + \sum_{k\geq 1} \frac{2\mathrm{Re}(\eta_k)}{\mathrm{Re}(\eta_k)^2 + (v - \mathrm{Im}(\eta))^2} dv| \\ &\qquad \leq 4\pi N_{T,\mathcal{C},\mu}(\mu + |\rho|^2 + 1) + (4t|\rho| + 5\pi)d. \end{aligned}$$

If $q \geq 1$, then $0 \leq \log q \leq 2(t+1)d+1$, where we have used the fact that $a \leq 0$, and hence

$$
\begin{aligned}
&|\int_{-1}^{1} \sum_{k\geq 1} \frac{2\mathrm{Re}(\eta_k)}{\mathrm{Re}(\eta_k)^2+(v-\mathrm{Im}(\eta))^2} dv| \\
(1) \qquad &\leq 4\pi N_{T,\mathcal{C},\mu}(\mu+|\rho|^2+1) \\
&\quad + (4t|\rho|+4t+4+5\pi)d+2.
\end{aligned}
$$

If $0 < q < 1$, then $\log q < 0$. Since every term

$$\int_{-1}^{1} \frac{2\mathrm{Re}(\eta_k)}{\mathrm{Re}(\eta_k)^2+(v-\mathrm{Im}(\eta))^2} dv$$

is negative, the same inequality (3) holds.

Let $m$ be the number of poles of $\det C(z)$ in the interval $[-V,0]$. Then

$$
\begin{aligned}
m\int_{-1/V}^{1/V} \frac{1}{1+v^2} dv &\leq \sum_{\eta_k\in\mathbb{R}, -V\leq \eta_k\leq 0} \int_{-1}^{1} \frac{|\eta_k|}{\eta_k^2+v^2} dv \\
&\leq \sum_{k\geq 1} \int_{-1}^{1} \frac{2|\mathrm{Re}(\eta_k)|}{\mathrm{Re}(\eta_k)^2+(v-\mathrm{Im}(\eta))^2} dv.
\end{aligned}
$$

Together with the above inequality (3), we get

$$m \leq b(N_{T,\mathcal{C},\mu}(\mu+|\rho|^2+1)+(t+t|\rho|)\dim \mathcal{E}_{cus}(\mathcal{C},\mu)+1),$$

where $b$ is a constant depending on $V$. This completes the proof.

**6.4.** Next we estimate the poles of the scattering matrix

$$C(z) : \mathcal{E}_{cus}(\mathcal{C},\mu) \to \mathcal{E}_{cus}(\mathcal{C},\mu)$$

in $[-V,V]$ counted with multiplicity and the rank of the residue.

The functional equation in Equation 5.2.(4) shows that for $v \in [-V,0]$, $C(v)$ is Hermitian, and hence there exist meromorphic functions $\lambda_1(v), \cdots, \lambda_d(v)$, which are eigenvalues of $C(v)$, where $d = \dim \mathcal{E}_{cus}(\mathcal{C},\mu)$.

**Proposition 6.4.1.** *For any $V > 0$, the number of the poles of $\lambda_1(v), \cdots, \lambda_d(v)$ in $[-V,0]$, counted with multiplicity, is bounded by $b'(N_{T,\mathcal{C},\mu}(\mu+|\rho|^2+1)+(t+t|\rho|)d+1)$, where $b'$ is a constant depending on $V$, and hence the number of poles of the scattering matrix $C(s)$ in $[-V,V]$ is bounded by*

$$b''(N_{T,\mathcal{C},\mu}(\mu+|\rho|^2+1)+(t+t|\rho|)d+1),$$

*where $b''$ is a constant depending on $V$.*

*Proof.* By definition, det $C(v) = \prod_{j=1}^{d} \lambda_j(v)$. For any pole $\eta_k$ of $C(v)$ in $[-V, 0]$, if none of $\lambda_j(v)$ vanishes at $\eta_k$, then the order of $C(v)$ at $\eta_k$ is equal to the sum of the order of poles of $\lambda_i(v)$ at $\eta_k$. On the other hand, if some $\lambda_j(\eta_k) = 0$, then the functional equation of $C(z)$ implies that $\lambda_j(v)$ has a pole at $-\eta_k > 0$. The pole $-\eta_k$ of $\lambda_j(u)$ is simple and gives rise to a residual eigenfunction $\mathrm{Res}_{-\eta_k} E(\Phi_j(\eta_k), v)$ in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$. This implies that $\lambda_j(v)$ has a simple zero at $\eta_k$. Since different such poles $-\eta_k$ produce linearly independent residual eigenfunctions, the sum of the number of the poles of $\lambda_j(v)$, $j = 1, \cdots, d$, in $[-V, 0]$ is bounded by the number of the poles of the determinant $\det C(v)$ in $[-V, 0]$ and the dimension of the residual discrete spectrum in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$. By Corollary 5.2.7, the latter is bounded by

$$N_{T,\mathcal{C},\mu}(\mu + |\rho|^2) + \dim \mathcal{E}_{cus}(\mathcal{C}, \mu). \tag{1}$$

Combined with Proposition 6.3.2, this gives the desired bound on the number of poles of $C(v)$ in $[-V, 0]$.

On the other hand, the number of the poles of $C(v)$ in $[0, V]$ is bounded by the dimension of the residual discrete spectrum, which is bounded by the number in Equation (1) as mentioned above. Combined with the previous paragraph, this completes the proof of the proposition.

## 7. Bounds on the higher rank residual discrete spectrum and proof of Theorem 1.1.3

**7.1.** In this section, we shall complete the proof of Theorem 1.1.3 by using the bounds on the number of poles of rank-1 scattering matrices in Proposition 6.4.1 and the factorization of the scattering matrices in §2.7 to bound the number of singular hyperplanes of the higher rank scattering matrices and hence the counting function of the higher rank residual discrete spectrum.

**7.2.** For any rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, denote the counting function of the cuspidal spectrum of $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ by $N_{cus,P}(\lambda)$, and the counting function of the pseudo-Laplacian $\Delta_T$ for $L^2(\Gamma_{M_P}\backslash X_P, \sigma_{M_P})$ by $N_{T,P}(\lambda)$. For an association class $\mathcal{C}$, the counting function of the cuspidal spectrum of $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ is denoted by $N_{cus,\mathcal{C}}(\lambda)$, and the corresponding counting function of the pseudo-laplacian $\Delta_T$ by $N_{T,\mathcal{C}}(\lambda)$.

Let $\mathcal{C}$ be an association class of rational parabolic subgroups of rank greater than or equal to 2. As in §2.4, let $\mathcal{C}_1, \cdots, \mathcal{C}_r$ be the $\mathbf{G}(\mathbb{Q})$-

conjugacy classes in $\mathcal{C}$. Then for every $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$, $i, j \in \{1, \cdots, r\}$, $s \in W(\mathfrak{a}_i, \mathfrak{a}_j)$, there is a scattering matrix

$$C_{ij}(s, \Lambda) : \mathcal{E}_{cus}(\mathcal{C}_i, \mu) \to \mathcal{E}_{cus}(\mathcal{C}_j, \mu), \quad \Lambda \in \mathfrak{a}_i^* \otimes \mathbb{C}.$$

**Proposition 7.2.1.** *For every $\lambda > 0$, and an association $\mathcal{C}$ of rank greater than or equal to 2 the total number of singular hyperplanes of all the scattering matrices $C_{ij}(s, \Lambda)$ acting on $\mathcal{E}_{cus}(\mathcal{C}_i, \mu)$ with $\mu \leq \lambda$ which meet the bounded region $\{\Lambda \in \mathfrak{a}_i^* \otimes \mathbb{C} \mid |\Lambda| \leq |\rho_P|\}$, is bounded by*

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

*where $\mathbf{Q}$ runs over a set of representatives of $\Gamma$-conjugacy 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 for every $\mathbf{Q}$ in the first sum $'\mathbf{P}$ runs over a set of representatives of $\Gamma_{M_Q}$-conjugacy classes of rank-1 rational parabolic subgroups of $\mathbf{M_Q}$, and $c$ is a constant depending only on $\mathbf{G}$ and $T$.*

*Proof.* For any $\mu \in \mathrm{Spec}_{cus}(\mathcal{C})$ with $\mu \leq \lambda$, we first bound the singular hyperplanes of the scattering matrices on $\mathcal{E}_{cus}(\mathcal{C}, \mu)$. We use the notation in §2.7. Let $A$ be the split component of a parabolic subgroup $\mathbf{P} \in \mathcal{C}$. Then Lemma 2.7.4 shows that to bound the number of poles of the scattering matrices for the parabolic subgroups in $\mathcal{C}$, it suffices to bound the poles of $C_{ji}(1, \Lambda)$ in the region $\{\Lambda \in \mathfrak{a}^* \otimes \mathbb{C} \mid |\Lambda| \leq |\rho_P|\}$, where $C_i, C_j$ are adjacent chambers of $\mathfrak{a}$.

Let $\mathbf{P}_i, \mathbf{P}_j$ be the rational parabolic subgroups corresponding to $C_i, C_j$, and $\mathbf{Q}$ the parabolic subgroup containing both $\mathbf{P}_i, \mathbf{P}_j$ with

$$\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) = \mathrm{rank}_{\mathbb{Q}}(\mathbf{P}_i) - 1.$$

Then $\mathbf{P}_i, \mathbf{P}_j$ determine parabolic subgroups $'\mathbf{P}_i, '\mathbf{P}_j$ of $\mathbf{M_Q}$ as in §2.7. Lemma 2.7.5 implies that the poles of $C_{ji}(1, \Lambda)$ are contained in the poles of $C_{'j,'i}(s, '\Lambda)$.

Since the association class $'\mathcal{C}$ is of rank-one, from Proposition 6.4.1 it follows that the number of poles of $C_{'j,'i}(s, '\Lambda)$ in the region $\{'\Lambda \in {'\mathfrak{a}^*} \otimes \mathbb{C} \mid |'\Lambda| \leq |\rho_P|\}$, counted with the multiplicity and the rank of the residue, is bounded by

$$b(N_{T,'\mathcal{C},\mu}(\mu + |\rho_{'P}|^2 + 1) + \dim \mathcal{E}_{cus}('\mathcal{C}, \mu) + 1),$$

where $b$ is a constant depending only on $\mathbf{G}$ and the truncation parameter $T$, and $N_{T,'\mathcal{C},\mu}(\lambda)$ is the counting function of $\Delta_{T,'\mathcal{C},\lambda}$ (§3.8), $'\mathbf{P} \in {}'\mathcal{C}$ . Since

$$\begin{aligned}
\sum_{\mu\leq\lambda} N_{T,'\mathcal{C},\mu}(\mu + |\rho_{'P}|^2 + 1) \leq & \sum_{\mu\leq\lambda} N_{T,'\mathcal{C},\mu}(\lambda + |\rho_{'P}|^2 + 1) \\
= & N_{T,'\mathcal{C}}(\lambda + |\rho_{'P}|^2 + 1) \\
\leq & N_{T,Q}(\lambda + |\rho_P|^2 + 1), \\
\sum_{\mu\leq\lambda} \dim \mathcal{E}_{cus}('\mathcal{C}, \mu) = & N_{cus,'\mathcal{C}}(\lambda),
\end{aligned}$$

we get that for the two adjacent chambers $C_i, C_j$ above, the total number of poles in the region $\{'\Lambda \in {}'\mathfrak{a}^* \otimes \mathbb{C} \mid |'\Lambda| \leq |\rho_P|\}$ of the rank-one scattering matrices $C_{'i,'j}(1, '\Lambda)$ for all $\mu \in \mathrm{Spec}_{cus}('\mathcal{C})$ with $\mu \leq \lambda$ is bounded by

$$\begin{aligned}
b(N_{T,'\mathcal{C}}(\lambda + |\rho_P|^2 + 1) + N_{cus,'\mathcal{C}}(\lambda) + 1) \leq & b(\sum_{\mathbf{Q}} N_{T,Q}(\lambda + |\rho_P|^2 + 1) \\
& + \sum_{\mathbf{Q}} \sum_{'\mathbf{P}} N_{cus,'P}(\lambda) + 1),
\end{aligned}$$

where $\mathbf{Q}, '\mathbf{P}$ run over the same set of parabolic subgroups as in the proposition.

By Lemma 2.7.3, for any $C_{ji}(s, \Lambda)$, the number of the above rank-one scattering matrices $C_{'i,'j}(1, '\Lambda)$ which appear in the factorization of $C_{ji}(s, \Lambda)$ in Lemma 2.7.4 is bounded by a constant which only depends on $\mathbf{G}$. Thus Proposition 7.2.1 follows from the above bound for the number of the poles of $C_{'i,'j}(1, '\Lambda)$.

**Proposition 7.2.2.** *For an association class $\mathcal{C}$ of rational parabolic subgroups $\mathbf{P}$ with $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P}) \geq 2$, the counting function $N_{res,\mathcal{C}}(\lambda)$ of the residual discrete spectrum of $\Delta$ in $L^2_{\mathcal{C}}(\Gamma\backslash X, \sigma)$ is bounded as follows:*

$$N_{res,\mathcal{C}}(\lambda) \leq c(\sum_{\mathbf{Q}} N_{T,Q}(\lambda + |\rho_P|^2 + 1) + \sum_{\mathbf{Q}} \sum_{'\mathbf{P}} N_{cus,'P}(\lambda) + 1)^{\mathrm{rank}_{\mathbb{Q}}(\mathbf{P})},$$

*where $\mathbf{Q}$ runs over a set of representatives of $\Gamma$-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 for every $\mathbf{Q}$ in the first sum $'\mathbf{P}$ runs over a set of representatives of $\Gamma_{M_Q}$-conjugacy classes of rank-1 rational* *parabolic subgroups of* $\mathbf{M_Q}$*, and* $c$ *is a constant depending only on* $\mathbf{G}$ *and* $T$*. In particular, as* $\lambda \to +\infty$,

$$N_{res,\mathcal{C}}(\lambda) = O(1)\lambda^{m/2},$$

*where* $m$ *is the maximum of* $(\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q})+1)\dim \Gamma_{M_Q}\backslash X_Q$ *for all rational parabolic subgroups* $\mathbf{Q} \subset \mathbf{G}$ *with* $\mathrm{rank}_{\mathbb{Q}}(\mathbf{Q}) \leq \mathrm{rank}_{\mathbb{Q}}(\mathbf{G}) - 1$.

*Proof.* First we note that if $\mu > \lambda$, the residual eigenvalues in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ are greater than $\lambda$ and hence do not contribute to $N_{res,\mathcal{C}}(\lambda)$. Proposition 2.6.1 shows that the number of residual eigenvalues in $L^2_{\mathcal{C},\mu}(\Gamma\backslash X, \sigma)$ is bounded by the number of complete flags of singular hyperplanes of the corresponding scattering matrices meeting the bounded region $\{\Lambda \in \mathfrak{a}^* \otimes \mathbb{C} \mid |\Lambda| \leq |\rho_P|\}$. Then the bound on the number of the singular hyperplanes of the scattering matrices in Proposition 7.2.1 gives a bound on the number of complete flags by raising it to the power $\mathrm{rank}_{\mathbb{Q}}(\mathbf{P})$ and hence proves the first bound on $N_{res,\mathcal{C}}(\lambda)$ in Proposition 7.2.2.

To prove the second bound on $N_{res,\mathcal{C}}(\lambda)$, we notice that

$$N_{T,Q}(\lambda + |\rho_P|^2 + 1) = O(1)\lambda^{\frac{1}{2}\dim \Gamma_{M_Q}\backslash X_Q}$$

by Theorem 3.3.2, and

$$N_{cus,'P}(\lambda) = O(1)\lambda^{\frac{1}{2}\dim \Gamma_{M_{'P}}\backslash X_{'P}}$$

by Lemma 2.3.2. Since $\dim \Gamma_{M_{'P}}\backslash X_{'P} \leq \dim \Gamma_{M_Q}\backslash X_Q$ and

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

we get that as $\lambda \to +\infty$

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

q.e.d.

*Proof of Theorem 1.1.3.* Let $\mathcal{C}_1, \cdots, \mathcal{C}_k$ be representatives of association classes of rational parabolic subgroups of rank greater than or equal to 2. Let $N'_d(\lambda)$ be the counting function of combination of the cuspidal spectrum and the rank-one residual spectrum as in Proposition 5.2.9. Then

$$N_d(\lambda) = N'_d(\lambda) + \sum_{i=1}^{k} N_{res,\mathcal{C}_i}(\lambda),$$

and the bounds in Propositions 5.2.9 and 7.2.2 give the bound on $N_d(\lambda)$ stated in Theorem 1.1.3.

**Remark 7.2.3.** In an early version of this manuscript, it was mistakenly claimed that $N_{res,\mathcal{C}}(\lambda)$ is bounded by the total number of singular hyperplanes of the scattering matrices which meet the bounded region $\{\Lambda \in \mathfrak{a}^* \otimes \mathbb{C} \mid |\Lambda| \leq |\rho_P|\}$ above for all the cuspidal pairs $(\mathcal{C}, \mu)$, $\mu \leq \lambda$. Since the bound for the latter in Proposition 7.2.1 is of smaller order than $\lambda^{n/2}$, if the rank of $\mathcal{C}$ is strictly greater than 1, it would imply that $N_{res,\mathcal{C}}(\lambda)$ is of smaller order than $\lambda^{n/2}$, and hence the Weyl upper bound is satisfied by $N_d(\lambda)$. The same problem occurred in [35, p. 523]. Because of this, the bound in [35, Theorem 0.1] should be replaced by the weaker one:

$$N_d(\lambda) \leq c(1 + \lambda^{\frac{n}{2} + \frac{3n}{2}\mathrm{rank}_{\mathbb{Q}}(\mathbf{G})})$$

as mentioned in §1.1.

## References

[1] J. Arthur, *The trace formula and Hecke operators*, Number theory, trace formulas and discrete groups, (ed. by 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] A.Borel, *Some metric properties of arithmetic quotients of symmetric spaces and an extension theorem*, J. Differential Geom. **6** (1972) 543–560.

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

[5] A. Borel & N.Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, Princeton Univ. Press, 1980.

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

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

[8] A. Deitmar & W. Hoffmann, *Spectral estimates for towers of noncompact quotients*, Preprint 1998.

[9] H. Donnelly, *On the point spectrum for finite volume symmetric spaces of negative curvature*, Comm. Partial Differential Equations **6** (1981) 963–992.

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

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

[12] J. Duistermaat, J. Kolk & V. Varadarajan, *Spectra of compact locally symmetric manifolds of negative curvature*, Invent. Math. **52** (1979) 27–93.

[13] I. Efrat, *The Selberg trace formula for* $\mathrm{PSL}_2(\mathbb{R})^n$, Mem. Amer. Math. Soc. **65** (1987) No. 359.

[14] ———, *Cusp forms and higher rank*, unpublished manuscript.

[15] ———, *On the discrete spectrum of certain discrete groups*, Bull. Amer. Math. Soc. **24** (1991) 125–129.

[16] J. Franke, *Harmonic analysis in weighted* $L_2$*-spaces*, preprint, First revision, 1997.

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

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

[19] D. Hejhal, *The Selberg trace formula*, Vol. 2, Lecture Notes in Math. Vol. 1001, Springer, 1983.

[20] M. Huxley, *Scattering matrices for congruence subgroups*, Modular forms, (ed. by R.A.Rankin), Horwood, Chichester, 1984, 141–156.

[21] L.Ji, *The trace class conjecture for arithmetic groups*, J. Differential Geom. **48** (1988) 165–203.

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

[23] ———, *Spectral degeneration of hyperbolic Riemann surfaces*, J. Differential Geom. **38** (1993) 263–313.

[24] L. Ji & R. MacPherson, *Geometry of compactifications of locally symmetric spaces*, Preprint 1997.

[25] C. Judge, *On the existence of Maass cusp forms on hyperbolic surfaces with cone points*, J. Amer. Math. Soc. **8** (1995) 715–759.

[26] T. Kato, *Perturbation theory for linear operators*, Springer, second edition, 1976.

[27] R. Langlands, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Math. Vol. 544, Springer, 1976.

[28] ———, *Eisenstein series*, Algebraic groups and discontinuous subgroups, (ed. by A.Borel & G.Mostow), Proc. Sympos. Pure. Math., Vol. IX, 1966, 235–252.

[29] ______, *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.

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

[31] G. Margulis, *Discrete subgroups of semisimple Lie groups*, Springer, 1990.

[32] Y. Matsushima & S. Murakami, *On vector bundle valued harmonic forms and automorphic forms on symmetric Riemannian manifolds*, Ann. of Math. **78** (1963) 365–416.

[33] C. Moeglin & J. Waldspurger, *Le spectre résiduel de* GL*(n)*, Ann. Sci. École Norm. Sup. **22** (1989) 605–674.

[34] ______, *Spectral decomposition and Eisenstein series*, Cambridge Univ. Press, 1995.

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

[36] ______, *Eigenvalue estimates for locally symmetric spaces of finite volume*, Sympos. Partial Differential Equations, (Holzhau, 1988), Teubner–Texte Math., Vol.112, Teubner, Leipzig, 1989, 179–196.

[37] ______, *The trace class conjecture in the theory of automorphic forms II*, Geom. Funct. Anal. **8** (1998) 315–355.

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

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

[40] ______, *The theory of Eisenstein systems*, Academic Press, New York, 1981.

[41] R. Phillips & P. Sarnak, *On cusp forms for co-finite subgroups of* PSL*(2,*$\mathbb{R}$*)*, Invent. Math. **80** (1985) 339–364.

[42] M. Reed & B. Simon, *Methods of modern mathematical physics IV: Analysis of operators*, Academic Press, New York, 1978.

[43] A. Reznikov, *Eisenstein matrix and existence of cusp forms in rank-one symmetric spaces*, Geom. Funct. Anal. **3** (1993) 79–105.

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

[45] P. Sarnak, *On cusp forms*, The Selberg trace formula and related topics, (ed. by D.Hejhal et al.), Contemp. Math. Vol 53, Amer. Math. Soc., 1984, 393–407.

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

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

[48] N. Wallach, *On the distribution of eigenvalues of the Laplacian of a locally symmetric space*, Differential Geometry, (ed. by B.Lawson), Pitman Monographs Surveys Pure Appl. Math. Vol. 52, Longman Sci. Tech. Harlow, 1991, 337–350.

[49] ______, *On the constant term of a square integrable automorphic form*, Operator algebras and group representations. II, (ed. by G.Arsene et al.), Monographs Stud. Math. Vol. 18, Pitman, Boston, 1984.

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

[51] S. Wolpert, *Disappearance of cusp forms in special families*, Ann. of Math. **139** (1994) 239-291.

UNIVERSITY OF MICHIGAN