ON THE GLOBAL SUP-NORM OF GL(3) CUSP FORMS*

BY

VALENTIN BLOMER

Mathematisches Institut, Georg-August Universität Göttingen
Bunsenstraße 3-5, D-37073 Göttingen, Germany
e-mail: vblomer@math.uni-goettingen.de

AND

GERGELY HARCOS

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences
POB 127, Budapest H-1364, Hungary
and
Central European University, Nador u. 9, Budapest H-1051, Hungary
e-mail: gharcos@renyi.hu, harcosg@ceu.edu

AND

PÉTER MAGA

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences
POB 127, Budapest H-1364, Hungary
e-mail: magapeter@gmail.com

* First author partially supported by the DFG-SNF lead agency program grant BL 915/2-1. Second and third author supported by NKFIH (National Research, Development and Innovation Office) grants NK 104183, ERC_HU_15 118946, K 119528. Second author also supported by ERC grant AdG-321104, and third author also supported by the Postdoctoral Fellowship of the Hungarian Academy of Sciences.

Received June 15, 2017 and in revised form March 24, 2018
ABSTRACT

Let $\phi$ be a spherical Hecke–Maaß cusp form on the non-compact space $\text{PGL}_3(\mathbb{Z}) \backslash \text{PGL}_3(\mathbb{R})$. We establish various pointwise upper bounds for $\phi$ in terms of its Laplace eigenvalue $\lambda_\phi$. These imply, for $\phi$ arithmetically normalized and tempered at the archimedean place, the bound

$$\|\phi\|_\infty \ll \lambda_\phi^{39/40 + \varepsilon}$$

for the global sup-norm (without restriction to a compact subset). On the way, we derive a new uniform upper bound for the $\text{GL}_3$ Jacquet–Whittaker function.

1. Introduction

Eigenfunctions of the Laplace–Beltrami operator $\Delta$ on a Riemannian manifold $X$ are the spectral building blocks of the Hilbert space $L^2(X)$. It is therefore a classical question to study their analytic properties asymptotically as the eigenvalue tends to infinity. In addition to its intrinsic interest in global analysis, a relation to quantum physics is provided by the fact that $e^{-it\Delta}$ is the time evolution operator of the Schrödinger equation describing a freely moving particle on $X$. Therefore, an $L^2$-eigenfunction of $\Delta$ is understood in quantum physics as a bound state, and its absolute square is interpreted as the probability density of the corresponding stationary wave. Somewhat unexpectedly, number theory also enters the scene, namely when the manifold $X$ possesses some additional arithmetic structure such as a commutative family of arithmetically defined Hecke operators (which are normal and commute with $\Delta$ as well). In this case, the most interesting Laplace eigenfunctions are those that are in addition eigenfunctions of the Hecke algebra, and as such are amenable to number theoretic tools. We will present an example of this kind in a moment.

A fundamental quantity associated to an $L^2$-normalized eigenfunction $\phi$ with eigenvalue $\lambda_\phi$ is its sup-norm. A good upper bound for $\|\phi\|_\infty$ can be seen as a basic measure of equidistribution of its mass on $X$. If $X$ is compact (or $X$ is non-compact but $\phi$ is restricted to a fixed compact subset $\Omega \subset X$), then we have the general bound proved by Hörmander [Hö]

$$\|\phi\|_\infty \ll \lambda^{(d-1)/4}, \quad d = \dim X.$$  

(1)

This bound is sharp, e.g., it is attained for the $d$-sphere $S^d$ for any $d \geq 1$ and for special eigenfunctions $\phi$. If $X$ is a compact locally symmetric space of dimension $d$ and rank $r$ (or we restrict to a compact subset $\Omega$ of such a space) and we
require that $\phi$ is not only a Laplace eigenfunction but an eigenfunction of all differential operators invariant under the group of isometries of the universal cover of $X$, then we have the stronger bound of Sarnak [Sar]

$$\|\phi\|_\infty \ll \lambda^{(d-r)/4}. \tag{2}$$

Even though neither (1) nor (2) are conjectured to be sharp for negatively curved manifolds, they provide a robust framework to work with.

The situation changes completely for the global sup-norm on non-compact manifolds, in which case (1) and (2) no longer need to be true. A typical example is a locally symmetric space $X = \Gamma \backslash G/K$, where $G$ is a non-compact semi-simple Lie group, $K \leq G$ is a maximal compact subgroup, and $\Gamma \leq G$ is a non-uniform lattice. It turns out that in this case the sup-norm of an eigenfunction $\phi$ is often determined by its behavior in the cuspidal regions of $X$, even though it may eventually decay very quickly. The simplest—and the only thoroughly explored—example is $G = \text{SL}_2(\mathbb{R})$, $K = \text{SO}_2(\mathbb{R})$, $\Gamma = \text{SL}_2(\mathbb{Z})$, so that $X = \Gamma \backslash G/K = \Gamma \backslash \mathcal{H}_2$ (where $\mathcal{H}_2 = \{x + iy : y > 0\}$) is the familiar modular surface. It is an arithmetic manifold in the above sense, as it admits the standard family of Hecke operators. A joint $L^2$-eigenfunction of the Laplace and the Hecke operators is often called a Hecke–Maaß cusp form. It decays exponentially as $y \to \infty$, and it is conjectured to satisfy $\|\phi|_\Omega\|_\infty \ll_{\varepsilon,\Omega} \lambda_\phi^{1/2 + o(1)}$ as $\lambda_\phi \to \infty$ for every compact subset $\Omega \subseteq X$ and every $\varepsilon > 0$, but in the cuspidal region it has considerable size. Precisely, we have $\phi(i\lambda_\phi^{1/2}/(2\pi)) = \lambda_\phi^{1/12 + o(1)}$, cf. [Sar], reflecting the very similar behavior of the normalized $GL_2$ Whittaker function

$$W_\nu(x) := \frac{\sqrt{\pi} K_\nu(2\pi x)}{|\Gamma(1/2 + \nu)|}, \quad \nu \in i\mathbb{R}. \tag{3}$$

Indeed, $W_\nu(x)$ decays exponentially as $x \to \infty$, but it has a large bump: $W_{it}(t/(2\pi)) \asymp t^{1/6}$ (see [Ba] for a uniform asymptotic expansion).

A systematic study of this behavior for Hecke–Maaß cusp forms on the locally symmetric space

$$X = X_n = \text{GL}_n(\mathbb{Z})\mathbb{Z}(\mathbb{R}) \backslash \text{GL}_n(\mathbb{R})/O_n(\mathbb{R})$$

and its congruence covers was initiated by Brumley and Templier [BT]. Here, $\mathbb{Z}(\mathbb{R})$ is the center of $\text{GL}_n(\mathbb{R})$ and $O_n(\mathbb{R})$ is the orthogonal subgroup. In particular, it was proved in [BT] that (2) fails for $X = X_n$ when $n \geq 5$, namely the global sup-norm on $X$ is significantly larger than the local sup-norm on a fixed
compact subset $\Omega \subseteq X$. Precisely, the lower bound by Brumley–Templier [BT] for the global sup-norm and the upper bound by Blomer–Maga [BM] for the local sup-norm can be contrasted as

$$\|\phi\|_{\infty} \gg \varepsilon \lambda_{\phi}^{n(n-1)(n-2)/24-\varepsilon} > \lambda_{\phi}^{n(n-1)/8-\delta_n} \gg_{\Omega} \|\phi|_{\Omega}\|_{\infty},$$

where $\delta_n > 0$ is a constant depending only on $n$. In addition, in the case $n = 3$, Brumley and Templier derived the upper bound [BT, Prop. 1.6]

(4) $$\|\phi\|_{\infty} \ll \lambda_{\phi}^{5/2} \quad \text{on } X = X_3,$$

by using the rapid decay of $\phi$ high in the cusp and making the dependence of (2) on the injectivity radius in the remaining piece of the manifold explicit. We note in passing that in the case $n = 3$, it is known that any savings $\delta_3 < 1/76$ is admissible\(^1\) for the upper bound of the local sup-norm [HRR], while in the case $n = 2$, the global sup-norm problem has been studied extensively (see [IS, Sah1, Sah2, BHMM] and the references therein).

Despite these important advances, the investigation of the global sup-norm of eigenfunctions on non-compact locally symmetric spaces of rank exceeding one is still in its infancy, and in this article we take a closer look at the rank two example $n = 3$ with the aim of proving considerably stronger bounds than (4) by a different technique. With this in mind, let $\phi$ be a Hecke–Maaß cusp form on $GL_3$ over $\mathbb{Q}$, which is spherical at every place and has trivial central character, regarded as a complex-valued function on the quotient space $X = X_3$. Alternatively, we may think of $\phi$ as a complex-valued function on $GL_3(\mathbb{R})$ satisfying

$$\phi(\gamma hgk) = \phi(g) \quad \text{for all } \gamma \in GL_3(\mathbb{Z}), \ h \in \mathbb{Z}(\mathbb{R}), \ k \in O_3(\mathbb{R}),$$

which is further an eigenfunction of the invariant differential operators and the Hecke operators [Go, Sections 6.1–6.4]. We recall from [Go, Sections 1.2–1.3] that the symmetric space $Z(\mathbb{R}) \backslash GL_3(\mathbb{R})/O_3(\mathbb{R})$ can be represented by matrices of the generalized upper half-plane

$$\mathcal{H}_3 := \left\{ \begin{pmatrix} 1 & x_2 & x_3 \\ 1 & x_1 \\ 1 \\ 1 & y_1y_2 & y_1 \end{pmatrix} : y_1, y_2 > 0, \ x_1, x_2, x_3 \in \mathbb{R} \right\},$$

\(^1\) By saying this we do not mean to imply that $1/76$ is a special threshold beyond which the bound fails.
and the quotient space $X = \text{GL}_3(\mathbb{Z}) \backslash \mathcal{H}_3$ has a fundamental domain lying in the Siegel set defined by

$$\begin{align*}
|x_1|, |x_2|, |x_3| &\leq 1/2 \quad \text{and} \quad y_1, y_2 \geq \sqrt{3}/2.
\end{align*}$$

In this paper, we provide upper bounds for $|\phi(z)|$ in terms of the height parameters $y_1, y_2$ (assuming they are at least $\sqrt{3}/2$) and the Laplace eigenvalue $\lambda_{\phi}$, and examine what they yield for the global sup-norm

$$\|\phi\|_\infty = \sup_{z \in \mathcal{H}_3} |\phi(z)|.$$  

We shall work with the following natural normalizations:

- $\phi$ is arithmetically normalized if it has leading Fourier coefficient $\lambda_{\phi}(1,1) = 1$ with respect to the standard Jacquet–Whittaker function (cf. [Go, Thm. 6.4.11] and (14) below);
- $\phi$ is $L^2$-normalized if it has $L^2$-norm 1 with respect to the measure on $X$ induced by the standard left-invariant probability measure on $\mathcal{H}_3$ (cf. [Go, Prop. 1.5.3]).

By [Bl, Lemma 1] and its proof, these two normalizations differ by a positive constant times $L(1, \phi, \text{Ad})^{1/2}$ when $\phi$ is tempered at the archimedean place, and this would also be true for non-tempered forms if we slightly renormalized the standard Jacquet–Whittaker function as in the display below [Bl, (2.13)] with the effect of correspondingly changing the notion of “arithmetically normalized”.

With these conventions, our main results are as follows (see also the remarks after the theorems):

**Theorem 1:** Let $\phi$ be an arithmetically normalized Hecke–Maaß cusp form on $X$. Assume that $\phi$ is tempered at the archimedean place. Then for any $z \in \mathcal{H}_3$ with $y_1, y_2 \geq \sqrt{3}/2$, and for any $\varepsilon > 0$, we have

$$\phi(z) \ll_{\varepsilon} \min(y_1, y_2) \left( \frac{\lambda_{\phi}^{1+\varepsilon}}{y_1y_2} + \frac{\lambda_{\phi}^{3/2+\varepsilon}}{(y_1y_2)^2} \right).$$

**Theorem 2:** Let $\phi$ be an $L^2$-normalized Hecke–Maaß cusp form on $X$. Then for any $z \in \mathcal{H}_3$ with $y_1, y_2 \geq \sqrt{3}/2$, we have

$$\phi(z) \ll \lambda_{\phi}^{3/4} + \lambda_{\phi}^{5/8} y_1y_2.$$
**Theorem 3:** Let $\phi$ be an arithmetically normalized Hecke–Maaß cusp form on $X$. Assume that $\phi$ is tempered at the archimedean place. Then for any $\varepsilon > 0$, we have

$$
\| \phi \|_\infty \ll \varepsilon \lambda^{39/40-\varepsilon}.
$$

**Remarks:** Assume that $\phi$ is tempered at the archimedean place. If $\phi$ is arithmetically normalized, then (7) holds with the extra factor $L(1, \phi, \text{Ad})^{1/2} \ll \varepsilon \lambda^{1/2} \phi$ on the right-hand side by the work of Brumley [Br1, Cor. 2] or Li [Li, Thm. 2]. Similarly, if $\phi$ is $L^2$-normalized, then (6) and (8) hold with the extra factor $L(1, \phi, \text{Ad})^{-1/2} \ll \varepsilon \lambda^{1/2+\varepsilon}$ on the right-hand side by the work of Brumley [Br2, Thm. 3] (see also [La, Appendix]). If $\phi$ is self-dual, i.e., $\phi$ is the symmetric square of a classical (even or odd) Hecke–Maaß cusp form on $\text{SL}_2(\mathbb{Z}) \backslash \mathcal{H}_2$ (cf. [Ra, Thm. A]), we even know that

$$
L(1, \phi, \text{Ad})^{-1/2} \ll \varepsilon \lambda^{\phi} \phi
$$

by a result of Ramakrishnan and Wang [RW, Cor. C], so that in this case no adjustment to (6) and (8) is necessary. Finally, we note that for $\phi$ self-dual, the exponent $39/40$ in (8) cannot be lowered below $3/8$, as follows from the work of Brumley and Templier [BT, Cor. 1.10].

We prove our results by employing two very different methods. On the one hand, we estimate the Fourier–Whittaker expansion of $\phi$ termwise, which eventually leads to Theorem 1. Unlike in the rank one case $n = 2$, the fact that the group of upper-triangular unipotent $3 \times 3$ matrices is not commutative leads to an additional sum over an infinite subset of $\text{SL}_2(\mathbb{Z})$ which requires careful treatment. As a result of independent interest and as a contribution to the analytic theory of higher rank Whittaker functions, we provide in Lemma 2 a new uniform upper bound for the $\text{GL}_3$ Jacquet–Whittaker function.

On the other hand, as in earlier approaches on compact spaces, we use a pre-trace formula, but we make the analysis of the geometric side uniform in the height of the considered point $z \in \mathcal{H}_3$; this is familiar for the group $\text{GL}_2$, but seems to have not yet been worked out in higher rank. This leads to Theorem 2. A key ingredient of the proof is Lemma 3, which states a simple but efficient bound for the norm of the Cartan projection of upper-triangular matrices. As we do not use amplification, the proof is independent of Hecke operators and consequently the result also holds for all $L^2$-normalized Maaß forms, not just the Hecke eigenforms.
Our final Theorem 3 is a combination of Theorems 1 and 2. There is nothing particularly special about the exponent $39/40$, except that the corresponding bound (8) is considerably stronger than (4) and already fairly close to (2). Marginal improvements are possible, for instance by inserting an amplifier into the pre-trace formula.

ACKNOWLEDGEMENTS. We thank Gergő Nemes for valuable discussions concerning the special functions that appear in this paper. We also thank the referee whose careful reading and detailed comments helped us to improve the exposition.

2. Archimedean parameters and the Weyl group

Associated to every Hecke–Maaß cusp form $\phi$ on $X$ (in fact, the Hecke property is not needed for this discussion), there is an $S_3$-orbit of (archimedean) Langlands parameters

$$(\mu_1, \mu_2, \mu_3) \in \mathbb{C}^3 \text{ satisfying } \mu_1 + \mu_2 + \mu_3 = 0$$

and a corresponding $S_3$-orbit of (archimedean) spectral parameters

$$(\nu_0, \nu_1, \nu_2) \in \mathbb{C}^3 \text{ satisfying } \nu_0 = \nu_1 + \nu_2.$$  

We follow the conventions of [Bu] except that we shift $\nu_1$ and $\nu_2$ there by $-1/3$, so that (cf. [Bu, (8.3)])

$$\lambda_{\phi} = 1 - \frac{1}{2}(\mu_1^2 + \mu_2^2 + \mu_3^2) = 1 - 3\nu_1^2 - 3\nu_1\nu_2 - 3\nu_2^2,$$

and $(1/3 + \nu_2, 1/3 + \nu_1) \in \mathbb{C}^2$ are the spectral parameters in [Go, Section 6].

The Weyl group $S_3$ acts by permutations on the Langlands parameters, hence an $S_3$-orbit of spectral parameters takes the shape

$$\{(\nu_0, \nu_1, \nu_2), (\nu_2, -\nu_1, \nu_0), (-\nu_1, \nu_2, -\nu_0),$$

$$(-\nu_0, -\nu_2, -\nu_1), (\nu_2, -\nu_0, \nu_1), (\nu_1, \nu_0, -\nu_2)\}.$$

Unfortunately, the relevant literature has many inconsistencies that are hard to track. For example, in Goldfeld’s book [Go], Definition 6.5.2 is inconsistent with Definition 9.4.3, and hence Theorem 6.5.15 contradicts Theorem 10.8.6 and the subsequent discussion. In fact $\nu_1$ and $\nu_2$ should be flipped in [Go, Thm. 6.5.15], in harmony with [Bu, (8.3)]. Compare also the functions $I_\nu$ on [Bu, p. 19] and [Go, p. 154].
If \( \phi \) is tempered at the archimedean place (as assumed in Theorems 1 and 3), the Langlands parameters and the spectral parameters are purely imaginary. Moreover, each \( S_3 \)-orbit of spectral parameters contains a unique triple with nonnegative imaginary parts

\[(\nu_0, \nu_1, \nu_2) \in (i\mathbb{R}_{\geq 0})^3 \text{ satisfying } \nu_0 = \nu_1 + \nu_2.\]

From now on we regard this triple as the spectral parameter of \( \phi \). For convenience, we also introduce

\[(T_0 := \max(2, |\nu_0|, |\nu_1|, |\nu_2|) = \max(2, |\nu_0|) \asymp \lambda_{\phi}^{1/2}).\]

### 3. The Fourier–Whittaker expansion

The restriction of \( \phi \) to \( H_3 \) admits a Fourier–Whittaker expansion (cf. [Go, (6.2.1)])

\[
\phi(z) = \sum_{m_1=1}^{\infty} \sum_{m_2 \neq 0} \frac{\lambda_{\phi}(m_1, m_2)}{|m_1m_2|} \times \sum_{\gamma \in U_2(\mathbb{Z}) \setminus SL_2(\mathbb{Z})} W^{\text{sign}(m_2)}_{\nu_1, \nu_2} \left( \begin{array}{c} |m_1m_2| \\ m_1 \\ 1 \end{array} \right) \left( \begin{array}{c} \gamma \\ 1 \\ z \end{array} \right),
\]

where \( U_2 \) is the subgroup of upper-triangular unipotent matrices \((1^*1)\) and \( W^{\pm}_{\nu_1, \nu_2} : \text{GL}_3(\mathbb{R}) \rightarrow \mathbb{C} \) is the standard Jacquet–Whittaker function (cf. [Go, (6.1.2)] and [Bl, (2.10)–(2.11)]). The function \( W^{\pm}_{\nu_1, \nu_2} \) is invariant under \( Z(\mathbb{R}) \), right-invariant under \( O_3(\mathbb{R}) \), and its restriction to \( H_3 \) has the following integral representation by a result of Vinogradov–Takhtadzhyan [VT] (cf. [Bl, (2.12)]):

\[
\left( \prod_{j=0}^{2} \frac{\Gamma \left( \frac{1}{2} + \frac{3}{2}\nu_j \right)}{\pi^{\frac{3}{2} + \frac{3}{2}\nu_j}} \right) W^{\pm}_{\nu_1, \nu_2} \left( \begin{array}{c} x_2 \\ x_3 \\ 1 \end{array} \right) \left( \begin{array}{c} y_1 \\ y_2 \\ y_1 \end{array} \right) = 4e(x_1 \pm x_2) y_1^{1 + \frac{\nu_1 - \nu_2}{\nu_2}} y_2^{1 + \frac{\nu_2 - \nu_1}{\nu_1}} \times \int_0^{\infty} K_{\frac{3}{2}\nu_0} (2\pi y_1 \sqrt{1 + u}) K_{\frac{3}{2}\nu_0} (2\pi y_2 \sqrt{1 + u^{-1}}) u^{\frac{3}{2}(\nu_1 - \nu_2)} \frac{du}{u},
\]

where \( K_{\nu} \) is the usual \( K \)-Bessel function. This formula is a special case of [St1, Thm. 2.1] with the constant \( 2^{n-1} \) corrected to \( 2^{2n-3} \) there, as recorded by [St2, (4.3) & p. 132]. Indeed, (15) follows readily from the first display of
Vol. 229, 2019  GLOBAL SUP-NORM OF GL(3) CUSP FORMS  365

[St1, p. 318], by updating the constant 4 to 8, then substituting $u^{1/2}$ for $u$, and finally noting that $(\nu_1, \nu_2)$ and $(y_1, y_2)$ there equal $(1/3 + \nu_1, 1/3 + \nu_2)$ and $(y_2, y_1)$ in our notation.\(^3\) An independent verification of (15) can be obtained by comparing carefully [BB, (5.8)] with [Bu, (10.1)] (or more directly with [Bu, (10.2)], after inserting a missing factor of $\pi^{-2}$ on the right-hand side there).

We remark in passing that by [Go, Prop. 5.5.2, Eq. (5.5.5), Thm. 5.9.8], the left-hand side of (15) is the same over the whole spectral $S_3$-orbit (11), hence in the tempered case (i.e., when $\nu_j \in i\mathbb{R}$) the corresponding Fourier coefficients $\lambda_\phi(m_1, m_2)$ only differ from each other by six constants on the unit circle. At any rate, since the unipotent part (i.e., the entries $x_1, x_2, x_3$) acts on the right hand side of (15) via a rotation, it is convenient to introduce also

$$\tilde{W}_{\nu_1, \nu_2}(y_1, y_2) := 4\pi^{2} \prod_{j=0}^{2} |\Gamma\left(\frac{1}{2} + \frac{3}{2} \nu_j\right)|^{-1} y_1 y_2 \left(\frac{y_1}{y_2}\right)^{\nu_1 - \nu_2}$$

(16)

$$\times \int_{0}^{\infty} K_{\frac{3}{2} \nu_0}(2\pi y_1 \sqrt{1+u}) K_{\frac{3}{2} \nu_0}(2\pi y_2 \sqrt{1+u-1}) u^{\frac{3}{2}(\nu_1 - \nu_2)} \frac{du}{u}.$$

4. An upper bound for the Jacquet–Whittaker function

In this section we estimate, under the assumption (12), the function $W_{\nu_1, \nu_2}$ given by (15), or equivalently the function $\tilde{W}_{\nu_1, \nu_2}$ given by (16). We start with a bound involving the normalized GL$_2$ Whittaker function introduced in (3):

**Lemma 1:** Assume that $\nu \in i\mathbb{R}$, and put $T := \max(1/2, |\nu|)$. Then, for any $t, A > 0$, we have

$$\int_{t}^{\infty} \frac{W_{\nu}^2(x)}{\sqrt{x^2 - t^2}} \, dx \ll_A (\log(3T)) \left(1 + \frac{t}{T}\right)^{-A}.$$

**Proof.** Applying the crude bound (cf. [HM, Prop. 9])

$$W_{\nu}(x) \ll e^{-\pi x}, \quad x > T,$$

we obtain

$$\int_{\max(t,T)}^{\infty} \frac{W_{\nu}^2(x)}{\sqrt{x^2 - t^2}} \, dx \ll \frac{1}{\sqrt{t}} \int_{t}^{\infty} \frac{e^{-2\pi x}}{\sqrt{x - t}} \, dx \ll e^{-2\pi t},$$

\(^3\) Accordingly, in [Go, (6.1.3)], the constant 4 should be 8, while $\nu_1$ and $\nu_2$ should be flipped.
hence it suffices to prove that
\[
\int_t^T \frac{W_\nu^2(x)}{\sqrt{x^2 - t^2}} \, dx \ll \log(3T), \quad 0 < t < T.
\]
Observing also that (cf. [GR, 6.576.4])
\[
\int_0^\infty \frac{\sqrt{x}}{\sqrt{x^2 - t^2}} \, dx \ll \int_0^\infty \frac{W_\nu^2(x)}{x} \, dx = \frac{1}{8},
\]
we are left with proving
\[
\int_t^{2t} \frac{W_\nu^2(x)}{\sqrt{x^2 - t^2}} \, dx \ll \log(3T), \quad 0 < t < T.
\]
To reformulate, we need to show
\[
\int_t^{2t} \frac{W_\nu^2(x)}{\sqrt{x - t}} \, dx \ll t^{1/2} \log(3T), \quad 0 < t < T,
\]
and we shall accomplish this by using the bound (cf. [BH, p. 679] and [HM, Prop. 9])
\[
W_\nu(x) \ll \min(T^{1/6}, T^{1/4}|2\pi x - T|^{-1/4}), \quad 0 < x < 2T.
\]
For $0 < t \leq T/20$ this bound readily yields that
\[
\int_t^{2t} \frac{W_\nu^2(x)}{\sqrt{x - t}} \, dx \ll \int_t^{2t} \frac{1}{\sqrt{x - t}} \, dx \ll t^{1/2},
\]
hence we can assume that $T/20 < t < T$. Then, our task reduces to showing that
\[
\int_I \frac{T^{1/3}}{\sqrt{x - t}} \, dx + \sum_M \int_{I(M)} \frac{T^{1/2}M^{-1/2}}{\sqrt{x - t}} \, dx \ll T^{1/2} \log(3T),
\]
where
\[
I := \{x \in [t, 2t] : |2\pi x - T| \leq T^{1/3}\},
\]
\[
I(M) := \{x \in [t, 2t] : M \leq |2\pi x - T| < 2M\},
\]
and $M$ runs through the numbers $M \in [T^{1/3}, 20T]$ of the form $M = 2^kT^{1/3}$ with $k \in \mathbb{N}$. Note that $I \subset [t, 2t]$ is either empty or an interval of length less than $T^{1/3}$, while $I(M) \subset [t, 2t]$ is a disjoint union of at most two intervals of
lengths less than $M$. Therefore,
\[ \int \frac{T^{1/3}}{\sqrt{x - t}} \, dx \ll T^{1/3}T^{1/6} = T^{1/2}, \]
\[ \int \frac{T^{1/2}M^{-1/2}}{\sqrt{x - t}} \, dx \ll T^{1/2}M^{-1/2}M^{1/2} = T^{1/2}, \]
and we are done upon noting that the number of dyadic parameters $M$ is $\ll \log(3T)$. □

**Lemma 2:** For any $y_1, y_2, A > 0$ we have, under the assumption (12) and with the notation (13),
\[ \tilde{W}_{\nu_1, \nu_2}(y_1, y_2) \ll A (\log T_0) \sqrt{y_1y_2} \left( 1 + \frac{y_1}{T_0} \right)^{-A} \left( 1 + \frac{y_2}{T_0} \right)^{-A}. \]

**Proof.** From Stirling’s approximation and (12), it follows that
\[ \prod_{j=0}^{2} \left| \Gamma \left( \frac{1}{2} + \frac{3}{2} \nu_j \right) \right|^{-1} \ll \left| \Gamma \left( \frac{1}{2} + \nu \right) \right|^{-2} \text{ with } \nu := \frac{3}{2} \nu_0. \]

Then we see from (16) and the Cauchy–Schwarz inequality that
\[ \tilde{W}_{\nu_1, \nu_2}(y_1, y_2) \ll \sqrt{y_1y_2} \int_0^\infty \frac{|W_{\nu}(y_1 \sqrt{1 + u})W_{\nu}(y_2 \sqrt{1 + u^{-1}})|}{(1 + u)^{1/4}(1 + u^{-1})^{1/4}} \frac{du}{u} \]
\[ \ll \sqrt{y_1y_2} \left( \int_0^\infty \frac{W_{\nu}^2(y_1 \sqrt{1 + u})}{\sqrt{1 + u^{-1}}} \frac{du}{u} \right)^{1/2} \]
\[ \times \left( \int_0^\infty \frac{W_{\nu}^2(y_2 \sqrt{1 + u^{-1}})}{\sqrt{1 + u}} \frac{du}{u} \right)^{1/2}. \]

Applying the change of variables $x := y_1 \sqrt{1 + u}$ (resp. $x := y_2 \sqrt{1 + u^{-1}}$) in the last two integrals, we get
\[ \tilde{W}_{\nu_1, \nu_2}(y_1, y_2) \ll \sqrt{y_1y_2} \left( \int_{y_1}^\infty \frac{W_{\nu}^2(x)}{\sqrt{x^2 - y_1^2}} \, dx \right)^{1/2} \left( \int_{y_2}^\infty \frac{W_{\nu}^2(x)}{\sqrt{x^2 - y_2^2}} \, dx \right)^{1/2}. \]

Finally, estimating these two integrals by Lemma 1, we are done. □

**5. Proof of Theorem 1**

Let $\phi$ be an arithmetically normalized Hecke–Maaß cusp form on $X$ as in Theorem 1. Then the Fourier coefficients
\[ \lambda_\phi(m_1, m_2) = \lambda_\phi(m_1, |m_2|) \]
in (14) are actual eigenvalues of various Hecke operators (cf. [Go, Thm. 6.4.11]). Moreover, summing over the representatives \( \gamma = \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \) of \( U_2(\mathbb{Z}) \backslash SL_2(\mathbb{Z}) \) in (14) is the same as summing over the coprime pairs \( (c, d) \in \mathbb{Z}^2 \). Computing, as in [Go, (6.5.4)], the \( GL_2 \) Iwasawa decomposition in the upper-left \( 2 \times 2 \) block of

\[
\left( \begin{array}{cc} \gamma & \\
1 & 1 \end{array} \right) zO_3(\mathbb{R}) = \left( \begin{array}{cccc}
a & b & x_1 & x_2 \\
c & d & y_1 & y_2 \\
 & & x_3 & y_3 \\
1 & 1 & 1 & 1 \end{array} \right) O_3(\mathbb{R}),
\]

where \( z = x_2 + iy_2 \in \mathcal{H}_2 \), we see from (14)–(16) that

\[
\phi(z) \ll \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{|\lambda_\phi(m_1, m_2)|}{m_1 m_2} \sum_{(c, d) \in \mathbb{Z}^2, \gcd(c, d)=1} |\tilde{W}_{\nu_1, \nu_2}(m_1 y_1 |cz_2 + d|, m_2 y_2 |cz_2 + d|^{-2})|.
\]

Taking any \( \varepsilon > 0 \), and applying the Cauchy–Schwarz inequality, we obtain

\[
\phi(z) \ll \left( \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{|\lambda_\phi(m_1, m_2)|^2}{m_1^{2+\varepsilon} m_2^{1+\varepsilon}} \right)^{1/2}
\]

\[
\times \left( \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{m_1^{2\varepsilon}}{m_2^{1-\varepsilon}} F(m_1 y_1, m_2 y_2)^2 \right)^{1/2},
\]

where

\[
F(s_1, s_2) := \sum_{(c, d) \in \mathbb{Z}^2, \gcd(c, d)=1} |\tilde{W}_{\nu_1, \nu_2}(s_1 |cz_2 + d|, s_2 |cz_2 + d|^{-2})|.
\]

It remains to estimate the two factors on the right hand side of (17). The first factor (the arithmetic part) will be estimated in Subsection 5.1. The second factor (the analytic part) will be estimated in Subsection 5.2.

5.1. Estimating the arithmetic part. By [Go, Sections 6.3–6.4], the dual form

\[
\tilde{\phi}(z) := \phi((z^{-1})^t), \quad z \in \mathcal{H}_3,
\]

\[
\tilde{\phi}(z) := \phi((z^{-1})^t), \quad z \in \mathcal{H}_3,
\]
is an arithmetically normalized Hecke–Maaß cusp form on $X = \text{GL}_3(\mathbb{Z}) \backslash \mathcal{H}_3$ with spectral parameters $(\nu_0, \nu_2, \nu_1)$ and Hecke eigenvalues

$$\lambda_{\tilde{\phi}}(m_1, m_2) = \lambda_{\phi}(m_2, m_1) = \lambda_{\phi}(m_1, m_2).$$

Introducing the Rankin–Selberg $L$-function (cf. [Go, Section 7.4])

$$L(s, \phi \times \tilde{\phi}) = \zeta(3s) \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{|\lambda_{\phi}(m_1, m_2)|^2}{m_1^{2s} m_2^s}, \quad \Re s > 1,$$

the first double sum on the right hand side of (17) can be estimated by the result of Brumley [Br1, Cor. 2] or Li [Li, Thm. 2]:

$$\sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{|\lambda_{\phi}(m_1, m_2)|^2}{m_1^{2+2\varepsilon} m_2^{1+\varepsilon}} = \frac{L(1 + \varepsilon, \phi \times \tilde{\phi})}{\zeta(3 + 3\varepsilon)} \ll \varepsilon T_0^{\varepsilon}.$$

5.2. Estimating the analytic part. We decompose (18) into dyadic sub-sums: for any $k, l \in \mathbb{N}$, we define

$$D(k, l) := \left\{(c, d) \in \mathbb{Z}^2 : \gcd(c, d) = 1, 2^k < 1 + \frac{s_1|cz_2 + d|}{T_0} \leq 2^{k+1}, \right.$$  

$$2^l < 1 + \frac{s_2|cz_2 + d|^{-2}}{T_0} \leq 2^{l+1}\right\}$$

and

$$F_{k, l}(s_1, s_2) := \sum_{(c,d) \in D(k,l)} |\tilde{W}_{\nu_1, \nu_2}(s_1|cz_2 + d|, s_2|cz_2 + d|^{-2})|.$$

Clearly,

$$F(s_1, s_2) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} F_{k, l}(s_1, s_2),$$

which implies, via the Cauchy–Schwarz inequality, that

$$F(s_1, s_2)^2 \ll \varepsilon \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} 2^{\varepsilon(k+l)} F_{k, l}(s_1, s_2)^2$$

holds for any $\varepsilon > 0$.

We estimate $F_{k, l}(s_1, s_2)$ on the right-hand side of (21) via Lemma 2 as

$$F_{k, l}(s_1, s_2) \ll \varepsilon A \frac{2^{-A(k+l)} T_0 \varepsilon}{s_1 s_2} \sum_{(c,d) \in D(k,l)} |cz_2 + d|^{-1/2}. $$
From the two conditions on the range of \( s_1, s_2 \) in the definition of \( D(k, l) \), it is immediate that \( F_{k,l}(s_1, s_2) \) vanishes unless \( s_1^2 s_2 \leq 2^{2k+1} T_0^3 \), which means that

\[
F_{k,l}(s_1, s_2) \ll_{\epsilon, A} 2^{-A(k+l)l} T_0^3 \sqrt{s_1 s_2} \cdot 1_{s_1^2 s_2 \leq 2^{2k+1} T_0^3} \sum_{(c,d) \in D(k,l)} |cz_2 + d|^{-1/2}.
\]

For the \((c,d)\)-summation, we only use the estimate \(|cz_2 + d| \leq 2^{k+1} T_0 s_1^{-1}\), which amounts to saying that we have to sum over the lattice \( \mathbb{Z} z_2 + \mathbb{Z} \subset \mathbb{C} \) in a given disk around the origin. The lattice has first successive minimum at least \( \sqrt{3}/2 \) and covolume \( 3 z_2 = y_2 \) also at least \( \sqrt{3}/2 \). Applying again a dyadic decomposition and considering the pairs \((c, d)\) satisfying \( 2^m \leq |cz_2 + d| < 2^{m+1} \), we obtain by [BHM, Lemma 1]

\[
\sum_{(c,d) \in D(k,l)} |cz_2 + d|^{-1/2} \ll \sum_{m \in \mathbb{Z}} \frac{2^{-m/2} (2^m + 2^{2m} y_2^{-1})}{1/2 \leq 2^m \leq 2^{k+1} T_0 s_1^{-1}} \ll 2^{k/2} T_0^{1/2} s_1^{-1/2} + 2^{3k/2} T_0^{3/2} s_1^{-3/2} y_2^{-1}.
\]

Plugging this into (22), we conclude

\[
F_{k,l}(s_1, s_2) \ll_{\epsilon, A} T_0^\epsilon \sum_{s_1^2 s_2 \leq 2^{2k+1} T_0^3} (2^{(1/2-A)k-A} T_0^{1/2} s_2^{1/2} + 2^{(3/2-A)k-A} T_0^{3/2} s_1^{-1/2} s_2 y_2^{-1} + 2^{(3/2-A)k-A} T_0^{5/2} s_1^{-1/2} s_2 y_2^{-1}).
\]

Using the previous bound in (21), then substituting \( s_1 = m_1 y_1 \) and \( s_2 = m_2 y_2 \), and finally summing with respect to \( m_1 \) and \( m_2 \) as in (17), we obtain

\[
\sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{m_2^{2\epsilon}}{m_1^{-\epsilon}} F(m_1 y_1, m_2 y_2)^2 \ll_{\epsilon, A} T_0^{2\epsilon} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} 1_{m_1^2 m_2 \leq 2^{2k+1+3 T_0^3} y_2^{-1}}
\]

\[
\times \frac{m_1^{2\epsilon}}{m_2^{-\epsilon}} (2^{(\epsilon+1-2A)k+(\epsilon-2A)l} T_0 m_2 y_2 + 2^{(\epsilon+3-2A)k+(\epsilon-2A)l} T_0^3 m_1^{-2} m_2 y_1^{-2} y_2^{-1}).
\]

The first term in the last two lines of (23) contributes (after the summation in the second line there)

\[
\ll_{\epsilon, A} T_0^{(2\epsilon+2) y_2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} 2^{(\epsilon+1-2A)k+(\epsilon-2A)l} \sum_{m_1=1}^{\infty} m_1^{2\epsilon} \sum_{1 \leq m_2 \leq 2^{2k+1+3 T_0^3} m_1^{-2} y_1^{-2} y_2^{-1}} m_2^{\epsilon}.
\]
Here, the $m_2$-sum is at most $(2^{2k+l+3}T_0^3m_1^{-2}y_1^{-2}y_2^{-1})^{1+\varepsilon}$. Then clearly the resulting $m_1$-sum is convergent, which altogether means that this first term gives rise to

$$\ll_{\varepsilon,A} T_0^{4+5\varepsilon} y_1^{-2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} 2^{(3\varepsilon+3-2A)k+(2\varepsilon+1-2A)l}. \quad (24)$$

Similarly, the second term in the last two lines of (23) contributes (after the summation in the second line there)

$$\ll_{\varepsilon,A} T_0^{3+2\varepsilon} y_1^{-2} y_2^{-1} \times \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} 2^{(\varepsilon+3-2A)k+(\varepsilon-2A)l} \sum_{m_1=1}^{\infty} m_1^{2\varepsilon-2} \sum_{1 \leq m_2 \leq 2^{2k+l+3}T_0^3m_1^{-2}y_1^{-2}y_2^{-1}} m_2^{\varepsilon}. \quad (25)$$

Here, the $m_2$-sum is again at most $(2^{2k+l+3}T_0^3m_1^{-2}y_1^{-2}y_2^{-1})^{1+\varepsilon}$, and the resulting $m_1$-sum is convergent as above. In the end, the second term gives rise to

$$\ll_{\varepsilon,A} T_0^{6+5\varepsilon} y_1^{-4} y_2^{-2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} 2^{(3\varepsilon+5-2A)k+(2\varepsilon+1-2A)l}. \quad (25)$$

We see that in both (24) and (25), the double sums over $k$ and $l$ are convergent upon choosing $A := 3$, say. Returning to (23), this means that

$$\sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{m_1^{2\varepsilon}}{m_2^{1-\varepsilon}} F(m_1 y_1, m_2 y_2)^2 \ll_{\varepsilon} \frac{T_0^{4+5\varepsilon}}{y_1^2} + \frac{T_0^{6+5\varepsilon}}{y_1^4 y_2^2}. \quad (26)$$

5.3. CONCLUSION. The inequalities (17), (20) and (26) altogether give

$$\phi(z) \ll_{\varepsilon} \frac{T_0^{2+3\varepsilon}}{y_1} + \frac{T_0^{3+3\varepsilon}}{y_1^2 y_2} \ll_{\varepsilon} \frac{\lambda_\phi^{1+2\varepsilon}}{y_1} + \frac{\lambda_\phi^{3/2+2\varepsilon}}{y_1^2 y_2},$$

and we write this in the more symmetric form

$$\phi(z) \ll_{\varepsilon} y_2 \left( \frac{\lambda_\phi^{1+\varepsilon}}{y_1 y_2} + \frac{\lambda_\phi^{3/2+\varepsilon}}{(y_1 y_2)^2} \right). \quad (27)$$

Finally, we utilize the dual form introduced in (19) to show that (27) remains valid when we interchange $y_1$ and $y_2$ on the right hand side. Indeed, we can
express $\phi(z)$ as

$$\phi(z) = \tilde{\phi}((z^{-1})^t) = \tilde{\phi}(hw(z^{-1})^tw),$$

$$h := \begin{pmatrix} y_1y_2 & & \\ & y_1y_2 & \\ y_1y_2 & & \end{pmatrix}, \quad w := \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix},$$

where we have the Iwasawa decomposition (see [Go, (6.3.2)] for a similar computation)

$$hw(z^{-1})^tw = \begin{pmatrix} 1 & -x_1 & x_1x_2 - x_3 \\ & 1 & -x_2 \\ & & 1 \end{pmatrix} \begin{pmatrix} y_1y_2 & & \\ & y_2 & \\ & & 1 \end{pmatrix}.$$ 

Applying (27) and noting that $\lambda_{\tilde{\phi}} = \lambda_{\phi}$ by (10), we infer

$$\phi(z) = \tilde{\phi}(hw(z^{-1})^tw) \ll \varepsilon y_1 \left( \frac{\lambda_{\phi}^{1+\varepsilon}}{y_1y_2} + \frac{\lambda_{\phi}^{3/2+\varepsilon}}{(y_1y_2)^2} \right).$$

Together with (27), this proves (6). The proof of Theorem 1 is complete.

6. Proof of Theorem 2

Let $\phi$ be an $L^2$-normalized Maass cusp form on $X$ as in Theorem 2. Note that $\phi$ is allowed to be non-tempered at the archimedean place, hence its Langlands parameters $\mu := (\mu_1, \mu_2, \mu_3)$ introduced in (9) are not necessarily purely imaginary, but they satisfy by unitarity and the standard Jacquet–Shalika bounds [JS]

$$\max(\lvert \Re \mu_1 \rvert, \lvert \Re \mu_2 \rvert, \lvert \Re \mu_3 \rvert) \leq \frac{1}{2} \quad \text{and} \quad \{\mu_1, \mu_2, \mu_3\} = \{-\mu_1, -\mu_2, -\mu_3\}.$$ 

Note also that we can assume (5) without loss of generality, because we are already assuming $y_1, y_2 \geq \sqrt{3}/2$, and replacing $z$ by $\gamma z$ for an arbitrary unipotent integral matrix $\gamma \in U_3(\mathbb{Z})$ leaves both sides of (7) unchanged.

We shall establish (7) with the help of Selberg’s pre-trace formula. As a preparation, we denote by

$$\mathfrak{a} := \{\text{diag}(\alpha_1, \alpha_2, \alpha_3) \in M_3(\mathbb{R}) : \alpha_1 + \alpha_2 + \alpha_3 = 0\}$$

the Lie algebra of the diagonal torus of $\text{PGL}_3(\mathbb{R})$, and by $C : \text{PGL}_3(\mathbb{R}) \to \mathfrak{a}/S_3$ the Cartan projection induced from the Cartan decomposition for $\text{GL}_3(\mathbb{R})$:

$$g = hk_1 \exp(C(g))k_2 \quad \text{with} \quad h \in Z(\mathbb{R}), \quad k_1, k_2 \in O_3(\mathbb{R}).$$
We identify the complexified dual $a_C^*$ with the set of triples (cf. (9))

$$(\kappa_1, \kappa_2, \kappa_3) \in \mathbb{C}^3 \text{ satisfying } \kappa_1 + \kappa_2 + \kappa_3 = 0,$$

namely such a triple acts on $a_C := a \otimes_R \mathbb{C}$ by the $\mathbb{C}$-linear map

$$\text{diag}(\alpha_1, \alpha_2, \alpha_3) \mapsto \kappa_1 \alpha_1 + \kappa_2 \alpha_2 + \kappa_3 \alpha_3.$$ 

6.1. The pre-trace formula approach. Following [BM, Section 2] and [BP, Sections 2 & 6], we can construct a smooth, bi-$O_3(\mathbb{R})$-invariant function $f_\mu : \text{PGL}_3(\mathbb{R}) \to \mathbb{C}$ supported in a fixed compact subset $K$ (which is independent of $\mu \in a_C^*$) with the following properties. On the one hand, the function obeys (cf. [BM, (2.3)–(2.4)])

$$(28) \quad f_\mu(g) \ll \lambda_{\phi}^{3/2} (1 + \lambda_{\phi}^{1/2} \|C(g)\|)^{-1/2}, \quad g \in \text{PGL}_3(\mathbb{R}),$$

where $\| \cdot \|$ stands for a fixed $S_3$-invariant norm on $a \cong \mathbb{R}^2$. On the other hand, its spherical transform $\tilde{f}_\mu : a_C^*/S_3 \to \mathbb{C}$, defined as in [He, (17) of Section II.3] or [BP, (2.3)], satisfies

$$\tilde{f}_\mu(\mu) \geq 1, \quad \tilde{f}_\mu(\kappa) \geq 0$$

for all Langlands parameters $\kappa \in a_C^*$ occurring in $L^2(X)$, including possibly non-tempered parameters. Then, using positivity in Selberg’s pre-trace formula (see [BP, (6.1)]), or more directly by a Mercer-type pre-trace inequality (cf. [BHMM, (3.15)]), we obtain

$$(29) \quad |\phi(z)|^2 \leq \sum_{\gamma \in \text{PGL}_3(\mathbb{Z})} f_\mu(z^{-1} \gamma z).$$

We can now derive quickly a somewhat weaker version of (7). By (28), we clearly have

$$(30) \quad f_\mu(g) \ll \lambda_{\phi}^{3/2}, \quad g \in \text{PGL}_3(\mathbb{R}),$$

whence by (29),

$$(31) \quad |\phi(z)|^2 \ll \lambda_{\phi}^{3/2} \sum_{\gamma \in \text{PGL}_3(\mathbb{Z})} 1 \text{ for } \|C(z^{-1} \gamma z)\| \ll 1.$$

Here, the uniform bound $\|C(z^{-1} \gamma z)\| \ll 1$ depends on the fixed compact subset $K$ in which $f_\mu$ is supported. Labeling the entries of $\gamma$ as

$$(32) \quad \gamma = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$
a straightforward calculation gives

\[(33) \quad z^{-1} \gamma z = \left( \begin{array}{ccc}
    a - dx_2 + gx_1 x_2 - gx_3 & b h (a - e + h x_1 - dx_2 + gx_1 x_2) x_2 - (h + gx_2) x_3 \\
    (d - gx_1) y_2 & e + dx_2 - (h + gx_2) x_1 & f + cx_1 + dx_3 - (i + h x_1 + gx_2) x_3 \\
    (h + gx_2) y_1 & -gx_1 & i + h x_1 + gx_3
  \end{array} \right)\]

with

\[(34) \quad * = \frac{c + bx_1 - fx_2 + (-e + i + h x_1) x_1 x_2 + ax_3 - (i + h x_1 + dx_2 - gx_1 x_2) x_3 - gx_3}{y_1 y_2}.
\]

Now \(\|C(z^{-1} \gamma z)\| \ll 1\) combined with (5) implies, in this order,

\[(35) \quad g \ll y_1^{-1} y_2^{-1}, \quad h \ll y_1^{-1}, \quad d \ll y_2^{-1},
\]

\[a, e, i \ll 1, \quad f \ll y_1, \quad b \ll y_2, \quad c \ll y_1 y_2.
\]

Therefore, in (31), the number of relevant matrices \(\gamma\) is \(\ll y_1^2 y_2^2\), and we obtain readily

\[(36) \quad \phi(z) \ll \lambda_\phi^{3/4} y_1 y_2.
\]

6.2. An interlude on the Cartan projection. To improve upon the bound (36), we shall make full use of (28) instead of applying its weaker version (30). With this aim in mind, we prove the following estimate on the Cartan projection, which works in general for \(\mathrm{PGL}_n(\mathbb{R})\) or \(\mathrm{SL}_n(\mathbb{R})\).

**Lemma 3:** Let \(g \in \mathrm{SL}_n(\mathbb{R})\) be an upper-triangular matrix with positive diagonal entries, and let \(1_n \in \mathrm{SL}_n(\mathbb{R})\) denote the \(n \times n\) identity matrix. If \(g\) lies in a fixed compact subset of \(\mathrm{SL}_n(\mathbb{R})\), then

\[\|C(g)\| \asymp \|g^t g - 1_n\| \asymp \|g - 1_n\|.
\]

**Proof.** The first relation is standard and follows from the Cartan decomposition \(g = k_1 \exp(C(g)) k_2\) with \(k_1, k_2 \in \mathrm{SO}_n(\mathbb{R})\). So let us focus on the second relation. As \(g\) lies in a fixed compact subset, we have that \(\|g^t g - 1_n\| \asymp \|g^t - g^{-1}\|\), so it suffices to prove that \(\|g^t - g^{-1}\| \asymp \|g - 1_n\|\). Here, \(g^t\) is lower-triangular with some positive diagonal \(a\), and \(g^{-1}\) is upper-triangular with diagonal \(a^{-1}\).
Therefore,
\[
\|g^t - g^{-1}\| \asymp \|g^t - a\| + \|a - a^{-1}\| + \|a^{-1} - g^{-1}\|
\]
\[
\asymp \|g^t - a\| + \|a - 1_n\| + \|1_n - a^{-1}\| + \|a^{-1} - g^{-1}\|
\]
\[
\asymp \|g^t - 1_n\| + \|1_n - g^{-1}\|
\]
\[
\asymp \|g - 1_n\|,
\]
where in the last step we used again that \(g\) lies in a fixed compact subset. We are done. \(\Box\)

We are now ready to improve (36). In the light of (35), it is convenient to rewrite (29) as
\[
|\phi(z)|^2 \leq \sum_{j=1}^{4} \sum_{\gamma \in M_j} f_\mu(z^{-1} \gamma z),
\]
where, in the notation of (32),
\[
M_1 := \{\gamma \in \text{PGL}_3(\mathbb{Z}) : g = 0, h = 0, d = 0\},
\]
\[
M_2 := \{\gamma \in \text{PGL}_3(\mathbb{Z}) : g = 0, h \neq 0, d = 0\},
\]
\[
M_3 := \{\gamma \in \text{PGL}_3(\mathbb{Z}) : g = 0, h = 0, d \neq 0\},
\]
\[
M_4 := \{\gamma \in \text{PGL}_3(\mathbb{Z}) : g \neq 0 \text{ or } hd \neq 0\}.
\]

6.3. The contribution of \(M_1\). In this case, \(g = h = d = 0\) forces \(a, e, i \in \{\pm 1\}\). Then (33)–(34) simplify to
\[
kz^{-1} \gamma z = \begin{pmatrix} 1 & \pm \frac{b+(\pm 1\pm 1)x_2}{y_2} & \pm \frac{c+bx_1-fx_2+(\pm 1\pm 1)x_1x_2+(\pm 1\pm 1)x_3}{y_1} \\ \pm \frac{y_1y_2}{y_1} & \pm \frac{f+(\pm 1\pm 1)x_1}{y_1} & 1 \end{pmatrix},
\]
where \(k\) abbreviates \(\text{diag}(a, e, c) \in O_3(\mathbb{R})\), and all combinations of the \(\pm\) signs are allowed.

For a dyadic parameter \(\lambda_\phi^{-1/2} \ll K \ll 1\), let us examine the contribution to (37) of the matrices \(\gamma \in M_1\) that satisfy \(\|C(z^{-1} \gamma z)\| \asymp K\). Applying (28), we obtain
\[
\sum_{\gamma \in M_1 \atop K \leq \|C(z^{-1} \gamma z)\| < 2K} f_\mu(z^{-1} \gamma z) \ll \lambda_\phi^{5/4} K^{-1/2} \sum_{\gamma \in M_1 \atop K \leq \|C(z^{-1} \gamma z)\| < 2K} 1.
\]
The matrix count on the right-hand side is given by the number of choices for the integral triple \((f, b, c) \in \mathbb{Z}^3\) in (38). Applying Lemma 3 in the form 

\[ \|kz^{-1} \gamma z - 1_3\| \ll K, \]

we infer

\[
\sum_{\gamma \in M_1} 1 \ll \#(f, b, c) \ll (1 + y_1 K)(1 + y_2 K)(1 + y_1 y_2 K).
\]

Keeping in mind that \(y_1, y_2 \gg 1\), we can now estimate the right-hand side of (39) as

\[
\lambda_\phi^{5/4}(K^{-1/2} + y_1 y_2 K^{1/2} + y_1 y_2 K^{3/2} + y_1^2 y_2 K^{3/2} + y_1^2 y_2^2 K^{5/2}).
\]

Summing up over dyadic parameters satisfying \(\lambda_\phi^{-1/2} \ll K \ll 1\), and using \(y_1, y_2 \gg 1\) again, we get

\[
\sum_{\gamma \in M_1} f_\mu(z^{-1} \gamma z) \ll \lambda_\phi^{3/2} + \lambda_\phi^{5/4} y_1 y_2^2.
\]

We estimate the remaining contribution of the matrices \(\gamma \in M_1\) with

\[ \|C(z^{-1} \gamma z)\| < \lambda_\phi^{-1/2} \]

similarly, this time applying (30). In the end, we obtain

\[ \sum_{\gamma \in M_1} f_\mu(z^{-1} \gamma z) \ll \lambda_\phi^{3/2} + \lambda_\phi^{5/4} y_1 y_2^2. \]

6.4. The Contribution of \(M_2\) and \(M_3\). We work out the relevant bound only for \(M_2\), as the argument for \(M_3\) is very similar. As above, \(g = d = 0\) forces \(a = \pm 1\). In this case, since \(h \neq 0\), we see from (35) that \(y_1 \ll 1\) and there are only \(O(1)\) choices for the bottom right \(2 \times 2\) block \((e \ b \atop h \ i) \in \text{GL}_2(\mathbb{Z})\) of \(\gamma\). For each such block, we multiply \(z^{-1} \gamma z\) by an orthogonal matrix \(k \in (O_2(\mathbb{R}))\) on the left to arrive at an upper-triangular matrix with positive diagonal entries.\(^4\)

This multiplies the top row by \(a = \pm 1\), i.e., (33)–(34) become

\[
kz^{-1} \gamma z = \begin{pmatrix}
1 \pm \frac{b+(\pm 1-c+hx_1)x_2-hx_3}{y_2} & \pm \frac{c+bx_1-fx_2+(-e+i+hx_1)x_1x_2\pm x_3-(i+hx_1)x_3}{y_1y_2} \\
\ast & \ast
\end{pmatrix}.
\]

\(^4\) In the case of \(M_3\), we would multiply \(z^{-1} \gamma z\) by an orthogonal matrix \(k \in (O_2(\mathbb{R}))\) on the right, to achieve the same.
For a dyadic parameter $\lambda^{-1/2} \ll K \ll 1$ we estimate, similarly as in (39),

$$\sum_{\gamma \in \mathcal{M}_2, K \leq \|C(z^{-1}\gamma z)\| < 2K} f_\mu(z^{-1}\gamma z) \ll \lambda^{5/4}K^{-1/2} \sum_{\gamma \in \mathcal{M}_2, K \leq \|C(z^{-1}\gamma z)\| < 2K} 1.$$  

Applying Lemma 3 again, we can bound the matrix count on the right hand side as

$$\sum_{\gamma \in \mathcal{M}_2, K \leq \|C(z^{-1}\gamma z)\| < 2K} 1 \ll \#(b,c) \ll (1 + y_2 K)(1 + y_1 y_2 K) \ll 1 + y_2^2 K^2,$$

and a similar calculation as before leads to the bound

$$\sum_{\gamma \in \mathcal{M}_2} f_\mu(z^{-1}\gamma z) \ll \lambda^{3/2} + \lambda^{5/4} y_2^2.$$  

The analogous argument applied to $\mathcal{M}_3$ yields

$$\sum_{\gamma \in \mathcal{M}_3} f_\mu(z^{-1}\gamma z) \ll \lambda^{3/2} + \lambda^{5/4} y_2^2.$$  

6.5. **THE CONTRIBUTION OF $\mathcal{M}_4$.** We see from the first three entries of (35) that $\mathcal{M}_4$ can only contribute when $y_1 \ll 1$ and $y_2 \ll 1$, and then from (35) and (30) we obtain readily that

$$\sum_{\gamma \in \mathcal{M}_4} f_\mu(z^{-1}\gamma z) \ll \lambda^{3/2}.$$  

6.6. **CONCLUSION.** By (37)–(43), we arrive at (7), and the proof of Theorem 2 is complete.

7. **Proof of Theorem 3**

Let $\phi$ be an arithmetically normalized Hecke–Maaß cusp form on $X$ as in Theorem 3. Then (6), and also (7) with an extra factor of $\lambda_\phi^5$, hold in a suitable fundamental domain for $X = \text{GL}_3(\mathbb{Z})\backslash \mathcal{H}_3$ in the light of (5), Theorems 1–2, and the remarks made after Theorem 3. We balance between these two bounds as follows.

If $y_1 y_2 > \lambda_\phi^{7/20}$, then we apply (6) to obtain

$$\phi(z) \ll \varepsilon \min(y_1, y_2) \left( \frac{\lambda_\phi^{1+\varepsilon}}{y_1 y_2} + \frac{\lambda_\phi^{3/2+\varepsilon}}{(y_1 y_2)^{3/2}} \right) \leq \frac{\lambda_\phi^{1+\varepsilon}}{(y_1 y_2)^{1/2}} + \frac{\lambda_\phi^{3/2+\varepsilon}}{(y_1 y_2)^{3/2}} \ll \lambda_\phi^{39/40+\varepsilon}. $$
Else, if $y_1y_2 \leq \lambda_{\phi}^{7/20}$, we apply (7) with an extra factor of $\lambda_{\phi}^\varepsilon$ to obtain
\[ \phi(z) \ll \varepsilon \lambda_{\phi}^{3/4+\varepsilon} + \lambda_{\phi}^{5/8+\varepsilon} y_1y_2 \ll \lambda_{\phi}^{39/40+\varepsilon}. \]

The proof of Theorem 3 is complete.

References

[Ba] C. B. Balogh, Asymptotic expansions of the modified Bessel function of the third kind of imaginary order, SIAM Journal on Applied Mathematics 15 (1967), 1315–1323.

[Bl] V. Blomer, Applications of the Kuznetsov formula on $GL(3)$, Inventiones Mathematicae 194 (2013), 673–729.

[BB] V. Blomer and J. Buttcane, On the subconvexity problem for $L$-functions on $GL(3)$, submitted, arXiv:1504.02667

[BHM] V. Blomer, G. Harcos and D. Miličević, Bounds for eigenforms on arithmetic hyperbolic 3-manifolds, Duke Mathematical Journal 165 (2016), 625–659.

[BHMM] V. Blomer, G. Harcos, P. Maga and D. Miličević, The sup-norm problem for $GL(2)$ over number fields, Journal of the European Mathematical Society, to appear, arXiv:1605.09360

[BH] V. Blomer and R. Holowinsky, Bounding sup-norms of cusp forms of large level, Inventiones Mathematicae 179 (2010), 645–681.

[BM] V. Blomer and P. Maga, Subconvexity for sup-norms of cusp forms on $PGL(n)$, Selecta Mathematica 22 (2016), 1269–1287.

[BP] V. Blomer and A. Pohl, The sup-norm problem on the Siegel modular space of rank two, American Journal of Mathematics 136 (2016), 999–1027.

[Br1] F. Brumley, Second order average estimates on local data of cusp forms, Archiv der Mathematik 87 (2006), 19–32.

[Br2] F. Brumley, Effective multiplicity one on $GL_N$ and narrow zero-free regions for Rankin-Selberg $L$-functions, American Journal of Mathematics 128 (2006), 1455–1474.

[BT] F. Brumley and N. Templier, Large values of cusp forms on $GL_n$, arXiv:1411.4317

[Bu] D. Bump, Automorphic Forms on $GL(3, \mathbb{R})$, Lecture Notes in Mathematics, Vol. 1083, Springer, Berlin, 1984.

[Go] D. Goldfeld, Automorphic Forms and $L$-functions for the Group $GL(n, \mathbb{R})$, Cambridge Studies in Advanced Mathematics, Vol. 99, Cambridge University Press, Cambridge, 2006.

[GR] I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, Elsevier/Academic Press, Amsterdam, 2015.

[HM] G. Harcos and P. Michel, The subconvexity problem for Rankin-Selberg $L$-functions and equidistribution of Heegner points. II, Inventiones Mathematicae 163 (2006), 581–655.

[He] S. Helgason, Geometric Analysis on Symmetric Spaces, Mathematical Surveys and Monographs, Vol. 39, American Mathematical Society, Providence, RI, 2008.
[HRR] R. Holowinsky, G. Ricotta and E. Royer, On the sup-norm of $SL_3$ Hecke-Maass cusp form, arXiv:1404.3622

[Hö] L. Hörmander, The spectral function of an elliptic operator, Acta Mathematica 121 (1968), 193–218.

[IS] H. Iwaniec and P. Sarnak, $L^\infty$ norms of eigenfunctions of arithmetic surfaces, Annals of Mathematics 141 (1995), 301–320.

[JS] H. Jacquet and J. A. Shalika, On Euler products and the classification of automorphic representations. I, American Journal of Mathematics 103 (1981), 499–558.

[La] E. Lapid, On the Harish-Chandra Schwartz space of $G(F)\backslash G(\mathbb{A})$, in Automorphic Representations and $L$-functions, Tata Institute of Fundamental Research Studies in Mathematics, Vol. 22, Tata Institute of Fundamental Research, Mumbai, 2013, pp. 335–377.

[Li] X. Li, Upper bounds on $L$-functions at the edge of the critical strip, International Mathematics Research Notices 4 (2010), 727–755.

[Ra] D. Ramakrishnan, An exercise concerning the selfdual cusp forms on $GL(3)$, Indian Journal of Pure and Applied Mathematics 45 (2014), 777–785.

[RW] D. Ramakrishnan and S. Wang, On the exceptional zeros of Rankin–Selberg $L$-functions, Compositio Mathematica 135 (2003), 211–244.

[Sah1] A. Saha, Large values of newforms on $GL(2)$ with highly ramified central character, International Mathematics Research Notices 13 (2016), 4103–4131.

[Sah2] A. Saha, Hybrid sup-norm bounds for Maass newforms of powerful level, Algebra & Number Theory 11 (2017), 1009–1045.

[Sar] P. Sarnak, Letter to Cathleen Morawetz, August 2004, available at http://publications.ias.edu/sarnak

[St1] E. Stade, On explicit integral formulas for $GL(n, \mathbb{R})$-Whittaker functions, Duke Mathematical Journal 60 (1990), 313–362.

[St2] E. Stade, Mellin transforms of $GL(n, \mathbb{R})$ Whittaker functions, American Journal of Mathematics 123 (2001), 121–161.

[VT] I. Vinogradov and L. Takhtadzhyan, Theory of Eisenstein Series for the group $SL(3, \mathbb{R})$ and its application to a binary problem, Journal of Soviet Mathematics 18 (1982), 293–324.