



HAL
open science

LARGE VALUES OF CUSP FORMS ON GL_n

Farrell Brumley, Nicolas Templier

► **To cite this version:**

Farrell Brumley, Nicolas Templier. LARGE VALUES OF CUSP FORMS ON GL_n . 2014. hal-01408448

HAL Id: hal-01408448

<https://hal.science/hal-01408448v1>

Preprint submitted on 4 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LARGE VALUES OF CUSP FORMS ON GL_n

FARRELL BRUMLEY AND NICOLAS TEMPLIER

ABSTRACT. We establish lower bounds on the sup norm of Hecke–Maass cusp forms on congruence quotients of $GL_n(\mathbb{R})$. The argument relies crucially on uniform estimates for Jacquet-Whittaker functions. These purely local results are of independent interest, and are valid in the more general context of split semi-simple Lie groups. Furthermore, we undertake a fine study of self-dual Jacquet-Whittaker functions on $GL_3(\mathbb{R})$, showing that their large values are governed by the Pearcey function.

CONTENTS

1. Introduction	1
2. Sketch of proofs	7
3. Notation and preliminaries	10
4. Reduction to local estimates	14
5. Rapid decay estimates and the proof of Theorem 1.3 for $GL_n(\mathbb{R})$	18
6. Proof of Theorem 1.5	21
7. Non-degeneracy of phase functions and proof of Theorem 1.3	27
8. Proof of Theorem 1.7	28
9. Proof of Corollary 1.8	36
References	42

1. INTRODUCTION

Let M be a compact Riemannian manifold with non-negative Laplacian Δ . A classical problem in semiclassical analysis asks for sharp local and global bounds on the sup norm of L^2 -normalized Laplacian eigenfunctions. For M of dimension d , Hörmander [32] showed, by purely local considerations, that a Laplacian eigenfunction f of eigenvalue λ satisfies

$$(1.1) \quad \|f\|_\infty \ll \lambda^{\frac{d-1}{4}} \|f\|_2.$$

This bound is sharp for the zonal spherical harmonics on the sphere but for negatively curved M one expects a strong delocalization of eigenstates which would result in power savings.

When M is a non-compact complete Riemannian manifold, the above local bound continues to hold as long as one restricts both the L^∞ and L^2 norms to sufficiently nice bounded subsets of M , such as geodesic balls. Very little is known, however, if one asks for upper bounds (of any quality) on the sup norm over the entire manifold.

The suggestion that the local bound (1.1) might no longer hold in the general non-compact setting can be traced to the work of Donnelly [21], where it is shown that, once the dimension of M is fixed, the implied constant depends on an upper bound on the absolute value of the

Key words and phrases. Sup norms, Maass forms, Whittaker functions, oscillatory integrals, Lagrangian mappings, Pearcey function.

2010 Mathematics Subject Classification. Primary: 11F70. Secondary: 58K55.

sectional curvature and a lower bound on the injectivity radius. Beyond that, no general quantitative results seem to be available.

In this article we are interested in non-compact finite volume Riemannian locally symmetric spaces – the traditional setting of automorphic forms. Thus $M = \Gamma \backslash S$, where S is a Riemannian globally symmetric space and Γ is a nonuniform lattice in the group G of orientation preserving isometries of S . For spaces $\Gamma \backslash S$ where Γ is arithmetic (for example, when the rank G is at least 2), it is known [44] that there are an infinite number of linearly independent bounded and square-integrable Laplacian eigenfunctions, thus a quantitative comparison of their L^∞ and L^2 norms is well-defined. In this degree of generality, there are no known *a priori* bounds; certainly the proof of (1.1) breaks down, for although the curvature of locally symmetric spaces is bounded, the global injectivity radius is zero.

1.1. Sup norm benchmarks in the cusp. Let G be a split real reductive group with Riemannian globally symmetric space S . Let Γ be a non-uniform lattice in G . We consider functions f on $\Gamma \backslash S$ which are eigenfunctions of all intrinsic operators, namely the full ring of G -invariant differential operators $\mathcal{D}_G(S)$ on S . Then f has an infinitesimal character, which we assume to be regular and going to infinity in a dilated sector. We assume furthermore that f is a cusp form, since these are known to decay rapidly at infinity and are thus bounded.

We introduce a constant associated with G that will help us to establish a benchmark *lower* bound for sup norms of (generic) cusp forms on $\Gamma \backslash S$. Let B be a Borel subgroup of G with unipotent radical U . Let $\text{ht}(G)$ be the sum of the heights of the positive roots. Then we define the non-negative half-integer

$$c(G) = (\text{ht}(G) - \dim U)/2.$$

Note that $c(G) = 0$ if and only if G is a product of rank one groups. This constant arises from the theory of oscillatory integrals; we will comment on it shortly.

We shall be particularly interested in the group $\text{GL}_n(\mathbb{R})$, for $n \geq 2$. Write

$$S_n = \text{GL}_n(\mathbb{R})/Z_+ \text{O}(n)$$

for the associated Riemannian globally symmetric space, where Z_+ is the connected component of the identity of the center of $\text{GL}_n(\mathbb{R})$. For Γ a congruence subgroup of $\text{GL}_n(\mathbb{Z})$, our first main theorem furnishes a lower bound on sup norms of cusp forms on $\Gamma \backslash S_n$, with respect to the above constant $c(n) = c(\text{GL}_n(\mathbb{R}))$.

We are only interested in cusp forms f which are eigenfunctions of the Hecke operators. We shall make the additional technical assumption that the infinitesimal character is tempered. For simplicity, we shall refer to functions f satisfying the above conditions (including the conditions of regular and tempered infinitesimal character constrained to a sector) as *Hecke cusp forms*. In the case of $G = \text{GL}_n(\mathbb{R})$, the existence of an infinite number of linearly independent Hecke cusp forms is a well-known result of Müller [50].

Theorem 1.1. *For any Hecke cusp form f on $\Gamma \backslash S_n$ as above, with Laplacian eigenvalue λ , we have*

$$\|f\|_\infty \gg_\varepsilon \lambda^{\frac{c(n)}{2} - \varepsilon} \|f\|_2.$$

As we show in the next section, the lower bounds in Theorem 1.1 are not sharp. Nevertheless, from the explicit values of the exponents

$$\frac{c(n)}{2} = \frac{n(n-1)(n-2)}{24} \quad \text{and} \quad \frac{d-1}{4} = \frac{n^2+n-4}{8},$$

one sees that $\frac{c(n)}{2} > \frac{d-1}{4}$ for all $n \geq 6$. From this we deduce the following corollary.

Corollary 1.2. *For $n \geq 6$, the inequality (1.1) does not hold for any Hecke cusp form f on $\Gamma \backslash S_n$. In fact, for any bounded subset \mathcal{B} of $\Gamma \backslash S_n$, all but finitely many Hecke cusp forms f satisfy*

$$\|f\|_\infty > \sup_{g \in \mathcal{B}} |f(g)|.$$

We may express the latter statement of the corollary as saying that for $n \geq 6$, Hecke cusp forms on $\Gamma \backslash S_n$ achieve their maximum in the cusp, not in the bulk. We conjecture this to be true for all $n \geq 2$. It holds for $n = 5$ by combining Theorem 1.1 with the recent upper bound of Blomer-Maga [12] and Marshall [45].

The exceedingly large values of cusp forms in Theorem 1.1 can be viewed as the semiclassical expression of a result of Kleinbock-Margulis [37], according to which almost all geodesics penetrate the cusp at logarithmic speed $1/\text{ht}(G)$. This reflects the small volume carried by the cusps, creating a bottleneck phenomenon as standing waves transition from an oscillatory to a decay regime. Finally we also would like to mention the result of [47] that for $\Gamma = \text{GL}_n(\mathbb{Z})$, the first eigenvalue satisfies $\lambda \geq (n^3 - n)/24$ which also exhibits a cubic growth in n .

We emphasize that the lower bounds of Theorem 1.1 are of a very different nature than those of Rudnick-Sarnak [54], Milićević [46], or Lapid-Offen [42], all of which show power growth of sup norms of Hecke eigenfunctions. These latter results stem from the functorial (in the sense of Langlands) origin of these eigenfunctions, and their proofs involve compact periods. The behavior of such eigenfunctions in the cusp is thus not reflected in these bounds.

1.2. Lower bounds on Whittaker functions. Theorem 1.1 is deduced from corresponding lower bounds of Whittaker functions, through the Fourier-Whittaker period of f along U in a similar way to [60]. This passage makes use of some special features of the group GL_n , but the bounds on Whittaker functions themselves are valid in wider generality. We thus return to the setting of a split semisimple real Lie group G with associated symmetric space S .

Recall that a Whittaker function on S is a $\mathcal{D}_G(S)$ -eigenfunction W of moderate growth which transforms under the U -action by a non-degenerate additive character ψ . One can think of W as a section of a line bundle defined by ψ over the quotient $U \backslash S$. Whittaker functions on $U \backslash S$ vanish at 0 and are of exponential decay at infinity, so are bounded. They are not, however, square-integrable for the natural quotient measure on $U \backslash S$. Nevertheless, one may normalize W in a natural way using its expression as an oscillatory integral. The existence of such an expression for an arbitrary infinitesimal character (where the integral may not converge absolutely) was proved by Jacquet [35], and the uniqueness by Shalika [57]. Given an infinitesimal character, tools from representation theory then allow one to canonically define a unique Whittaker function on S , which we call the Jacquet-Whittaker function. We shall investigate its L^∞ norm.

Just as we did for the Hecke cusp form f , we suppose that the infinitesimal character of W on S is tempered, regular, and goes to infinity along a ray. This assumption is then implicit in the following result.

Theorem 1.3. *The Jacquet-Whittaker function W on S , with Laplacian eigenvalue λ , satisfies*

$$\|W\|_\infty \gg \lambda^{\frac{c(G)}{2}}.$$

Our methods in fact give more uniform versions of Theorem 1.1 and 1.3 that are valid along the walls of the positive Weyl chamber. To ease the notation, we do not explore this here. In a different context it is interesting to mention [9, Corollary 12.4] which establishes lower bounds for matrix coefficients when the K -types vary.

Remark 1.4. We make two remarks on the general linear group in the formulation and proofs of Theorems 1.1 and 1.3.

i) In the case of $G = \mathrm{GL}_n(\mathbb{R})$, there is a naturally defined inner product (using the mirabolic subgroup) with respect to which any Whittaker function on S is L^2 -integrable and which, moreover, assigns the Jacquet-Whittaker function L^2 -norm 1. One can then express Theorem 1.3 in the usual scale-invariant way as

$$\|W\|_\infty \gg \lambda^{\frac{c(n)}{2}} \|W\|_2.$$

One reflection of this special feature of GL_n is the existence of a formula (due to Stade [59]) relating the L^2 -norm of the Whittaker function to local Rankin-Selberg L -functions. We exploit this fact to give an alternative proof of Theorem 1.3 for $\mathrm{GL}_n(\mathbb{R})$ in §5.3.

ii) The restriction to $G = \mathrm{GL}_n(\mathbb{R})$ in Theorem 1.1 is in part due to the genericity of Hecke cusp forms on $\Gamma \backslash S_n$. This property is used to reduce lower bounds on f to those on any given (non-degenerate) Fourier-Whittaker coefficient. It is well known that cusp forms on other groups may fail to be generic. To extend the statement of Theorem 1.1 to such a setting (using Theorem 1.3), one might either wish to use different special functions and investigate their size, or retain the Fourier-Whittaker coefficients and simply restrict one's attention to the generic spectrum. In either approach, one must be able to control the relation between the L^2 normalization of the cusp form and that of the special function. For Whittaker functions on GL_n this is provided by Rankin-Selberg theory and known bounds on L -functions (see §4). Outside the context of GL_n , recent conjectures of Lapid-Mao [41] may be relevant.

The constant $c(n)$ in Theorem 1.3 arises from the representation of Whittaker functions as oscillatory integrals over U . The $\mathrm{ht}(G)/2$ term can be thought of as the asymptotics of a half-density, while $-\dim(U)/2$ is square-root cancellation over U . The next subsections provide a deeper study of these oscillatory integrals, by examining the regimes where square-root cancellation fails (in which case the lower bound can be improved slightly) due to degeneracies.

1.3. Lagrangian mappings associated with Whittaker functions. We return to the general setting of sup norms on compact Riemannian manifolds.

It is a general principle in semiclassical analysis (see [58, 62]) and the theory of Fourier Integral Operators (see [33, 22]) that eigenfunctions which exhibit extremal L^p growth, if they exist, should concentrate in phase space $T^*(M)$ along certain Lagrangian submanifolds Λ which are invariant under the action of the underlying Hamiltonian dynamics. For example, the zonal spherical harmonics on the sphere saturating the L^∞ bound (1.1) concentrate on the meridian torus Λ consisting of geodesics joining the poles (the antipodal points of the fixed rotation axis). The zonal spherical harmonics achieve their largest values at the poles, which are precisely the singularities of the projection $\Lambda \rightarrow M$.

Similarly, a Whittaker function W , since it can be represented as an oscillatory integral, gives rise to a Lagrangian submanifold Λ . We call $\mathrm{Im}(\Lambda \rightarrow U \backslash S)$ the *essential support* of the Whittaker function W . The singularities of the Lagrangian mapping $\Lambda \rightarrow U \backslash S$ produce large values of W . More generally, the singularities of the Lagrangian mapping $\Lambda \rightarrow U \backslash S$ induce a stratification of Λ according to the degeneracy of the fibers. The type of degeneracy determines (via its singularity index) the corresponding bump in the asymptotics for the Whittaker function W .

There is a certain quantum integrable system whose eigenstates are the spherical Whittaker functions; see for example [38]. The classical integrable system is the Toda lattice [49] which we take to be defined on the symplectic space \mathcal{J}^* of linear functionals in \mathfrak{p}^* vanishing on $[\mathfrak{u}, \mathfrak{u}]$. Here \mathfrak{p} is the tangent space at the origin in S and \mathfrak{u} is the Lie algebra of U . Let $\mathcal{L} \subset \mathcal{J}^*$ be

the isospectral submanifold corresponding to the infinitesimal character of W ; it is a compact Lagrangian submanifold of \mathcal{J}^* .

One of the tools we develop in this paper is an explicit description of $\Lambda \rightarrow U \backslash S$ for symmetric spaces S associated with split semisimple real Lie groups G . We use in an essential way the symplectic reduction of the Hamiltonian action of U on $T^*(S)$. See Theorem 6.2 for a more precise statement.

Theorem 1.5. *The Lagrangian Λ of a spherical Whittaker function embeds as an open subvariety of the Toda Lagrangian \mathcal{L} .*

The complement of the essential support $\text{Im}(\Lambda \rightarrow U \backslash S)$ describes the classically forbidden region of the Toda flow. The corresponding quantum eigenstates – the Whittaker functions – then decay rapidly in this region. So while the archimedean Whittaker functions are not of compact support, the essential support provides a substitute. This is parallel to the theory of Fourier Integral Operators, where we could view W as a distribution whose microlocal support is the Lagrangian Λ . From the above description of \mathcal{L} as an isospectral variety, we may immediately deduce from Theorem 1.5 that all the simple roots of an element in $\text{Im}(\Lambda \rightarrow U \backslash S)$ have size at most $\sqrt{\lambda}$. Information of this sort is a crucial input for the proof of Theorem 1.3.

1.4. Applications to GL_3 . Reduction theory allows us convert the rapid decay of W into that of (generic) Hecke cusp forms. We carry this out for $\text{GL}_3(\mathbb{R})$ and thereby quantify the threshold distance into the cusp beyond which a cusp form on $\Gamma \backslash S_3$ must decay rapidly. Then, by truncating $\Gamma \backslash S_3$ at this threshold, we can establish quantitative *upper* bounds on the sup norm. We obtain the following result.

Proposition 1.6. *In a Siegel domain, any Hecke cusp form f on $\Gamma \backslash S_3$ with Laplacian eigenvalue λ decays rapidly at height greater than λ . Moreover,*

$$(1.2) \quad \|f\|_\infty \ll \lambda^{5/2} \|f\|_2.$$

We expect the bound (1.2) to be very far from the truth. See Remark 4.3 for a discussion of why we have limited the scope of the above Proposition to GL_3 .

We now turn to a refinement of Theorem 1.1 for GL_3 . As was mentioned at the end of §1.2, despite the surprising large exponent $c(n)$ for large n , the proof of Theorem 1.1 does not take into account possible singularities of the underlying oscillatory integrals of Whittaker functions. Let Λ be the Lagrangian submanifold associated to a self-dual spherical Whittaker function on S_3 . In this setting, the following theorem allows us to improve the lower bound of Theorem 1.1 for $n = 3$ by the singularity index of A_3 -type degeneracies.

Theorem 1.7. *The Lagrangian mapping $\Lambda \rightarrow U \backslash S$ induces a Whitney stratification*

$$\Lambda^{(3)} \subset \overline{\Lambda^{(2)}} \subset \overline{\Lambda^{(1)}} = \Lambda$$

where $\Lambda^{(1)}$ is the open dense submanifold of regular points, and

$$\Lambda^{(k)} := \{x \in \Lambda : x \text{ is a type } A_k \text{ singularity}\}.$$

The most singular stratum $\Lambda^{(3)}$ is closed and consists of two points on a single fiber.

Using the method of normal forms of degenerate phase functions we obtain the following consequence of Theorem 1.7. This kind of analysis goes back to [11] where each generic singularity of corank 1 and 2 is studied.

Corollary 1.8. *For any non-zero Whittaker function W on S_3 as above, with self-dual infinitesimal character and Laplacian eigenvalue λ , we have*

$$(1.3) \quad \|W\|_\infty \gg \lambda^{3/8} \|W\|_2,$$

We return to the existence of extremal eigenfunctions on compact Riemannian manifolds. For M of negative curvature, one does not expect strong localisation behavior along Lagrangian submanifolds in phase space. For example, the quantum ergodicity theorem establishes the existence of a density 1 subsequence of an orthonormal basis of eigenfunctions for $L^2(M)$ which do not localise on any proper subvariety of $T^*(M)$.

Nevertheless, for non-compact Riemannian manifolds, there is a sense in which this non-localisation feature of negative curvature asymptotically fails near infinity (see also [16] for a different work in this direction). The idea is that if a cusp form f on $\Gamma \backslash S$ is well-approximated by its Fourier-Whittaker expansion, then f localizes where W does. In particular, this is true of Hecke cusp forms on $\Gamma \backslash S_3$; the large values of W in Theorem 1.7 created by their localization along Λ then transfer to those of f . (This transfer principle from W to f is also used in the proof of Theorem 1.1.)

Corollary 1.9. *For any Hecke cusp form f on $\Gamma \backslash S_3$, with self-dual infinitesimal character and Laplacian eigenvalue λ , we have*

$$\|f\|_\infty \gg_\epsilon \lambda^{3/8-\epsilon} \|f\|_2.$$

A result of Iwaniec-Sarnak [55] states that for Hecke-Maass cusp forms f on the modular surface $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}$ one has

$$(1.4) \quad \|f\|_\infty \gg_\epsilon \lambda^{1/12-\epsilon} \|f\|_2.$$

This is derived from classical estimates on K -Bessel functions in the transition range [7], which in particular imply (for non-zero Whittaker functions W on S_2 , the Poincaré upper half plane)

$$(1.5) \quad \|W\|_\infty \gg \lambda^{1/12} \|W\|_2.$$

The lower bound (1.3) is then the appropriate generalization of (1.5) to S_3 , and Corollary 1.9 is the generalization of (1.4) to $\Gamma \backslash S_3$.

Observe that (1.3) and (1.5) may be rewritten as

$$(1.6) \quad \|W\|_\infty \gg \sqrt{\lambda}^{c(n)+\beta_n} \|W\|_2$$

for $n = 2$ and 3 , where $\beta_n = (1/2) - 1/(n+1)$ is the A_n singularity index in the classification of simple singularities. The A_2 singularity for the $\mathrm{GL}_2(\mathbb{R})$ Whittaker function arises from the “turning point” of the projection of the circle \mathcal{L}_λ centered at the origin in \mathfrak{p}^* of radius $\sqrt{\lambda - 1/4}$ onto the \mathfrak{u}^* -axis. At the fold, the spherical $\mathrm{GL}_2(\mathbb{R})$ -Whittaker function is modelled by the Airy function, the prototypical function which transitions from an oscillatory to a decay regime. Similarly, the Pearcey function is associated with A_3 -type singularities (see Figure 1 and [11]) and it models the peak behavior of spherical $\mathrm{GL}_3(\mathbb{R})$ Whittaker functions.¹

Finally, we remark that the existence of tempered GL_3 cusp forms satisfying the self-dual condition at infinity of Corollary 1.9 can be seen by taking symmetric square lifts (and character twists thereof) of tempered GL_2 Hecke-Maass cusp forms. The restriction to such f should be unnecessary and we have assumed it solely to simplify certain local calculations. Note that locally self-dual at infinity does not imply globally self-dual, as for example is shown by twisting a globally self-dual form by a non-quadratic Dirichlet character.

¹The Airy function is responsible for some natural thresholds encountered in analytic number theory, especially problems having to do with the bounding of periods such as in the work of Bernstein and Reznikov [10]. It will be interesting to see what role the Pearcey function plays in the $\mathrm{GL}(3)$ automorphic theory.

Acknowledgments. We would like to thank Michael Berry, Bill Casselman, Erez Lapid, Simon Marshall, Andre Reznikov, and Peter Sarnak for many enlightening discussions. Some of the results of this paper were first announced at the Oberwolfach workshop on the analytic theory of automorphic forms and further presented at various other meetings, e.g. the Banff workshop on Whittaker functions and Physics, and the 17th Midrasa Mathematicae at Jerusalem. We thank the organizers for these invitations and the participants for their helpful comments.

2. SKETCH OF PROOFS

We now provide a brief sketch of the proofs of the results announced in the introduction. For the reader's benefit, we follow the same subsection structure of the introduction. We have tried to keep the notational overhead to a minimum; for any unexplained notation the reader should consult §3.

2.1. Proof sketch of results in §1.1.

2.1.1. *Reduction of Theorem 1.1 to Theorem 1.3.* The proof of Theorem 1.1 begins by considering the integral of the cusp form f over a closed unipotent orbit against a non-degenerate character. We obtain in this way the global Whittaker function

$$W_f(g) = \int_{(\Gamma \cap U) \backslash U} f(ug) \overline{\psi(u)} du.$$

Since the cycle $(\Gamma \cap U) \backslash U$ is compact, we can deduce lower bounds for f from those of W_f .

The multiplicity one of the spherical Whittaker space and recently proven bounds on Rankin-Selberg L -functions then allow us to replace W_f by the Jacquet-Whittaker function W_ν . As the notation suggests, this latter function is of purely local nature: it sees only the infinitesimal character ν but not the global automorphic form f . This then reduces the proof of Theorem 1.1 to Theorem 1.3.

2.1.2. *Sketch of proof of Theorem 1.3 for $\mathrm{GL}_n(\mathbb{R})$.* In the special case of $G = \mathrm{GL}_n(\mathbb{R})$ we may prove Theorem 1.3 as follows. We consider the zeta integral

$$\Psi(s, W_\nu, \overline{W_\nu}) = \int_{U_{n-1} \backslash \mathrm{GL}_{n-1}(\mathbb{R})} |W_\nu(g)|^2 |\det(g)|^{s-1} dg.$$

Measure identifications and transformation properties of W_ν allow one to write

$$\Psi(s, W_\nu, \overline{W_\nu}) = \int_A |W_\nu(a)|^2 \det(a)^s \delta(a)^{-1} da,$$

up to non-zero absolute constant depending on volume normalization, and the Stade formula (see (5.2)) states that the above integral is equal to the local Rankin-Selberg L -function

$$L(s, \pi_\nu \times \tilde{\pi}_\nu) / L(1, \pi_\nu \times \tilde{\pi}_\nu).$$

Specializing to $s = 1$ we obtain the L^2 -norm squared of W_ν , and we see that it is normalized to be equal to 1. The idea is to take $\mathrm{Re}(s)$ large which puts a greater weight on the region where W_ν has large values, and comparing this with the volume of the region will yield the bound of Theorem 1.3.

Carrying out this strategy, we see from Stirling's formula that $\Psi(\sigma, W_{t\nu}, \overline{W_{t\nu}})$ has size $t^{(\sigma-1)\dim U}$ as $t \rightarrow \infty$. On the other hand, Theorem 1.5 implies that $\Psi(\sigma, W_{t\nu}, \overline{W_{t\nu}})$ is majorized by

$$\max_{a \in A} |W_{t\nu}(a)|^2 \int_{\mathrm{Im}(\Lambda_{t\nu} \rightarrow U \backslash S)} \det(a)^\sigma \delta(a)^{-1} da.$$

For $\sigma > n - 1$, the integral converges to a constant times $t^{(\sigma-1)\dim U - c(n)}$. We deduce the bound $\max_{a \in A} |W_{t\nu}(a)| \gg t^{c(n)}$, as desired.

2.1.3. *Sketch of proof of Theorem 1.3 for general G .* A spherical Whittaker function with infinitesimal character ν is a constant multiple of the oscillatory integral

$$(2.1) \quad W_\nu(a) = \delta(a)^{1/2} \nu(a)^{-1} \int_U \delta(\mathbf{w}u)^{1/2} e^{i(B(H_\nu, H(\mathbf{w}u)) - \ell_1(aua^{-1}))} du,$$

where $a \in A$. See §3 for the notation used in the above expression. The size of the $\delta(a)^{1/2}$ factor is easy to determine; that of the oscillatory integral is more subtle, for the phase function depends on both parameters ν and a .

The method of stationary phase states that the asymptotic of this integral is determined by the critical set of the phase function $B(H_\nu, H(\mathbf{w}u)) - \ell_1(aua^{-1})$ measuring the interaction of the Iwasawa projection $H(\mathbf{w}u)$ (tested by ν) with non-degenerate characters $u \mapsto e^{2\pi i \ell_1(aua^{-1})}$. If there are no critical points, then the integral decays rapidly, overwhelming the polynomial growth of $\delta(a)^{1/2}$. If there do exist critical points, then the asymptotic size of the above integral is governed by local contributions around each one. A non-degenerate critical point makes a contribution of size $t^{-\dim U/2}$. A degenerate critical point will make a larger contribution, of size $t^{-\dim U/2 + \beta}$ for a certain rational number β which is a numerical invariant of the degeneracy.

To prove Theorem 1.3 we show in §7 that for every ν there exists a such that the above phase function admits non-degenerate critical points whose local contributions do not cancel. For this, we follow the approach of Duistermaat [22] and Hörmander [33], as suggested to us by Simon Marshall. Now, in standard presentations in the FIO literature, the emphasis is on upper bounds for L^p estimates, so the symbol is chosen to be transverse to the Lagrangian Λ . For our purpose of establishing a lower bound we make the opposite choice of a symbol which is tangent to Λ . The modified phase $B(H_\nu, H(\mathbf{w}u)) - \ell_1(aua^{-1}) - \langle \xi, a \rangle$ is then Morse-Bott. This produces a *lower bound* (not necessarily sharp, since at degeneracies the lower bound could be stronger) on the oscillatory integral of size $t^{-\dim U/2}$. When the size of half-density $\delta(a)^{1/2}$ is taken into account, this yields the exponent $c(n)$.

2.2. **Theorem 1.5 and the method of co-adjoint orbits.** We now give an intuitive explanation for why one should expect to realize the Whittaker Lagrangian Λ_ν in \mathcal{L}_ν , as stated in Theorem 1.5. Our inspiration is the geometric setting of the *method of co-adjoint orbits*.

Consider the action of G on the space of linear functionals \mathfrak{g}^* given by the co-adjoint action. For $g \in G$ and $\lambda \in \mathfrak{g}^*$ this is defined as $\text{Ad}_g^* \lambda = \lambda \circ \text{Ad}_{g^{-1}}$, where $\text{Ad} : G \rightarrow \text{Aut}(\mathfrak{g})$ is the adjoint representation. The orbits under this action are endowed with a natural G -invariant symplectic form, which at a point λ is given by the formula $\Omega_\lambda(X, Y) = -\lambda([X, Y])$. The action of G on an orbit \mathcal{O} is Hamiltonian with corresponding moment map the inclusion $\Phi_G : \mathcal{O} \hookrightarrow \mathfrak{g}^*$.

We are particularly interested in co-adjoint orbits attached to $\xi \in \mathfrak{p}^*$. A natural way of obtaining them is to first consider the cotangent bundle $T^*(S)$. This receives a G -action inherited from the natural G -action on S by isometries. We make the equivariant identification $T^*(S) = G \times_K \mathfrak{p}^*$ under which the moment map $T^*(S) \rightarrow \mathfrak{g}^*$ for the G -action is described by $[g, \xi] \mapsto \text{Ad}_g^* \xi$. Then the image of any G -orbit in $T^*(S)$ is a coadjoint orbit in \mathfrak{g}^* associated with some $\xi \in \mathfrak{p}^*$.

The method of co-adjoint orbits roughly states that, in favorable circumstances, irreducible unitary representations π of G will be in finite-to-one correspondence with co-adjoint orbits \mathcal{O} . The association of a unitary representation with the Hamiltonian system of the symplectic G -manifold \mathcal{O} is referred to as *geometric quantization*. Moreover, operations in the unitary

dual (e.g. induction, restriction) should in principle correspond to operations on corresponding orbits (e.g. intersection, projection). Other rough parallels exist; for example, the uncertainty principle is expressed in this set-up as a correspondence between the vectors in the unitary representation π and balls of unit volume in the co-adjoint orbit $\mathcal{O} = \mathcal{O}_\pi$.

Of greatest interest to us is the following situation. For a subgroup K of G with Lie algebra \mathfrak{k} , the level sets of the corresponding moment map $\Phi_K : \mathcal{O} \hookrightarrow \mathfrak{g}^* \rightarrow \mathfrak{k}^*$ should correspond to phase states with quantities conserved by K . For example, taking K to be a maximal compact subgroup, spherical Whittaker functions W are associated with K -fixed vectors of irreducible unitary unramified representations of G . From the tempered hypothesis on W , these representations are obtained by induction from some $\nu \in i\mathfrak{a}^* \subset i\mathfrak{p}^*$. Letting \mathcal{O} be the coadjoint orbit of $\text{Im}(\nu) \in \mathfrak{p}^*$, isolating $\Phi_K^{-1}(0)$ in \mathcal{O} then corresponds to picking out K -fixed vectors in π .

Furthermore, given two subgroups $U, K < G$, one can hope to understand the U -isotypic distribution of a K -fixed vector in π via the projection map $\Phi_K^{-1}(0) \rightarrow \mathfrak{u}^*$. Theorem 1.5 carries out this yoga for K a maximal compact subgroup of G and U the unipotent radical of a Borel.

On one hand, the Toda Lagrangian \mathcal{L}_ν is the intersection $\Phi_K^{-1}(0) \cap \Phi_{U_{\text{der}}}^{-1}(0)$ in the coadjoint orbit \mathcal{O} (see §6.2), where $U_{\text{der}} = [U, U]$ is the commutator subgroup. This intersection then admits a Lagrangian mapping to $\mathfrak{u}_{\text{ab}}^*$, with $\mathfrak{u}_{\text{ab}}^*$ denoting the characters of \mathfrak{u} vanishing on $\mathfrak{u}_{\text{der}}$. When $G = \text{GL}_2(\mathbb{R})$, for example, one obtains the projection from the circle of radius ξ to the \mathfrak{u}^* -axis. On the other, the Whittaker Lagrangian Λ_ν admits a similar description with respect to the moment maps arising from the natural G -action on the cotangent bundle $T^*S \rightarrow S$ (see Proposition 6.1). Reducing Λ_ν by the U -action then defines an open embedding from $\Lambda_\nu^{\text{red}} \rightarrow U \backslash S$ into $\mathcal{L}_\nu \rightarrow \mathfrak{u}_{\text{ab}}^*$. This is the statement of Theorem 6.2, which makes more precise Theorem 1.5 from the introduction.

2.3. Proof sketch of results in §1.4.

2.3.1. *Sketch of proof of Proposition 1.6.* To establish the rapid decay of f high in the cusp, one first expands f in its Fourier-Whittaker expansion. One must then check that every term in the expansion is itself evaluated high enough into the cusp for the decay estimates of Theorem 1.5 to apply; this is an exercise in reduction theory, which we carry out for $\text{GL}_3(\mathbb{R})$. In this way, the decay estimate on Whittaker functions of Theorem 1.5 transfers, at least for $n = 3$, to the cusp form f .

To deduce an upper bound on the sup norm of f from a quantitative estimate of its essential support, we argue as follows. First recall a result of Sarnak [55] which states that a cusp form f of eigenvalue λ on a compact locally symmetric space of dimension d and rank r satisfies

$$(2.2) \quad \|f\|_\infty \ll \lambda^{(d-r)/4} \|f\|_2.$$

In fact, this holds for non-compact locally symmetric spaces as well, as long as one restricts to nice enough bounded subsets, such as geodesic balls. The key is that the quantitative dependence of the implied constant on the injectivity radius in (2.2) is rather easy to explicate. So we simply go through Sarnak's proof of (2.2) on the *truncation* of $\Gamma \backslash S_3$ to the essential support of f , since it has positive calculable global injectivity radius.

2.3.2. *Proof sketch of Theorem 1.7 and Corollary 1.9.* The description of Λ_ν given in Theorem 1.5 is convenient for computations: roughly speaking, the equations defining the fiber over $a \in U \backslash S = A$ are the tridiagonal symmetric matrices with off-diagonals the positive simple roots of a and characteristic polynomial agreeing with that of ν . For $G = \text{PGL}_3(\mathbb{R})$ and ν self-dual, this boils down to the following problem.

Let $t > 1$. Let \mathcal{J} denote the real tridiagonal symmetric 3×3 matrices. Determine the intersection configuration of the solutions $s \in \mathcal{J}$ having fixed non-zero off-diagonal entries to the cubic equation $\det(s) = 0$ and the quadratic equation $\|s\| = t^2$.

§8 is dedicated to the solution of this problem. In particular, the A_3 singularities are created when the two equations have two intersection points, both with multiplicity 3. Stationary phase asymptotics for A_3 singularities then produce the $\lambda^{3/8}$ lower bound for the corresponding spherical Whittaker function. Finally, to deduce the bounds on the cusp form f as stated in Corollary 1.9 one follows the argument sketched in §2.1.1.

3. NOTATION AND PRELIMINARIES

In this section establish basic notation that we'll need for later calculations. We will take G to be a split semi-simple real Lie group throughout this section.

3.1. Basic notation on roots. Let Θ denote a Cartan involution on G . Denote by θ the differential of Θ on \mathfrak{g} , the (real) Lie algebra of G . One has an orthogonal direct sum decomposition $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$ into the -1 and $+1$ eigenspaces of θ . Then \mathfrak{k} is the Lie algebra of K , the group of fixed points of Θ .

Choose a maximal abelian subalgebra \mathfrak{a} of \mathfrak{p} preserved by θ . The Weyl group of G is $W = W(\mathfrak{g}, \mathfrak{a}) = N_K(\mathfrak{a})/Z_K(\mathfrak{a})$, the quotient of the normalizer by the centralizer of the adjoint action of K on \mathfrak{a} . Denote by $A = \exp(\mathfrak{a})$ the associated closed connected subgroup of G . Then A is a maximal split torus of G preserved by Θ . Let $\mathfrak{a}^* = \text{Hom}(\mathfrak{a}, \mathbb{R})$ be the dual of \mathfrak{a} and $\mathfrak{a}_{\mathbb{C}}^* = \mathfrak{a}^* \otimes \mathbb{C} = \text{Hom}(\mathfrak{a}, \mathbb{C}) = \mathfrak{a}^* + i\mathfrak{a}^*$ its complexification. We agree to the notational convention for which $\langle \nu, X \rangle$ is the evaluation of $\nu \in \mathfrak{a}_{\mathbb{C}}^*$ at $X \in \mathfrak{a}$. Moreover, when $a \in A$ we write $\nu(a)$ for $e^{\langle \nu, \log a \rangle}$.

Let $\Delta = \Delta(\mathfrak{g}, \mathfrak{a})$ denote the set of (restricted) roots. We have $\mathfrak{g} = \mathfrak{a} \oplus \bigoplus_{\alpha \in \Delta} \mathfrak{g}_{\alpha}$, with each \mathfrak{g}_{α} of dimension one. For $\alpha \in \Delta$ let H_{α} be the corresponding co-root; this, by definition, is the unique element in $\mathfrak{a}_{\alpha} = [\mathfrak{g}_{\alpha}, \mathfrak{g}_{-\alpha}] \subseteq \mathfrak{a}$ such that $\langle \alpha, H_{\alpha} \rangle = 2$. For a root $\alpha \in \Delta$ let $X_{\alpha} \in \mathfrak{g}_{\alpha}$ be defined by $[X_{-\alpha}, X_{\alpha}] = H_{\alpha}$. The choice of a system of simple roots Π determines a set of positive roots Δ_+ . Let $\mathfrak{u} = \bigoplus_{\alpha \in \Delta_+} \mathfrak{g}_{\alpha}$ and $\bar{\mathfrak{u}} = \bigoplus_{\alpha \in \Delta_+} \mathfrak{g}_{-\alpha}$. Let $\rho \in \mathfrak{a}^*$ be half the sum of the positive roots; thus $\langle \rho, \cdot \rangle$ is half the trace of the adjoint action on \mathfrak{u} .

Let U and \bar{U} be the connected closed subgroups of G whose Lie algebras are \mathfrak{u} and $\bar{\mathfrak{u}}$, respectively. We have $\bar{U} = \Theta U$. Let B be the unique Borel subgroup of G containing A and U . Then U is the unipotent radical of B , the Lie algebra of B is $\mathfrak{b} = \mathfrak{a} \oplus \mathfrak{u}$, and one has the Langlands decomposition $B = MAU$, where $M = B \cap K$.

Denote by $\text{Ad} : G \rightarrow \text{Aut}(\mathfrak{g})$ the adjoint representation. For $g \in G$ and $X \in \mathfrak{g}$ we will often use the shorthand X^g to denote $\text{Ad}_{g^{-1}}X$. (The inverse in the latter notation is there for the right-action rule $X^{gh} = (X^g)^h$ to hold.) Similarly, for $g, z \in G$ we write $z^g = g^{-1}zg$. For $g \in G$ and $\lambda \in \mathfrak{g}^*$ we let $\text{Ad}_g^*\lambda := \lambda \circ \text{Ad}_{g^{-1}}$.

Fix a choice $B(\cdot, \cdot) : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathbb{R}$ of Ad-invariant non-degenerate symmetric bilinear form, normalized to be positive definite on \mathfrak{p} . Then $-B(X, \theta Y)$ is positive definite on \mathfrak{g} ; let $\|X\|^2 = -B(X, \theta X)$ be the associated norm on elements of \mathfrak{g} . The restriction of $B(\cdot, \cdot)$ to \mathfrak{a} defines a positive definite bilinear form. We use $B(\cdot, \cdot)$ to identify \mathfrak{a}^* with \mathfrak{a} . Furthermore, we can extend $B(\cdot, \cdot)$ to a hermitian scalar product on $\mathfrak{a}_{\mathbb{C}}$, allowing us to identify $\mathfrak{a}_{\mathbb{C}}^*$ with $\mathfrak{a}_{\mathbb{C}}$. For $\xi \in \mathfrak{a}^*$, we let H_{ξ} denote the unique element in \mathfrak{a} such that $\langle \xi, H \rangle = B(H_{\xi}, H)$ for every $H \in \mathfrak{a}$. If $\nu \in i\mathfrak{a}^*$ with $\xi = \text{Re } \nu$, we write H_{ν} for H_{ξ} .

The root hyperplane (or wall) associated to the element $\alpha \in \Delta$ is the linear subspace of \mathfrak{a} on which it vanishes. The Weyl chambers are the connected components of the complement of all walls in \mathfrak{a} . The union of all Weyl chambers is the set $\mathfrak{a}_{\text{reg}}$ of regular elements. Let \mathfrak{a}_+ (resp. \mathfrak{a}_+^*) denote the positive Weyl chamber in \mathfrak{a} (resp. \mathfrak{a}^*). The Weyl group acts simply

transitively on the Weyl chambers. An element H is regular if and only if $H^w = H$ for some $w \in W$ implies $w = e$. The long Weyl element, which we denote by w , sends \mathfrak{a}_+ to $-\mathfrak{a}_+$. We make once and for all a choice of a lift of the longest Weyl group element to an element in K and we continue to write it as w .

3.2. Iwasawa decomposition. The Iwasawa decomposition is $G = UAK$. We denote by $\kappa(g)$ the unique element in K such that $g\kappa(g)^{-1} \in AU$, and $\tau(g) = g\kappa(g)^{-1}$.

For $a \in A$ let $\delta(a) = |\det(\text{Ad}(a)|_U)|$, the Jacobian of the automorphism of U sending u to aua^{-1} . Thus, if du is any Haar measure on U , then $\int_U f(aua^{-1})du = \delta(a) \int_U f(u)du$. Since $a \in A$ acts on $X \in \mathfrak{g}_\alpha$ via the adjoint action by multiplication by $\langle \alpha, \log a \rangle$ we have $\delta(a) = \prod_{\alpha \in \Delta_+} \alpha(a) = \rho(a)^2$. For any choice of left-invariant Haar measures du, da, dk on U, A , and K , respectively, the product measure $dg = \delta(a)^{-1} du da dk$ defines a left-invariant Haar measure on G .

Recall the Iwasawa decomposition of the Lie algebra $\mathfrak{g} = \mathfrak{u} \oplus \mathfrak{a} \oplus \mathfrak{k}$. We denote by $E_\mathfrak{a}$ (resp., $E_\mathfrak{u}, E_\mathfrak{k}$) the projection from \mathfrak{g} onto \mathfrak{a} (resp., $\mathfrak{u}, \mathfrak{k}$). Note that unlike $E_\mathfrak{a}$, the projections $E_\mathfrak{u}, E_\mathfrak{k}$ are not orthogonal with respect to B .

The map $H : G \rightarrow \mathfrak{a}$ sending $g = ue^Xk$ ($u \in U, X \in \mathfrak{a}, k \in K$) to X is called the Iwasawa projection. Its derivative was computed in [24, Corollary 5.2]. We state and prove a consequence of this which will be useful for us in §6.

Lemma 3.1. *For $\xi \in \mathfrak{a}^*$, the directional derivative along $X \in \mathfrak{g}$ of the function $g \mapsto \langle \xi, H(g) \rangle$ is $\langle \xi, X^{\kappa(g)^{-1}} \rangle$.*

Proof. From the linearity of $\xi \mapsto \langle \xi, H \rangle$, we may pass the derivative inside the bracket. From [24, Lemma 5.1] the directional derivative along X of the Iwasawa projection $g \mapsto H(g)$ is $E_\mathfrak{a}(X^{\kappa(g)^{-1}})$. This gives $X \cdot \langle \xi, H(g) \rangle = \langle \xi, E_\mathfrak{a}(X^{\kappa(g)^{-1}}) \rangle$. As \mathfrak{a} is orthogonal to $\mathfrak{k} \oplus \mathfrak{u}$ we have $\langle \xi, E_\mathfrak{a}(X^{\kappa(g)^{-1}}) \rangle = \langle \xi, X^{\kappa(g)^{-1}} \rangle$, as desired. \square

If we apply the above lemma to the case when $g = wu$, for $u \in U$ then we recover the following result which appears in Cohn [17, Proposition 9.1]. Let $\xi \in \mathfrak{a}^*$ be given. Then u is a critical point of $\langle \xi, H(wu) \rangle$ if and only if $\xi^{\kappa(wu)} \in \mathfrak{a}^*$. In particular, if ξ is regular then the only critical point of $\langle \xi, H(wu) \rangle$ is the identity e .

3.3. Bruhat decomposition. Next we recall the Bruhat decomposition,

$$G = \bigsqcup_{w \in W} G_w,$$

where $G_w = BwU_w$ and $U_w = U \cap (w^{-1}\bar{U}w)$. The cell $G_w = BwU$ associated with the long Weyl element w is called the *big cell*; it is open and dense in G . For any $w \in W$ let \mathfrak{u}_w denote the Lie algebra of U_w . Note that $\mathfrak{u}^w = \mathfrak{u}$. We have

$$\mathfrak{u}_w = \bigoplus_{\alpha \in \Delta_+(w)} \mathfrak{g}_\alpha, \quad \text{where } \Delta_+(w) = \{\alpha \in \Delta_+ : -w\alpha \in \Delta_+\}.$$

Write \mathfrak{u}^w for the direct sum of the \mathfrak{g}_α for $\alpha \in \Delta_+ - \Delta_+(w)$, so that $\mathfrak{u} = \mathfrak{u}_w \oplus \mathfrak{u}^w$.

We call an element $\ell \in \mathfrak{u}^*$ *degenerate* if it vanishes identically on some simple root space \mathfrak{g}_α , $\alpha \in \Pi$. We call it *non-degenerate* otherwise. Since at least one of the roots in $\Delta_+ - \Delta_+(w)$ is simple, ℓ is degenerate if and only if it belongs to \mathfrak{u}_w^* for some $w \neq w$. The set of non-degenerate functionals is therefore equal to $\mathfrak{u}^* - \bigcup_{w \neq w} \mathfrak{u}_w^*$.

The Bruhat decomposition of G gives rise to a cellular decomposition on the flag variety $B \backslash G$. By definition, these cells are the orbits of the cosets Bw , where $w \in W$, under the natural right-action of U on $B \backslash G$. When we make the identification $B \backslash G = M \backslash K$, the action

of U on $B \backslash G$ induces a right-action of U on $M \backslash K$ given by $(k, u) \mapsto M\kappa(ku)$. The Bruhat cell $B \backslash BwU$ is then identified with the image S_w^+ of the map

$$U \rightarrow M \backslash K, \quad u \mapsto M\kappa(wu).$$

We thus obtain the following decomposition

$$M \backslash K = \bigsqcup_{w \in W} S_w^+.$$

We are borrowing the notation S_w^+ (for *stable manifold*) from [24, §3]. We note that S_w^+ are stable under right M -action. Moreover under the inversion $k \mapsto k^{-1}$, the cell S_w^+ is mapped bijectively to $S_{w^{-1}}^+$.

Compare the following result to [*loc. cit.*, Proposition 7.1].

Lemma 3.2. *For $w \in W$ the differential of the above map $U \rightarrow S_w^+$ is given by*

$$(3.1) \quad d\kappa(wu)(Y) = E_t(Y^{\kappa(wu)^{-1}})^{\kappa(wu)}.$$

The restriction to U_w induces a diffeomorphism of U_w onto S_w^+ .

Proof. We begin by writing $wu = \tau(wu)\kappa(wu)$; thus for any $t \in \mathbb{R}$ we have $wue^{tY} = \tau(wu)e^{tY^k}k^{-1}$, where we have set $k = \kappa(wu)^{-1}$. Then $d\kappa(wu)(Y)$ is equal to

$$\frac{d}{dt}\kappa(wue^{tY})|_{t=0} = \frac{d}{dt}\kappa(e^{tY^k}k^{-1})|_{t=0} = \frac{d}{dt}e^{tE_t(Y^k)}k^{-1}|_{t=0}.$$

Conjugating this by k , we obtain the desired formula.

For the second statement, it suffices to observe that the isotropy subgroup of the point Bw for the U -action on $B \backslash G$ is the analytic subgroup $U^w = U \cap w^{-1}Uw$ of G whose Lie algebra is \mathfrak{u}^w . Since $U = U_w U^w$ and $U_w \cap U^w = \{e\}$ the claim follows. \square

3.4. Bruhat cells. In preparation for §6 we collect some information about Bruhat cells and how they relate to coadjoint K -orbits.

For any $\xi \in \mathfrak{a}^*$ denote by K_ξ the centralizer of ξ in K . Note that $M \subset K_\xi$, and in fact $M = K_\xi$ for ξ regular. Thus $M \backslash K$ maps surjectively to $K_\xi \backslash K$, which itself can be identified with $\text{Ad}_K^*(\xi)$. We thereby obtain a decomposition

$$(3.2) \quad \text{Ad}_K^*(\xi) = \bigcup_{w \in W} \text{Ad}_{S_w^+}^*(\xi).$$

It is a disjoint union if ξ is regular which we however do not assume in this subsection.

Lemma 3.3. *Let $s \in \text{Ad}_K^*(\xi)$ for some $\xi \in \mathfrak{a}^*$. If the restriction of s to \mathfrak{u} is non-degenerate then $s \in \text{Ad}_{S_w^+}^*(\xi)$.*

Proof. In view of the decomposition (3.2) it suffices to show that $s \in \text{Ad}_{S_w^+}^*(\xi)$ implies that $w = w$. Assume that $s \in \text{Ad}_{S_{w^{-1}}^+}^*(\xi)$. Thus we may assume that $s = \text{Ad}_{k^{-1}}^*(\xi)$ where $k = \kappa(wu)$ for some $u \in U$. Then for any $X \in \mathfrak{u}^w$,

$$H(we^X u) = H(e^{\text{Ad}_w X} wu) = H(wu),$$

because $X^w \in \mathfrak{u}$. Therefore we have $X.H(wu) = 0$, and thus $X.\langle \xi, H(wu) \rangle = 0$. By Lemma 3.1 this is equivalent to $\langle \xi, \text{Ad}_k(X) \rangle = 0$, and therefore $\langle s, X \rangle = 0$ for all $X \in \mathfrak{u}^w$. In other words s vanishes on \mathfrak{u}^w , therefore the restriction of s to \mathfrak{u} belongs to \mathfrak{u}_w^* . But this restriction is non-degenerate by hypothesis; thus $w = w$ which concludes the proof. \square

It is instructive to compare this with [24]. Let $\Omega_w := wS_w^+$, the translate by $w \in W$ of the big cell. Then the Ω_w form an open covering of $M \setminus K$ (cf. [loc. cit., Corollary 3.8, Proposition 7.1]), and we have

$$\bigcap_{w \in W} \text{Ad}_{S_w^+}^*(\xi^w) = \bigcap_{w \in W} \text{Ad}_{\Omega_w}^*(\xi).$$

If the restriction of s to \mathfrak{u} is non-degenerate then s belongs to this intersection. Indeed this follows immediately from Lemma 3.3 since $\text{Ad}_K^*(\xi) = \text{Ad}_K^*(\xi^w)$ for all $w \in W$.

For example, when $n = 2$, such an s belongs to the circle $\text{Ad}_K^*(\xi)$ minus the two points $\{\xi, \xi^w\}$. Thus we have removed in this case the hyperplane \mathfrak{a}^* which is the kernel of the projection $\mathfrak{p}^* \rightarrow \mathfrak{u}^*$.

3.5. Spherical representations and invariants. Let $S = G/K$ be the globally Riemannian symmetric space associated with G , and $\mathcal{D}_G(S)$ of left G -invariant differential operators on S . The Harish-Chandra isomorphism identifies the differential eigencharacters $\text{Hom}(\mathcal{D}_G(S), \mathbb{C})$ with the space of (spherical) infinitesimal characters $\mathfrak{a}_{\mathbb{C}}^*/W$. For $\nu \in \mathfrak{a}_{\mathbb{C}}^*$ let λ_{ν} be the associated Laplacian eigenvalue given by evaluating the associated differential eigencharacter on Δ . The Laplacian being an order two differential operator, when we scale ν by $t > 1$ we obtain $\lambda_{t\nu} \asymp t^2 \lambda_{\nu}$.

For $\nu \in i\mathfrak{a}^*$ consider the representation of G by right-translation on the space of smooth functions $f : G \rightarrow \mathbb{C}$ satisfying

$$f(bg) = f(g)\delta(b)^{1/2}e^{\langle \nu, H(b) \rangle} \quad g \in G, b \in B.$$

The inner product $\int_K f_1(k)\overline{f_2(k)}dk$, where dk is the probability Haar measure on K is G -invariant. We denote by π_{ν} the completion of this space relative to this normalized inner product. Then π_{ν} is an irreducible unitary spherical tempered representation. We have $\pi_{\nu} \simeq \pi_{\nu'}$ if and only if $\nu = w\nu'$ for some $w \in W$. We shall only be interested in ν regular; so that $w\nu \neq \nu$ unless $w = e$. The isomorphism classes of irreducible unitary regular tempered spherical representations of G are parametrized by ν lying in the positive chamber $i\mathfrak{a}_+^*$.

We define the *height of G* to be

$$\text{ht}(G) = \sum_{\alpha \in \Delta^+} \text{ht}(\alpha),$$

where $\text{ht}(\alpha)$ is the sum of the coefficients of α when written as a linear combination of the positive simple roots. The height of G has the following property: for an element $a \in A$ and a positive real $t > 0$ let ta be the unique element in A whose simple roots satisfy $\alpha(ta) = t\alpha(a)$ for all $\alpha \in \Pi$. Then one easily deduces that

$$\delta(ta) = t^{\text{ht}(G)}\delta(a).$$

In particular, the height of G describes the size of the *spherical vector* in π_{ν} along directions ta . Recall that the spherical vector in π_{ν} is the unique K -fixed vector taking value 1 at the identity. It has L^2 -norm 1 and is given by the expression

$$f_{\nu}(g) = e^{\langle \rho + \nu, H(g) \rangle} = \delta(g)^{1/2}e^{\langle \nu, H(g) \rangle}.$$

Here and elsewhere, $\delta(g) = \delta(a)$ if $g = uak$; alternatively, $\delta(g) = e^{2\langle \rho, H(g) \rangle}$.

3.6. Whittaker models, phase functions, and associated Lagrangians. We now describe various Whittaker structures associated with the above representations π_{ν} .

Let ψ be a unitary character of U . Then ψ factors through $U_{\text{der}} = [U, U]$ and since the abelianization of U is $U_{\text{ab}} = U/U_{\text{der}} = \prod_{\alpha \in \Pi} U_{\alpha}$, where U_{α} is the analytic subgroup with Lie

algebra \mathfrak{g}_α , we may factorize ψ as $\psi = \prod_{\alpha \in \Pi} \psi_\alpha$. We call ψ *non-degenerate* if each ψ_α is non-trivial. We denote by $\psi_1 = \prod_{\alpha \in \Pi} \psi_{1,\alpha}$ the unique character of U such that $\psi_{1,\alpha}(u_\alpha) = e^{2\pi i u_\alpha}$ for all $\alpha \in \Pi$. We let ℓ_1 be the unique element in $\mathfrak{u}_{\text{ab}}^*$ such that $\psi_1(u) = e^{i\langle \ell_1, u \rangle}$.

Now consider the space $C^\infty(U \backslash G, \psi)$ of smooth functions W on G satisfying the transformation formula $W(ug) = \psi(u)W(g)$ for all $g \in G$ and $u \in U$. Then G acts on $C^\infty(U \backslash G, \psi)$ by right-translation. This is the Whittaker space associated to ψ ; it is the induction to G of the one-dimensional representation ψ of U .

For $\nu \in i\mathfrak{a}^*$, one can define [35, 56] a non-zero linear form on π_ν by setting

$$\mathbb{J}^\psi(f) = \int_U f(wu) \overline{\psi(u)} du,$$

a conditionally convergent integral. One readily verifies that for $u \in U$, $\mathbb{J}^\psi(\pi_\nu(u)f) = \psi(u)\mathbb{J}^\psi(f)$, so that $0 \neq \mathbb{J}^\psi \in \text{Hom}_U(\pi_\nu, \psi)$. It is known that $\dim_{\mathbb{C}} \text{Hom}_U(\pi_\nu, \psi) = 1$. Thus \mathbb{J}^ψ is the unique non-zero element up to scaling. We can replace f by its translate by a group element to form $\mathbb{J}^\psi(\pi_\nu(g)f)$, which as a function on G satisfies $\mathbb{J}^\psi(\pi_\nu(ug)f) = \psi(u)\mathbb{J}^\psi(\pi_\nu(g)f)$ for every $g \in G$ and $u \in U$. The assignment $f \mapsto \mathbb{J}^\psi(\pi_\nu(\cdot)f)$ is a non-zero intertwining from (the smooth subspace of) π_ν to $C^\infty(U \backslash G, \psi)$. We denote the image by $\mathcal{W}(\pi_\nu, \psi)$ and refer to it as the Whittaker model of π_ν .

Let W_ν^ψ denote the image of the spherical function $f_\nu \in \text{Ind}(\nu)^\infty$ under this intertwining: $W_\nu^\psi(g) = \mathbb{J}^\psi(\pi_\nu(g)f_\nu)$ for $g \in G$. This is the Jacquet-Whittaker function, given explicitly by

$$W_\nu^\psi(g) = \int_U \delta(wug)^{1/2} e^{iB(H_\nu, H(wug))} \overline{\psi(u)} du.$$

Clearly W_ν^ψ lies in $\mathcal{W}(\pi_\nu, \psi)^K$, the one-dimensional space of K -fixed vectors in $\mathcal{W}(\pi_\nu, \psi)$. When $\psi = \psi_1$ we simplify the notation and write W_ν in place of $W_\nu^{\psi_1}$. From the above integral we may extract the oscillatory dependence via

$$(3.3) \quad F_\nu(u, g) = B(H_\nu, H(wug)) - \langle \ell_1, u \rangle,$$

the *Whittaker phase function*. By the right K -invariance in the second variable we often view F_ν as a function on $U \times S$, and write $F_\nu(u, x)$ for $x = gK$. It is easily seen that a change of variables produces the alternative expression (2.1).

Denote by Σ_ν the fiber critical set of F_ν with respect to the natural projection $U \times S \rightarrow S$; thus

$$\Sigma_\nu = \{(u, x) \in U \times S : d_u F_\nu(u, x) = 0\}.$$

There is an associated fiber preserving immersion

$$(3.4) \quad \Sigma_\nu \rightarrow T^*(S), \quad (u, x) \mapsto (x, d_x F_\nu(u, x)),$$

into the cotangent bundle $T^*(S) \rightarrow S$ of S , whose image we denote by Λ_ν .

If ν is regular then F_ν is a non-degenerate phase function [39, Theorem 6.7.1], in the sense that Σ_ν is a smooth manifold of dimension $\dim S$ and Λ_ν is a Lagrangian submanifold of $T^*(S)$. In particular, $\Lambda_\nu \rightarrow S$ is a Lagrangian mapping. For more on the correspondence between phase functions and Lagrangian manifolds see [22] and [33].

4. REDUCTION TO LOCAL ESTIMATES

The purpose of this section is to reduce the proof of Theorem 1.1 to Theorem 1.3, and of Proposition 1.6 to Theorem 1.5.

4.1. Reduction of Theorem 1.1 to Theorem 1.3. Let $G = \mathrm{PGL}_n(\mathbb{R})$ and $K = \mathrm{PO}(n)$, and take f to be as in the statement of Theorem 1.1. We view f as a right K -invariant function in $L^2(\Gamma \backslash G)$. Choose a non-degenerate character ψ of U , trivial on $\Gamma_U = \Gamma \cap U$, and consider the Whittaker integral

$$W_f(g) = \int_{\Gamma_U \backslash U} f(ug) \overline{\psi(u)} du, \quad g \in G.$$

Since $\Gamma_U \backslash U$ is compact, we deduce that $\|f\|_\infty \geq \mathrm{vol}(\Gamma_U \backslash U)^{-1} \|W_f\|_\infty$.

From the Hecke assumption on f , we know that W_f is a non-zero vector belonging to the one-dimensional space $\mathcal{W}(\pi_\nu, \psi)^K$ of K -fixed vectors in the local Whittaker model of π_ν . There is a unique (up to scaling) G -invariant inner product on $\mathcal{W}(\pi_\nu, \psi)$ given by

$$\int_{U \backslash P_n} W_1(p) \overline{W_2(p)} dp,$$

where P_n the mirabolic subgroup of G consisting of (homothety classes of) matrices with $(0, \dots, 1)$ in the bottom row and dp is any choice of right-invariant Haar measure on $U \backslash P_n$. (Such a measure exists since U and P_n are unimodular.) This is a result of Baruch [8, Corollary 10.4], extending to the archimedean case the analogous result of Bernstein over non-archimedean local fields.

The unfolding of the Rankin-Selberg integral implies [25] that

$$\|f\|_2^2 = c \mathrm{Res}_{s=1} \Lambda(s, \pi \times \tilde{\pi}) \|W_f\|_2^2,$$

where π is the cuspidal automorphic representation generated by the Hecke eigenfunction f . Here $c > 0$ is a constant depending only on the volume normalization. Moreover, by Li [43] (see also [14, 48, 53]) we have

$$\mathrm{Res}_{s=1} L(s, \pi \times \tilde{\pi}) \ll_\varepsilon \lambda^\varepsilon, \quad \text{for all } \varepsilon > 0.$$

The implicit constant depends on Γ , but since we view the space $\Gamma \backslash S_n$ as being fixed, we will always drop the dependence on Γ . From this we deduce the lower bound

$$\|W_f\|_\infty \gg_\varepsilon \lambda^{-\varepsilon} \|W_f\|_\infty / \|W_f\|_2.$$

From its scale invariance and the multiplicity one of spherical Whittaker functions, this last quotient is unchanged under the substitution of the global Whittaker period W_f by any other non-zero vector $W \in \mathcal{W}(\pi_\nu, \psi)^K$. If we set

$$h(\nu) = \|W\|_\infty / \|W\|_2 \quad (0 \neq W \in \mathcal{W}(\pi_\nu, \psi)^K),$$

then we deduce that $\|f\|_\infty \gg_\varepsilon \lambda^{-\varepsilon} h(\nu)$. This completes the reduction of Theorem 1.1 to Theorem 1.3 (see also Remark 1.4). \square

Remark 4.1. In [27], Gelbart, Lapid, and Sarnak establish a lower bound on Langlands-Shahidi L -functions $L(1 + it, f, r)$ for generic automorphic cusp forms f and $|t| \rightarrow \infty$. Their method, like that of this paper, relies on lower bounds for Whittaker functions. To compare,

- in this paper, a lower bound for $\|W\|_\infty$ and the convexity upper bound for $L(1, f \times \tilde{f})$ together imply a lower bound for $\|f\|_\infty$;
- in [27], a lower bound for Whittaker functions [*loc. cit.*, Lem. 7] and an upper bound for $\|\Lambda^T E(\frac{1}{2} + it, f)\|_2$ [*loc. cit.*, Prop. 2] together imply a lower bound for $L(1 + it, f, r)$.

4.2. Reduction of Proposition 1.6 to Theorem 1.5. For simplicity we assume that $\Gamma = \mathrm{PGL}_3(\mathbb{Z})$; the general case for arbitrary congruence Γ is similar. We use the following subgroup notation: B_2 is the standard Borel subgroup of GL_2 , U_2 its unipotent radical, and A_2 the group of diagonal matrices. We view GL_2 as embedded in $G = \mathrm{PGL}_3(\mathbb{R})$ via $g \mapsto \begin{pmatrix} g & \\ & 1 \end{pmatrix}$. The analogous subgroups U , A , and K of G have the same meaning as in the introduction.

– **FOURIER EXPANSION:** The Fourier–Whittaker expansion of the L^2 -normalized cusp form f at the unique cusp for $\Gamma = \mathrm{PGL}_3(\mathbb{Z})$ is given by

$$f(g) = \sum_m \sum_{[\gamma]} \rho_f(m) W_\nu(d_m \begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g),$$

where $m = (m_1, m_2)$ ranges over all vectors in $\mathbb{Z}_{\neq 0}^2$, $d_m = \mathrm{diag}(m_1 m_2, m_2, 1)$, and $[\gamma]$ ranges over cosets $U_2(\mathbb{Z}) \backslash \mathrm{GL}_2(\mathbb{Z})$. The coefficients $\rho_f(m)$ are certain complex numbers satisfying $\rho_f(1, 1) \neq 0$. They grow at most polynomially in $t \max(|m_1|, |m_2|)$, a fact established in [13]. (Recall that the spectral parameter ν of f is in $t\Omega \subset i\mathfrak{a}_+^*$ with $t > 1$).

– **STAYING IN THE CUSP:** Writing $g \in \mathrm{PGL}_3(\mathbb{R})$ in its Iwasawa decomposition $g = uak$, we can clearly assume that $k = e$. The hypothesis of Proposition 1.6 is that $a = \mathrm{diag}(y_1 y_2, y_2, 1)$ satisfies $\min(y_1, y_2) \geq \sqrt{3}/2$ and $\max(y_1, y_2) \gg t$ for a large parameter t . Theorem 1.5 then states that such g lie in the rapid decay regime for $W_{t\nu}$. We would like to say that this is equally true for every translate $d_m \begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g$ appearing in the Fourier–Whittaker expansion above. Now since A normalizes U , the A -part of $d_m \begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g$ in the Iwasawa $\mathrm{PGL}_3(\mathbb{R}) = UAK$ decomposition is equal to d_m times the A -part of $\begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g$. For the latter matrix, we have the following lower bound on the maximum of the roots.

Lemma 4.2. *Let $g \in \mathrm{PGL}_3(\mathbb{R})$ be as above and let $\gamma \in \mathrm{GL}_2(\mathbb{Z})$. Let $a' = \mathrm{diag}(y'_1 y'_2, y'_2, 1)$ be the Iwasawa A -part of $\begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g$. Then*

$$\max(y'_1, y'_2) \gg \max(y_1, y_2).$$

Proof. If $\gamma \in B_2(\mathbb{Z})$ then there is $k' = \mathrm{diag}(\pm 1, \pm 1)$ such that $\gamma k' \in U_2(\mathbb{Z})$. We see then that $\begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g \in UaK$. Thus in this case we in fact have $y'_i = y_i$.

If $\gamma \notin B_2(\mathbb{Z})$ then we use the Bruhat decomposition of $\mathrm{GL}_2(\mathbb{Z})$ to write γ as bwu' , for some $b \in B_2(\mathbb{Q})$ and $u' \in U_2(\mathbb{Q})$, where $w = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}$. Since $B_2(\mathbb{Q}) = U_2(\mathbb{Q})A_2(\mathbb{Q})$ we can clearly assume that $b \in A_2(\mathbb{Q})$, say $b = \mathrm{diag}(\pm 1/q, q)$ for $q \in \mathbb{Q}^\times$. Since γ has integer entries one in fact has $q \in \mathbb{Z} - \{0\}$.

Now

$$\begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g = \begin{pmatrix} b & \\ & 1 \end{pmatrix} w \begin{pmatrix} u' & \\ & 1 \end{pmatrix} ua = \begin{pmatrix} b & \\ & 1 \end{pmatrix} \cdot waw \cdot wv,$$

where $v = a^{-1} \begin{pmatrix} u' & \\ & 1 \end{pmatrix} ua \in U$. The roots ${}^w y_i$ of waw are ${}^w y_1 = y_1^{-1}$ and ${}^w y_2 = y_1 y_2$. Moreover, if $v = \begin{pmatrix} 1 & x & * \\ & 1 & * \\ & & 1 \end{pmatrix}$ then wv has Iwasawa A -part $\mathrm{diag}(1/\sqrt{1+x^2}, \sqrt{1+x^2}, 1)$. Thus $y'_1 = q^{-2} y_1^{-1} (1+x^2)^{-1}$ and $y'_2 = |q| y_1 y_2 \sqrt{1+x^2}$. The first root y'_1 can be very small, but since $|q| \geq 1$, $\sqrt{1+x^2} \geq 1$, and $\min(y_1, y_2) \gg 1$ we have $y'_2 \gg \max(y_1, y_2)$ as desired. \square

– **CONCLUSION OF THE PROOF:** We continue with the reduction of Proposition 1.6 to Theorem 1.5. Note that the maximum of the roots of $d_m \begin{pmatrix} \gamma & \\ & 1 \end{pmatrix} g$ is equal to $\max(m_1 y'_1, m_2 y'_2)$.

Applying Theorem 1.5 to every term in the Fourier–Whittaker expansion of f we find

$$f(g) \ll_N \sum_m \sum_{[\gamma]} \max(m_1 y'_1, m_2 y'_2)^{-N}.$$

The sum over m converges and is easy to deal with thus we only consider the sum over $[\gamma]$.

We distinguish two ranges. In the first range we consider the set of elements $[\gamma]$ of bounded height $\text{ht}(\gamma) \leq \max(y_1, y_2)^M$ for a real $M > 0$ to be chosen below. This is a finite set of elements $[\gamma]$ of cardinality at most $\max(y_1, y_2)^{c_1 M}$ for some constant c_1 . We have that $\max(y'_1, y'_2) \gg \max(y_1, y_2)$ uniformly for all γ by Lemma 4.2. Thus the sum over all elements γ in this range is less than $\max(y_1, y_2)^{c_1 M - N}$. It becomes arbitrary small for N large.

In the second range we consider the tail of the sum which consists of elements $[\gamma]$ of height $\geq \max(y_1, y_2)^M$. We apply a Bruhat decomposition of the element $(\begin{smallmatrix} \gamma & \\ & 1 \end{smallmatrix})g$ and find that its A -part is the product of the A -part of $(\begin{smallmatrix} \gamma & \\ & 1 \end{smallmatrix})$ times the A -part of kg for some element $k \in K$ (which is the K -part of $(\begin{smallmatrix} \gamma & \\ & 1 \end{smallmatrix})$). At least one of the roots of the A -part of $(\begin{smallmatrix} \gamma & \\ & 1 \end{smallmatrix})$ is greater than $\max(y_1, y_2)^{c_2 M}$ for some $c_2 > 0$. We cannot gain a fine control of the corresponding root of the A -part of kg because k may be arbitrary, but fortunately we can easily say that it is at least greater than $\max(y_1, y_2)^{-c_3}$ for some $c_3 < \infty$. The contribution of this second range is thus bounded by

$$\sum_{[\gamma]} \text{ht}(\gamma)^{-c_2 MN} \max(y_1, y_2)^{c_3 N} \ll \max(y_1, y_2)^{N(c_3 - c_2 M)},$$

where we have essentially estimated the convergent sum of a geometric series. Choosing M large enough so that $c_3 < c_2 M$ this is negligible for N large.

We have obtained a rapid decay bound for each of the two ranges and the argument establishing the rapid decay is complete.

– PROOF OF (1.2): If $\Gamma \backslash S$ is a compact locally symmetric space of dimension d and rank r , Sarnak [55] showed that any $\mathcal{D}_G(S)$ -eigenfunction f , with Laplacian eigenvalue λ , satisfies

$$(4.1) \quad \|f\|_\infty \ll \lambda^{(d-r)/4} \|f\|_2.$$

Note the improvement by $r/4$ over the Hörmander bound (1.1). The proof of (1.2) is based on an explication of the dependence of the implied constant on the injectivity radius of $\Gamma \backslash S$ in Sarnak's bound (4.1).

Now let $\Gamma \backslash S$ be any non-compact locally symmetric space and let $p \in \Gamma \backslash S$ be arbitrary. (We will specialize to $\Gamma \backslash S_3$ momentarily.) For $R > 0$ smaller than the local injectivity radius about p let $B(p, R)$ denote the geodesic ball of radius R . A direct inspection of the proof of (4.1) yields

$$(4.2) \quad \max_{x \in B(p, R)} |f(x)| \leq C \left(\int_{B(p, R)} |\omega_\lambda(x)|^2 dx \right)^{-1/2} \left(\int_{B(p, R)} |f(x)|^2 dx \right)^{1/2},$$

where ω_λ be the unique spherical function on G about p having the same $\mathcal{D}_G(S)$ -eigenvalues as f (and thus of eigenvalue λ) and normalized so that $\omega_\lambda(p) = 1$. Going high in the cusp, we can find $p \in \Gamma \backslash S$ with arbitrarily small injectivity radius; in particular we can take $0 < R < 1/\sqrt{\lambda}$. On such balls, the spherical function ω_λ is $\asymp 1$ and one has

$$(4.3) \quad \int_{B(p, R)} |\omega_\lambda(x)|^2 dx \asymp \text{vol}(B(p, R)) \asymp R^d \quad (0 < R < 1/\sqrt{\lambda}).$$

Now let $(\Gamma \backslash S)^{\leq T}$ denote the truncation of $\Gamma \backslash S$ up to height T . We first claim that the injectivity radius on $(\Gamma \backslash S)^{\leq T}$ is at least $1/T^2$. Indeed, let $p = \Gamma uak$ lie in a truncated Siegel set for Γ cut out by the condition that $\max_i y_i \ll T$, and suppose that there is $g \in G$ with $\text{dist}(e, g) \ll 1/T^2$ and $\gamma \in \Gamma$ such that $pg = \gamma p$. Our goal is to prove that $\gamma = e$. We write the equality $pg = \gamma p$ as $a.kgk^{-1}.a^{-1} = u^{-1}\gamma u$ and observe that $\text{dist}(e, kgk^{-1}) \ll 1/T^2$ since $k \in K$ varies in a compact. The conjugation by a is described by its roots y_i ; by hypothesis, the largest dilation is T^2 . Since u also varies in a compact this implies $\text{dist}(e, \gamma) \ll 1$. Thus if the constant is chosen small enough, $\gamma = e$ as desired.

We may therefore bound the value at any point $p \in (\Gamma \backslash S)^{\leq T}$ by its maximum over the geodesic ball of radius $1/T^2$ about p . In particular, it follows from (4.2) and (4.3) that for any $p \in (\Gamma \backslash S)^{\leq \sqrt{\lambda}}$ we have

$$\max_{x \in B(p, 1/\lambda)} |f(x)| \ll \lambda^{d/2} \|f\|_2.$$

We now specialize to the case $S = S_3$, Γ a congruence subgroup of $\mathrm{GL}_3(\mathbb{Z})$ and f a Hecke cusp form. Since by the first half of Proposition 1.6 the size of f on the complement of $\Gamma \backslash S_3^{\leq \sqrt{\lambda}}$ is smaller than any power of λ , the bound (1.2) is proved, where we use $\dim S_3 = 5$.

Remark 4.3. Although Theorem 1.5 below is valid for arbitrary n , it does not seem sufficient to extend Proposition 1.6 to all n . The Fourier expansion of f on $\Gamma \backslash S_n$ still holds, but Lemma 4.2 is not true for $n \geq 4$. This can be seen in the following example.

Let $\gamma = \begin{pmatrix} 0 & 1 & 0 \\ 1 & N & -1 \\ 0 & N & -1 \end{pmatrix}$, and $g = a \in A$ with roots y_i . It can be verified that

$$y'_1 = y_1^{-1}(1 + N^2 y_2^2)^{-1/2}, \quad y'_2 = y_1(N^2 + y_2^{-2})^{-1/2}, \quad y'_3 = y_3(1 + N^2 y_2^2)^{1/2}.$$

Letting $y_2 = y_3 = 1$ and y_1 large of size about N^2 we see that the Lemma 4.2 is not valid in this case.

In view of this it would be interesting to investigate more the essential support of cusp forms in higher rank. It may be that one needs to take into account the different directions in which g can go to infinity into the cusp.

5. RAPID DECAY ESTIMATES AND THE PROOF OF THEOREM 1.3 FOR $\mathrm{GL}_n(\mathbb{R})$

In this section we establish several estimates for Whittaker functions with large eigenvalue. In the first two subsections, we give quantitative information on the rapid decay regime of spherical Whittaker functions in the general setting of split semisimple real Lie groups. In the third subsection, we use these results to prove Theorem 1.3 in the case of $\mathrm{GL}_n(\mathbb{R})$.

5.1. Rapid decay. Let W_ν be a spherical Whittaker function on a split semisimple real Lie group. The following proposition gives the rapid decay of $W_\nu(a)$ for a large with respect to ν . The proof is through repeated integration by parts and a convolution identity. This kind of argument is relatively standard, e.g. in estimates of Eisenstein series (see [5, §4]).

Proposition 5.1. *Fix a non-zero $\nu \in \mathfrak{ia}^*$ of norm 1. Then for t large enough and for all $a \in A$ with $\min_{\alpha \in \Pi} \alpha(a) \gg 1$ and $\max_{\alpha \in \Pi} \alpha(a) \gg t$ one has*

$$|W_{t\nu}(a)| \ll_N \left(\max_{\alpha \in \Pi} \alpha(a) \right)^{-N}$$

for every $N \geq 1$.

Proof. For any $\varphi \in C_c^\infty(K \backslash G/K)$ we have $\widehat{\varphi}(\nu)W_\nu = W_\nu \star \varphi$, where $\varphi \mapsto \widehat{\varphi}(\nu)$ is the spherical transform. We decompose the convolution integration using the Iwasawa coordinates to get

$$\begin{aligned} (W_\nu \star \varphi)(a) &= \int_G W_\nu(ag)\varphi(g)dg = \int_A \int_U W_\nu(aua_1)\varphi(ua_1)\delta(a_1)^{-1}du da_1 \\ &= \int_A \int_U \psi_a(u)W_\nu(aa_1)\varphi(ua_1)\delta(a_1)^{-1}du da_1, \end{aligned}$$

where $\psi_a(u) = \psi(aua^{-1})$. Since φ is compactly supported, $\varphi(ua_1)$ vanishes all $u \in U$ for $a_1 \in A$ outside some set $A_c \subset A$ defined by inequalities $|\alpha(a)| \leq c$ for $\alpha \in \Pi$. Thus

$$(W_\nu \star \varphi)(a) = \int_{A_c} W_\nu(aa_1)\delta(a_1)^{-1} \int_U \psi_a(u)\varphi(ua_1)du da_1.$$

An application of the Cauchy-Schwarz inequality to the integral over A_c yields

$$|\widehat{\varphi}(\nu)|^2 |W_\nu(a)|^2 \leq \int_{A_c} |W_\nu(aa_1)|^2 \delta(a_1)^{-1} da_1 \cdot \int_{A_c} \left| \int_U \psi_a(u) \varphi(ua_1) du \right|^2 \delta(a_1)^{-1} da_1.$$

Changing variables and using $\min_{\alpha \in \Pi} \alpha(a) \gg 1$ the first integral is

$$\delta(a) \int_{a^{-1}A_c} |W_\nu(a_1)|^2 \delta(a_1)^{-1} da_1 \leq \delta(a) \int_{A_1} |W_\nu(a_1)|^2 \delta(a_1)^{-1} da_1 \ll \delta(a),$$

where A_1 is contained in some fixed translate of the negative Weyl chamber $\exp(-\mathfrak{a}_+)$ because $\min_{\alpha \in \Pi} \alpha(a)$ is bounded below.

We bound the remaining integral for a specific choice of the function φ . For any $\varphi \in C_c^\infty(K \backslash G / K)$ put

$$\varphi_\nu(k_1 a k_2) = \nu(a) \varphi(a), \quad k_1, k_2 \in K, \quad a \in A.$$

From [23, Lemma 6.3 and (6.9)] there exists $\varphi \in C_c^\infty(K \backslash G / K)$ such that $|\widehat{\varphi}_\nu(\lambda)| \geq 1$ for all $\lambda, \nu \in i\mathfrak{a}^*$ such that $\|\lambda - \nu\| \leq 1$. We claim that for any $a \in A$ and any $B \geq 1$,

$$\int_U \psi_a(u) \varphi_\nu(ua_1) du \ll_B \|\nu\|^B (\max_{\alpha \in \Pi} \alpha(a))^{-B},$$

where the implied constant depends only on B and the choice of φ .

For any $\alpha \in \Pi$ and $0 \neq X_\alpha \in \mathfrak{u}_\alpha$ then the first derivative of the additive character is

$$X_\alpha \cdot \psi_a(u) = i\alpha(a) \psi_a(u).$$

For $a \in A$ if we let α_{\max} be such that $\alpha_{\max}(a) = \max_{\alpha \in \Pi} \alpha(a)$, then upon integrating by parts B -times we find that the integral is, up to a sign, equal to

$$\alpha_{\max}(a)^{-B} \int_U \psi_a(u) \varphi_\nu(X_{\max}^B; ua_1) du,$$

where X_{\max}^B is viewed as an element in $\mathcal{U}(\mathfrak{g})$. Now using the definition of φ_ν we have $\varphi_\nu(X_{\max}^B; ua_1) \ll \|\nu\|^B$ where the implied constant depends only on φ . The support conditions on φ then allow one to conclude. \square

5.2. Precise decay regime. In this paragraph we give an alternative description of the rapid decay regime of the Whittaker function relative to Proposition 5.1. We are again assuming here that G is an arbitrary split semisimple real Lie group.

The idea here is standard: the Whittaker function is given as an oscillatory integral, and where there are no critical points one has rapid decay (again by integration by parts). To make the link with later sections, we express the rapid decay regime in terms of the fibers of an associated Lagrangian mapping $\Lambda_\nu \rightarrow S$ introduced in §3.6.

Proposition 5.2. *Let $\nu \in i\mathfrak{a}^*$ be non-zero and $X \in \mathfrak{a}$ be such that $e^X \in S$ lies outside the image of $\Lambda_\nu \rightarrow S$. Then for t large enough we have*

$$W_{t\nu}(e^{(\log t)X}) \ll_{N,\nu,X} t^{-N}$$

for every $N \geq 1$.

Proof. The scaling of ν by t and of X by $\log t$ allows us to write the oscillatory factor in the Jacquet integral as $e^{itF_\nu(u, e^X)}$. Since e^X lies outside the image of Λ_ν , the phase function $u \mapsto F_\nu(u, e^X)$ has no critical points. One takes a smooth partition of unity over all of U (say over dyadic shells) and estimates the integral over each shell separately. Rapid decay follows from repeated integration by parts. Keeping track of the shell in the estimate, and

summing over all shells, then gives the global estimate. More details are provided in the proof of Theorem 9.1. \square

We shall see in §6.3 that (under a regularity assumption on ν) for e^X lying outside the image of $\Lambda_\nu \rightarrow S$ we have

$$(5.1) \quad \sum_{\alpha \in \Pi} e^{2\langle \alpha, X \rangle} \leq \|\nu\|^2.$$

This allows one to compare Proposition 5.1 and Proposition 5.2.

5.3. Proof of Theorem 1.3 for $G = \mathrm{GL}_n(\mathbb{R})$. We may now prove Theorem 1.3 in the special case of $\mathrm{PGL}_n(\mathbb{R})$. The argument will combine Proposition 5.1 (specialized to $\mathrm{PGL}_n(\mathbb{R})$) with the Stade formula (see (5.2) below).

The unramified principal series representation π_ν and the Whittaker model $\mathcal{W}(\pi_\nu, \psi)$ come equipped with canonically normalized G -invariant inner products (see §4.1). The Jacquet-Whittaker function W_ν is the image of a unitary intertwining of the L^2 -normalized K -fixed vector in π_ν . We deduce that $\|W_\nu\|_2 = 1$. (Here we are relying critically on the assumption that $G = \mathrm{PGL}_n(\mathbb{R})$.) It therefore suffices to provide a lower bound for

$$\|W_\nu\|_\infty = \sup_{g \in G} |W_\nu(g)| = \sup_{a \in A} |W_\nu(a)|.$$

Let $\Psi(s, W_\nu, \overline{W_\nu})$ denote the integral

$$\Gamma_{\mathbb{R}}(ns) \int_A |W_\nu(a)|^2 \det(a)^s \delta(a)^{-1} da.$$

By the Stade formula [59] we have

$$(5.2) \quad \Psi(s, W_\nu, \overline{W_\nu}) = \frac{L(s, \pi_\nu \times \widetilde{\pi}_\nu)}{L(1, \pi_\nu \times \widetilde{\pi}_\nu)}.$$

Applying Stirling's formula to the quotient of Gamma factors we obtain

$$\Psi(\sigma, W_\nu, \overline{W_\nu}) \asymp \prod_{i \neq j} (1 + |\mu_i - \mu_j|)^{(\sigma-1)/2} \asymp \lambda_\nu^{(\sigma-1) \dim U/2}.$$

It will be convenient to introduce explicit coordinates in the integral defining $\Psi(s, W_\nu, \overline{W_\nu})$ in order to extract the size of W_ν . Writing

$$a = \mathrm{diag}(y_1 \cdots y_{n-1}, y_2 \cdots y_{n-1}, \dots, y_{n-1}, 1) \in A,$$

we have

$$\delta(a) = \prod_{i=1}^{n-1} y_i^{(n-1-i)i}, \quad \det(a) = \prod_{i=1}^{n-1} y_i^i, \quad da = \frac{dy_1}{y_1} \cdots \frac{dy_{n-1}}{y_{n-1}}.$$

With these coordinates we may write $\Psi(s, W_\nu, \overline{W_\nu})$ as

$$\Gamma_{\mathbb{R}}(ns) \int_0^\infty \cdots \int_0^\infty |W_\nu(\mathrm{diag}(y_1 \cdots y_{n-1}, \dots, y_{n-1}, 1))|^2 \prod_{i=1}^{n-1} y_i^{i(s-i)} \frac{dy_i}{y_i}.$$

We now decompose $\Psi(s, W_\nu, \overline{W_\nu})$ as $I_1 + I_2$, where the integral I_1 is taken over the range $\max_i y_i \gg \sqrt{\lambda_\nu}$ and the integral I_2 over the complementary range. By Proposition 5.1, after passing to a one-dimensional integral, we have

$$I_2 \ll \int_{r \gg \sqrt{\lambda_\nu}} r^{-N} \frac{dr}{r} \ll \lambda_\nu^{-N}$$

for $N > 1$ large enough. On the other hand we have $I_1 \leq V \|W_\nu\|_\infty^2$, where for a large enough constant $C > 1$ we have put

$$V = \prod_{i=1}^{n-1} \int_0^{C\sqrt{\lambda_\nu}} y_i^{i(s-i)} \frac{dy_i}{y_i}.$$

As long as $\sigma \geq 1$, we deduce $\|W_\nu\|_\infty^2 \gg V^{-1} \lambda_\nu^{(\sigma-1) \dim U/2}$. For $\sigma > n-1$ we have

$$V \asymp \lambda_\nu^{\frac{1}{2} \sum_{i=1}^{n-1} (\sigma-i)(n-i)} = \lambda_\nu^{(\sigma-1-\frac{n-2}{3}) \dim U/2}.$$

Taking $\sigma = n-1 + o(1)$ concludes the argument. \square

The above argument can be refined to give lower bounds on W_ν even when ν is irregular. Indeed Proposition 5.1 (as well as Proposition 5.2) are valid for irregular ν , as is the Stade formula. We have included the regularity assumption in Theorem 1.3 to simplify notation and bring the idea of the proof to the forefront.

Remark 5.3. We speculate on the geometric significance of the exceptionally large exponent $c(n)$ in Theorem 1.1. For convenience, we restrict to the case $\Gamma = \mathrm{PGL}_n(\mathbb{Z})$ in this paragraph.

Using standard notation for Siegel sets we consider the cuspidal “end” $\mathfrak{S}^{>Y} := \omega A^{>Y} K$, where $\omega \subset U$ is a compact subset of U and

$$A^{>Y} = \{a = \mathrm{diag}(y_1 \cdots y_{n-1}, \dots, y_{n-1}, 1) \in A : \min_i y_i \geq \sqrt{3}/2, \max_i y_i > Y\},$$

for some parameter $Y \geq 1$. The right G -invariant measure, when expressed in the Iwasawa UAK coordinates, is given by $dg = \delta(a)^{-1} dudadk$. Then the volume of this collar is

$$\int_{\mathfrak{S}^{>Y}} dg \asymp \int_{A^{>Y}} \delta(a)^{-1} da \asymp Y^{-\mathrm{ht}(\mathrm{PGL}_n)}.$$

The relative volume of $\mathfrak{S}^{>Y}$ is therefore seen to decrease as n gets large, and this by a cubic power of n . In other words, the cuspidal regions of $\Gamma \backslash S_n$ become dramatically more “pinched” as n gets large. The narrower cusps of the higher rank spaces $\Gamma \backslash S_n$ create a bottleneck as the cusp forms to transition from the oscillatory to the decay regime. With so little space to do so they get exceedingly large, in a sort of automorphic Gibbs phenomenon.

As mentioned in the introduction, Kleinbock and Margulis proved in [37] that almost all geodesics penetrate the cusp at logarithmic speed $1/\mathrm{ht}(G)$. There, the collar plays the role of a moving target for the geodesic flow.

6. PROOF OF THEOREM 1.5

The goal in this section is to study the critical points of the Whittaker phase function and to deduce from this Theorem 1.5 from the introduction.

6.1. An explicit description of Λ_ν . Let $\nu \in i\mathfrak{a}^*$; in this subsection, we do not assume that ν is regular. We would like to calculate the equations of the image Λ_ν of the map $\Sigma_\nu \rightarrow T^*(S)$ of (3.4).

The main tool is the moment map for the Hamiltonian action of G on $T^*(S)$, which we now explicitly describe. Recall that $S = G/K$. Let \mathfrak{p}^* be the space of functionals on \mathfrak{g} that vanish on \mathfrak{k} . When $T^*(S)$ is identified with the fiber product $G \times_K \mathfrak{p}^*$, the moment map is $[g, \xi] \mapsto \mathrm{Ad}_g^*(\xi)$.

Proposition 6.1. *Let $\nu \in i\mathfrak{a}^*$ be arbitrary. Then Λ_ν consists of $[g, \xi] \in G \times_K \mathfrak{p}^*$ such that*

$$(6.1) \quad \xi \in \mathrm{Ad}_K^*(\mathrm{Im} \nu)$$

and

$$(6.2) \quad \text{Ad}_g^*(\xi) - \ell_1 \in \ker(\mathfrak{g}^* \rightarrow \mathfrak{u}^*),$$

where $\ell_1 \in \mathfrak{u}_{\text{ab}}^*$ is the non-degenerate abelian functional introduced in §3.6.

Proof. We use Lemma 3.1 to evaluate the derivative of the Whittaker phase function (3.3) with respect to $Y \in \mathfrak{p}$. We obtain

$$F_\nu(u, g; Y) = \langle \text{Ad}_{\kappa(wug)^{-1}}^*(\text{Im } \nu), Y \rangle.$$

We deduce from the definition (3.4) of the immersion $\Sigma_\nu \rightarrow T^*(S)$ that

$$\Lambda_\nu = \left\{ [g, \text{Ad}_{\kappa(wug)^{-1}}^*(\text{Im } \nu)] : (u, x) \in \Sigma_\nu, x = gK \right\}.$$

The condition (6.1) is thus satisfied. On the other hand, we may again use Lemma 3.1 to evaluate the derivative of $F_\nu(u, g)$ with respect to $Z \in \mathfrak{u}$. We obtain

$$F_\nu(u; Z, g) = \langle \text{Ad}_g^* \text{Ad}_{\kappa(wug)^{-1}}^*(\text{Im } \nu), Z \rangle - \langle \ell_1, Z \rangle.$$

We deduce that the fiber critical set Σ_ν consists of pairs (u, x) , with $x = gK$, such that for all $Z \in \mathfrak{u}$

$$\langle \text{Ad}_g^* \text{Ad}_{\kappa(wug)^{-1}}^*(\text{Im } \nu), Z \rangle = \langle \ell_1, Z \rangle,$$

showing that the condition (6.2) is also met.

Conversely let $[g, \xi] \in G \times_K \mathfrak{p}^*$ satisfy (6.1) and (6.2). We need to show that there exists $u \in U$ such that $\xi = \text{Ad}_{\kappa(wug)^{-1}}^*(\text{Im } \nu)$. From (6.2) it follows that the restriction of $\text{Ad}_{\kappa(g)}^*(\xi)$ to \mathfrak{u} is non-degenerate. Lemma 3.3 then implies that

$$\text{Ad}_{\kappa(g)}^*(\xi) \in \text{Ad}_{S_w^+}^*(\text{Im } \nu).$$

Thus there exists $k \in M \cdot S_w^+ \subset K$ such that $\text{Ad}_{\kappa(g)}^*(\xi) = \text{Ad}_{k^{-1}}^*(\text{Im } \nu)$; indeed recall that since $w^2 = 1$, the big Bruhat cell is invariant under $k \rightarrow k^{-1}$. By Lemma 3.2 there is a unique $v \in U$ such that $k = \kappa(wv)$. Furthermore letting $u = e^{H(g)}v\tau(g)^{-1} \in U$ we see that $k = \kappa(wv) = \kappa(wu\tau(g))$ and therefore $\text{Ad}_{\kappa(g)}^*(\xi) = \text{Ad}_{\kappa(wu\tau(g))^{-1}}^*(\text{Im } \nu)$ as desired. \square

Equation (6.2) is essentially the definition of the Peterson variety, see [40]. Clearly Λ_ν is U -invariant, and in the next subsection we shall investigate the quotient by the U -action.

6.2. Lagrangian equivalence. We would now like to state and prove a precise version of Theorem 1.5 from the introduction.

We begin by recalling the characteristic subvariety of the Toda lattice, which sits inside \mathcal{J}^* , the space of functionals in \mathfrak{p}^* vanishing on $[\mathfrak{u}, \mathfrak{u}]$. Note that for $G = \text{GL}_n(\mathbb{R})$ the dual of \mathcal{J}^* is the space of tridiagonal symmetric matrices, otherwise known as the (symmetric) *Jacobi matrices*. See [40, (5.4.2)] for a closely related construction, where it is also shown that \mathcal{J}^* is a symplectic space [*loc. cit.*, Prop 6.4]. One has a natural quotient map $\mathcal{J}^* \rightarrow \mathfrak{u}_{\text{ab}}^*$. We then define \mathcal{L}_ν set-theoretically as

$$\mathcal{L}_\nu = \text{Ad}_K^*(\text{Im } \nu) \cap \mathcal{J}^*.$$

Thus we have a triple

$$\mathcal{L}_\nu \longrightarrow \mathcal{J}^* \longrightarrow \mathfrak{u}_{\text{ab}}^*.$$

On the other hand, note that any abelian functional on \mathfrak{u} is fixed under the adjoint action of U on \mathfrak{u}^* and, if non-degenerate, is a regular value under the moment map $T^*(S) \rightarrow \mathfrak{g}^* \rightarrow \mathfrak{u}^*$. Thus the Hamiltonian action of U on $T^*(S)$ preserves the fiber over ℓ_1 , and on this fiber the action is free and proper. Let $M_1 = U \backslash\backslash_1 T^*(S)$ be the symplectic reduction of the U -action on $T^*(S)$ over ℓ_1 ; it is endowed with a natural symplectic structure. Moreover, the Lagrangian fibration $T^*(S) \rightarrow S$ also reduces under the U -action, and by the Iwasawa decomposition we

obtain a Lagrangian fibration $M_1 \rightarrow A$ which one can identify with $T^*(A) \rightarrow A$. Letting Λ_ν^{red} be the quotient $U \backslash \Lambda_\nu$, we obtain

$$\Lambda_\nu^{\text{red}} \longrightarrow M_1 \longrightarrow A.$$

We would like to relate these two triples, placing them into a commutative diagram. The next result accomplishes this and should be thought of as a more precise version of Theorem 1.5.

Theorem 6.2. *The map*

$$T^*(S) \rightarrow \mathfrak{p}^*, \quad [g, \xi] \mapsto \text{Ad}_{\kappa(g)}^* \xi,$$

descends to a map $M_1 \rightarrow \mathcal{J}^$, sending Λ_ν^{red} to \mathcal{L}_ν . The induced diagram*

$$(6.3) \quad \begin{array}{ccccc} \Lambda_\nu^{\text{red}} & \longrightarrow & M_1 & \longrightarrow & A \\ \downarrow & & \downarrow & & \downarrow \\ \mathcal{L}_\nu & \longrightarrow & \mathcal{J}^* & \longrightarrow & \mathfrak{u}_{\text{ab}}^* \end{array}$$

is commutative, where the vertical maps are open embeddings and both squares are Cartesian. Here, the last vertical map $A \rightarrow \mathfrak{u}_{\text{ab}}^$ is $a \mapsto \ell_a = \text{Ad}_a^* \ell_1$.*

Moreover, if $\nu \in \mathfrak{ia}^$ is regular, then Λ_ν^{red} is Lagrangian inside M_1 and \mathcal{L}_ν is Lagrangian in \mathcal{J}^* . Each of the two rows of the diagram defines a Lagrangian mapping.*

Proof. The preimage of ℓ_1 under the moment map is the set of $[g, \xi] \in G \times_K \mathfrak{p}^*$ satisfying (6.2). It follows from this and the fact that $[\mathfrak{u}, \mathfrak{u}]$ is stable under the adjoint action for the Borel subgroup B that for any ξ in the fiber over ℓ_1 and any $X \in [\mathfrak{u}, \mathfrak{u}]$ we have

$$\langle \text{Ad}_{\kappa(g)}^* \xi, X \rangle = \langle \text{Ad}_g^* \xi, \text{Ad}_{\tau(g)} X \rangle \in \langle \ell_1 + \ker(\mathfrak{g}^* \rightarrow \mathfrak{u}^*), [\mathfrak{u}, \mathfrak{u}] \rangle = 0.$$

Thus the restriction of the map $T^*(S) \rightarrow \mathfrak{p}^*$ to the fiber over ℓ_1 takes values in \mathcal{J}^* . From the left U -invariance of $g \rightarrow \kappa(g)$ we obtain a map from $M_1 = U \backslash \! \! \! \backslash_1 T^*(S)$ to \mathcal{J}^* .

Now let $\xi \in \Lambda_\nu$. Since ξ lies in $\text{Ad}_K^*(\text{Im } \nu)$ by (6.1) then so does $\text{Ad}_{\kappa(g)}^* \xi$, whence Λ_ν is mapped to $\mathcal{L}_\nu = \text{Ad}_K^*(\text{Im } \nu) \cap \mathcal{J}^*$. The same is therefore true of Λ_ν^{red} .

The vertical map $A \rightarrow \mathfrak{u}_{\text{ab}}^*$ is clearly an open embedding and it is also easy to verify that the right square is Cartesian. Inspecting the equation (6.2) we see that the vertical map $M_1 \rightarrow \mathcal{J}^*$ is an open embedding. As a consequence the map $\Lambda_\nu^{\text{red}} \rightarrow \mathcal{L}_\nu$ is an open embedding. That the left square is Cartesian is precisely the content of Proposition 6.1.

Now recall that for ν regular Λ_ν is Lagrangian in $T^*(S)$. Moreover, we have already observed after the proof of Proposition 6.1 that Λ_ν is invariant under the U -action. It follows that $\Lambda_\nu^{\text{red}} = U \backslash \Lambda(F_\nu)$ is Lagrangian inside M_1 , since the reduction of an invariant Lagrangian is again Lagrangian; see [26, Thm. 3.2]. The first row is a Lagrangian mapping because it is the reduction under the U -action of the Lagrangian mapping

$$\Lambda_\nu \longrightarrow T^*(S) \longrightarrow S.$$

It only remains to verify that \mathcal{L}_ν is a Lagrangian submanifold of \mathcal{J}^* . This can be deduced by Zariski density from the fact that Λ_ν^{red} is Lagrangian inside M_1 . \square

Remark 6.3. The map $M_1 \rightarrow \mathcal{J}^*$ of the above theorem is in fact a symplectomorphism onto its image. See [52, §4.5, Theorem 4] for this and more on the relation between the Toda system and the reduced geodesic flow on symmetric spaces. There are also deep connections [28, 29] with the cohomology of flag manifolds and mirror symmetry.

The fact that \mathcal{L}_ν is Lagrangian in \mathcal{J}^* when ν is regular was already proven in [39, Theorem 6.7.1]. We provide an independent proof of this, making our treatment self-contained.

Remark 6.4. The method of co-adjoint orbits described in §2.2 yields a natural interpretation of the construction of \mathcal{L}_ν . We intersect the coadjoint orbit $\text{Ad}_G^*(\text{Im } \nu)$ with $\text{Ker}(\mathfrak{g}^* \rightarrow \mathfrak{k}^*)$ to capture the spherical vector of the representation π_ν . This intersection is precisely the K -orbit $\text{Ad}_K^*(\text{Im } \nu)$ and geometrically $\text{Ad}_K^*(\text{Im } \nu)$ is the zero level set of the moment map of the K -action on $\text{Ad}_G^*(\text{Im } \nu)$.² We then project $\text{Ad}_K^*(\text{Im } \nu)$ via the natural projection³ map $\mathfrak{g}^* \rightarrow \mathfrak{u}^*$ and intersect with the subspace $\mathfrak{u}_{\text{ab}}^*$.

For $G = \text{SL}_2(\mathbb{R})$, $\mathcal{L}_\nu = \text{Ad}_K^*(\text{Im } \nu)$ is a circle. For general G and regular ν , it is a compact aspherical manifold of dimension the rank of G ; see [20, 61]. For example for $G = \text{SL}_3(\mathbb{R})$ it is known that \mathcal{L}_ν is a genus 2 surface.

Remark 6.5. It is also possible to reduce the fiber critical set Σ_ν under the U -action for example by noting that for any $v \in U$,

$$F_\nu(uv, v^{-1}g) = F_\nu(u, g) - \langle \ell_1, v \rangle.$$

Equivalently we can restrict the second parameter to belong to A and we denote by $\Sigma_\nu^{\text{red}} \subset U \times A$ the set of pairs (u, a) which are critical for $u \mapsto F_\nu(u, a)$. We also note that there is a natural section $\Lambda_\nu^{\text{red}} \rightarrow \Lambda_\nu$ obtained by taking the Iwasawa A -part of g .

We have a natural map $\Sigma_\nu^{\text{red}} \rightarrow \Lambda_\nu^{\text{red}}$ as before. Composing with the open embedding to \mathcal{L}_ν we obtain a map $\Sigma_\nu^{\text{red}} \rightarrow \mathcal{L}_\nu$ given by

$$(6.4) \quad (u, a) \mapsto \text{Ad}_{\kappa(\mathfrak{w}ua)^{-1}}^*(\text{Im } \nu).$$

The diagram (6.3) being a Cartesian square implies that for any $a \in A$ the set (u, a) of critical points in Σ_ν^{red} lying over a is sent bijectively to the isospectral fiber over ℓ_a .

In a first version of this paper our analysis revolved around this and the properties of the map (6.4). In the present version we have favored what seems like a more efficient treatment via Theorem 6.2.

6.3. Stratification by singularity type and associated asymptotics. For ν regular, we have thus defined three Lagrangian mappings

$$(6.5) \quad \Lambda_\nu \rightarrow S, \quad \Lambda_\nu^{\text{red}} \rightarrow A, \quad \text{and} \quad \mathcal{L}_\nu \rightarrow \mathfrak{u}_{\text{ab}}^*$$

and described their precise relationship in Theorem 6.2 and the discussion preceding it. In this subsection and the next, we would like to put Theorem 1.7 and Corollary 1.8 into a more general context.

Let $E \rightarrow B$ be a Lagrangian fibration and $L \rightarrow E$ a Lagrangian immersion. Then the composition $\pi : L \rightarrow B$ is a Lagrangian mapping to the base space B . We obtain a stratification of L via the fibers of π . Letting L^{sing} denote the set of singular points for π , we write

$$(6.6) \quad B = \text{C} \sqcup \text{L} \sqcup \text{S}$$

where

- the *caustic locus* C is the image of L^{sing} under π ;
- the *light zone* L is $\text{Im } \pi - \text{C}$;
- the *shadow zone* S is $B - \text{Im } \pi$.

One could refine the singular locus L^{sing} and hence C into singularity types. See [4, §2] for the general theory of stratifications via coranks of the first differential of a smooth mapping restricted to singular loci, and [*loc. cit.*, §21] for that same theory applied to the special case of Lagrangian mappings. For the Lagrangian mappings in (6.5) it would be interesting to

²One knows that $\text{Ad}_K^*(\text{Im } \nu)$ is Lagrangian inside $\text{Ad}_G^*(\text{Im } \nu)$; see [6]. We shall not use this fact.

³By contrast the Kostant convexity theorem says that the projection of $\text{Ad}_K^*(\text{Im } \nu)$ onto \mathfrak{a}^* is the convex hull of the Weyl group orbit of $\text{Im } \nu$.

have an intrinsic description of \mathbb{C} . The answer may be rather complicated for, as we shall see in §8, the example of $G = \mathrm{PGL}_3$ already exhibits a rich structure.

We highlight two ways in which the above decomposition of the base B yields information about the corresponding oscillatory integrals. We specialize to the case of the mappings in (6.5) associated with Whittaker functions and refer the reader to [22, 33] for the more general passage from Lagrangian mappings to Fourier integral operators.

– *Shadow zone and rapid decay:* The image of either of the first two Lagrangian mappings in (6.5) should be thought of as the “essential support” of the Whittaker function. For example, we showed in Proposition 5.2 that the Whittaker function decays rapidly in the shadow zone. In comparison, viewing the Whittaker function as an eigenfunction of the quantum Toda lattice, \mathcal{L}_ν is the characteristic variety of the system of linear partial differential equations. The image of $\mathcal{L}_\nu \rightarrow \mathfrak{u}_{\mathrm{ab}}^*$ corresponds to the classically allowed region.

Now \mathcal{L}_ν is closed in $\mathrm{Ad}_K^*(\mathrm{Im} \nu)$ and so is compact. It follows that the classically allowed region is also compact. In fact, since the projection $\mathfrak{g}^* \rightarrow \mathfrak{u}_{\mathrm{ab}}^*$ is orthogonal for the invariant scalar product (see §3.1), we see that the image of $\mathcal{L}_\nu \rightarrow \mathfrak{u}_{\mathrm{ab}}^*$ is included in the ball of radius $\|\nu\|$. In light of Theorem 6.2, the same inclusion holds for the image of $\Lambda_\nu^{\mathrm{red}} \rightarrow A$. This proves inequality (5.1).

– *Singularities and degenerate critical points:* Let

$$\Sigma_\nu \ni (u, x) \longmapsto [g, \xi] \in \Lambda_\nu.$$

Let $Q_\nu(u, x) = \nabla_u^2 F_\nu(u, x)$ be the fiber Hessian of F_ν at (u, x) and let $d_\nu([g, \xi])$ be the differential of the mapping $\Lambda_\nu \rightarrow S$ at $[g, \xi]$. Then one has an isomorphism (see [4, §19.3] or [33, Theorem 3.14])

$$(6.7) \quad \ker d_\nu([g, \xi]) \xrightarrow{\cong} \ker Q_\nu(u, x).$$

In particular, (u, x) is a degenerate critical point for $u \mapsto F_\nu(u, x)$ if and only if $[g, \xi]$ is singular for the mapping $\Lambda_\nu \rightarrow S$. In other words, (u, x) is non-degenerate if and only if the tangent space of Λ_ν at $[g, \xi]$ is transversal to the fiber of the projection $[g, \xi] \mapsto x$, where $x = gK$. This correspondence remains true for the reduced mapping $\Lambda_\nu^{\mathrm{red}} \rightarrow A$.

6.4. Numerical invariants of Lagrangian singularities. Let $\pi : L \rightarrow B$ be a Lagrangian mapping. A point $p \in L$ is singular for π if the differential $d\pi_p : T_p L \rightarrow T_{\pi(p)} B$ is not of full rank at p . In this subsection we discuss several of the numerical invariants one may associate with Lagrangian singularities, which are the map germs of such singular points, viewed up to Lagrangian equivalence. For more information on the theory of singularities, the reader is referred to the classic book by Arnol’d, Gusein-Zade, and Varchenko [4].

After the *corank*, which is the codimension of the image of $d\pi_p$, the first numerical invariant we discuss is the *multiplicity* of a singularity, usually denoted μ . Roughly speaking, it is the (maximum) number of non-degenerate critical points into which a singularity splits under a small perturbation. In effect, one can show (cf. [4, §6.3]) that a function having a critical point of finite multiplicity μ is equivalent, in a neighborhood of the point, to a polynomial of degree $\mu + 1$. For the precise definition of μ (also called the Milnor number), the reader can consult [30, Definition 2.1]

One of the subtler numerical invariants to apprehend is the *modality* of a singularity. This non-negative integer, traditionally denoted by m , counts the number of continuous parameters (or moduli) that enter into the definition of the associated normal form. We refer to [30, §2.4] or [4, p.184] for the exact definition. A singularity of modality 0 is called *simple*. Simple singularities, having no moduli, appear discretely.

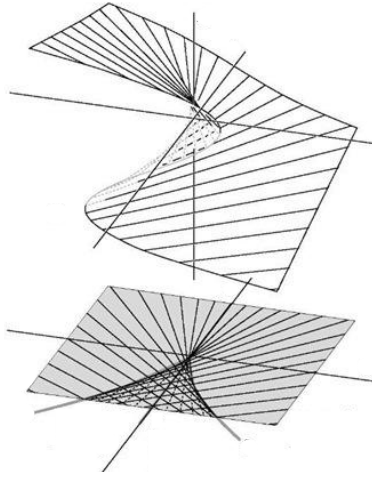


FIGURE 1. The Whitney pleat

Arnol'd has classified stable simple singularities. (Stable singularities are those which persist under small perturbations; they are the only ones visible “with the naked eye”.) Below we list the notation for the simple singularities, along with function germs representing each class:

- (A) $A_k: \pm x_1^{k+1} + x_2^2 + \cdots + x_n^2 \quad (k \geq 2);$
- (D) $D_k: x_2(x_1^2 \pm x_2^{k-2}) + x_3^2 + \cdots + x_n^2 \quad (k \geq 4);$
- (E) $E_6: x_1^3 \pm x_2^4 + x_3^2 + \cdots + x_n^2;$
 $E_7: x_1(x_1^2 + x_2^3) + x_3^2 + \cdots + x_n^2;$
 $E_8: x_1^3 + x_2^5 + x_3^2 + \cdots + x_n^2.$

In reference to their organizational structure – reminiscent of that of finite subgroups of $SU(2)$ – simple singularities are sometimes called *ADE singularities*.

Singularities of type *A* are of corank 1 and those of type *D* and *E* are of corank 2. Thus any simple singularity is of corank at most 2 (see [2, Lemma 4.2]). Moreover, any corank 1 singularity of finite multiplicity is necessarily simple. Thus the type *A* singularities (also called Morin singularities) can be characterized as those having corank 1 and finite multiplicity; these facts are summarized in [4, §11.1] or [30, Theorem 2.48]. The multiplicity of an ADE singularity is indicated in its subscript.

We shall be primarily interested in A_2 and A_3 singularities. An A_2 -type singularity is sometimes referred to as a *fold singularity*, and an A_3 -type singularity as a *cuspl singularity*. As an example of a fold singularity, consider the projection of the sphere to the horizontal plane touching the south pole. The singular points are the points of the equator; they are all fold singularities. They arise from a coalescence of two critical points. One can realize a cuspl singularity from the projection of the surface $z = x^3 + xy$ to the (y, z) -plane; the warp on one half of the surface is known as a Whitney pleat. For visualizations of both of these fundamental examples, see Figures 7 and 8 in Section 1 of [4]. We have also included a graph of the Whitney pleat in Figure 1. A famous theorem of Whitney (see [4, §1.5]) states that the stable singularities of a differentiable map between surfaces are either non-degenerate, or of type A_2 or A_3 .

Finally, there is yet another numerical invariant of a critical point, called the *singularity index* and denoted β . The singularity index is defined by the asymptotic behavior of associated oscillatory integrals [1, Definition 3], [3, Definition 4.2.1]. The index of singularity is β if the integral is of size $t^{-\frac{m}{2} + \beta}$ for generic choice of amplitude function. Arnol'd [1, 3] has calculated

the singularity index for all simple singularities and many others; they turn out to be rational numbers. For the simple singularities one has $\beta = 1/2 - 1/N$, where N is the corresponding Coxeter number $N(A_k) = k + 1$, $N(D_k) = 2k + 2$, $N(E_6) = 12$, $N(E_7) = 18$, $N(E_8) = 30$.

7. NON-DEGENERACY OF PHASE FUNCTIONS AND PROOF OF THEOREM 1.3

We assume in the whole section that $\nu \in i\mathfrak{a}^*$ is regular. We keep the same notation as in the previous §6.

7.1. Non-degenerate critical points.

Proposition 7.1. *The origin $0 \in \mathfrak{u}_{\text{ab}}^*$ is not a critical value of $\mathcal{L}_\nu \rightarrow \mathfrak{u}_{\text{ab}}^*$. The light zone $L_\nu \subset A$ contains a translate of the negative Weyl chamber*

$$\exp(-\mathfrak{a}_+) = \{a \in A, \alpha(a) < 1 \ \forall \alpha \in \Delta_+\}.$$

Proof. Under the map $a \mapsto \ell_a$, the preimage of a neighborhood of $0 \in \mathfrak{u}_{\text{ab}}^*$ contains a translate of the negative Weyl chamber inside A . Hence the second assertion about L_ν immediately follows from the first assertion because of Theorem 6.2.

Inside \mathcal{L}_ν the fiber above $0 \in \mathfrak{u}_{\text{ab}}^*$ consists of $\{\text{Im } \nu^w, w \in W\}$. In a neighborhood of any of these points we have that $\mathcal{L}_\nu \rightarrow \mathfrak{u}_{\text{ab}}^*$ is a local diffeomorphism. Indeed we compute that the tangent space of \mathcal{L}_ν at $\text{Im } \nu^w$ is $[\mathfrak{k}, \text{Im } \nu^w] \cap \mathcal{J}^*$ which surjects onto $\mathfrak{u}_{\text{ab}}^*$ because ν is regular. Equivalently we are computing the critical points of the Iwasawa projection $u \mapsto \langle \nu, H(wu) \rangle$ which is well-known [24] to be non-degenerate if ν is regular. \square

The proposition implies that for all g inside a certain explicit open set of S , the phase function $u \mapsto F_\nu(u, g)$ is Morse. This is the typical behavior of phase functions and more precisely it is true for any phase function with parameters that is nondegenerate in the sense of [22, §1], as follows from Sard's lemma. We have seen in the previous §6 that $F_\nu(u, g)$ is nondegenerate, which pertains to the smoothness of the Lagrangian Λ_ν inside T^*S .

7.2. Stationary phase approximation for Morse–Bott functions. We consider an oscillatory integral

$$(7.1) \quad \int_{\mathbb{R}^d} e^{itG(x)} \alpha(x) dx$$

where $\alpha, G \in C^\infty(\mathbb{R}^d)$ with α of compact support. The following is a generalization of the stationary phase approximation to the case of Morse–Bott functions, see e.g. [15].

Proposition 7.2. *Suppose that G is Morse–Bott and that the set of critical points of G contained in the support of α form a connected submanifold $W \subset \mathbb{R}^d$. Then the oscillatory integral (7.1) is asymptotic as $t \rightarrow \infty$ to*

$$\left(\frac{2\pi}{t}\right)^{\frac{d-e}{2}} e^{itG(W) - \frac{i\pi}{4}\sigma} \int_W \alpha(x) |\det_W G''(x)|^{-\frac{1}{2}} dx$$

where $e = \dim W$, $G(W)$ is the value of $G(x)$ at any point $x \in W$, and σ (resp. $\det_W G''$) is the signature (resp. determinant) of the Hessian of G in the direction transverse to W .

7.3. Proof of Theorem 1.3. By Proposition 7.1 we can choose an open set $V \subset A$ inside the translate of the negative Weyl chamber, which is small enough so that the restriction of $\Lambda_\nu \rightarrow A$ to V is an unramified covering. Let $\psi \in C^\infty(V)$ be a function such that the graph of $d\psi$ is entirely inside Λ_ν . Such a function exists because Λ_ν is transverse to the vertical fibers of $T^*V \rightarrow V$.

We test the Whittaker function against a symbol localized in phase-space inside a single sheet of $\Lambda_\nu \cap T^*V$. Namely we form the integral

$$\int_V W_{t\nu}(ta)\delta(ta)^{-1/2}\nu(ta)e^{-it\psi(a)}da,$$

which in view of (2.1) is equal to

$$\int_V \int_U \delta(\mathbf{w}u)^{1/2} e^{it(B(H_\nu, H(\mathbf{w}u)) - \langle \ell_1, aua^{-1} \rangle - \psi(a))} d\mathbf{u} da.$$

By construction the new phase function $(u, a) \mapsto B(H_\nu, H(\mathbf{w}u)) - \langle \ell_1, aua^{-1} \rangle - \psi(a)$ is Morse-Bott. Indeed it has a single connected manifold of critical points parametrized by $a \in V$.

We apply Proposition 7.2 which shows that up to non-zero constants the integral is asymptotic to $t^{-\dim(U)/2}$ as $t \rightarrow \infty$. Applying the triangle inequality we deduce that for each $t \geq 1$ there exists $a \in V$ such that $W_{t\nu}(ta)$ is asymptotically greater than $t^{(\text{ht}(G) - \dim(U))/2}$. \square

7.4. The Whittaker function as superposition of plane waves. In fact we can prove the more precise result that for a in a negative translate of the Weyl chamber inside the light zone L_ν , the Whittaker function $a \mapsto W_\nu(a)$ is asymptotically a linear superposition of $|W|$ plane waves, where $W = W(\mathfrak{g}, \mathfrak{a})$ is the Weyl group. This is because we have shown that $F_\nu(u, a)$ is Morse, and we can apply the stationary phase approximation in its uniform version with parameters which can be found from [34, Theorem 7.7.6] and [64, Theorem 2.9]. The fibers of the Lagrangian mapping $\Lambda_\nu \rightarrow A$ above a negative translate of the Weyl chamber, which have cardinality $|W|$, correspond to the momentum of the plane waves. Since by construction the momentum are distinct, these plane waves are linearly independent which implies the lower bound of Theorem 1.3.

We note that the other proof we have given in §7.3 above amounts to directly testing $W_\nu(a)$ against one of the plane wave. Some of the $|W|$ plane waves coalesce when a approaches the caustic C_ν , which will be studied in the next section for $G = \text{PGL}_3(\mathbb{R})$.

8. PROOF OF THEOREM 1.7

In this section, we impose a self-duality assumption on the spectral parameter $\nu \in i\mathfrak{a}_+^*$. This allows us to give a precise description of the critical set for $G = \text{PGL}_3(\mathbb{R})$. More precisely, we shall provide explicit equations defining the shadow zone, the light region, and the caustic locus defined in §6.3.

Now a uniform description of the asymptotic behavior of the Jacquet-Whittaker function depends on more than just this partition. One also needs information on the *configuration* of the critical points, which is encoded in the singularities of the Lagrangian mapping. Thus, in the main result of this section, Proposition 8.1, we shall decompose the caustic locus \mathbf{C} into strata according to the degeneracy type, and decompose the light region \mathbf{L} according to the size of the fibers.

All of this information will determine the asymptotic size of $W_\nu(a)$, uniformly in ν and a .

8.1. **Notation and hypotheses.** Let

$$i\mathfrak{a}_{\text{sd}}^* = \{\nu \in i\mathfrak{a}_+^* : \langle \nu, H_1 \rangle = \langle \nu, H_2 \rangle\}$$

be the center of the positive Weyl chamber $i\mathfrak{a}_+^*$. Unramified principal series representations π_ν are self-dual precisely for $\nu \in i\mathfrak{a}_{\text{sd}}^*$, whence the notation. Note that $\nu \in i\mathfrak{a}_{\text{sd}}^*$ is the positive ray generated by $\nu_0 = i(w_1 + w_2)$. Thus, we may write $\nu = 2\pi t\nu_0$ for $t > 0$. When studying the Lagrangian $\Lambda_{t\nu_0}^{\text{red}}$ we can, without loss of generality, restrict to $t = 1$; this follows from the scale invariance of the phase function in the (ν, a) parameters.

For notational simplicity, we shall work with Lie algebra structures rather than their duals. Thus instead of $i\mathfrak{a}^*$ we work with \mathfrak{a} , using the identification between the two given by the form $B(X, Y) = \text{tr}(XY)$. Thus, the matrix in \mathfrak{a} corresponding to $\nu_0 \in i\mathfrak{a}^*$ is $H = \text{diag}(1, 0, -1)$, and we shall work with $\text{Ad}_K(H)$ rather than $\text{Ad}_K^*(\text{Im } \nu_0)$.

Similarly, we shall work with the traceless symmetric matrices \mathfrak{p} and the 4-dimensional subspace of tridiagonals \mathcal{J} , rather than \mathfrak{p}^* and \mathcal{J}^* . We denote by \mathcal{J}_+ the open cone with positive entries on the first diagonal and we coordinatize \mathcal{J}_+ as

$$(8.1) \quad \mathcal{J}_+ = \left\{ \begin{pmatrix} \frac{1}{3}(2x_1 + x_2) & y_1 & 0 \\ y_1 & \frac{1}{3}(x_2 - x_1) & y_2 \\ 0 & y_2 & -\frac{1}{3}(x_1 + 2x_2) \end{pmatrix} : x_1, x_2 \in \mathbb{R}, y_1, y_2 \in \mathbb{R}_{>0} \right\}.$$

We systematically (and without further comment) use the coordinates on A given by the positive simple roots $y_1 = \alpha_1(a)$ and $y_2 = \alpha_2(a)$. Let $\mathcal{A}(a)$ denote the fiber over $a \in A$ under the map $\mathcal{J}_+ \rightarrow A$ (it is a 2-dimensional affine space).

There is a Lagrangian mapping $\Lambda_H^{\text{red}} \rightarrow A$ as described in (6.5). Let

$$\mathcal{F}(a) \subset \Lambda_H^{\text{red}}$$

denote the fiber over $a \in A$. According to (6.6) we have $A = \mathbf{S} \sqcup \mathbf{L} \sqcup \mathbf{C}$, according to whether $\mathcal{F}(a)$ is empty, consists entirely of non-singular points, has at least one singular point, respectively.

8.2. **Statement of result.** According to Theorem 6.2 there is a canonical bijection between $\text{Ad}_K(H) \cap \mathcal{J}_+$ and Λ_H^{red} , and this bijection commutes with the projection maps to A . Moreover Theorem 6.2 provides an explicit description of $\mathcal{F}(a)$ inside $\mathcal{A}(a)$. Namely, if

$$\chi_{\det}(a) = \{s \in \mathcal{A}(a) : \det(s) = 0\} \quad \text{and} \quad \chi_{\text{tr}}(a) = \{s \in \mathcal{A}(a) : \text{Tr}(s^2) = 2\},$$

then

$$(8.2) \quad \mathcal{F}(a) = \chi_{\det}(a) \cap \chi_{\text{tr}}(a).$$

This is the starting point for studying $\mathcal{F}(a)$ and the partition $A = \mathbf{S} \sqcup \mathbf{L} \sqcup \mathbf{C}$.

We begin by defining certain subsets of A which are represented graphically in Figure 2. Let

$$\mathbf{L}_1 = \{27y_1^4 y_2^4 - 18y_1^2 y_2^2 + 4y_2^2 + 4y_1^2 < 1\}$$

and

$$\mathbf{L}_2 = \{27y_1^4 y_2^4 - 18y_1^2 y_2^2 + 4y_2^2 + 4y_1^2 > 1 \text{ and } y_1^2 + y_2^2 < 1\}.$$

Let a_{cusp} be the unique point in A given by

$$(y_1, y_2) = (1/\sqrt{3}, 1/\sqrt{3}).$$

Finally put

$$\mathbf{C}_1 = \{y_1^2 + y_2^2 = 1\},$$

and

$$\mathbf{C}_2 = \left\{ (y_1, y_2) \neq (1/\sqrt{3}, 1/\sqrt{3}) : 27y_1^4 y_2^4 - 18y_1^2 y_2^2 + 4y_2^2 + 4y_1^2 = 1 \right\}.$$

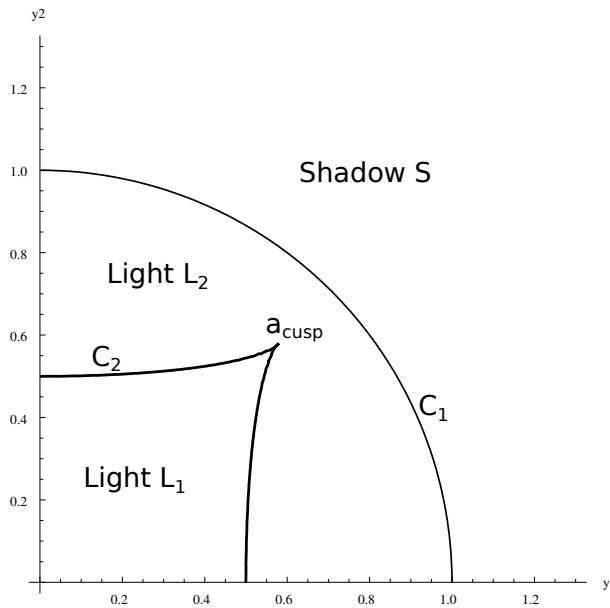


FIGURE 2. Critical point configuration for self-dual spectral parameter.

In this section we prove the following result.⁴

Theorem 8.1. *We have*

$$S = \{y_1^2 + y_2^2 > 1\}, \quad L = L_1 \sqcup L_2, \quad C = C_1 \sqcup \{a_{\text{cusp}}\} \sqcup C_2.$$

Moreover, we have the following critical point configurations:

- (1) for all $a \in L_1$ we have $|\mathcal{F}(a)| = 6$;
- (2) for all $a \in L_2$ we have $|\mathcal{F}(a)| = 2$;
- (3) for all $a \in C_1$ we have $|\mathcal{F}(a)| = 1$, consisting of a point of fold type;
- (4) for all $a \in C_2$ we have $|\mathcal{F}(a)| = 4$, two of which are non-degenerate, and two of which are degenerate of fold type;
- (5) we have $|\mathcal{F}(a_{\text{cusp}})| = 2$, and the two points are of cuspidal type.

We note the above varieties and equations are invariant under involution $(y_1, y_2) \mapsto (y_2, y_1)$ which is the reflection across the diagonal. This is explained by the equivariant action of Ad_w on $\Lambda_H^{\text{red}} \rightarrow A$.

8.3. Idea of proof. From (8.1) we get

$$\begin{aligned} \chi_{\det}(a) &= \{9y_1^2(x_1 + 2x_2) - 9y_2^2(2x_1 + x_2) = 2(x_2^3 - x_1^3) + 3(x_1x_2^2 - x_1^2x_2)\} \\ \chi_{\text{tr}}(a) &= \{x_1^2 + x_1x_2 + x_2^2 = 3(1 - y_1^2 - y_2^2)\}. \end{aligned}$$

The idea of the proof of Theorem 8.1 is to study the “intersection configuration” of $\chi_{\det}(a)$ with $\chi_{\text{tr}}(a)$ – affine curves in $\mathcal{A}(a)$ of degree 3 and 2, respectively – as a varies throughout A . For a near the origin they will intersect (transversally) in 6 points, and for a large they will not intersect at all; these are the regions L_1 and S . For intermediate ranges of

⁴In the early stages of the elaboration of this paper, the authors’ intuition was that the caustic set C should be precisely equal to the boundary arc C_1 . It was therefore quite surprising to discover that there are interior points C_2 which also contribute to the caustic set (and in fact account for the largest values of the Whittaker function).

a , transversal intersections will coalesce into points of tangency, before disappearing. This can happen in a few different ways, roughly corresponding to the ways in which a degree 6 polynomial can factorize over the reals. On the other hand, symmetry constraints will limit which combinations can arise. Once the intersection configuration has been mapped out, one can then read off the underlying singularity type by the numerical invariants recalled in §6.4.

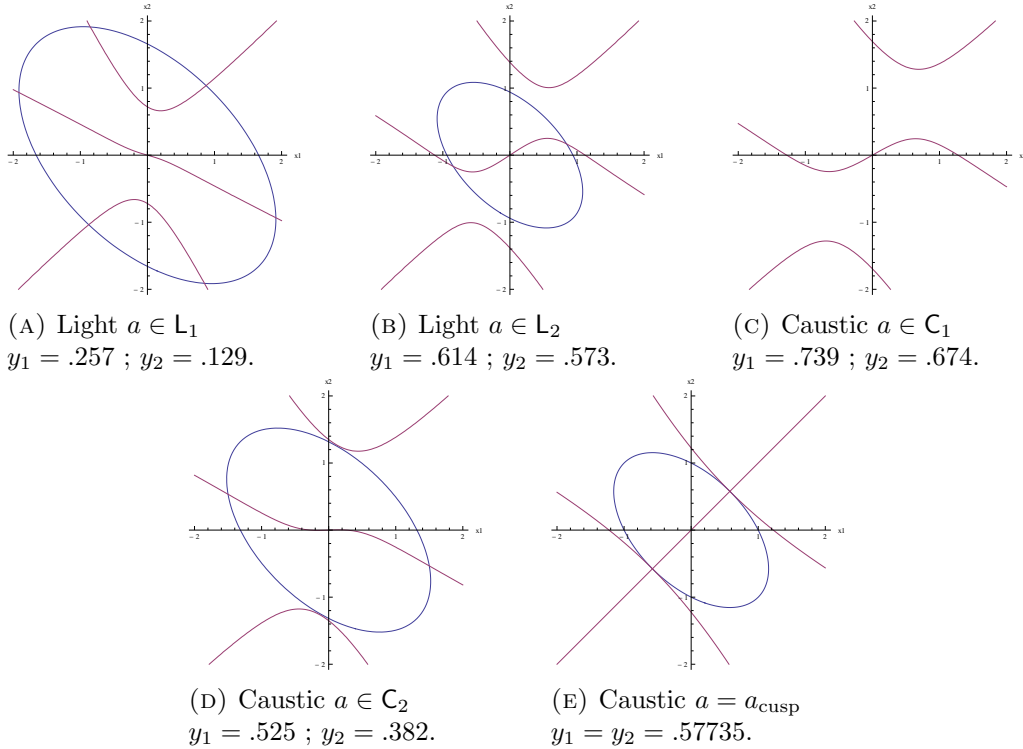


FIGURE 3. The curves $\chi_{\det}(a)$ and $\chi_{\text{tr}}(a)$ for different values of a

In Figure 3, we show five different intersection configurations corresponding to the five cases of Theorem 8.1. They can be mapped onto the corresponding strata of Figure 2. Note that in the configuration (C) representing the outer caustic C_1 , the ellipse in (B) has collapsed to a single point; in the shadow zone S (not pictured) this point has disappeared. Compare Figure 3 to the classical bifurcation diagram of cuspidal singularities, as given, for example, in [18, Figure 4].

We illustrate the argument by carrying it out along the ray $y_1 = y_2 = y$ with $y > 0$. In this case the equation for $\chi_{\det}(a)$ simplifies. Indeed the linear term $x_1 - x_2$ factors, making $\chi_{\det}(a)$ the union of the line $x_1 = x_2$ and the quadric hyperbola with equation

$$9y^2 = 2x_1^2 + 5x_1x_2 + 2x_2^2.$$

The intersection with $\chi_{\text{tr}}(a)$ can be easily computed and we find that the different zones $a \in L_1$, $a \in C_2$, $a \in L_2$, $a \in C_1$, $a \in S$ are given by the intervals

$$0 < y < \frac{1}{\sqrt{3}}, \quad y = \frac{1}{\sqrt{3}}, \quad \frac{1}{\sqrt{3}} < y < \frac{1}{\sqrt{2}}, \quad y = \frac{1}{\sqrt{2}}, \quad \text{and} \quad y > \frac{1}{\sqrt{2}},$$

respectively, a result which agrees with Figure 2 and Theorem 8.1.

8.4. The shadow zone. In this section we establish the first statement in Theorem 8.1 regarding shadow zone. We also establish the cardinality of the fibers in C_1 and a lower bound in the fibers in the other regions.

Proposition 8.2. *We have*

$$\mathbf{S} = \{y_1^2 + y_2^2 > 1\}.$$

Moreover, when $y_1^2 + y_2^2 = 1$ there is one unique critical point, and if $y_1^2 + y_2^2 < 1$ there are at least two distinct critical points.

Proof. By (5.1), it suffices to prove that $\mathbf{S} \subseteq \{y_1^2 + y_2^2 > 1\}$. According to (8.2), we must show that if $a \in A$ verifies $y_1^2 + y_2^2 \leq 1$ then $\chi_{\det}(a) \cap \chi_{\text{tr}}(a) \neq \emptyset$. Any element in $\mathcal{A}(a)$ has norm-squared $2y_1^2 + 2y_2^2 + d^2 + e^2 + f^2$, for some diagonal entries d, e, f . From the hypothesis $y_1^2 + y_2^2 \leq 1$ we deduce that $\chi_{\text{tr}}(a)$ is not empty. It contains, say, the elements

$$s_{\pm} = \pm \text{diag}(\alpha, -\alpha, 0) + \Upsilon(a), \quad \text{where} \quad \Upsilon(a) = \begin{pmatrix} 0 & y_1 & 0 \\ y_1 & 0 & y_2 \\ 0 & y_2 & 0 \end{pmatrix},$$

for some $\alpha \geq 0$.

Note that $y_1^2 + y_2^2 = 1$ if and only if $\chi_{\text{tr}}(a) = \{\Upsilon(a)\}$. As $\Upsilon(a)$ has determinant 0, we have $\Upsilon(a) \in \chi_{\det}(a)$ as required.

If $y_1^2 + y_2^2 < 1$ then the points s_{\pm} are distinct and $\det(s_{\pm}) = \pm \alpha y_2^2$ are of opposite sign. Now $\chi_{\text{tr}}(a)$, being an circle, is connected. By the intermediate value theorem, there is $s \in \chi_{\text{tr}}(a)$ such that $\det(s) = 0$. As the same is true of the antipode of s , there are at least two distinct points lying in $\chi_{\det}(a) \cap \chi_{\text{tr}}(a)$. \square

We illustrate the argument of the proof above with a graph in the (x_1, x_2) -plane of the two curves $\chi_{\det}(a)$ and $\chi_{\text{tr}}(a)$. Let $s_i := \alpha d_i + \Upsilon(a)$ where

$$\begin{aligned} d_1 &:= \text{diag}(-1, 1, 0) & d_2 &:= \text{diag}(0, 1, -1) & d_3 &:= \text{diag}(1, 0, -1) \\ d_4 &:= \text{diag}(1, -1, 0) & d_5 &:= \text{diag}(0, -1, 1) & d_6 &:= \text{diag}(-1, 0, 1). \end{aligned}$$

Thus in particular $s_1 = s_+$ and $s_4 = s_-$. We have the property that $s_i \in \chi_{\text{tr}}(a)$ for all $i = 1, \dots, 6$ and the points are cyclically ordered. Suppose by symmetry that $y_1 > y_2$. Then it can be verified that $\det(s_i) > 0$ if $i \in \{1, 2, 3\}$ while $\det(s_i) < 0$ if $i \in \{4, 5, 6\}$. Thus there are at least two intersection points in $\chi_{\det}(a) \cap \chi_{\text{tr}}(a)$ which was how we established Proposition 8.2. We consider the following numerical values $y_1 = .614$ and $y_2 = .573$ in Figure 4.

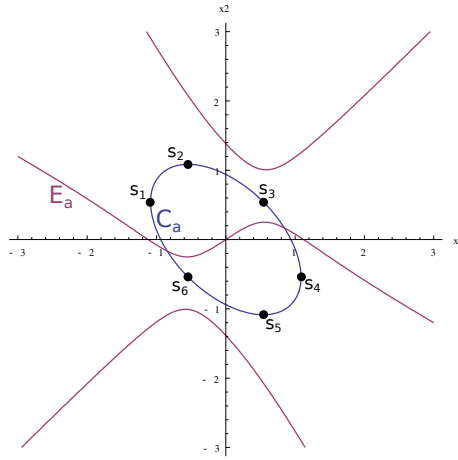


FIGURE 4. The curves $\chi_{\det}(a)$ and $\chi_{\text{tr}}(a)$ for $a \in \mathbf{L}_2$.

The case $y_1^2 + y_2^2 = 1$ is even simpler since the ellipse collapses into a single point $\{\Upsilon(a)\}$ and the resulting configuration is shown in Figure 3 on Page 31 with the numerical values $y_1 = .739$ and $y_2 = .673$.

8.5. Determination of the caustic locus. For $a \in A$ consider the polynomials

$$\begin{aligned} C_a(x) &= x_1^2 + x_1x_2 + x_2^2 - 3(1 - y_1^2 - y_2^2), \\ E_a(x) &= 2x_1^3 + 3x_1^2x_2 + 9x_1y_1^2 + 18x_2y_1^2 - 2x_2^3 - 3x_1x_2^2 - 9x_2y_2^2 - 18x_1y_2^2, \\ D_a(x) &= x_1y_1^2 + x_2y_2^2 - x_1^2x_2 - x_1x_2^2. \end{aligned}$$

The first two are the defining equations for the curves $\chi_{\text{tr}}(a)$ and $\chi_{\text{det}}(a)$, respectively. The role of the last one will be explained presently.

Proposition 8.3. *We have $a \in \mathbb{C}$ if and only if there exists $x \in \mathbb{R}^2$ satisfying*

$$C_a(x) = E_a(x) = D_a(x) = 0.$$

Moreover, any singularity for $\Lambda_H^{\text{red}} \rightarrow A$ (and hence any degenerate critical point of the Whittaker phase function F_H) is of corank 1.

Proof. Using the coordinates (8.1), the existence of a solution to $C_a(x) = E_a(x) = 0$ is equivalent to the fiber $\mathcal{F}(a)$ being non-empty. To characterize $a \in \mathbb{C}$ we must then, in view of (6.7), determine the a for which there is $s \in \mathcal{F}(a)$ whose tangent space along Λ_H^{red} fails to be transverse to the fiber. If $\mathfrak{t}(s)$ denotes this tangent space, then this condition is equivalent to $\mathfrak{a} \cap \mathfrak{t}(s) \neq 0$. We shall show, again using the coordinates (8.1), that this is the same as the existence of a solution to $D_a(x) = 0$.

Let $s \in \text{Ad}_K(H)$ and write $T(s)$ for the tangent space of s along the whole adjoint orbit $\text{Ad}_K(H)$. Then we may identify $T(s)$ with $\{[k, s] : k \in \mathfrak{k}\}$. Now if $s \in \text{Ad}_K(H) \cap \mathcal{J}$ then $\mathfrak{t}(s)$ may be identified with $T(s) \cap \mathcal{J}$. To compute this intersection explicitly we denote matrices in \mathfrak{k} as

$$k = \begin{pmatrix} 0 & b & c \\ -b & 0 & a \\ -c & -a & 0 \end{pmatrix} \quad (a, b, c \in \mathbb{R}).$$

Taking $s \in \Lambda_H^{\text{red}}$, viewed as an element of $\text{Ad}_K(H) \cap \mathcal{J}_+$ via diagram (6.3) and with the coordinates of (8.1), and setting the upper right-hand entry of $[k, s]$ to zero, we find that $\mathfrak{t}(s)$ is the subspace of $T(s)$ cut out by the equation $-ay_1 + by_2 - c(x_1 + x_2) = 0$.

Having computed $\mathfrak{t}(s)$, one then finds $\mathfrak{a} \cap \mathfrak{t}(s)$ by setting the off-diagonals of $\mathfrak{t}(s)$ to zero. A short calculation produces the system

$$\begin{pmatrix} y_1y_2 & y_2^2 - x_1(x_1 + x_2) \\ -y_1^2 + x_2(x_1 + x_2) & -y_1y_2 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = 0.$$

Thus $\mathfrak{t}(s) \cap \mathfrak{a}$ is not reduced to 0 if and only if the determinant $(x_1 + x_2)D_a(x)$ of the above square matrix vanishes. Now it can be directly checked that no $x \in \mathbb{R}^2$ with $x_1 + x_2 = 0$ can satisfy both $C_a(x) = 0$ and $E_a(x) = 0$. This establishes the first claim.

To see the second claim, note that solutions to the above matrix equation precisely describe the kernel of the differential of the map $\Lambda_H^{\text{red}} \rightarrow A$ at s . The corank 1 property of singularities for this map is then evident since $y_1y_2 \neq 0$ and so the above matrix is never 0. The link to the Whittaker phase function is made via (6.7). \square

Lemma 8.4. *We have $\mathbb{C}_1 \cup \{a_{\text{cusp}}\} \subset \mathbb{C}$.*

Proof. We first deduce from Proposition 8.2 that every point of \mathbb{C}_1 is critical. To show that every $a \in \mathbb{C}_1$ is in fact degenerate, we note that the corresponding s has vanishing diagonal elements, so that equation $D_a(x) = 0$ is trivially true.

Note that the symmetric matrices

$$s_{\text{cusp}}^+ = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \quad \text{and} \quad s_{\text{cusp}}^- = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

lie in $\text{Ad}_K(H)$, for their characteristic polynomial $x - x^3$ is the same as that of H . Moreover, both s_{cusp}^+ and s_{cusp}^- lie in the affine subspace $\mathcal{A}(a_{\text{cusp}})$. Thus both s_{cusp}^+ and s_{cusp}^- are critical points of F_H over a_{cusp} . Finally, s_{cusp}^+ and s_{cusp}^- verify the equation $D_a(x) = 0$, which shows that they are degenerate. \square

Now observe that the equation $C_a(x) = 0$ is that of a conic, which, if $y_1^2 + y_2^2 < 1$, is not reduced to a point. We may therefore choose a birational map from $\mathbb{P}^1(\mathbb{R})$ to its solution locus. We make the substitution

$$(8.3) \quad x_1 = \frac{1-t^2}{1+t+t^2} \sqrt{3(1-y_1^2-y_2^2)}, \quad x_2 = \frac{t(t+2)}{1+t+t^2} \sqrt{3(1-y_1^2-y_2^2)}.$$

With this parametrization, the polynomials $E_a(x)$ and $D_a(x)$ become

$$\begin{aligned} E_a(t) &= y_1^2(t^2 + 4t + 1)^3 + y_2^2(t^2 - 2t - 2)^3 + 2t^6 + 6t^5 - 15t^4 - 40t^3 - 15t^2 + 6t + 2, \\ D_a(t) &= y_1^2(1-t^2)(t^2 + 4t + 1)^2 + y_2^2(2t+t^2)(t^2 - 2t - 2)^2 + 6t^5 + 15t^4 - 15t^2 - 6t, \end{aligned}$$

again under the hypothesis that $y_1^2 + y_2^2 < 1$.

Proposition 8.5. *We have $\mathcal{C} \subset \mathcal{C}_1 \sqcup \{a_{\text{cusp}}\} \sqcup \mathcal{C}_2$.*

Proof. It suffices to show that if a is in \mathcal{C} but not in \mathcal{C}_1 then a is in $\{a_{\text{cusp}}\} \sqcup \mathcal{C}_2$.

We see that $a \in \mathcal{C} - \mathcal{C}_1$ satisfies $y_1^2 + y_2^2 < 1$ and moreover there is $t \in \mathbb{P}^1(\mathbb{R})$ such that $E_a(t) = D_a(t) = 0$. This system has a *complex* solution $t \in \mathbb{P}^1(\mathbb{C})$ if, and only if, the resultant $R(a) = \text{Res}(E_a(t), D_a(t))$ vanishes. One computes

$$R(a) = (y_1^2 + y_2^2 - 1)^4 (27y_1^4 y_2^4 - 18y_1^2 y_2^2 + 4y_2^2 + 4y_1^2 - 1)^2.$$

The set of $a \in A$ such that $y_1^2 + y_2^2 < 1$ and $R(a) = 0$ is precisely $\{a_{\text{cusp}}\} \sqcup \mathcal{C}_2$. \square

Note that we have the relation

$$D_a(t) = \left(\frac{-2t-1}{3} \right) E_a(t) + \left(\frac{t^2+t+1}{9} \right) E'_a(t).$$

From this it follows that

$$(8.4) \quad \text{the real solutions of } E_a(t) = D_a(t) = 0 \text{ are precisely those of } E_a(t) = E'_a(t) = 0.$$

This latter system is slightly more convenient, since it allows us to characterize degenerate critical points in terms of the multiplicities of roots of the polynomial $E_a(t)$. Note that the discriminant of $E_a(t)$ is proportional by an absolute constant to

$$(y_1^2 + y_2^2 - 1)^2 (27y_1^4 y_2^4 - 18y_1^2 y_2^2 + 4y_2^2 + 4y_1^2 - 1)^2,$$

whose zero set agrees with the expression $R(a)$ above.

8.6. Light configuration. In this section we finish the proof of the light zone configuration in Theorem 8.1.

Proposition 8.6. *We have the following critical point configurations:*

- (1) for any $a \in \mathcal{L}_1$ one has $|\mathcal{F}(a)| = 6$;
- (2) for any $a \in \mathcal{L}_2$ one has $|\mathcal{F}(a)| = 2$.

Proof. Another way to state the proposition is that for $a \in L_1$ (resp., $a \in L_2$) there are 6 (resp., 2) distinct real solutions to $E_a(t) = 0$.

Note that, for $i = 1, 2$, it is enough to show the stated root configuration for *some* value of $a \in L_i$. Indeed, the root configuration cannot change within L_i , since changing to any other root configuration would require hitting the caustic C . By Proposition 8.5, this is impossible.

For (1) we can, for example, use $(y_1, y_2) = (1/2, 1/2)$. In this case, equation $E_a(t) = 0$ becomes $10t^6 + 30t^5 - 3t^4 - 56t^3 - 21t^2 + 12t + 1 = 0$, which has 6 distinct real roots. For (2) we can use the point $(y_1, y_2) = (\sqrt{3}/2\sqrt{2}, \sqrt{3}/2\sqrt{2})$. In this case, we obtain $(2t^2 + 2t - 1)(11t^4 + 22t^3 + 9t^2 - 2t + 5) = 0$, which has two distinct real roots. \square

As a corollary, we deduce the following result.

Corollary 8.7. *If $a \in \{a_{\text{cusp}}\} \sqcup C_2$, then any solution t to $E_a(t) = 0$ is real. In particular, we have $C = C_1 \sqcup \{a_{\text{cusp}}\} \sqcup C_2$.*

Proof. Suppose that for some $a \in \{a_{\text{cusp}}\} \sqcup C_2$ there is a pair of non-real, complex conjugate roots of $E_a(t) = 0$. Then there exists a neighborhood U of a such that the same is true for every $a' \in U$. But this neighborhood necessarily intersects L_1 , where Proposition 8.6 assures us that there are no complex roots. Contradiction.

The second statement follows from the proof of Proposition 8.5. \square

We note that $|\mathcal{F}(a)|$ is even for $a \in L_1 \sqcup L_2$. This is explained by the involution $(x_1, x_2) \rightarrow (-x_1, -x_2)$ which preserves $\chi_{\det}(a)$ and $\chi_{\text{tr}}(a)$ above, and thus acts on the fibers $\mathcal{F}(a)$ for any $a \in A$. The only fixed points of the involution are $x_1 = x_2 = 0$ which project to the caustic C_1 . In fact we shall see below that $|\mathcal{F}(a)| = 1$ for every $a \in C_1$ which is the only case where the multiplicity is odd.

8.7. Degeneracy types. Having obtained the caustic locus in Corollary 8.7, we now look at the fibers $\mathcal{F}(a)$ over caustic points. We first determine their multiplicities, which will be of great help in identifying their degeneracy type.

Proposition 8.8. *We have the following critical point configurations:*

- (1) for any $a \in C_1$ one has $|\mathcal{F}(a)| = 1$, of multiplicity 2;
- (2) one has $|\mathcal{F}(a_{\text{cusp}})| = 2$, each of multiplicity 3;
- (3) for any $a \in C_2$ one has $|\mathcal{F}(a)| = 4$, two of multiplicity 2, two of multiplicity 1.

Proof of (1): We have already proved (1) in Proposition 8.2.

Proof of (2): By (8.4) we must show that $E_{a_{\text{cusp}}}(t) = 0$ admits two distinct real roots, each of multiplicity 3. Inserting $(y_1, y_2) = (1/\sqrt{3}, 1/\sqrt{3})$ into the formula for $E_a(t) = 0$ we obtain $(2t^2 + 2t - 1)^3 = 0$.

Proof of (3): Let $a \in C_3$. By (8.4) we must show that $E_a(t) = 0$ admits four distinct real roots, of which two are double and two are simple.

We will make use of the symmetry of the solution locus $C_a(x) = E_a(x) = D_a(x) = 0$ given by $x \mapsto -x$. In the parametrization (8.3), this corresponds to $\sigma(t) = (t + 2)/(-2t - 1)$. We deduce that if $t \in \mathbb{P}^1$ satisfies $E_a(t) = D_a(t) = 0$, then so does $\sigma(t)$. We deduce from (8.4) that the system $E_a(t) = E'_a(t) = 0$ is also invariant under σ . In other words, σ sends roots of $E_a(t)$ to roots of $E_a(t)$, and conserves their multiplicities.

By Corollary 8.7, $E_a(t)$ admits 6 real roots, when counted with multiplicity. Since $C_2 \subset C$, one of these roots must have multiplicity strictly greater than 1. Since $a \notin C_1$, any solution x to $E_a(x) = 0$ is non-zero, so that the map $x \mapsto -x$, and hence σ , has no fixed points. From these observations we deduce that either two roots are of multiplicity 2 and the others are

non-degenerate (as is stated in the proposition) or that there are 2 distinct real solutions, each with multiplicity 3. We must show that for $a \in \mathbf{C}_2$ the latter cannot occur.

Recall from [19] the notion of the principal subresultant coefficients $\text{PSPC}_\ell(P, Q)$. These can be used to characterize the exact number of roots a given polynomial has. For example, a degree 6 polynomial P has exactly 2 distinct complex roots if, and only if, $\text{PSPC}_4(P, P') \neq 0$ and $\text{PSPC}_\ell(P, P') = 0$ for $\ell = 0, 1, 2, 3$. If we show that the vanishing locus of $\text{PSPC}_3(E_a, E'_a)$ does not intersect \mathbf{C}_3 , then this effectively eliminates this root configuration. This in turn will follow from the fact that \mathbf{C}_2 is strictly contained in the square $\blacksquare = \{(y_1, y_2) : \max_i |y_i| < 1/\sqrt{3}\}$, while $\text{PSPC}_3(E_a, E'_a) = 0$ lies outside.

To see the first inclusion, recall that $a \in \mathbf{C}_2$ satisfies $27y_1^4y_2^4 - 18y_1^2y_2^2 + 4y_2^2 + 4y_1^2 = 1$. Since this equation is unchanged under $(y_1, y_2) \mapsto (y_2, y_1)$, it is enough to prove that $y_1 < 1/\sqrt{3}$. Fixing y_1 , the discriminant of the resulting quadratic equation (in the variable y_2^2) is $16(1 - 3y_1^2)^3$, which is positive precisely for $y_1 < 1/\sqrt{3}$.

On the other hand, we compute

$$\begin{aligned} \text{PSPC}_3(E_a, E'_a) = & 80(y_1^2 + y_2^2) - 50(y_1^4 + y_2^4) + 7(y_1^6 + y_2^6) - 51(y_1^4y_2^2 + y_1^2y_2^4) \\ & + 57(y_1^6y_2^2 + y_1^2y_2^6) + 249y_1^4y_2^4 - 166y_1^2y_2^2 - 25. \end{aligned}$$

Note that $(0, \sqrt{5})$, which lies outside \blacksquare , is a root of the above equation, so we have only to show that the solution locus of $\text{PSPC}_3(E_a, E'_a) = 0$ doesn't cross \blacksquare . Setting $y_2^2 = 1/3$ we obtain $(3y_2^2 - 1)(117y_2^4 - 138y_2^2 + 49) = 0$, which has no roots in $|y_2| < 1/\sqrt{3}$. The same is true with the roles of y_1 and y_2 reversed. \square

We may now determine the degeneracy type of each of the degenerate singularities lying over a caustic point.

Corollary 8.9. *We have the following description of the degeneracy types in the critical locus:*

- (1) *For any $a \in \mathbf{C}_1$, the unique critical point of $\mathcal{F}(a)$ is degenerate of type A_2 .*
- (2) *The two distinct critical points of $\mathcal{F}(a_{\text{cusp}})$ are degenerate of type A_3 .*
- (3) *For any $a \in \mathbf{C}_2$, the two degenerate critical points of $\mathcal{F}(a)$ are of type A_2 .*

Proof of (1). In Proposition 8.8 it was shown that the multiplicity is 2. This is enough to pinpoint A_2 as the degeneracy type, since a singularity of type A_k has multiplicity k . \square

Proof of (2). In Lemma 8.4, we found the two critical points u_{cusp}^\pm and showed in Proposition 8.3 that the corresponding Hessians $\nabla^2 F_H(u_{\text{cusp}}^\pm, a_{\text{cusp}})$ are both of corank 1. It follows that u_{cusp}^\pm are of degeneracy type A . By Proposition 8.8 the multiplicity of both u_{cusp}^\pm is 3. Hence the critical points are of type A_3 . \square

Proof of (3). If $a \in \mathbf{C}_2$, then according to Proposition 8.8 among the four distinct critical points of $\mathcal{F}(a)$ two are non-degenerate and two are degenerate of multiplicity 2. Since A_2 is the unique singularity class with multiplicity 2, we deduce that the two degenerate critical points in $\mathcal{F}(a)$ are fold singularities. \square

Remark 8.10. In the proof of (2) above we could bypass the use of Proposition 8.3 and only use the fact that the multiplicity of the singularity is 3. Indeed this implies that the singularity is simple [2, Lemma 4.2], and the classification theorem of Arnol'd then shows that it is of type A_3 .

9. PROOF OF COROLLARY 1.8

In this section we continue to assume that ν is self-dual and retain the notation from §8.1. We derive from the critical point configuration described in Theorem 8.1 a lower bound on

the $\mathrm{PGL}_3(\mathbb{R})$ Jacquet-Whittaker functions $W_\nu(a)$ in the vicinity of the cuspidal point. Since $\|W_\nu\|_2 = 1$, the following result will complete the proof of Corollary 1.8.

Theorem 9.1. *For all $t > 1$ we have*

$$W_{t\nu_0}(a) \gg t^{3/4}$$

for $a = \mathrm{diag}(y_1 y_2, y_2, 1) \in A$ such that

$$y_1 + y_2 = \frac{2}{\sqrt{3}}t + O(t^{1/4}) \quad \text{and} \quad y_1 - y_2 = O(t^{1/2}).$$

9.1. Asymptotics associated to cuspidal singularities. We begin by giving an alternative definition of an A_k singularity which is useful in determining the asymptotics of associated oscillatory integrals.

Definition 9.2 ([31], §7.7-7.9, or [34], Theorem 7.7.19). *A critical point (x_c, y_c) of a function $\varphi \in C^\infty(\mathbb{R}^m \times N, \mathbb{R})$ is a singularity of type*

$$\underbrace{S_1, \dots, 1}_k,$$

(also called a k -iterated S_1 singularity) where $k \geq 2$, if

- (1) $Q(x_c, y_c)$ has corank 1;
- (2) there is a non-zero vector Y of \mathbb{R}^m such that for every non-zero X in the line $\ker Q(x_c, y_c) \subset \mathbb{R}^m$ the $(k+1)$ -st partial derivative $\langle X, Y \rangle^{k+1} \varphi$ is non-zero at (x_c, y_c) .

The equivalence of k -iterated S_1 singularities with the description of type A_k singularities given in the previous subsection should be evident.

The Airy function is the first in a series of special functions associated to singularities of type A_k , for $k \geq 2$. The *generalized Airy function of order k* is defined as

$$Ai_k(y_1, \dots, y_{k-1}) = \int_{\mathbb{R}} \exp \left(i \left(y_{k-1} x + \dots + y_1 \frac{x^{k-1}}{k-1} + \frac{x^{k+1}}{k+1} \right) \right) dx,$$

the integral is improper (converges in the limit but not absolutely). For $k = 2$ we recover the Airy function: $Ai_2(y) = (2\pi) \mathrm{Ai}(y)$. In general, the order k Airy function governs the asymptotic behavior of oscillatory integrals whose phase functions have A_k type singularities. For more information on generalized Airy functions we refer the reader to [31, §7.9].

In this section we shall be interested in cusp singularities. The order 3 Airy function bears a special name: one calls

$$\mathrm{Pe}(y_1, y_2) = Ai_3(y_1, y_2) = \int_{\mathbb{R}} \exp \left(i \left(y_2 x + y_1 \frac{x^2}{2} + \frac{x^4}{4} \right) \right) dx$$

the *Pearcey function*. It was first introduced (and numerically computed) in [51]. Unlike the Airy function $\mathrm{Ai}(y)$, the Pearcey function (and indeed all higher order Airy functions) is a complex valued function. The phase function

$$\varphi(x, y) = y_2 x + y_1 \frac{x^2}{2} + \frac{x^4}{4}, \quad (x \in \mathbb{R}, y = (y_1, y_2) \in \mathbb{R}^2),$$

has critical set $\Sigma_\varphi = \{(x, y) \in \mathbb{R} \times \mathbb{R}^2 : y_2 + xy_1 + x^3 = 0\}$, whose horizontal projection is singular at $y = 0$. which embeds via λ_φ into 4-space as the Lagrangian plane $\Lambda_\varphi = \{(y, x^2/2, x) \in T^*\mathbb{R}^2 : y_2 + xy_1 + x^3 = 0\}$

The following lemma allows one to reduce the asymptotic behavior of an oscillatory integral whose phase function has a cusp singularity to the asymptotic behavior of the Pearcey function. For notational convenience, we set $\mathrm{Pe}_0 := \mathrm{Pe}$, $\mathrm{Pe}_1 := \partial_{y_1} \mathrm{Pe}$, and $\mathrm{Pe}_2 := \partial_{y_2} \mathrm{Pe}$.

Lemma 9.3. *Let (x_c, y_c) be a singularity of type A_3 for φ .*

Let Q denote the Hessian quadratic form $Q(x_c, y_c)$. Let V_0 denote the orthogonal complement inside \mathbb{R}^m of $\ker Q$. For any $(x, y) \in V_0 \times N$ let $Q_0(x, y)$ denote the matrix of second order partial derivatives of φ relative to V_0 . Write σ_0 for the signature of the non-degenerate rank $m - 1$ quadratic form $Q_0 = Q_0(x_c, y_c)$.

Then there exist

- (1) *a compact neighborhood $K \times K' \subset \mathbb{R}^m \times N$ of (x_c, y_c) ,*
- (2) *real valued functions $r_1, r_2, s \in C_c^\infty(N)$ supported in K' and satisfying $r_1(y_c) = r_2(y_c) = 0$ and $s(y_c) = \varphi(x_c, y_c)$,*

such that for all $\alpha \in C_c^\infty(\mathbb{R}^m)$ with support in K and all $t \geq 1$ the integral

$$\left(\frac{t}{2\pi}\right)^{\frac{m}{2}} \int_{\mathbb{R}^m} \alpha(x) e^{it\varphi(x,y)} dx$$

is equal to

$$e^{i\pi\sigma_0/4} e^{its(y)} t^{\frac{1}{4}} \left(\sum_{j=0,1,2} t^{-\frac{j}{4}} \alpha_j(y) \text{Pe}_j \left(t^{3/4} r_1(y), t^{1/2} r_2(y) \right) + O(t^{-1} \text{Sob}_{2,\infty}(\alpha)) \right),$$

for functions $\alpha_j \in C_c^\infty(N)$ supported around y_c , with α_0 satisfying

$$\alpha_0(y_c) = (2\pi)^{-1/2} |\det Q_0|^{-1/2} \alpha(x_c).$$

Proof. The proof proceeds in the same way as the analogous result [34, Theorem 7.7.19] for fold singularities.

One writes the integration domain as $\mathbb{R}^m = \mathbb{R}^{m-1} \times \exp(\ker Q)$, where $\mathbb{R}^{m-1} = \exp(V_0)$. We write the generic element of \mathbb{R}^{m-1} as v and the generic element of $\exp(\ker Q)$ as w . The coordinates of the critical point $x_c \in U$ are denoted $(v_c, w_c) \in \mathbb{R}^{m-1} \times \exp(\ker Q)$.

Keeping $w \in \exp(\ker Q)$ fixed, the leading term asymptotic of the integral

$$\left(\frac{t}{2\pi}\right)^{\frac{m-1}{2}} \int_{\mathbb{R}^{m-1}} \alpha(v, w) e^{it\varphi(v,w,y)} dv$$

is given by

$$e^{i\pi\sigma_0/4} e^{it\varphi(\bar{v}(y), w, y)} |\det Q_0(\bar{v}(y), w, y)|^{-1/2} \alpha(\bar{v}(y), w),$$

where $\bar{v}(y_c) = v_c$. (For this, see [34, Theorem 7.7.6] or [63, Theorem 2.9]). One then applies the A_3 stationary phase lemma to the remaining one-dimensional integral

$$\left(\frac{t}{2\pi}\right)^{\frac{1}{2}} \int_{\exp(\ker Q)} \beta(w, y) e^{it\phi(w,y)} dw.$$

Here we have written $\beta(w, y) = |\det Q_0(\bar{v}(y), w, y)|^{-1/2} \alpha(\bar{v}(y), w)$ and $\phi(w, y) = \varphi(\bar{v}(y), w, y)$. For the A_3 asymptotic, see [31, §7, Theorem 9.1] or [36, (3.12)]; moreover, one can easily adapt [34, Theorem 7.7.18] to the case of cusp singularities.

The result is a leading term asymptotic of the form specified in Lemma 9.3, *but without the explicit expression for $\alpha_0(y_c)$* . Indeed, in these sources no formula for $\alpha_0(y_c)$ is given. One can, however, extract this value from the proof of [34, Theorem 7.7.18]. We indicate how to do so now.

The Malgrange preparation theorem [34, Theorem 7.5.13], when applied to our phase function, shows the existence of a real valued function $W \in C^\infty(\exp(\ker Q) \times N)$ satisfying

$W(w_c, y_c) = 0$, $\partial_w W(w_c, y_c) > 0$, and real valued functions $r_1, r_2, s \in C^\infty(N)$ satisfying $r_1(y_c) = r_1(y_c) = 0$, $s(y_c) = \phi(w_c, y_c)$, such that

$$\phi(w, y) = \frac{W^4}{4} + r_2(y) \frac{W^2}{2} + r_1(y)W + s(y)$$

in a neighborhood of (w_c, y_c) . Concerning our amplitude function, a slightly different version of the Malgrange preparation theorem [34, Theorem 7.5.6] shows the existence of functions $q \in C^\infty(\exp(\ker Q) \times N)$ and $A_0, A_1, A_2 \in C^\infty(N)$, verifying

$$\beta(w, y) = (W^3 + r_2(y)W + r_1(y))q(w, y) + A_2(y)W^2 + A_1(y)W + A_0(y)$$

in a neighborhood of (w_c, y_c) . Following the argument of Hörmander in [34, Theorem 7.7.18], one sees that the leading term asymptotics for the \mathbb{R}^m -integral are given by

$$\begin{aligned} & \sum_{j=0,1,2} (2\pi)^{-1/2} e^{i\pi\sigma_0/4} e^{its(y)} A_j(y) \int_{\exp(\ker Q)} W^j e^{it\left(\frac{W^4}{4} + r_2(y) \frac{W^2}{2} + r_1(y)W\right)} dW \\ &= \sum_{j=0,1,2} (2\pi)^{-1/2} e^{i\pi\sigma_0/4} e^{its(y)} A_j(y) t^{\frac{1-j}{4}} \text{Pe}_j(t^{3/4}r_1(y), t^{1/2}r_2(y)). \end{aligned}$$

The functions $(2\pi)^{-1/2}A_j$ are the α_j appearing in the statement of Lemma 9.3.

One computes the value of each $\alpha_j(y_c)$ by evaluating $\partial_y^j \beta(w_c, y_c)$. For example,

$$\begin{aligned} \alpha_0(y_c) &= (2\pi)^{-1/2} \beta(w_c, y_c) \\ &= (2\pi)^{-1/2} |\det Q_0(\bar{v}(y_c), w_c, y_c)|^{-1/2} \alpha(\bar{v}(y_c), w_c) \\ &= (2\pi)^{-1/2} |\det Q_0|^{-1/2} \alpha(x_c). \end{aligned}$$

This proves the lemma. □

9.2. Proof of Theorem 9.1. Let $K'(t)$ be the set of $a \in A$ such that

$$y_1 + y_2 = 2/\sqrt{3} + O(t^{-3/4}) \quad \text{and} \quad y_1 - y_2 = O(t^{-1/2}).$$

The sets $K'(t)$ form a sequence of shrinking neighborhoods of the cuspidal point a_{cusp} given by $(y_1, y_2) = (1/\sqrt{3}, 1/\sqrt{3})$. The theorem asks for a lower bound for $W_{t\nu_0}(ta)$ for $a \in K'(t)$. We put $K'_0 = K'(t_0)$ for t_0 sufficiently large.

Since $\delta(a)^{1/2} = y_1 y_2$, the formula (2.1) reads

$$|W_{t\nu_0}(ta)| = t^2 y_1 y_2 \left| \int_U \delta(wu)^{1/2} e^{it\varphi(u,a)} du \right|,$$

where $\varphi(u, a) = B(H, H(wu)) - \ell_1(aua^{-1})$. (Recall from §8.1 that $H = \text{diag}(1, 0, -1)$.) Since for all $a \in K'_0$ we have $y_1 y_2 \asymp 1$, it is enough to show that

$$(9.1) \quad \left| \int_U \delta(wu)^{1/2} e^{it\varphi(u,a)} du \right| \gg t^{-5/4}$$

for $a \in K'(t)$.

By Proposition 8.9, part (2), the points $(u_{\text{cusp}}^\pm, a_{\text{cusp}})$ are cusp singularities. Let $\chi^+, \chi^-, \{\chi_n\}_{n \geq 1}$ be a smooth partition of unity of U , identified with \mathbb{R}^3 . Put

$$K^\pm = \text{supp}(\chi^\pm), \quad K_n = \text{supp}(\chi_n),$$

and assume that

- (a) $\chi^+(u_{\text{cusp}}^+) = \chi^-(u_{\text{cusp}}^-) = 1$;
- (b) $K_n \subseteq \{x \in \mathbb{R}^3 : N/2 \leq |x| \leq 2N\}$ for n large enough, where $N = 2^n$;
- (c) $\text{Sob}_{k,\infty}(\chi_n) \ll_k N^{-k}$, $\forall n \geq 1, \forall k \geq 1$.

Let

$$\alpha^\pm(u) = \delta(\mathbf{w}u)^{1/2} \chi^\pm(u), \quad \alpha_n(u) = \delta(\mathbf{w}u)^{1/2} \chi_n(u),$$

and recall that

$$(9.2) \quad \delta(\mathbf{w}u)^{1/2} = e^{\langle \rho, H(\mathbf{w}u) \rangle} = e^{\langle w_1 + w_2, H(\mathbf{w}u) \rangle}.$$

We put

$$I^\pm(t, a) = \int_U \alpha^\pm(u) e^{it\varphi(u, a)} du, \quad I_n(t, a) = \int_U \alpha_n(u) e^{it\varphi(u, a)} du,$$

and examine each of these integrals individually. Clearly

$$\int_U \delta(\mathbf{w}u)^{1/2} e^{it\varphi(u, a)} du = I^+(t, a) + I^-(t, a) + \sum_{n \geq 1} I_n(t, a).$$

We shall show that

$$(9.3) \quad \sum_{n \geq 1} I_n(t, a) \ll_k t^{-k}$$

for any $k > 1$, uniformly for $a \in K'_0$, and

$$(9.4) \quad I^+(t, a) + I^-(t, a) \gg t^{-5/4},$$

uniformly for $a \in K'(t)$. These two estimates imply (9.1).

For $n \geq 1$, the set $\mathcal{K}(n) := K_n \times K'_0 \subset U \times A$ is disjoint from the critical manifold Σ of $\varphi(u, a)$. We have

$$I_n(t, a) \ll_k t^{-k} \text{vol}(K_n) \|d\varphi\|_{\mathcal{K}(n)}^{-k} \text{Sob}_{k, \infty}(\alpha_n),$$

for all $t > 1$, uniformly for all $a \in K'_0$. Now $\|d\varphi\|_{\mathcal{K}(n)}^{-1} = O(1)$, uniformly in n , since $\mathcal{K}(n)$ is bounded uniformly away from Σ . Moreover, $\text{vol}(K_n) \asymp N^3$ by Property (b) of χ_n , and $\text{Sob}_{k, \infty}(\alpha_n) \ll_k N^{-k-2}$ by (9.2) and Properties (b) and (c) of χ_n . Thus $I_n(t, a) \ll_k N^{-k+1} t^{-k}$ for all $t > 1$, uniformly for $a \in K'_0$. We then take $k \geq 2$ and sum over n to obtain (9.3).

We now come to the integrals $I^\pm(t, a)$. Shrinking K^\pm if necessary, Lemma 9.3 asserts that there exist real valued functions $r_1^\pm, r_2^\pm, s^\pm \in C_c^\infty(A)$ supported in K'_0 , with r_i^\pm vanishing to zero order at a_{cusp} and $s^\pm(a_{\text{cusp}}) = \varphi(u_{\text{cusp}}^\pm, a_{\text{cusp}})$, such that $(\frac{t}{2\pi})^{3/2} I^\pm(t, a)$ is equal to

$$\pm i e^{its^\pm(p)} t^{\frac{1}{4}} \sum_{j=0,1,2} t^{-\frac{j}{4}} \alpha_j^\pm(p) \text{Pe}_j \left(t^{3/4} r_1^\pm(p), t^{1/2} r_2^\pm(p) \right) + O(t^{-3/4})$$

for all $a \in K'_0$ and all $t > 1$. The functions $\alpha_j^\pm \in C_c^\infty(A)$ are supported around a_{cusp} , and α_0^\pm satisfies

$$\alpha_0^\pm(a_{\text{cusp}}) = (2\pi)^{-1/2} |\det Q_0^\pm|^{-1/2} \delta(\mathbf{w}u_{\text{cusp}}^\pm)^{1/2}.$$

We have used the notation Q_0^\pm from Lemma 9.4, as well as the fact that $e^{i\pi\sigma_0^\pm/4} = \pm i$.

Now we know from §8.2 that the local coordinates $r^\pm(a) := (r_1^\pm(a), r_2^\pm(a))$ about a_{cusp} are given by

$$r^\pm(a) = \left(y_1 + y_2 - \frac{2}{\sqrt{3}}, y_1 - y_2 \right) + O \left(\max \left\{ \left| y_1 + y_2 - \frac{2}{\sqrt{3}} \right|^2, |y_1 - y_2|^2 \right\} \right).$$

Thus $K'(t)$ is precisely the neighborhood of a_{cusp} given by

$$t^{3/4} r_1^\pm(a) = O(1) \quad \text{and} \quad t^{1/2} r_2^\pm(a) = O(1).$$

This puts us in position to use Lemma 9.3, which asserts that

$$\left(\frac{t}{2\pi} \right)^{3/2} I^\pm(t, a) \sim t^{1/4} c^\pm(t, a) \text{Pe} \left(t^{3/4} r_1^\pm(a), t^{1/2} r_2^\pm(a) \right),$$

uniformly for $a \in K'(t)$, where $c^\pm(t, a) = \pm i e^{its^\pm(a)} \alpha_0^\pm(a)$.

We have shown that, for all $a \in K'_0$,

$$I^+(t, a) + I^-(t, a) \sim (2\pi)^{3/2} (C^+(t, a) + C^-(t, a)) t^{-5/4}$$

as $t \rightarrow \infty$, where

$$C^\pm(t, a) = c^\pm(t, a) \text{Pe} \left(t^{3/4} r_1^\pm(a), t^{1/2} r_2^\pm(a) \right).$$

Now the Pearcey function does not vanish in a neighborhood of the origin. Moreover, neither of $c^\pm(t, a_{\text{cusp}})$ vanishes (for any t). We write

$$C^+(t, a) + C^-(t, a) = \left(1 + \frac{C^+(t, a)}{C^-(t, a)} \right) C^-(t, a),$$

with the aim of showing that

$$\left| \frac{C^+(t, a)}{C^-(t, a)} \right| = \frac{\alpha_0^+(a)}{\alpha_0^-(a)} \cdot \left| \frac{\text{Pe} \left(t^{3/4} r_1^+(a), t^{1/2} r_2^+(a) \right)}{\text{Pe} \left(t^{3/4} r_1^-(a), t^{1/2} r_2^-(a) \right)} \right|$$

is strictly less than 1.

The value of α_0^+/α_0^- at the cuspidal point a_{cusp} is, by Lemma 9.4, equal to ≈ 0.02 ; hence, for any sufficiently small $\delta > 0$ there exists a fixed (independent of t) neighborhood of a_{cusp} in which α_0^+/α_0^- is at most $1 - \delta$. Furthermore, the value of the second quotient (involving the Pearcey function) at a_{cusp} is 1. We can adjust the implied constants in the definition of $K'(t)$ if necessary so that the quotient of Pearcey functions is at most $\frac{1}{1-\delta/2}$ for all $a \in K'(t)$. We deduce that

$$\left| \frac{C^+(t, a)}{C^-(t, a)} \right| \leq \frac{1 - \delta}{1 - \delta/2} < 1$$

for all t and all $a \in K'(t)$.

Since $C^-(t, a)$ is non zero in $K'(t)$, we have $C^+(t, a) + C^-(t, a) \neq 0$ for all t and all $a \in K'(t)$; in particular, (9.4) holds.

It remains to calculate some of the numerical invariants associated to the cuspidal critical points $(u_{\text{cusp}}^\pm, a_{\text{cusp}})$.

Lemma 9.4. *Let Q_0^\pm denote the restriction of Q^\pm to the orthogonal complement of $\ker Q^\pm$ inside \mathfrak{u} . Then*

$$\frac{|\det Q_0^+|^{-1/2} \delta(\mathfrak{w}u_{\text{cusp}}^+)^{1/2}}{|\det Q_0^-|^{-1/2} \delta(\mathfrak{w}u_{\text{cusp}}^-)^{1/2}} = \frac{121}{2767 + 1596\sqrt{3}} \approx 0.02.$$

Moreover if σ_0^\pm denotes the signature of Q_0^\pm , then $\sigma_0^\pm = \pm 2$.

Proof. Let \mathfrak{u}_0^\pm denote the orthogonal complement of $\ker Q^\pm$ inside \mathfrak{u} ; then

$$\mathfrak{u}_0^\pm = \{ {}^t(x, y, z) \in \mathbb{R}^3 : \epsilon_\pm(x - y) \mp z = 0 \}.$$

Let $\mathcal{B}^\pm = \{ e_1^\pm, e_2^\pm \}$ be the orthonormal basis of \mathfrak{u}_0^\pm given by

$$e_1^\pm = \frac{1}{2\sqrt{\epsilon_\pm}} \begin{pmatrix} \pm 1 \\ 0 \\ \epsilon_\pm \end{pmatrix}, \quad e_2^\pm = \frac{1}{2\sqrt{\epsilon_\pm}} \begin{pmatrix} 0 \\ \mp 1 \\ \epsilon_\pm \end{pmatrix}.$$

We use \mathcal{B}^\pm to identify \mathfrak{u}_0^\pm with \mathbb{R}^2 via $x e_1^\pm + y e_2^\pm \mapsto {}^t(x, y) \in \mathbb{R}^2$. With respect to \mathcal{B}^\pm the Gram matrix of $Q_0^\pm = \text{Res}_{\mathfrak{u}_0^\pm}(Q^\pm)$ is

$$A_0^\pm = \begin{pmatrix} {}^t e_1^\pm Q^\pm e_1^\pm & {}^t e_1^\pm Q^\pm e_2^\pm \\ {}^t e_2^\pm Q^\pm e_1^\pm & {}^t e_2^\pm Q^\pm e_2^\pm \end{pmatrix} = \frac{1}{72} \begin{pmatrix} -23 \pm 26\sqrt{3} & -40 \pm 23\sqrt{3} \\ -40 \pm 23\sqrt{3} & -47 \pm 30\sqrt{3} \end{pmatrix}.$$

We deduce that $|\det Q_0^\pm| = \frac{1}{288}(13 \mp 4\sqrt{3})$.

From (9.2) and the values

$$u_{\text{cusp}}^+ = \begin{pmatrix} 1 & -1-\sqrt{3} & 2+\sqrt{3} \\ 0 & 1 & -1-\sqrt{3} \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad u_{\text{cusp}}^- = \begin{pmatrix} 1 & 1-\sqrt{3} & 2-\sqrt{3} \\ 0 & 1 & 1-\sqrt{3} \\ 0 & 0 & 1 \end{pmatrix}$$

we calculate

$$\delta(wu_{\text{cusp}}^\pm)^{1/2} = (12 \pm 6\sqrt{3})^{-1}.$$

The proposition follows from the above numerical calculations. \square

REFERENCES

- [1] V. I. Arnol'd. Integrals of rapidly oscillating functions, and singularities of the projections of Lagrangian manifolds. *Funkcional. Anal. i Priložen.*, 6(3):61–62, 1972. 26
- [2] V. I. Arnol'd. Normal forms of functions near degenerate critical points, the Weyl groups A_k, D_k, E_k and Lagrangian singularities. *Funkcional. Anal. i Priložen.*, 6(4):3–25, 1972. 26, 36
- [3] V. I. Arnol'd. Remarks on the method of stationary phase and on the Coxeter numbers. *Uspehi Mat. Nauk*, 28(5(173)):17–44, 1973. 26
- [4] V. I. Arnol'd, S. M. Gusein-Zade, and A. N. Varchenko. *Singularities of differentiable maps. Volume 1*. Modern Birkhäuser Classics. Birkhäuser/Springer, New York, 2012. Reprint of the 1985 edition. 24, 25, 26
- [5] J. G. Arthur. A trace formula for reductive groups. I. Terms associated to classes in $G(\mathbf{Q})$. *Duke Math. J.*, 45(4):911–952, 1978. 18
- [6] H. Azad, E. van den Ban, and I. Biswas. Symplectic geometry of semisimple orbits. *Indag. Math. (N.S.)*, 19(4):507–533, 2008. 24
- [7] C. B. Balogh. Asymptotic expansions of the modified Bessel function of the third kind of imaginary order. *SIAM J. Appl. Math.*, 15:1315–1323, 1967. 6
- [8] E. M. Baruch. A proof of Kirillov's conjecture. *Ann. of Math. (2)*, 158(1):207–252, 2003. 15
- [9] J. Bernstein and B. Krötz. Smooth Frechet globalizations of Harish-Chandra modules. 199(1):45–111, 2014. 3
- [10] J. Bernstein and A. Reznikov. Subconvexity bounds for triple L -functions and representation theory. *Ann. of Math. (2)*, 172(3):1679–1718, 2010. 6
- [11] M. Berry. Waves and Thom's theorem. *Advances in Physics*, 25:1–26, 1976. 5, 6
- [12] V. Blomer and P. Maga. Subconvexity for sup-norms of automorphic forms on $\text{PGL}(n)$. 3
- [13] F. Brumley. Effective multiplicity one on GL_N and narrow zero-free regions for Rankin-Selberg L -functions. *Amer. J. Math.*, 128(6):1455–1474, 2006. 16
- [14] F. Brumley. Second order average estimates on local data of cusp forms. *Arch. Math. (Basel)*, 87(1):19–32, 2006. 15
- [15] J. Chazarain. Formule de Poisson pour les variétés riemanniennes. *Invent. Math.*, 24:65–82, 1974. 27
- [16] A. Cianchi and V. G. Maz'ya. Bounds for eigenfunctions of the Laplacian on noncompact Riemannian manifolds. *Amer. J. Math.*, 135(3):579–635, 2013. 6
- [17] L. Cohn. *Analytic theory of the Harish-Chandra C -function*. Lecture Notes in Mathematics, Vol. 429. Springer-Verlag, Berlin, 1974. 11
- [18] Y. Colin De Verdière. Singular Lagrangian manifolds and semiclassical analysis. *Duke Math. J.*, 116(2):263–298, 2003. 31
- [19] M. Coste. An Introduction to Semialgebraic Geometry. *Dip. Mat. Univ. Pisa*, 2000. 36
- [20] M. W. Davis. Some aspherical manifolds. *Duke Math. J.*, 55(1):105–139, 1987. 24
- [21] H. Donnelly. Bounds for eigenfunctions of the Laplacian on compact Riemannian manifolds. *J. Funct. Anal.*, 187(1):247–261, 2001. 1
- [22] J. J. Duistermaat. Oscillatory integrals, Lagrange immersions and unfolding of singularities. *Comm. Pure Appl. Math.*, 27:207–281, 1974. 4, 8, 14, 25, 27
- [23] J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan. Spectra of compact locally symmetric manifolds of negative curvature. *Invent. Math.*, 52(1):27–93, 1979. 19
- [24] J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan. Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups. *Compositio Math.*, 49(3):309–398, 1983. 11, 12, 13, 27
- [25] B. Feigon, E. Lapid, and O. Offen. On representations distinguished by unitary groups. *Publ. Math. Inst. Hautes Études Sci.*, pages 185–323, 2012. 15

- [26] E. García-Toraño Andrés, E. Guzmán, J. C. Marrero, and T. Mestdag. Reduced dynamics and Lagrangian submanifolds of symplectic manifolds. *J. Phys. A*, 47(22):225203, 24, 2014. 23
- [27] S. S. Gelbart and E. M. Lapid. Lower bounds for L -functions at the edge of the critical strip. *Amer. J. Math.*, 128(3):619–638, 2006. 15
- [28] A. A. Gerasimov, D. R. Lebedev, and S. V. Oblezin. New integral representations of Whittaker functions for classical Lie groups. *Uspekhi Mat. Nauk*, 67(1(403)):3–96, 2012. 23
- [29] A. Givental and B. Kim. Quantum cohomology of flag manifolds and Toda lattices. *Comm. Math. Phys.*, 168(3):609–641, 1995. 23
- [30] G.-M. Greuel, C. Lossen, and E. Shustin. *Introduction to singularities and deformations*. Springer Monographs in Mathematics. Springer, Berlin, 2007. 25, 26
- [31] V. Guillemin and S. Sternberg. *Geometric asymptotics*. American Mathematical Society, Providence, R.I., 1977. Mathematical Surveys, No. 14. 37, 38
- [32] L. Hörmander. The spectral function of an elliptic operator. *Acta Math.*, 121:193–218, 1968. 1
- [33] L. Hörmander. Fourier integral operators. I. *Acta Math.*, 127(1-2):79–183, 1971. 4, 8, 14, 25
- [34] L. Hörmander. *The analysis of linear partial differential operators. I*. Classics in Mathematics. Springer-Verlag, Berlin, 2003. Distribution theory and Fourier analysis, Reprint of the second (1990) edition. 28, 37, 38, 39
- [35] H. Jacquet. Fonctions de Whittaker associées aux groupes de Chevalley. *Bull. Soc. Math. France*, 95:243–309, 1967. 3, 14
- [36] E. Kalligiannaki, T. Katsaounis, and G. N. Makrakis. High frequency waves near cusp caustics. *Quart. Appl. Math.*, 61(1):111–129, 2003. 38
- [37] D. Y. Kleinbock and G. A. Margulis. Logarithm laws for flows on homogeneous spaces. *Invent. Math.*, 138(3):451–494, 1999. 3, 21
- [38] B. Kostant. On Whittaker vectors and representation theory. *Invent. Math.*, 48(2):101–184, 1978. 4
- [39] B. Kostant. The solution to a generalized Toda lattice and representation theory. *Adv. in Math.*, 34(3):195–338, 1979. 14, 23
- [40] B. Kostant. Flag manifold quantum cohomology, the Toda lattice, and the representation with highest weight ρ . *Selecta Math. (N.S.)*, 2(1):43–91, 1996. 22
- [41] E. Lapid and Z. Mao. A conjecture on Whittaker-Fourier coefficients of cusp forms. *J. Number Theory*, 146:448–505, 2015. 4
- [42] E. Lapid and O. Offen. Compact unitary periods. *Compos. Math.*, 143(2):323–338, 2007. 3
- [43] X. Li. Upper bounds on L -functions at the edge of the critical strip. *Int. Math. Res. Not. IMRN*, (4):727–755, 2010. 15
- [44] E. Lindenstrauss and A. Venkatesh. Existence and Weyl’s law for spherical cusp forms. *Geom. Funct. Anal.*, 17(1):220–251, 2007. 2
- [45] S. Marshall. Sup norms of maass forms on semisimple groups. *arXiv:1405.7033*. 3
- [46] D. Milićević. Large values of eigenfunctions on arithmetic hyperbolic 3-manifolds. *Geom. Funct. Anal.*, 21(6):1375–1418, 2011. 3
- [47] S. D. Miller. The highest lowest zero and other applications of positivity. *Duke Math. J.*, 112(1):83–116, 2002. 3
- [48] G. Molteni. Upper and lower bounds at $s = 1$ for certain Dirichlet series with Euler product. *Duke Math. J.*, 111(1):133–158, 2002. 15
- [49] J. Moser. Finitely many mass points on the line under the influence of an exponential potential—an integrable system. In *Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pages 467–497. Lecture Notes in Phys., Vol. 38. Springer, Berlin, 1975. 4
- [50] W. Müller. Weyl’s law for the cuspidal spectrum of SL_n . *Ann. of Math. (2)*, 165(1):275–333, 2007. 2
- [51] T. Pearcey. The structure of an electromagnetic field in the neighbourhood of a cusp of a caustic. *Philos. Mag. (7)*, 37:311–317, 1946. 37
- [52] A. Perelomov. *Generalized coherent states and their applications*. Texts and Monographs in Physics. Springer-Verlag, Berlin, 1986. 23
- [53] D. Ramakrishnan and S. Wang. On the exceptional zeros of Rankin-Selberg L -functions. *Compositio Math.*, 135(2):211–244, 2003. 15
- [54] Z. Rudnick and P. Sarnak. The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Comm. Math. Phys.*, 161(1):195–213, 1994. 3
- [55] P. Sarnak. Letter to Morawetz (2004); Letter to Reznikov on restrictions of eigenfunctions (2008). 6, 9, 17
- [56] F. Shahidi. Whittaker models for real groups. *Duke Math. J.*, 47(1):99–125, 1980. 14
- [57] J. A. Shalika. The multiplicity one theorem for GL_n . *Ann. of Math. (2)*, 100:171–193, 1974. 3

- [58] C. D. Sogge and S. Zelditch. Riemannian manifolds with maximal eigenfunction growth. *Duke Math. J.*, 114(3):387–437, 2002. 4
- [59] E. Stade. Archimedean L -factors on $GL(n) \times GL(n)$ and generalized Barnes integrals. *Israel J. Math.*, 127:201–219, 2002. 4, 20
- [60] N. Templier. Large values of modular forms. *Cambridge J. Math.*, 2(1):91–116, 2014. 3
- [61] C. Tomei. The topology of isospectral manifolds of tridiagonal matrices. *Duke Math. J.*, 51(4):981–996, 1984. 24
- [62] J. A. Toth and S. Zelditch. Riemannian manifolds with uniformly bounded eigenfunctions. *Duke Math. J.*, 111(1):97–132, 2002. 4
- [63] V. S. Varadarajan. Oscillatory integrals and their applications to harmonic analysis on semisimple Lie groups. In *Representation theory of reductive groups (Park City, Utah, 1982)*, volume 40 of *Progr. Math.*, pages 207–222. Birkhäuser Boston, Boston, MA, 1983. 38
- [64] V. S. Varadarajan. The method of stationary phase and applications to geometry and analysis on Lie groups. In *Algebraic and analytic methods in representation theory (Sønderborg, 1994)*, volume 17 of *Perspect. Math.*, pages 167–242. Academic Press, San Diego, CA, 1997. 28

INSTITUT GALILÉE, UNIVERSITE PARIS 13, 93430 VILLETANEUSE FRANCE
E-mail address: `brumley@math.univ-paris13.fr`

DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY, MALOTT HALL, ITHACA NY 14853-4201.
E-mail address: `templier@math.cornell.edu`