GEODESIC RESTRICTIONS OF ARITHMETIC EIGENFUNCTIONS

SIMON MARSHALL

Abstract

Let X be an arithmetic hyperbolic surface arising from a quaternion division algebra over $\mathbb{Q}$. Let ψ be a Hecke–Maass form on X, and let ℓ be a geodesic segment. We obtain a power saving over the local bound of Burq, Gérard, and Tzvetkov for the L^2-norm of ψ restricted to ℓ, by extending the technique of arithmetic amplification developed by Iwaniec and Sarnak. We also improve the local bounds for various Fourier coefficients of ψ along ℓ.

1. Introduction

If X is a compact Riemannian manifold and ψ is a Laplace eigenfunction on X satisfying $\Delta\psi + \lambda^2\psi = 0$, it is an interesting problem to study the extent to which ψ can concentrate on small subsets of X. Two well-studied formulations of this problem are to normalize ψ by $\|\psi\|_2 = 1$, and either bound $\|\psi\|_p$ for $2 \le p \le \infty$ or bound the L^p-norms of ψ restricted to some submanifold. We shall be interested in both of these problems in the case where X is two-dimensional and the submanifold we restrict to is a geodesic segment ℓ. The basic upper bound for $\|\psi\|_p$ in this case was proven by Sogge [30] (see also Avakumović [1] and Levitan [21] when $p = \infty$) and is

$$\|\psi\|_p \ll \lambda^{\delta(p)}, \tag{1}$$

where $\delta(p)$ is given by

$$\delta(p) = \begin{cases} 1/2 - 2/p, & p \ge 6, \\ 1/4 - 1/2p, & 2 \le p \le 6. \end{cases}$$

The standard bound for $\|\psi|_\ell\|_p$ is due to Burq, Gérard, and Tzvetkov [11] and is

$$\|\psi|_\ell\|_p \ll \lambda^{\delta'(p)}, \tag{2}$$

DUKE MATHEMATICAL JOURNAL
Vol. 165, No. 3, © 2016 DOI 10.1215/00127094-3166736
Received 16 July 2013. Revision received 11 March 2015.
First published online 10 December 2015.
2010 *Mathematics Subject Classification*. Primary 35P20; Secondary 11F25, 11F41.

where $\delta'(p)$ is given by

$$\delta'(p) = \begin{cases} 1/2 - 1/p, & p \geq 4, \\ 1/4, & 2 \leq p \leq 4. \end{cases}$$

Both of these bounds are sharp when X is the round 2-sphere, but can be strengthened under extra geometric assumptions on X such as negative curvature (see, e.g., [12], [13], [31]–[33]). It should be noted that all such improvements in the negatively curved case are by at most a power of $\log \lambda$.

We now let X be a compact congruence arithmetic hyperbolic surface arising from a quaternion division algebra over $\mathbb{Q}$. We let ψ be a Hecke–Maass form on X, which we shall always assume to be L^2-normalized. In this context, it is more natural to let $\lambda \in \mathbb{C}$ be the spectral parameter of ψ, which satisfies $\Delta\psi + (1/4 + \lambda^2)\psi = 0$. As we are considering large-eigenvalue asymptotics, we will also assume that $\lambda \in \mathbb{R}$, in which case the bounds (1) and (2) remain unchanged. For these X, Iwaniec and Sarnak [20] have shown that the bound $\|\psi\|_\infty \ll \lambda^{1/2}$ given by (1) may be strengthened by a power to $\|\psi\|_\infty \ll_\epsilon \lambda^{5/12+\epsilon}$. Their approach, known as arithmetic amplification, is to construct a projection operator onto ψ by using the Hecke operators as well as the wave group. It has been adapted by other authors to study the pointwise norms of arithmetic eigenfunctions in various aspects (see, e.g., [6]–[9], [19], [34] as well as the alternative approach taken in [4]). In this paper we apply amplification to a new kind of semiclassical problem, namely, improving the exponent in the bound (2) for $\|\psi|_\ell\|_2$. Our main result is as follows.

THEOREM 1.1
Let ψ be a Hecke–Maass form on X with spectral parameter λ. For any geodesic segment ℓ of unit length we have

$$\|\psi|_\ell\|_2 \ll_\epsilon \lambda^{3/14+\epsilon}, \tag{3}$$

where the implied constant is independent of ℓ.

We may combine Theorem 1.1 with the main theorem of [2] to obtain an improvement over the local bound $\|\psi\|_4 \ll \lambda^{1/8}$.

COROLLARY 1.2
We have $\|\psi\|_4 \ll_\epsilon \lambda^{1/8-1/56+\epsilon}$.

This is much weaker than the bound $\|\psi\|_4 \ll_\epsilon \lambda^\epsilon$ announced by Sarnak and Watson [28, Theorem 3] in the case of $X = \mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H}$, although their result may be conditional on the Ramanujan conjecture. See also [5] for results in the case of holomorphic eigenforms. The results of [10] give an equivalence (up to factors of λ^{ϵ}) between a sublocal bound for $\|\psi\|_4$ and one for $\|\psi|_{\ell}\|_2$ that is uniform in ℓ, and so the bound of Sarnak and Watson implies Theorem 1.1 with an exponent of $1/8$. However, we feel that our method is of interest as it does not rely on special value identities or summation formulas, and we hope to apply it to restriction problems on other groups by combining it with the techniques of [23]; see, for instance, [22] for the case of restricting an SL_3 Maass form to a maximal flat subspace.

1.1. Fourier coefficients along geodesics

The methods we use to prove Theorem 1.1 also allow us to prove bounds for Fourier coefficients of ψ along ℓ. We shall describe the bounds that one expects on a general Riemannian surface before stating our improvements in the arithmetic case. As before, let Y be a compact Riemannian surface, let ψ be a Laplace eigenfunction on Y, and let $\ell \subset Y$ be a unit geodesic segment. We let $\ell : [0,1] \to X$ be an arc length parameterization of ℓ, and let $b \in C_0^{\infty}(\mathbb{R})$ be a function with $\mathrm{supp}(b) \subset [0,1]$. For $t \in \mathbb{R}$, define

$$\langle \psi, be^{itx} \rangle = \int_{-\infty}^{\infty} b(x)e^{-itx}\psi\big(\ell(x)\big)\,dx.$$

The bound for $\langle \psi, be^{itx} \rangle$ that one expects depends on both λ and t. In the "classical range" when $|t| \le \lambda$, it is

$$\langle \psi, be^{itx} \rangle \ll \lambda^{1/2}\big(1+\lambda-|t|\big)^{-1/2}. \tag{4}$$

This is proved in [14, Theorem 1.1] and [36, Section 2] when $t=0$. Moreover, the arguments used there should imply (4) in full. When Y is hyperbolic, (4) follows from Lemma 3.8; see also [27] for the case when Y is hyperbolic, ℓ is closed, and $t=0$. If Y is assumed to be negatively curved and $t=0$, (4) was improved to $\langle \psi, b \rangle = o(1)$ in [14].

Now assume that ψ is a Hecke–Maass form on an arithmetic surface X as before. We have the following improvement over (4).

THEOREM 1.3

Let $1/2 > \delta > 0$, and let $I_\delta = [-1+\delta, -\delta] \cup [\delta, 1-\delta]$.

(a) *We have $\langle \psi, b \rangle \ll_\epsilon \lambda^{-1/12+\epsilon}$.*

(b) *If $t/\lambda \in I_\delta$, then we have $\langle \psi, be^{itx} \rangle \ll_{\epsilon,\delta} \lambda^{-1/18+\epsilon}$.*

(c) *Define $\beta = \min|\lambda \pm t|$. If $\beta \le \lambda^{2/3}$, then we have $\langle \psi, be^{itx} \rangle \ll_\epsilon \lambda^{5/24+\epsilon}(1+\beta)^{1/24}$.*

All of these bounds are uniform in t and ℓ.

Remark
The bound $\beta \leq \lambda^{2/3}$ in Theorem 1.3 could be replaced with $\lambda^{1-\delta}$ for any $\delta > 0$; however, when $\beta \geq \lambda^{1/7+\epsilon}$ the bound $\langle \psi, be^{itx} \rangle \ll_\epsilon \lambda^{5/24+\epsilon}(1+\beta)^{1/24}$ is weaker than the local bound of Lemma 3.8.

As in [20], Theorems 1.1 and 1.3 can both be strengthened under the assumption that the Hecke eigenvalues of ψ are not small. In the case of Theorem 1.1 and Theorem 1.3(c), this assumption allows us to employ an amplifier of sufficient length that it becomes profitable to estimate the Hecke recurrence by using spectral methods, rather than the standard diophantine ones. We state the result we obtain in this way as a separate theorem. Let $\lambda(n)$ be the automorphically normalized Hecke eigenvalues of ψ, and assume that they satisfy the bounds

$$\sum_{N \leq p \leq 2N} \left|\lambda(p)\right| \gg_\epsilon N^{1-\epsilon} \tag{5}$$

for all $N \geq 2$ and

$$\left|\lambda(p)\right| \leq 2p^\theta \tag{6}$$

for some $\theta < 1/2$ and p prime. Note that (6) is known with $\theta = 7/64$ (see [3]). We then prove the following.

THEOREM 1.4
Suppose that the normalized Hecke eigenvalues $\lambda(n)$ satisfy (5) and (6). We have

$$\|\psi|_\ell\|_2 \ll_\epsilon \lambda^{1/(8-8\theta)+\epsilon}.$$

If $\beta = \min|t \pm \lambda|$ and $\beta \leq \lambda^{2/3}$, then we have

$$\langle \psi, be^{itx} \rangle \ll_\epsilon \lambda^{\theta/2+\epsilon}(1+\beta)^{1/4-\theta/2}. \tag{7}$$

Both bounds are uniform in t and ℓ.

In particular, Theorem 1.4 gives $\langle \psi, be^{itx} \rangle \ll_\epsilon \lambda^\epsilon$ when $|t-\lambda| \ll \lambda^\epsilon$ under the assumption that $\theta = 0$. We note that (7) becomes weaker than the local bound of Proposition 3.8 when $\beta \geq \lambda^{1/2+\epsilon}$.

1.2. *Relations with L-values and subconvexity*

When ℓ is a closed geodesic, the integrals $\langle \psi, e^{itx} \rangle$ are the Fourier coefficients of ψ along ℓ, and they may be expressed in terms of the L-values $L(1/2, \psi \otimes \theta_\chi)$ by a formula of Waldspurger [35]. Here, χ is a Grossencharacter of a real quadratic field that corresponds to the chosen complex exponential on ℓ, and θ_χ is the associated theta series on GL_2. When $|t| \leq (1-\delta)\lambda$, the bound $\langle \psi, e^{itx} \rangle \ll 1$ corresponds to the convex bound for $L(1/2, \psi \otimes \theta_\chi)$, as one sees from explicit forms of Waldspurger's formula such as [24]. As a result, Theorem 1.3 gives a subconvex bound for $L(1/2, \psi \otimes \theta_\chi)$ in this range. In the other direction, Michel and Venkatesh have proven the strong bound $\langle \psi, e^{itx} \rangle \ll \lambda^{-\delta}$ uniformly for $t \in \mathbb{R}$, by combining Waldspurger's formula with their general solution of subconvexity for GL_2 (see [25, Section 1.4], in particular equation (1.6)).

In the case when $X = SL(2,\mathbb{Z})\backslash\mathbb{H}$ and ℓ is the infinite vertical geodesic from i to $i\infty$, or a compact segment thereof, Ghosh, Reznikov, and Sarnak [18, Theorems 1.1, 1.3] have proved the essentially optimal bound $1 \ll \|\psi|_\ell\|_2 \ll_\epsilon \lambda^\epsilon$ by using L-values. They prove a similar bound for restrictions to a closed horocycle.

1.3. Outline of the proof

We first give the rough idea behind the amplification method. The bound $\|\psi|_\ell\|_2 \ll \lambda^{1/4}$ is proved by taking an approximate spectral projector T onto eigenfunctions with frequency near λ, defining R to be the operator of restriction to ℓ, and bounding the mapping norm $RT : L^2(X) \to L^2(\ell)$. We obtain our improvement by letting $\mathcal{T}$ be a Hecke operator that approximately projects onto functions with the same system of Hecke eigenvalues as ψ and bounding the norm of $RT\mathcal{T}$. The asymptotic orthogonality of systems of Hecke eigenvalues means that we expect $RT\mathcal{T}$ to have small norm on Maass forms other than ψ. Correspondingly, we may prove a bound for the norm of $RT\mathcal{T}$ by using geometric methods. Combining this with the fact that ψ is an eigenfunction of $T\mathcal{T}$ with large eigenvalue gives the result.

We now describe our method in more detail. The basic amplification inequality we use is Proposition 3.1. We prove this in Section 3.1 by taking an amplified pretrace formula, but instead of evaluating at a point on the diagonal as in [20], we integrate it against a test function on $\ell \times \ell$. The resulting identity may be thought of as a relative trace formula. We then drop all terms on the spectral side except the one in which we are interested.

We must bound the geometric side of this inequality, which requires solving a diophantine problem and an analytic problem. The diophantine problem is estimating the number of times a Hecke operator maps ℓ back close to itself, in the sense that $\gamma\ell$ must be contained in a small tubular neighborhood of ℓ. We do this in Section 3.2 by adapting the proof of the corresponding lemma [20, Lemma 1.3], and in Section 4 we prove a stronger bound by using spectral methods, conditional on (5) and (6). The analytic problem is to estimate an oscillatory integral over $\gamma\ell \times \ell$. This is the main new ingredient and occupies Sections 5–7 of the paper.

By combining these two ingredients, we prove Theorem 1.3(a),(b) in Section 3.3, Theorem 1.1 and Theorem 1.3(c) in Section 3.4, and Theorem 1.4 in Section 4.

2. Notation

Throughout the paper, the notation $A \ll B$ will mean that there is a positive constant C such that $|A| \le CB$, and $A \sim B$ will mean that there are positive constants C_1 and C_2 such that $C_1 B \le A \le C_2 B$.

2.1. *Quaternion algebras*

Let $A = (\frac{a,b}{\mathbb{Q}})$ be a quaternion division algebra over $\mathbb{Q}$, where we assume that $a, b \in \mathbb{Z}$ are square-free with $a > 0$. We choose a basis $1, \omega, \Omega, \omega\Omega$ for A over $\mathbb{Q}$ that satisfies $\omega^2 = a$, $\Omega^2 = b$, and $\omega\Omega + \Omega\omega = 0$. We denote the norm and trace by $N(\alpha) = \alpha\overline{\alpha}$ and $\operatorname{tr}(\alpha) = \alpha + \overline{\alpha}$. We let R be a maximal order in A (or more generally an Eichler order, see [15]), and for $m \ge 1$ let

$$R'(m) = \{\alpha \in R \mid N(\alpha) = m\}.$$

$R'(1)$ is the group of elements of norm 1; it acts on $R'(m)$ by multiplication on the left, and $R'(1)\backslash R'(m)$ is known to be finite (see [15]). Fix an embedding $\phi : A \to M_2(F)$, the (2×2)-matrices with entries in $F = \mathbb{Q}(\sqrt{a})$ by

$$\phi(\alpha) = \begin{pmatrix} \overline{\xi} & \eta \\ b\overline{\eta} & \xi \end{pmatrix},$$

where

$$\alpha = x_0 + x_1\omega + (x_2 + x_3\omega)\Omega = \xi + \eta\Omega.$$

For $m > 0$, we let $R(m) \subset PSL(2, \mathbb{R})$ be the image of $\phi(R'(m))$ under projection. We define the lattice $\Gamma = R(1)$, which is cocompact as we assumed A to be a division algebra and let $X = \Gamma\backslash\mathbb{H}$.

2.2. *Hecke operators*

We define the Hecke operators $T_n : L^2(X) \to L^2(X)$, $n \ge 1$, by

$$T_n f(z) = \frac{1}{\sqrt{n}} \sum_{\alpha \in R(1)\backslash R(n)} f\big(\phi(\alpha)z\big).$$

There is a positive integer q (depending on R) such that for $(n, q) = 1$, T_n has the following properties (see [15]):

$$T_n = T_n^*, \quad \text{that is, } T_n \text{ is self-adjoint,}$$

$$T_m T_n = \sum_{d \mid (m,n)} T_{mn/d^2}.$$

2.3. Lie groups and algebras

We let $G = PSL(2,\mathbb{R})$. We let K, A, and N be the standard subgroups of G, with parameterizations

$$k(\theta) = \begin{pmatrix} \cos\theta/2 & \sin\theta/2 \\ -\sin\theta/2 & \cos\theta/2 \end{pmatrix}, \qquad a(y) = \begin{pmatrix} e^y & 0 \\ 0 & 1 \end{pmatrix}, \qquad n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

In particular, $k(\theta)$ represents an anticlockwise rotation by θ about the point i. We denote the Lie algebra of G by $\mathfrak{g}$ and equip $\mathfrak{g}$ with the norm

$$\|\cdot\| : \begin{pmatrix} X_1 & X_2 \\ X_3 & -X_1 \end{pmatrix} \mapsto \sqrt{X_1^2 + X_2^2 + X_3^2}. \tag{8}$$

This norm defines a left-invariant metric on G, which we denote by d. We denote the Lie algebras of K, A, and N by $\mathfrak{k}$, $\mathfrak{a}$, and $\mathfrak{n}$ and write the Iwasawa decomposition as

$$g = n(g)\exp\big(A(g)\big)k(g) = \exp\big(N(g)\big)\exp\big(A(g)\big)k(g). \tag{9}$$

We define

$$H = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix} \in \mathfrak{a}, \qquad X_{\mathfrak{n}} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{n}, \qquad X_{\mathfrak{k}} = \begin{pmatrix} 0 & 1/2 \\ -1/2 & 0 \end{pmatrix}. \tag{10}$$

We identify $\mathfrak{a} \simeq \mathbb{R}$ under the map $H \mapsto 1$ and consider $A(g)$ as a function $A : G \to \mathbb{R}$ under this identification, and likewise for $\mathfrak{n}$ and $N(g)$. We let φ_s denote the standard spherical function with spectral parameter s on $\mathbb{H}$ or G, depending on the context.

We let dg be the Haar measure on G that is the extension of the usual hyperbolic volume by the measure of mass 1 on K.

2.4. Maass forms

We let $\psi \in L^2(X)$ be a Hecke–Maass form that is an eigenfunction of Δ and the operators T_n with $(n,q) = 1$. We let $\lambda(n)$ be the Hecke eigenvalues of ψ and λ be its spectral parameter, so that

$$T_n\psi = \lambda(n)\psi,$$
$$\Delta\psi + (1/4 + \lambda^2)\psi = 0.$$

We assume that $\|\psi\|_2 = 1$ with respect to the hyperbolic volume on X. Note that because Δ and T_n with $(n,q) = 1$ are self-adjoint, we may assume that ψ is real-valued.

3. Amplification

This section contains all the arithmetic methods used in the proof of Theorems 1.1 and 1.3. The amplification inequality that we derive from the pretrace formula is given in Proposition 3.1. The estimation of Hecke returns is carried out in Section 3.2. Theorems 1.3 and 1.1 are proven in Section 3.3 and Section 3.4, respectively, by combining these ingredients with bounds for oscillatory integrals proven in Sections 5–7.

We choose $g_0 \in G$ and let our geodesic segment ℓ be $\{g_0 a(x) : 0 \leq x \leq 1\}$. We fix a function $b \in C_0^\infty(0,1)$, which we may assume to be real-valued. If $\phi \in L^1_{\mathrm{loc}}(\mathbb{R})$, we shall study the integral

$$\langle \psi, b\phi \rangle = \int_{-\infty}^{\infty} \overline{b\phi}(x) \psi\big(g_0 a(x)\big)\, dx.$$

3.1. An amplification inequality

We fix a real-valued function $h \in C^\infty(\mathbb{R})$ of Paley–Wiener type that is nonnegative and satisfies $h(0) = 1$. Define h_λ^0 by $h_\lambda^0(s) = h(s-\lambda) + h(-s-\lambda)$, and let k_λ^0 be the K-biinvariant function on $\mathbb{H}$ with Harish-Chandra transform h_λ (see [17] or [29] for definitions). The Paley–Wiener theorem of Gangolli [16] implies that k_λ^0 is of compact support that may be chosen arbitrarily small. Define $k_\lambda = k_\lambda^0 * k_\lambda^0$, which has Harish-Chandra transform $h_\lambda = (h_\lambda^0)^2$. If $\phi \in L^1_{\mathrm{loc}}(\mathbb{R})$ and $g \in G$, we define

$$I(\lambda, \phi, g) = \iint_{-\infty}^{\infty} \overline{b\phi}(x_1) b\phi(x_2) k_\lambda\big(a(-x_1) g a(x_2)\big)\, dx_1\, dx_2.$$

We assume that the supports of b and k_λ are small enough that $I(\lambda, \phi, g) = 0$ unless $d(g, e) \leq 1$. Let $N \geq 1$ be an integer, and let $\{\alpha_n\}_{n=1}^N$ be complex numbers, both to be chosen later. Define the Hecke operator

$$\mathcal{T} = \sum_{n \leq N} \alpha_n T_n.$$

Our main amplification inequality is the following.

PROPOSITION 3.1

We have

$$\big|\langle \mathcal{T}\psi, b\phi \rangle\big|^2 \leq \sum_{m,n \leq N} |\alpha_m \alpha_n| \sum_{d|(m,n)} \frac{d}{\sqrt{mn}} \sum_{\gamma \in R(mn/d^2)} \big|I(\lambda, \phi, g_0^{-1} \gamma g_0)\big|.$$

Proof

Consider the kernel function

$$K(x, y) = \sum_{\gamma \in \Gamma} k_\lambda(x^{-1} \gamma y)$$

on $\Gamma \backslash G$. Let $\{\psi_i\}$ be an orthonormal basis for $L^2(X)$ consisting of Hecke–Maass forms with $\psi \in \{\psi_i\}$. Let μ_i be the spectral parameter of ψ_i. Then $K(x, y)$ has the spectral expansion

$$K(x,y) = \sum_i h_\lambda(\mu_i)\psi_i(x)\overline{\psi_i}(y).$$

If we apply T_n to $K(x,y)$ in the x variable and equate the geometric and spectral expansions, we obtain

$$\frac{1}{\sqrt{n}} \sum_{\gamma \in R(n)} k_\lambda(x^{-1}\gamma y) = \sum_i h_\lambda(\mu_i) T_n \psi_i(x)\overline{\psi_i}(y).$$

Applying $\mathcal{T}\mathcal{T}^*$ therefore gives

$$\sum_{m,n \le N} \alpha_m \overline{\alpha}_n \sum_{d|(m,n)} \frac{d}{\sqrt{mn}} \sum_{\gamma \in R(mn/d^2)} k_\lambda(x^{-1}\gamma y) = \sum_i h_\lambda(\mu_i) \mathcal{T}\psi_i(x)\overline{\mathcal{T}\psi_i}(y).$$

If we integrate this identity against $\overline{b\phi} \times b\phi$ on $g_0A \times g_0A$, we obtain

$$\sum_{m,n \le N} \alpha_m \overline{\alpha}_n \sum_{d|(m,n)} \frac{d}{\sqrt{mn}} \sum_{\gamma \in R(mn/d^2)} I(\lambda, \phi, g_0^{-1}\gamma g_0)$$
$$= \sum_i h_\lambda(\mu_i) \big| \langle \mathcal{T}\psi_i, b\phi \rangle \big|^2.$$

We have $h_\lambda^0(\mu_i) \in \mathbb{R}$; hence $h_\lambda(\mu_i) \ge 0$, for all i. We may therefore drop all terms on the right-hand side except ψ, which completes the proof. □

3.2. Estimation of Hecke returns

We now estimate how many times the Hecke operators map the geodesic g_0A close to itself, which amounts to giving a bound for the counting function

$$M(g,n,\kappa) = \big| \{ \eta \in R(n) \mid d(g^{-1}\eta g, e) \le 1, d(g^{-1}\eta g, A) \le \kappa \} \big|.$$

Our bound for $M(g,n,\kappa)$ is Lemma 3.3 below. We first need a diophantine lemma.

LEMMA 3.2

Let $A, B > 0$, $y \ge A$, and $B \ge \delta \ge 0$. Then we have

$$\big| \{ r,s \in \mathbb{Z} : |r^2 - ys^2 - n| \le \delta n, r^2 + s^2 \le Bn \} \big| \ll (n/\delta)^\epsilon (n\sqrt{\delta} + 1)$$

for all $\epsilon > 0$, where the implied constant depends only on A, B, and ϵ.

Proof

Let $Q \ge 1$. We can find relatively prime integers p, q with $1 \le q \le Q$ such that

$$\Big| \frac{p}{q} - y \Big| \le \frac{1}{qQ}.$$

The condition $|r^2 - ys^2 - n| \leq \delta n$ then implies that

$$\left|r^2 - (p/q)s^2 - n\right| \leq \delta n + \frac{s^2}{qQ},$$
$$|qr^2 - ps^2 - qn| \leq \delta n q + \frac{Bn}{Q}.$$

Choose $Q = \delta^{-1/2} + A^{-1}$, which implies that $p \geq 1$. We then have

$$\begin{aligned}|qr^2 - ps^2 - qn| &\leq \delta n(\delta^{-1/2} + A^{-1}) + Bn\delta^{1/2} \\ &\leq n\delta^{1/2}(1 + B + \sqrt{B}/A).\end{aligned}$$

If we let $C = 1 + B + \sqrt{B}/A$, the number of r and s satisfying $|qr^2 - ps^2 - qn| \leq Cn\delta^{1/2}$ and $r^2 + s^2 \leq Bn$ is equal to

$$\sum_{|m-qn|\leq Cn\delta^{1/2}} \left|\{qr^2 - ps^2 = m, r^2 + s^2 \leq Bn\}\right|.$$

Let $\mathcal{O}$ denote the ring of integers in the field $\mathbb{Q}(\sqrt{pq})$, and let $|\cdot|_1$ and $|\cdot|_2$ denote its two archimedean valuations. Every solution of $qr^2 - ps^2 = m$ determines an element $z = qr + s\sqrt{pq} \in \mathcal{O}$ with $Nz = mq$. We therefore have

$$\begin{aligned}&\left|\{qr^2 - ps^2 = m, r^2 + s^2 \leq Bn\}\right| \\ &\quad\leq \left|\left\{z \in \mathcal{O} : Nz = mq, |z|_1 + |z|_2 \leq 2\sqrt{Bn}(q + \sqrt{pq})\right\}\right|.\end{aligned}$$

The set on the right-hand side maps to the set of ideals in $\mathcal{O}$ with norm mp, and the number of such ideals is bounded by $\tau(mq) \ll_\epsilon (mq)^\epsilon \ll (n/\delta)^\epsilon$. The fibers of this map are orbits under multiplication by units, and the condition

$$|z|_1 + |z|_2 \leq 2\sqrt{Bn}(q + \sqrt{pq}) \ll n\delta^{-1/2}$$

and the fact that the regulators of real quadratic fields are bounded away from 0 implies that the fibers have size $\ll_\epsilon (n/\delta)^\epsilon$. Summing this bound over m completes the proof. □

LEMMA 3.3
If $\kappa \leq 1$, we have the bound

$$M(g, n, \kappa) \ll_\epsilon (n/\kappa)^\epsilon (n\sqrt{\kappa} + 1)$$

uniformly in g.

Proof
We follow the proof of the corresponding lemma (see [20, Lemma 1.3]). We first give a parameterization of the group gAg^{-1} of translations along the geodesic gA. Let $[\alpha, \beta, \gamma]$ be the quadratic form associated to gA, so that

$$\beta^2 - 4\alpha\gamma = 1 \tag{11}$$

and the roots of $\alpha z^2 + \beta z + \gamma$ are the endpoints of gA. Note that the case when one endpoint is at $i\infty$ corresponds to $\alpha = 0$. We may parameterize gAg^{-1} as

$$gAg^{-1} = \left\{ \begin{bmatrix} t - \beta u & -2\gamma u \\ 2\alpha u & t + \beta u \end{bmatrix} \middle| \, t^2 - u^2 = 1, t > 0 \right\}. \tag{12}$$

As Γ is cocompact, we may assume that g lies in a fixed compact set Ω. This implies that $\alpha, \beta, \gamma \ll 1$. If $d(g^{-1}\eta g, e) \leq 1$, we have

$$d(g^{-1}\eta g, A) \leq \kappa \to \eta = y + O(\kappa) \quad \text{with } y \in gAg^{-1}. \tag{13}$$

Because $\eta \in R(n)$, η is the image of an element $z \in R'(n)$. We have

$$\phi(z) = \begin{bmatrix} x_0 - x_1\sqrt{a} & x_2 + x_3\sqrt{a} \\ bx_2 - bx_3\sqrt{a} & x_0 + x_1\sqrt{a} \end{bmatrix} \tag{14}$$

with $\det(\phi(z)) = n$ and $Ex_i \in \mathbb{Z}$ for some $E \in \mathbb{Z}$ depending on the order R. Combining (12), (13), and (14) gives

$$\frac{x_0}{\sqrt{n}} = t + O(\kappa), \tag{15}$$

$$\frac{x_1\sqrt{a}}{\sqrt{n}} = \beta u + O(\kappa), \tag{16}$$

$$\frac{x_2}{\sqrt{n}} = -\Big(\gamma - \frac{\alpha}{b}\Big)u + O(\kappa),$$

$$\frac{x_3\sqrt{a}}{\sqrt{n}} = -\Big(\gamma + \frac{\alpha}{b}\Big)u + O(\kappa).$$

Equation (11) implies that one of $|\beta|$ and $|\gamma \pm \alpha/b|$ must be $\geq 1/4|b|$. We shall assume that $\beta \geq 1/4|b|$, as the other cases are similar. Because $d(g^{-1}\eta g, e) \leq 1$, the entries of $\phi(z)/\sqrt{n}$ must be bounded in terms of Ω, which gives $x_i \ll \sqrt{n}$ for all i. Combining this with equations (15) and (16) gives

$$1 = t^2 - u^2 = \frac{x_0^2}{n} - \frac{x_1^2 a}{n\beta^2} + O(\kappa),$$

or

$$x_0^2 - \frac{x_1^2 a}{\beta^2} - n \ll \kappa n.$$

Combining this with $x_i \ll \sqrt{n}$, Lemma 3.2 implies that the number of choices for x_0 and x_1 is $\ll_\epsilon (n/\kappa)^\epsilon (n\sqrt{\kappa}+1)$. If we fix x_0 and x_1, then x_2 and x_3 must satisfy

$$-bx_2^2 + abx_3^2 = n - x_0^2 + ax_1^2 \ll n$$

and $x_i \ll \sqrt{n}$. By working with the field $\mathbb{Q}(\sqrt{a})$ as in the proof of Lemma 3.2, we see that the number of choices for x_2 and x_3 is $\ll_\epsilon n^\epsilon$. This completes the proof. □

3.3. Bounds for geodesic periods

We now prove Theorem 1.3(a),(b). As ψ and b are real-valued, we may assume that $t \geq 0$. We begin with a specialization of Proposition 3.1.

PROPOSITION 3.4

If $t/\lambda \in [\delta, 1-\delta]$, we have

$$\left|\langle \mathcal{T}\psi, be^{itx}\rangle\right|^2 \ll_{\delta,\epsilon} N^\epsilon \lambda^\epsilon \Big(\sum_{n\leq N} |\alpha_n|^2 + N\lambda^{-1/3} \Big(\sum_{n\leq N} |\alpha_n| \Big)^2 \Big).$$

We also have

$$\left|\langle \mathcal{T}\psi, b\rangle\right|^2 \ll_{\epsilon} N^\epsilon \lambda^\epsilon \Big(\sum_{n\leq N} |\alpha_n|^2 + N\lambda^{-1/2} \Big(\sum_{n\leq N} |\alpha_n| \Big)^2 \Big).$$

Proof

If we apply Proposition 3.1 with $\phi(x) = e^{itx}$ and denote $I(\lambda, e^{itx}, g)$ by $I(\lambda, t, g)$, we have

$$\left|\langle \mathcal{T}\psi, be^{itx}\rangle\right|^2 \leq \sum_{m,n\leq N} |\alpha_m \alpha_n| \sum_{d|(m,n)} \frac{d}{\sqrt{mn}} \sum_{\gamma \in R(mn/d^2)} \left| I(\lambda, t, g_0^{-1}\gamma g_0)\right|. \quad (17)$$

We shall estimate the right-hand side of (17) by combining Lemma 3.3 with the following bounds for $I(\lambda, t, g)$, proven in Section 7. We let δ be as in Theorem 1.3.

PROPOSITION 3.5

If $t/\lambda \in [\delta, 1-\delta]$, we have

$$I(\lambda, t, g) \ll_\delta \begin{cases} (1+\lambda d(g,A))^{-1/2}, & d(g,A) \leq \lambda^{-1/3}, \\ \lambda^{-1/3}, & d(g,A) \geq \lambda^{-1/3}. \end{cases}$$

We also have

$$I(\lambda,0,g) \ll \big(1+\lambda d(g,A)\big)^{-1/2}.$$

We shall only consider the case $t/\lambda \in [\delta, 1-\delta]$, as the case $t=0$ is similar. The dependence of implied constants on the arbitrarily small $\epsilon > 0$ will be omitted for the remainder of the proof. If we assume that $d(g_0^{-1}\gamma g_0, e) \leq 1$, then we have $d(g_0^{-1}\gamma g_0, A) \leq 1$, and we cover $[0,1]$ with the intervals $I_0 = [0, \lambda^{-1}]$, $I_k = [e^{k-1}\lambda^{-1}, e^k\lambda^{-1}]$ for $1 \leq k \leq \frac{2}{3}\log\lambda$, and $I_\infty = [e^{-1}\lambda^{-1/3}, 1]$. When $d(g_0^{-1}\gamma g_0, A) \in I_0$ we apply the bounds

$$\begin{aligned} I(\lambda,t,g_0^{-1}\gamma g_0) &\ll_\delta 1, \\ M(g_0,n,\lambda^{-1}) &\ll n^\epsilon \lambda^\epsilon (n\lambda^{-1/2}+1) \end{aligned}$$

from Proposition 3.5 and Lemma 3.3 to obtain

$$\begin{aligned} &\sum_{m,n\leq N} \sum_{d\,|(n,m)} \frac{d\,|\alpha_m\alpha_n|}{\sqrt{mn}} \sum_{\substack{\gamma\in R(mn/d^2),\\ d(g_0^{-1}\gamma g_0, A)\in I_0}} \big|I(\lambda,t,g_0^{-1}\gamma g_0)\big| \\ &\ll_\delta N^\epsilon\lambda^\epsilon \sum_{m,n\leq N} |\alpha_m\alpha_n| \sum_{d\,|(n,m)} \frac{d}{\sqrt{mn}}\Big(\lambda^{-1/2}\frac{mn}{d^2}+1\Big) \\ &\ll_\delta N^\epsilon\lambda^\epsilon \sum_{m,n\leq N} |\alpha_m\alpha_n| \sum_{d\,|(n,m)} \Big(\frac{\sqrt{mn}}{d}\lambda^{-1/2} + \frac{d}{\sqrt{mn}}\Big). \end{aligned}$$

When $d(g_0^{-1}\gamma g_0, A) \in I_k$ we have

$$\begin{aligned} I(\lambda,t,g_0^{-1}\gamma g_0) &\ll_\delta e^{-k/2}, \\ M(g_0,n,e^k\lambda^{-1}) &\ll n^\epsilon\lambda^\epsilon(n\lambda^{-1/2}e^{k/2}+1), \end{aligned}$$

which gives

$$\begin{aligned} &\sum_{m,n\leq N} \sum_{d\,|(n,m)} \frac{d\,|\alpha_m\alpha_n|}{\sqrt{mn}} \sum_{\substack{\gamma\in R(mn/d^2),\\ d(g_0^{-1}\gamma g_0, A)\in I_k}} \big|I(\lambda,t,g_0^{-1}\gamma g_0)\big| \\ &\ll_\delta N^\epsilon\lambda^\epsilon \sum_{m,n\leq N} |\alpha_m\alpha_n| \sum_{d\,|(n,m)} \frac{d}{\sqrt{mn}}\Big(\lambda^{-1/2}\frac{mn}{d^2}+e^{-k/2}\Big) \\ &\ll_\delta N^\epsilon\lambda^\epsilon \sum_{m,n\leq N} |\alpha_m\alpha_n| \sum_{d\,|(n,m)} \Big(\frac{\sqrt{mn}}{d}\lambda^{-1/2} + \frac{d}{\sqrt{mn}}e^{-k/2}\Big). \end{aligned}$$

When $d(g_0^{-1}\gamma g_0, A) \in I_\infty$ we have

$$I(\lambda, t, g_0^{-1}\gamma g_0) \ll_\delta \lambda^{-1/3},$$
$$M(g_0, n, 1) \ll n^{1+\epsilon},$$

so that

$$\sum_{m,n\le N} \sum_{d|(n,m)} \frac{d|\alpha_m\alpha_n|}{\sqrt{mn}} \sum_{\substack{\gamma\in R(mn/d^2),\\ d(g_0^{-1}\gamma g_0, A)\in I_\infty}} \left|I(\lambda, t, g_0^{-1}\gamma g_0)\right|$$
$$\ll_\delta N^\epsilon \lambda^\epsilon \sum_{m,n\le N} |\alpha_m\alpha_n| \sum_{d|(n,m)} \frac{\sqrt{mn}}{d}\lambda^{-1/3}.$$

Combining these, and noting that we are summing over $\ll \log\lambda$ values of k, we obtain

$$\left|\langle \mathcal{T}\psi, be^{itx}\rangle\right|^2 \ll_\delta N^\epsilon\lambda^\epsilon \sum_{m,n\le N} |\alpha_m\alpha_n| \sum_{d|(n,m)} \left(\frac{\sqrt{mn}}{d}\lambda^{-1/3} + \frac{d}{\sqrt{mn}}\right). \tag{18}$$

As in [20, p. 310], we have

$$\sum_{m,n\le N} \sum_{d|(n,m)} \frac{\sqrt{mn}}{d}|\alpha_n\alpha_m| \ll N^{1+\epsilon}\Big(\sum_{n\le N}|\alpha_n|\Big)^2, \tag{19}$$

and

$$\sum_{m,n\le N} \sum_{d|(m,n)} \frac{d}{\sqrt{mn}}|\alpha_n\alpha_m| \ll N^\epsilon \sum_{n\le N}|\alpha_n|^2. \tag{20}$$

Combining (18) with (19) and (20) completes the proof. □

To deduce Theorem 1.3(b) from Proposition 3.4, choose $\{\alpha_n\}$ to be the amplifier used in [20]. It follows as on [20, p. 311] that

$$\left|\langle \psi, be^{itx}\rangle\right|^2 \ll_{\delta,\epsilon} N^\epsilon\lambda^\epsilon(N^{-1/2} + N\lambda^{-1/3}),$$

and choosing $N = \lambda^{2/9}$ completes the proof. Case (a) follows by using the same amplifier with $N = \lambda^{1/3}$.

3.4. *Bounds for L^2-norms*

To prove Theorem 1.1, it suffices to bound the L^2-norm of $b(x)\psi(g_0a(x)) \in L^2(\mathbb{R})$ for $b \in C_0^\infty(\mathbb{R})$ real-valued with $\text{supp}(b) \subseteq [0,1]$, provided the bound is uniform in g_0. If $f \in C_0^\infty(\mathbb{R})$, define its Fourier transform $\widehat{f}$ by

$$\widehat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i\xi x}\, dx,$$

and extend this to an operator on $L^2(\mathbb{R})$. Let β be a parameter satisfying $1 \le \beta \le \lambda$. Define $H_\beta^+, H_\beta^- \subset L^2(\mathbb{R})$ to be the spaces of functions whose Fourier support lies in $[\pm\lambda - \beta, \pm\lambda + \beta]$, and define $H_\beta = H_\beta^+ + H_\beta^-$. Let Π_β be the orthogonal projection onto H_β, and likewise for $\Pi_\beta^\pm$ and $H_\beta^\pm$.

We shall bound $\Pi_\beta b\psi$ and $(1 - \Pi_\beta) b\psi$ separately, by applying Proposition 3.1 with $\phi \in H_\beta$ and $\phi \in H_\beta^\perp$. As ψ and b are real-valued it suffices to bound $\Pi_\beta^+ b\psi$. We shall need the following bounds for $I(\lambda, \phi, g)$.

PROPOSITION 3.6
If $\phi \in H_\beta^\perp$ satisfies $\|\phi\|_2 = 1$ and $e \in G$ is the identity, we have

$$I(\lambda, \phi, e) \ll \lambda^{1/2} \beta^{-1/2}.$$

PROPOSITION 3.7
Suppose that $\beta \le \lambda^{2/3}$ and that $\phi \in H_\beta^+$ satisfies $\|\phi\|_2 = 1$. We have

$$\bigl| I(\lambda, \phi, g) \bigr| \ll_\epsilon \lambda^{1/2+\epsilon} \tag{21}$$

for all g, while if $\epsilon_0 > 0$ and $d(g, A) \ge \lambda^{-1/2+\epsilon_0} \beta^{1/2}$ we have

$$\bigl| I(\lambda, \phi, g) \bigr| \ll_{\epsilon_0, A} \lambda^{-A}. \tag{22}$$

We shall prove Proposition 3.6 in Section 5, and Proposition 3.7 in Section 6. Using these, we may prove the following bounds for $(1 - \Pi_\beta) b\psi$ and $\Pi_\beta^+ b\psi$. Theorem 1.1 follows from combining these with $\beta = \lambda^{1/7}$, and Lemma 3.9 implies Theorem 1.3(c).

LEMMA 3.8
We have $\|(1 - \Pi_\beta) b\psi\|_2 \ll \lambda^{1/4} \beta^{-1/4}$.

Proof
Let $\phi \in H_\beta^\perp$ with $\|\phi\|_2 = 1$. We pass to a finite-index sublattice $\Gamma' \subset \Gamma$ such that $\Gamma' \backslash \mathbb{H}$ has injectivity radius ≥ 10 and apply Proposition 3.1 with $\mathcal{T}$ the identity operator. Then only $\gamma = e$ contributes to the geometric side, and we obtain $|\langle b\psi, \phi \rangle|^2 \le |I(\lambda, \phi, e)|$. Proposition 3.6 then gives $\|(1 - \Pi_\beta) b\psi\|_2 \ll \lambda^{1/4} \beta^{-1/4}$. □

LEMMA 3.9
If $\beta \le \lambda^{2/3}$, we have $\|\Pi_\beta^+ b\psi\|_2 \ll_\epsilon \lambda^{5/24+\epsilon} \beta^{1/24}$ uniformly in β and ℓ.

Proof
We again have

$$\left|\langle \mathcal{T}\psi, b\phi\rangle\right|^2 \leq \sum_{m,n\leq N} |\alpha_m\alpha_n| \sum_{d|(m,n)} \frac{d}{\sqrt{mn}} \sum_{\gamma\in R(mn/d^2)} \left|I(\lambda,\phi,g_0^{-1}\gamma g_0)\right|. \tag{23}$$

Choose $\epsilon_0 > 0$. Proposition 3.7 implies that we only need to consider the terms in (23) with $d(g_0^{-1}\gamma g_0, e) \leq 1$ and $d(g_0^{-1}\gamma g_0, A) \leq \lambda^{-1/2+\epsilon_0}\beta^{1/2}$. Lemma 3.3 gives

$$M(g,n,\lambda^{-1/2+\epsilon_0}\beta^{1/2}) \ll_\epsilon n^\epsilon \lambda^{\epsilon_0+\epsilon}(n\lambda^{-1/4}\beta^{1/4}+1),$$

and so we have

$$\begin{aligned}
&\sum_{m,n\leq N}\sum_{d|(n,m)} \frac{d|\alpha_m\alpha_n|}{\sqrt{mn}} \sum_{\gamma\in R(mn/d^2)} \left|I(\lambda,\phi,g_0^{-1}\gamma g_0)\right| \\
&\quad \ll_\epsilon N^\epsilon\lambda^\epsilon \sum_{m,n\leq N}|\alpha_m\alpha_n| \sum_{d|(n,m)} \frac{d}{\sqrt{mn}} \lambda^{1/2} M(g,n,\lambda^{-1/2+\epsilon_0}\beta^{1/2}) \\
&\qquad + O_{A,\epsilon_0}(\lambda^{-A}) \\
&\quad \ll_\epsilon N^\epsilon \lambda^{\epsilon_0+\epsilon} \sum_{m,n\leq N}|\alpha_m\alpha_n| \sum_{d|(n,m)} \left(\frac{\sqrt{mn}}{d}\lambda^{1/4}\beta^{1/4} + \frac{d}{\sqrt{mn}}\lambda^{1/2}\right) \\
&\qquad + O_{A,\epsilon_0}(\lambda^{-A}).
\end{aligned} \tag{24}$$

Combining (24) with (19) and (20) and choosing ϵ_0 small gives

$$\left|\langle \mathcal{T}\psi, b\phi\rangle\right|^2 \ll_\epsilon N^\epsilon\lambda^\epsilon\Big(\lambda^{1/2}\sum_{n\leq N}|\alpha_n|^2 + N\lambda^{1/4}\beta^{1/4}\Big(\sum_{n\leq N}|\alpha_n|\Big)^2\Big) + O_A(\lambda^{-A}),$$

and Lemma 3.9 now follows as in Section 3.3 by choosing $N = \lambda^{1/6}\beta^{-1/6}$. □

4. Spectral estimation of Hecke returns

We now prove Theorem 1.4 by improving the amplifier used in Lemma 3.9. Our new ingredient is a spectral method for estimating $M(g,n,\kappa)$, which works by estimating the sum of $M(g,n,\kappa)$ over a range of n rather than one n at a time. The bound we obtain is Proposition 4.2, which allows us to prove the following result.

PROPOSITION 4.1
Assume that ψ satisfies (5) and (6), and that $1 \leq \beta \leq \lambda^{2/3}$. We have the bound

$$\|\Pi_\beta^+ b\psi\|_2 \ll_\epsilon \lambda^{\theta/2+\epsilon}\beta^{1/4-\theta/2}.$$

Theorem 1.4 follows by choosing $\beta = \lambda^{(1-2\theta)/(2-2\theta)}$ and combining this with Lemma 3.8.

Proof
We maintain the notations of Section 3. Let $\epsilon_0 > 0$, and let $N = \lambda^{1/2+\epsilon_0}\beta^{-1/2}$. Define $\mathcal{T}_1$ to be the operator

$$\mathcal{T}_1 = \sum_{N/2<p<N} \lambda(p) T_p.$$

Proposition 3.1 gives

$$\begin{aligned} \left|\langle \mathcal{T}_1 \psi, b\phi\rangle\right|^2 \le & \sum_{N/2<p<N} \sum_{\gamma\in\Gamma} \lambda(p)^2 \left|I(\lambda,\phi,g_0^{-1}\gamma g_0)\right| \\ & + \sum_{N/2<p_1,p_2<N} \frac{|\lambda(p_1)\lambda(p_2)|}{\sqrt{p_1p_2}} \sum_{\gamma\in R(p_1p_2)} \left|I(\lambda,\phi,g_0^{-1}\gamma g_0)\right|. \end{aligned} \tag{25}$$

We apply (21) and our assumption that $|\lambda(p)| \le 2p^\theta$, which gives

$$\begin{aligned} \left|\langle \mathcal{T}_1 \psi, b\phi\rangle\right|^2 \ll_\epsilon & \, N^{1+2\theta}\lambda^{1/2+\epsilon} \\ & + N^{2\theta} \sum_{N/2<p_1,p_2<N} \frac{1}{\sqrt{p_1p_2}} \sum_{\gamma\in R(p_1p_2)} \left|I(\lambda,\phi,g_0^{-1}\gamma g_0)\right|. \end{aligned}$$

Enlarging the sum to one over all $N^2/4 < n < N^2$ with $(n,q) = 1$, where q is the integer defined in Section 2.2, gives

$$\begin{aligned} \left|\langle \mathcal{T}_1 \psi, b\phi\rangle\right|^2 \ll_\epsilon & \, N^{1+2\theta}\lambda^{1/2+\epsilon} \\ & + N^{2\theta} \sum_{\substack{n\sim N^2 \\ (n,q)=1}} \frac{1}{\sqrt{n}} \sum_{\gamma\in R(n)} \left|I(\lambda,\phi,g_0^{-1}\gamma g_0)\right|. \end{aligned} \tag{26}$$

By Proposition 3.7, we only need to consider the terms in the second sum with $d(g_0^{-1}\gamma g_0, e) \le 1$ and $d(g_0^{-1}\gamma g_0, A) \le \lambda^{-1/2+\epsilon_0}\beta^{1/2}$, which gives

$$\begin{aligned} & \sum_{\substack{n\sim N^2 \\ (n,q)=1}} \frac{1}{\sqrt{n}} \sum_{\gamma\in R(n)} \left|I(\lambda,\phi,g_0^{-1}\gamma g_0)\right| \\ & \quad \ll_\epsilon \sum_{\substack{n\sim N^2 \\ (n,q)=1}} \frac{\lambda^{1/2+\epsilon}}{\sqrt{n}} M(g_0,n,\lambda^{-1/2+\epsilon_0}\beta^{1/2}) + O_{A,\epsilon_0}(\lambda^{-A}). \end{aligned}$$

The assumption that $N = \lambda^{1/2+\epsilon_0}\beta^{-1/2}$ implies that we may apply Proposition 4.2 with $\kappa = \lambda^{-1/2+\epsilon_0}\beta^{1/2}$ and $M = N^2$, which gives

$$\sum_{\substack{n\sim N^2\\(n,q)=1}} \frac{\lambda^{1/2+\epsilon}}{\sqrt{n}} M(g_0, n, \lambda^{-1/2+\epsilon_0}\beta^{1/2}) \ll_{\epsilon_0,\epsilon} N^3\lambda^{-1/2+2\epsilon_0+\epsilon}\beta.$$

Substituting this into (26) gives

$$\left|\langle \mathcal{T}_1\psi, b\phi\rangle\right|^2 \ll_{\epsilon_0,\epsilon} N^{1+2\theta}\lambda^{1/2+\epsilon} + N^{3+2\theta}\lambda^{-1/2+2\epsilon_0+\epsilon}\beta.$$

If we estimate the action of $\mathcal{T}_1$ on ψ by using our assumption (5) and substitute $N \sim \lambda^{1/2+\epsilon_0}\beta^{-1/2}$, and choose ϵ_0 sufficiently small, we obtain

$$\left|\langle \psi, b\phi\rangle\right| \ll_\epsilon \lambda^{\theta/2+\epsilon}\beta^{1/4-\theta/2},$$

as required. □

PROPOSITION 4.2
If M, $\delta > 0$, and $1 > \kappa > 0$ satisfy $M \geq \kappa^{-2-\delta}$, we have

$$\sum_{\substack{M/2<m<M\\(m,q)=1}} \frac{1}{\sqrt{m}} M(g_0, m, \kappa) \ll_{q,\delta} \kappa^2 M^{3/2},$$

where q is the integer defined in Section 2.2.

Proof
Let $b \in C_0^\infty(\mathfrak{g})$ be a real nonnegative function that is supported in the ball of radius 2 about the origin with respect to the norm $\|\cdot\|$ defined in (8), and equal to 1 on the ball of radius 1. Let $C_1 > 0$ be a constant to be chosen later. Define $b_\kappa \in C_0^\infty(\mathfrak{g})$ by $b_\kappa(X) = b(\kappa^{-1}C_1X)$, and let $\widetilde{b_\kappa} \in C_0^\infty(G)$ be the pushforward of b_κ under exp. Let $A > 0$, and define $f \in C_0^\infty(G)$ by

$$f(g) = \int_{-A}^{A} \kappa^{-2}\widetilde{b_\kappa}\big(a(-x)g_0^{-1}g\big)\,dx$$

so that f is an L^1-normalized bump function around the geodesic segment $\{g_0a(x) : -A \leq x \leq A\}$ at scale κ. We choose C_1 small enough and A large enough that the conditions $d(g_0^{-1}gg_0, e) \leq 1$ and $d(g_0^{-1}gg_0, A) \leq \kappa$ imply that $\langle f, gf\rangle \gg 1$, where the implied constant is independent of κ and g_0. If we define $\overline{f} \in L^2(\Gamma\backslash G)$ by

$$\overline{f}(g) = \sum_{\gamma\in\Gamma} f(\gamma g),$$

then $\|\overline{f}\|_2 \sim 1$ in $L^2(\Gamma\backslash G)$.

Choose $h \in C_0^\infty(0,\infty)$ to be real, nonnegative, and satisfy $h(x) = 1$ for $1/2 \leq x \leq 1$. If we define

$$\mathcal{S} = \sum_{(m,q)=1} h(m/M) T_m,$$

then we have

$$\sum_{\substack{M/2<m<M \\ (m,q)=1}} \frac{1}{\sqrt{m}} M(g_0, m, \kappa) \ll \sum_{(m,q)=1} \frac{h(m/M)}{\sqrt{m}} \sum_{\gamma \in R(m)} \langle f, \gamma f \rangle = \langle \overline{f}, \mathcal{S}\overline{f} \rangle,$$

and we may estimate the right-hand side spectrally. Expand $\overline{f}$ with respect to a decomposition of $L^2(\Gamma\backslash G)$ into automorphic representations as

$$\overline{f} = \sum_i \alpha_i \psi_i,$$

where ψ_i is an L^2-normalized vector in an automorphic representation with eigenvalue μ_i under the Casimir operator C. We have

$$\|C^n \overline{f}\|_2 \ll_n \kappa^{-2n},$$

and if n is even and $T > 0$, this implies

$$\begin{aligned} \sum_{|\mu_i| \geq T} \mu_i^n |\alpha_i|^2 \; &\leq \; \langle C^n \overline{f}, \overline{f} \rangle, \\ \sum_{|\mu_i| \geq T} |\alpha_i|^2 &\ll_n \kappa^{-2n} T^{-n}. \end{aligned}$$

If we apply this with $T = \kappa^{-2-\delta/2}$, we obtain

$$\overline{f} = \frac{\langle \overline{f}, 1 \rangle}{\mathrm{Vol}(X)} + \sum_{|\mu_i| \leq \kappa^{-2-\delta/2}}{}' \alpha_i \psi_i + r,$$

where $r \in L^2(\Gamma\backslash G)$ satisfies $\|r\|_2 \ll_{A,\delta} \kappa^A$. Here, Σ' denotes the sum over the nontrivial representations. Substituting this into $\langle \overline{f}, \mathcal{S}\overline{f} \rangle$ gives

$$\begin{aligned} \langle \overline{f}, \mathcal{S}\overline{f} \rangle = \frac{\langle \overline{f}, 1 \rangle^2}{\mathrm{Vol}(X)^2} \sum_{(m,q)=1} h(m/M) \frac{\sigma(m)}{\sqrt{m}} \\ + \sum_{|\mu_i| \leq \kappa^{-2-\delta}} |\alpha_i|^2 \sum_{(m,q)=1} h(m/M) \lambda_i(m) + O_{A,\delta}(\kappa^A) \sum_{m \ll M} \frac{\sigma(m)}{\sqrt{m}}, \end{aligned}$$

where $\lambda_i(m)$ are the Hecke eigenvalues of ψ_i. The result now follows from Lemma 4.3 below, and the bounds $\langle \overline{f}, 1\rangle \ll \kappa$ and

$$\sum_{n\le N} \frac{\sigma(n)}{\sqrt{n}} \sim N^{3/2}.$$

Note that our assumptions that $M \ge \kappa^{-2-\delta}$ and $|\mu_i| \le \kappa^{-2-\delta/2}$ guarantee that the hypothesis of the lemma is satisfied. □

LEMMA 4.3
If $\delta > 0$ and $M \ge |\mu_i|^{1+\delta}$, we have

$$\sum_{(m,q)=1} h(m/M)\lambda_i(m) \ll_{A,\delta} M^{-A},$$

where the implied constant is uniform in ψ_i.

Proof
We shall drop the subscript i, and assume that ψ is a vector in a principal series representation as the discrete series case is similar. We first consider the case $q = 1$.

Let r be the spectral parameter of ψ, so that $\mu = 1/4 + r^2$. By applying the functional equation and Stirling's formula, we see that the L-function $L(s, \psi)$ satisfies the estimate

$$L(-A + it, \psi) \ll_{A,\epsilon} (t^2 + r^2 + 1)^{A+1/2+\epsilon} \tag{27}$$

for A sufficiently large. If we let $\widehat{h}(s)$ be the Mellin transform of h, which is entire and decays rapidly in vertical strips, we obtain

$$\sum_m h(m/M)\lambda(m) = \int_{(2)} L(s, \psi)\widehat{h}(s)M^s\, ds.$$

If we shift the line of integration to $\sigma = -A$ and apply (27) and the rapid decay of $\widehat{h}$, we have

$$\begin{aligned}\sum_m h(m/M)\lambda(m) &\ll_{A,\epsilon} M^{-A}(1 + r^2)^{A+1/2+\epsilon} \\ &\ll_{A,\epsilon} M^{-A}\mu^{A+1/2+\epsilon} \\ &\ll_{A,\epsilon} M^{-A}M^{(1-\delta)(A+1/2+\epsilon)} \\ &\ll_{B,\delta} M^{-B},\end{aligned}$$

as required. In the case when $q > 1$, we apply the same argument to the incomplete L-function obtained by removing the local factors at primes dividing q from $L(s, \psi)$. □

Remark

The method we have used of estimating Hecke recurrences spectrally is unlikely to work in other situations. It requires us to choose an amplifier that makes the sums of eigenvalues in Proposition 4.2 longer than the relevant analytic conductors, and in other cases (such as higher rank or when using the operators T_{p^2} on GL_2 to give an unconditional theorem) this gives the amplifier so much mass that the "off-diagonal" term is worse than the trivial bound. The method also depends on the rapid decay of $I(\lambda, \phi, g)$ as $d(g, A)$ grows, and fails to improve the L^∞-bound of [20] under the assumption (5) because the rate of decay in the corresponding lemma [20, Lemma 1.1] is slower.

5. Bounding $I(\lambda, \phi, e)$

We now prove Proposition 3.6. If we define $p_\lambda(x) = k_\lambda(a(x))$, we have $I(\lambda, \phi, e) = \langle b\phi, p_\lambda * b\phi\rangle$. Define

$$I_\beta = [-\lambda - \beta/2, -\lambda + \beta/2] \cup [\lambda - \beta/2, \lambda + \beta/2],$$

and write $b\phi = \phi_1 + \phi_2$, where the Fourier transform of ϕ_2 is supported on I_β and the transform of ϕ_1 is supported on $\mathbb{R} \setminus I_\beta$. Because b was a fixed smooth function, we have $\|\phi_2\|_2 \ll_A \beta^{-A}$. We have

$$\begin{aligned}\langle b\phi, p_\lambda * b\phi\rangle &= \langle \phi_1, p_\lambda * \phi_1\rangle + \langle \phi_2, p_\lambda * \phi_2\rangle \\ &\leq \sup_{t \notin I_\beta} \big|\widehat{p_\lambda}(t)\big| + O_A(\beta^{-A}) \|\widehat{p_\lambda}\|_\infty.\end{aligned}$$

By [23, Lemma 2.7] (see also [11, Lemma 4.1]) we have

$$p_\lambda(x) \ll \lambda\big(1 + \lambda|x|\big)^{-1/2}, \tag{28}$$

and this implies that $\|\widehat{p_\lambda}\|_\infty \ll \lambda^{1/2}$. It therefore suffices to prove the following estimate.

LEMMA 5.1

We have $|\widehat{p_\lambda}(t)| \ll \lambda^{1/2}\beta^{-1/2}$ *for* $t \notin I_\beta$.

Proof

Let $b_1 \in C_0^\infty(\mathbb{R})$ be a cutoff function that is equal to 1 on $[-1, 1]$ and zero outside $[-2, 2]$. We have

$$\widehat{p_\lambda}(t) = \int_{-\infty}^{\infty} b_1(x) k_\lambda\big(a(x)\big) e^{-itx}\, dx.$$

We shall estimate this by inverting the Harish-Chandra transform and applying asymptotics for φ_s from [23]. The first is [23, Theorem 1.3], which implies that

$$b_1(x)\varphi_s(x) \ll \big(1+|sx|\big)^{-1/2}. \tag{29}$$

By [23, Theorem 1.5], there are functions $f_\pm \in C^\infty((0,3)\times\mathbb{R}_{\geq 0})$ such that

$$\Big(\frac{\partial}{\partial x}\Big)^n f_\pm(x,s) \ll_n x^{-n}(sx)^{-1/2} \tag{30}$$

and

$$\varphi_s\big(a(x)\big) = f_+(x,s)e^{isx} + f_-(x,s)e^{-isx} + O_A\big((sx)^{-A}\big) \tag{31}$$

for $x \in (0,3)$. The bound (29) gives

$$\int_{-\infty}^{\infty} b_1(x)\varphi_s\big(a(x)\big)e^{-itx}\,dx \ll \big(1+|s|\big)^{-1/2}$$

for all s and t.

After inverting the Harish-Chandra transform and applying the rapid decay of h, it suffices to prove the bound

$$\int_{-\infty}^{\infty} b_1(x)\varphi_s\big(a(x)\big)e^{-itx}\,dx \ll \lambda^{-1/2}\beta^{-1/2}$$

uniformly for $t \notin I_\beta$ and $|s-\lambda| \leq \beta/4$. We decompose the integral as

$$\int_{-\infty}^{\infty} b_1(x)b_1(\beta x)\varphi_s(x)e^{-itx}\,dx \\ + \sum_{n=1}^{\infty}\int_{-\infty}^{\infty} b_1(x)\big(b_1(2^{-n}\beta x) - b_1(2^{-n+1}\beta x)\big)\varphi_s(x)e^{-itx}\,dx.$$

The bound (29) implies that the first integral is $\ll \lambda^{-1/2}\beta^{-1/2}$. We shall estimate the integrals in the sum by applying (31). The error term in (31) makes a contribution of

$$(\lambda\beta^{-1}2^n)^{-A}\beta^{-1}2^n \ll \lambda^{-2}\beta 2^{-n} \leq \lambda^{-1/2}\beta^{-1/2}2^{-n}$$

to the nth term in the sum, which may be ignored. We shall only consider the integral over $x > 0$ and the main term involving f_+, as the other terms are similar. We must then estimate

$$\int_0^{\infty} b_1(x)\big(b_1(2^{-n}\beta x) - b_1(2^{-n+1}\beta x)\big) f_+(x,s)e^{i(s-t)x}\,dx.$$

If we replace x with $2^n\beta^{-1}x$, this becomes

$$\int_0^{\infty} 2^n\beta^{-1}b_1(2^n\beta^{-1}x)\big(b_1(x) - b_1(2x)\big) f_+(2^n\beta^{-1}x,s)e^{i(s-t)2^n\beta^{-1}x}\,dx. \tag{32}$$

The conditions $t \notin I_\beta$ and $|s-\lambda| \leq \beta/4$ imply that $|(s-t)2^n\beta^{-1}| \geq 2^{n-2}$, and (30) implies that the kth derivative of the amplitude factor in (32) is $\ll_k \lambda^{-1/2}2^{n/2}\beta^{-1/2}$. Integrating by parts once gives

$$\int_0^\infty 2^n\beta^{-1}b_1(2^n\beta^{-1}x)\big(b_1(x)-b_1(2x)\big)f_+(2^n\beta^{-1}x,s)e^{i(s-t)2^n\beta^{-1}x}\,dx \\ \ll \lambda^{-1/2}2^{-n/2}\beta^{-1/2},$$

and summing over n completes the proof. □

6. Oscillatory integrals when $t \sim \lambda$

In this section, we prove Proposition 3.7 by building up the integral $I(\lambda,\phi,g)$ in several steps. We begin with three calculations that we shall use repeatedly in this section and in Section 7.

LEMMA 6.1
Fix $g \in G$, and define $\theta : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{R}/2\pi\mathbb{Z}$ by $k(\sigma)g \in NAk(\theta(\sigma))$. Then θ is a diffeomorphism.

Proof
By using the Cartan decomposition, we may reduce to the case where $g = a(y)$. Taking inverses gives $a(-y)k(-\sigma) \in k(-\theta(\sigma))AN$, and applying both sides to the point at infinity gives

$$e^{-y}\cot(\sigma/2) = \cot\big(\theta(\sigma)/2\big). \tag{33}$$

This proves that θ is a bijection, and a diffeomorphism everywhere except at $\sigma = 0$. Rewriting the equation as $e^y\tan(\sigma/2) = \tan(\theta(\sigma)/2)$ proves it at $\sigma = 0$ also. □

LEMMA 6.2
Let $y, z \in G$, let $\sigma \in \mathbb{R}/2\pi\mathbb{Z}$, and let θ be determined by $k(\sigma)y^{-1} \in NAk(\theta)$. Then we have

$$A\big(k(\sigma)y^{-1}z\big) = A\big(k(\theta)z\big) - A\big(k(\theta)y\big).$$

Proof
Let $k(\sigma)y^{-1} = nak(\theta)$. We have

$$\begin{aligned} A\big(k(\sigma)y^{-1}z\big) &= A\big(nak(\theta)z\big) \\ &= A(na) + A\big(k(\theta)z\big) \end{aligned}$$

$$
\begin{aligned}
&= A\big(k(\theta)z\big) - A\big(n^{-1}a^{-1}k(\sigma)\big) \\
&= A\big(k(\theta)z\big) - A\big(k(\theta)y\big),
\end{aligned}
$$

as required. □

LEMMA 6.3
Let $g \in G$ have Iwasawa decomposition $g = nak(\theta)$. Then

$$\frac{\partial}{\partial t} A\big(ga(t)\big)\Big|_{t=0} = \cos\theta.$$

Proof
Let H be as in (10), and let $H^* \in \mathfrak{g}^*$ be dual to H. We have

$$
\begin{aligned}
A\big(ga(t)\big) &= A(a) + A\big(k(\theta)\exp(tH)k(-\theta)\big) \\
&= A(a) + A\big(\exp\big(t\,\mathrm{Ad}\big(k(\theta)\big)H\big)\big),
\end{aligned}
$$

and therefore

$$\frac{\partial}{\partial t} A\big(ga(t)\big)\Big|_{t=0} = H^*\big(\mathrm{Ad}\big(k(\theta)\big)H\big) = \cos\theta.$$ □

6.1. A uniformization result
We shall need the following uniformization lemma for the function A.

LEMMA 6.4
Let $D > 0$. There exists $\delta, \sigma > 0$ depending on D, and a real analytic function ξ : $(-\delta,\delta)\times(-D,D)^2 \to \mathbb{R}$ such that

$$\frac{\partial}{\partial y} A\big(k(\theta)n(x)a(y)\big) = 1 - \theta^2\xi(\theta,x,y),$$

and

$$
\begin{aligned}
\big|\xi(\theta,x,y)\big| &\geq \sigma, \qquad (34)\\
\Big|\frac{\partial^n \xi}{\partial y^n}(\theta,x,y)\Big| &\ll_{D,n} 1,
\end{aligned}
$$

for $(\theta,x,y) \in (-\delta,\delta)\times(-D,D)^2$.

Proof
Define the function $\alpha(\theta,x,y) : \mathbb{R}/2\pi\mathbb{Z}\times(-2D,2D)^2 \to \mathbb{R}/2\pi\mathbb{Z}$ by requiring that

$$k(\theta)n(x)a(y) \in NAk\big(\alpha(\theta,x,y)\big).$$

The analyticity of the Iwasawa decomposition implies that α is analytic as a function of (θ, x, y). Lemma 6.3 implies that

$$1 - \frac{\partial}{\partial y} A\big(k(\theta)n(x)a(y)\big) = 1 - \cos\alpha$$
$$= 2\sin^2(\alpha/2).$$

We choose δ such that $\sin(\alpha/2)$ vanishes on $(-2\delta, 2\delta) \times (-2D, 2D)^2$ if and only if $\theta = 0$. Lemma 6.1 implies that $\partial\alpha/\partial\theta$ never vanishes on $\{0\} \times (-2D, 2D)^2$, and so because α was analytic we see that there is a real analytic function ξ_0 on $(-2\delta, 2\delta) \times (-2D, 2D)^2$ such that $\sin(\alpha/2) = \theta\xi_0$. Defining $\xi = 2\xi_0^2$ and restricting the domain to $(-\delta, \delta) \times (-D, D)^2$ gives the result. □

6.2. *Constituent integrals of $I(\lambda, \phi, g)$*

We now estimate two one-dimensional integrals that appear in $I(\lambda, \phi, g)$.

PROPOSITION 6.5

Let $C, D > 0$ and $\epsilon_0 > 0$ be constants, and let $b \in C_0^\infty(\mathbb{R})$ be a function supported in $[0, 1]$. If $x, y \in [-D, D]$ and

$$|\theta| \geq C s^{-1/2+\epsilon_0} \beta^{1/2} \qquad \text{and} \qquad |t - s| \leq \beta \tag{35}$$

for some $s, \beta \geq 1$ satisfying $\beta \leq 3s^{2/3}$, then

$$\int_{-\infty}^{\infty} b(z) \exp\big(itz - isA\big(k(\theta)n(x)a(y+z)\big)\big)\, dz \ll s^{-A}. \tag{36}$$

The implied constant depends only on A, C, D, ϵ_0, and the size of the first n derivatives of b, where n depends on ϵ_0 and A.

Proof

By applying Lemma 6.4, we see that there is some $\delta > 0$ and a nonvanishing real analytic function ξ on $(-\delta, \delta) \times (-D-2, D+2)^2$ such that

$$\frac{\partial}{\partial z} A\big(k(\theta)n(x)a(y+z)\big) = 1 - \theta^2 \xi(\theta, x, y+z)$$

when $\theta \in (-\delta, \delta)$, $x, y \in [-D, D]$, and $z \in [0, 1]$. If $Z(\theta, x, y)$ is an antiderivative of ξ with respect to y that is smooth as a function of (θ, x, y), we may integrate this to obtain

$$A\big(k(\theta)n(x)a(y)\big) = y - \theta^2 Z(\theta, x, y) + c(x, \theta)$$

for some function $c(x,\theta)$. If $\theta \in (-\delta,\delta)$, we may use this to rewrite the integral (36) as

$$\begin{aligned}\int_{-\infty}^{\infty} & b(z)\exp\bigl(itz - isA\bigl(k(\theta)n(x)a(y+z)\bigr)\bigr)\,dz \\ &= e^{-is(y+c(\theta,x))}\int_{-\infty}^{\infty} b(z)\exp\bigl(i(t-s)z + is\theta^2 Z(\theta,x,y+z)\bigr)\,dz \\ &= e^{-is(y+c(\theta,x))}\int_{-\infty}^{\infty} b(z)\exp\bigl(is\theta^2\Psi(z)\bigr)\,dz, \qquad (37)\end{aligned}$$

where we define $\Psi(z) = Z(\theta,x,y+z) + s^{-1}\theta^{-2}(s-t)z$.

Our assumption (35) implies that

$$\bigl|s^{-1}\theta^{-2}(s-t)\bigr| \le s^{-1}\theta^{-2}\beta \le C^{-2}s^{-2\epsilon_0},$$

so that

$$\begin{aligned}\Psi &= Z(\theta,x,y+z) + O(s^{-2\epsilon_0})z \qquad \text{and} \\ \frac{\partial\Psi}{\partial z} &= \xi(\theta,x,y+z) + O(s^{-2\epsilon_0}).\end{aligned} \qquad (38)$$

It follows from (34) and (38) that for s sufficiently large, $|\partial\Psi/\partial z| > \sigma/2$ for all $\theta \in (-\delta,\delta)$, $x,y \in [-D,D]$ and $z \in [0,1]$. In addition, all other derivatives of Ψ are bounded from above. As (35) implies that $s\theta^2 \ge C^2 s^{2\epsilon_0}\beta \ge C^2 s^{2\epsilon_0}$, the bound (36) follows by integration by parts in (37).

In the case where $\theta \notin (-\delta,\delta)$, Lemma 6.3 implies that $(\partial/\partial z)A(k(\theta)n(x)a(y+z)) \le 1 - c_1$ for some $c_1 > 0$ depending only on δ, which gives

$$\frac{\partial}{\partial z}\bigl(its^{-1}z - A\bigl(k(\theta)n(x)a(y+z)\bigr)\bigr) \gg 1.$$

The result now follows by integration by parts. $\square$

The second one-dimensional integral that we shall estimate is as follows.

PROPOSITION 6.6
Let $C, D > 0$ and $\epsilon_0 > 0$ be constants, and let $b \in C_0^\infty(\mathbb{R})$ be a function supported in $[0,1]$. *If $x, y \in [-D,D]$ and*

$$|x| \ge Cs^{-1/2+\epsilon_0}\beta^{1/2} \qquad \text{and} \qquad |t-s| \le \beta \qquad (39)$$

for some $s, \beta \ge 1$ satisfying $\beta \le 3s^{2/3}$, then

$$\int_{-\infty}^{\infty} b(z)e^{itz}\varphi_s\bigl(n(x)a(y+z)\bigr)\,dz \ll s^{-A}. \qquad (40)$$

The implied constant depends only on A, C, D, ϵ_0, and the size of the first n derivatives of b, where n depends on ϵ_0 and A.

Proof
If we apply the functional equation $\varphi_s = \varphi_{-s}$, and substitute the formula for φ_{-s} as an integral of plane waves into the left-hand side of (40), it becomes

$$\int_{-\infty}^{\infty}\int_0^{2\pi} b(z)\exp\big(itz + (1/2 - is)A\big(k(\theta)n(x)a(y+z)\big)\big)\, d\theta\, dz.$$

Let $f(x) \in C_0^\infty(\mathbb{R})$ be a function with $\operatorname{supp}(f) \subseteq [-2, 2]$ and $f(x) = 1$ on $[-1, 1]$. Let C_1 be a positive constant to be chosen later. Define b_1 by $b_1(x) = f(C_1^{-1} \times s^{1/2-\epsilon_0}\beta^{-1/2}x)$ and set $b_2 = 1 - b_1$, so that $1 = b_1(\theta) + b_2(\theta)$ is a smooth partition of unity on $\mathbb{R}/2\pi\mathbb{Z}$ with

$$\begin{aligned}&\operatorname{supp}(b_1) \subseteq [-2C_1 s^{-1/2+\epsilon_0}\beta^{1/2}, 2C_1 s^{-1/2+\epsilon_0}\beta^{1/2}],\\ &\operatorname{supp}(b_2) \subseteq \mathbb{R}/2\pi\mathbb{Z} \setminus [-C_1 s^{-1/2+\epsilon_0}\beta^{1/2}, C_1 s^{-1/2+\epsilon_0}\beta^{1/2}].\end{aligned}$$

Proposition 6.5 implies that

$$\int_{-\infty}^{\infty}\int_0^{2\pi} b_2(\theta)b(z)\exp\big(itz + (1/2 - is)A\big(k(\theta)n(x)a(y+z)\big)\big)\, d\theta\, dz \ll_A s^{-A},$$

so that it suffices to estimate

$$\int_{-\infty}^{\infty}\int_0^{2\pi} b_1(\theta)b(z)\exp\big(itz + (1/2 - is)A\big(k(\theta)n(x)a(y+z)\big)\big)\, d\theta\, dz.$$

We shall do this by estimating the integrals

$$\int_0^{2\pi} b_1(\theta)\exp\big(-isA\big(k(\theta)n(x)a(y)\big)\big)\, d\theta \tag{41}$$

in θ, where now $D \geq |x| \geq Cs^{-1/2+\epsilon_0}\beta^{1/2}$ and $y \in [-D, D+1]$.

If $X \in \mathfrak{g}$, we let X^* be the vector field on $\mathbb{H}$ whose value at p is $\frac{\partial}{\partial t}\exp(tX)p|_{t=0}$. It may be shown that these vector fields satisfy $[X^*, Y^*] = -[X, Y]^*$, where the first Lie bracket is on $\mathbb{H}$ and the second is in $\mathfrak{g}$. We recall the vectors $X_{\mathfrak{n}}$ and $X_{\mathfrak{k}}$ defined in (10). It may be easily seen that the subset of $\mathbb{H}$ where the function $X_{\mathfrak{k}}^* A$ vanishes is exactly A, and the following lemma implies that it vanishes to first-order there.

LEMMA 6.7
We have $X_{\mathfrak{n}}^* X_{\mathfrak{k}}^* A(a(y)) = e^y$ for all y.

Proof
We have

$$X_{\mathfrak{n}}^* X_{\mathfrak{k}}^* A = X_{\mathfrak{k}}^* X_{\mathfrak{n}}^* A + [X_{\mathfrak{n}}^*, X_{\mathfrak{k}}^*] A.$$

It may be seen that the first term vanishes, and we have

$$[X_{\mathfrak{n}}^*, X_{\mathfrak{k}}^*] = -[X_{\mathfrak{n}}, X_{\mathfrak{k}}]^* = H^*,$$

which implies the lemma. □

Lemma 6.7 implies that there exist $\sigma, \delta > 0$ depending on D such that if $|x| < \sigma$ and $y \in [-D-1, D+2]$, then we have $|X_{\mathfrak{k}}^* A(n(x)a(y))| \geq \delta |x|$. Define

$$B = \{ n(x)a(y) \mid |x| \leq \sigma/2, y \in [-D, D+1] \},$$
$$B' = \{ n(x)a(y) \mid |x| < \sigma, y \in [-D-1, D+2] \}.$$

Let $p = n(x)a(y)$, where x and y are as in (41), and assume that $p \in B$. We recall the function $N(p)$ defined in (9), and assume without loss of generality that $1/8 > \epsilon_0$ so that $s^{-1/2+\epsilon_0}\beta^{1/2}$ decays as s grows. If s is sufficiently large and C_1 sufficiently small, depending on C, D, and σ, and $|\theta| \leq 2C_1 s^{-1/2+\epsilon_0}\beta^{1/2}$, we have $k(\theta)p \in B'$ and $|N(k(\theta)p)| \gg_{C,D} s^{-1/2+\epsilon_0}\beta^{1/2}$. It follows that

$$\left| \frac{\partial}{\partial \theta'} A\big(k(\theta')p\big) \Big|_{\theta'=\theta} \right| = \big| X_{\mathfrak{k}}^* A\big(k(\theta)p\big) \big| \geq \delta \big| N\big(k(\theta)p\big) \big| \gg_{C,D} s^{-1/2+\epsilon_0}\beta^{1/2}$$

when $|\theta| \leq 2C_1 s^{-1/2+\epsilon_0}\beta^{1/2}$. The proposition now follows from Lemma 6.8.

If $p \notin B$, then there is $\kappa > 0$ depending on D such that $|X_{\mathfrak{k}}^* A(k(\theta)p)| \geq \kappa$ when $|\theta| \leq 2C_1 s^{-1/2+\epsilon_0}\beta^{1/2}$ and s is sufficiently large (again using our assumption $\epsilon_0 < 1/8$). The proposition again follows by integration by parts. □

LEMMA 6.8
If $b \in C_0^\infty(\mathbb{R})$ is a cutoff function at scale $1 \geq \delta > 0$, and $\phi \in C^\infty(\mathbb{R})$ satisfies

$$\phi'(x) \gg \delta \qquad \phi^{(n)}(x) \ll_n 1$$

for $x \in \operatorname{supp}(b)$ and $n > 1$, then

$$\int b(x) e^{it\phi(x)}\, dx \ll_A \delta (t\delta^2)^{-A}.$$

Proof
We have

$$\int b(x) e^{it\phi(x)}\, dx = \delta \int b(\delta x) e^{i(t\delta^2)\delta^{-2}\phi(\delta x)}\, dx.$$

The function $b(\delta x)$ is now a cutoff function at scale 1, and $\delta^{-2}\phi(\delta x)$ satisfies

$$\frac{d}{dx}\delta^{-2}\phi(\delta x) \gg 1 \qquad \Big(\frac{d}{dx}\Big)^n \delta^{-2}\phi(\delta x) \ll_n \delta^{n-2} \le 1$$

for $x \in \mathrm{supp}(b(\delta\cdot))$ and $n > 1$. The lemma now follows by integration by parts. □

6.3. Proof of Proposition 3.7

If $b_0 \in C_0^\infty(0,1)$, $t_1, t_2 \in \mathbb{R}$, and $\phi \in L^1_{\mathrm{loc}}(\mathbb{R})$, we define

$$J(s,t_1,t_2,g) = \iint_{-\infty}^{\infty} b_0(x_1)b_0(x_2)e^{-i(t_1x_1-t_2x_2)}\varphi_s\big(a(-x_1)ga(x_2)\big)\,dx_1\,dx_2,$$

$$J(s,\phi,g) = \iint_{-\infty}^{\infty} \overline{b_0\phi}(x_1)b_0\phi(x_2)\varphi_s\big(a(-x_1)ga(x_2)\big)\,dx_1\,dx_2.$$

We now combine Propositions 6.5 and 6.6 to bound $J(s,t_1,t_2,g)$, which will imply Proposition 3.7 after integrating in the various spectral parameters.

PROPOSITION 6.9

Let $\epsilon_0 > 0$ be given. If $g \in G$ satisfies

$$d(g,e) \le 1 \qquad \textit{and} \qquad d(g,A) \ge s^{-1/2+\epsilon_0}\beta^{1/2}$$

for some $s, \beta \ge 1$ satisfying $\beta \le 3s^{2/3}$, and $t_1, t_2 \in [s-\beta, s+\beta]$, then

$$J(s,t_1,t_2,g) \ll_{A,\epsilon_0} s^{-A}. \tag{42}$$

Proof

Define $\theta(\sigma,x_1)$ by $k(\sigma)a(-x_1) \in NAk(\theta(\sigma,x_1))$. If we apply the functional equation $\varphi_s = \varphi_{-s}$, write φ_{-s} as an integral over K, and applying Lemmas 6.1 and 6.2, we obtain

$$\varphi_s\big(a(-x_1)ga(x_2)\big) = \int \exp\Big((1/2-is)\Big(A\big(k(\theta)ga(x_2)\big) - A\big(k(\theta)a(x_1)\big)\Big)\Big)\frac{d\sigma}{d\theta}\,d\theta. \tag{43}$$

Substituting this into the definition of $J(s,t_1,t_2,g)$ gives

$$\iint_{-\infty}^{\infty}\int_0^{2\pi} b_0(x_1)b_0(x_2)e^{-i(t_1x_1-t_2x_2)} \times \exp\Big((1/2-is)\Big(A\big(k(\theta)ga(x_2)\big) - A\big(k(\theta)a(x_1)\big)\Big)\Big)\frac{d\sigma}{d\theta}\,d\theta\,dx_1\,dx_2. \tag{44}$$

Let $g = k(\theta')n(x')a(y')$. The condition $d(g,e) \le 1$ implies that x' and y' are bounded. We then have $k(\theta)ga(x_2) = k(\theta+\theta')n(x')a(x_2+y')$. Choose a constant $C > 0$. If $\theta \notin [-Cs^{-1/2+\epsilon_0}\beta^{1/2}, Cs^{-1/2+\epsilon_0}\beta^{1/2}]$, then integrating (44) in x_1 and applying Proposition 6.5 shows that the integral of (44) over x_1 and x_2 with this value of θ is $\ll_{C,A,\epsilon_0} s^{-A}$, and if $\theta+\theta' \notin [-Cs^{-1/2+\epsilon_0}\beta^{1/2}, Cs^{-1/2+\epsilon_0}\beta^{1/2}]$ we obtain the same conclusion by integrating in x_2 . Combining these, we see that (44) will be $\ll_{C,A,\epsilon_0} s^{-A}$ unless $|\theta'| \le 2Cs^{-1/2+\epsilon_0}\beta^{1/2}$, and we assume that this is the case.

If C is chosen sufficiently small, the condition $|\theta'| \le 2Cs^{-1/2+\epsilon_0}\beta^{1/2}$ and our assumption that $d(g,A) \ge s^{-1/2+\epsilon_0}\beta^{1/2}$ imply that $|x'| \ge C_1 s^{-1/2+\epsilon_0}\beta^{1/2}$ for some absolute $C_1 > 0$. It follows that if C is sufficiently small, we have $N(a(-x_1 g)) \ge C_2 s^{-1/2+\epsilon_0}\beta^{1/2}$ for some absolute C_2 and all $0 \le x_1 \le 1$. The result now follows by applying Proposition 6.6 to the integral in x_2 defining $J(s,t_1,t_2,g)$ for each fixed x_1. □

Proof of Proposition 3.7

To prove the bound (21), observe that equation (28) implies that

$$\int_{-\infty}^{\infty} \left|k_\lambda\big(ga(x)\big)\right|^2 dx \ll \lambda \log \lambda$$

uniformly for $g \in G$. It follows that $b(x_1)b(x_2)k_\lambda(a(-x_1)ga(x_2))$ has norm $\ll_\epsilon \lambda^{1/2+\epsilon}$ as an element of $L^2(\mathbb{R}^2)$, and the result follows by Cauchy–Schwarz.

We now prove (22). If we invert the Fourier transform of ϕ and note that $\|\widehat{\phi}\|_1 \le \|\widehat{\phi}\|_2(2\beta)^{1/2} = (2\pi)^{1/2}(2\beta)^{1/2}$, we may apply Proposition 6.9 with 2β in place of β to obtain

$$J(s,\phi,g) \ll_{A,\epsilon_0} s^{-A} \tag{45}$$

for $|s-\lambda| \le \beta$. Inverting the Harish-Chandra transform of k_λ gives

$$I(\lambda,\phi,g) = \frac{1}{2\pi}\int_0^\infty J(s,\phi,g)h_\lambda(s)s\tanh(\pi s)\,ds. \tag{46}$$

We may assume without loss of generality that $\beta > \lambda^\epsilon$ for some $\epsilon > 0$. We estimate the integral of (46) over $|s-\lambda| > \beta$ by using the rapid decay of h and the integral over $|s-\lambda| \le \beta$ by using (45), which completes the proof. □

7. Oscillatory integrals when $t < \lambda$

We now prove Proposition 3.5. We will use the geometry of $\mathbb{H}$ to a greater extent than in previous sections, and in particular we will distinguish between elements of G and their images in $\mathbb{H}$. We assume that all geodesics in $\mathbb{H}$ carry an orientation. When we refer to the unit tangent vector to a geodesic at a point, we shall always mean in the direction of its orientation. If ℓ_1 and ℓ_2 are two intersecting geodesics, we shall denote by $\angle(\ell_1, \ell_2)$ the angle between their unit tangent vectors at the point of intersection measured in the counterclockwise direction from ℓ_1 to ℓ_2. We let ℓ be the geodesic corresponding to A, and let $\ell : x \mapsto a(x)K$ be the standard parameterization. We give ℓ the upward-pointing orientation, which we transfer to $g\ell$ for any g.

As in the proof of Proposition 3.7, it suffices to bound $J(s,t,g) := J(s,t,t,g)$. After substituting the expression (43) for $\varphi_s(a(-x_1)ga(x_2))$ into the integral defining $J(s,t,g)$, we obtain an oscillatory integral in the variables θ, x_1, and x_2 with phase $e^{-is\phi}$, where

$$\phi(x_1, x_2, \theta, g, \rho) = \rho(x_1 - x_2) - A\big(k(\theta)a(x_1)\big) + A\big(k(\theta)ga(x_2)\big)$$

and $\rho = t/s \geq 0$. We first assume that $\rho \in [\delta, 1-\delta]$ for some $1/2 > \delta > 0$; the case $\rho = 0$ is simpler and treated at the end of Section 7.5. Define $\alpha \in [0, \pi/2]$ to be the solution to $\cos\alpha = \rho$, which is bounded away from 0 and $\pi/2$. We shall study the critical points of ϕ in Sections 7.1–7.4, before deriving a bound for $I(s,t,g)$ from our results in Section 7.5. We shall write $\phi(x_1, x_2, \theta)$ when g and ρ are not varying.

7.1. The critical points of ϕ

LEMMA 7.1

The phase function ϕ has a critical point at $(x_1, x_2, \theta, g, \rho)$ exactly when $k(\theta)\ell(x_1)$ and $k(\theta)g\ell(x_2)$ lie on the same vertical geodesic v, which we give the upward-pointing orientation, and we have $\angle(v, k(\theta)\ell), \angle(v, k(\theta)g\ell) \in \{\pm\alpha\}$.

Proof

Suppose that (x_1', x_2', θ') is a critical point of ϕ. Define the functions $x(\theta)$, $y(\theta)$, and $\beta(\theta)$ by

$$k(\theta)a(x_1') = n\big(x(\theta)\big)a\big(y(\theta)\big)k\big(\beta(\theta)\big),$$

and let

$$n' = n\big(x(\theta')\big), \qquad a' = a\big(y(\theta')\big), \qquad \text{and} \qquad \beta' = \beta(\theta').$$

It may be seen that $v := n'\ell$ is the upward-pointing geodesic through $k(\theta')\ell(x_1')$ and that $\beta' = \angle(v, k(\theta')\ell)$. Lemma 6.3 then implies that $\beta' = \pm\alpha$. The calculation in the case of $\partial/\partial x_2$ is identical.

Lemma 6.2 gives

$$-A\big(k(\theta)a(x_1')\big) + A\big(k(\theta)ga(x_2')\big) = A\big(k\big(\beta(\theta)\big)a(x_1')^{-1}ga(x_2')\big),$$

so that

$$\begin{aligned}\frac{\partial\phi}{\partial\theta}(x_1', x_2', \theta') &= \frac{\partial}{\partial\theta}A\big(k\big(\beta(\theta)\big)a(x_1')^{-1}ga(x_2')\big)\Big|_{\theta=\theta'} \\ &= \frac{\partial\beta}{\partial\theta}(\theta')\frac{\partial}{\partial\beta}A\big(k(\beta)a(x_1')^{-1}ga(x_2')\big)\Big|_{\beta=\beta'}. \end{aligned} \tag{47}$$

Because $\partial\beta/\partial\theta$ does not vanish by Lemma 6.1, and

$$\frac{\partial}{\partial\theta}A\big(k(\theta)g\big)\Big|_{\theta=0} = 0$$

if and only if $g \in AK$, we have $\partial\phi/\partial\theta = 0$ if and only if $k(\beta')a(x_1')^{-1}ga(x_2') \in AK$; that is, $k(\beta')a(x_1')^{-1}g\ell(x_2')$ lies on ℓ. Because $k(\beta')a(x_1)^{-1} = a'^{-1}n'^{-1}k(\theta')$, this is equivalent to the condition that $k(\theta')ga(x_2') \in n'AK$, or that $k(\theta')g\ell(x_2')$ lies on the vertical geodesic v.

We finish with an observation that will be useful in calculating the Hessian of ϕ. We have $k(\beta')a(x_1')^{-1}ga(x_2') \in a(h)K$ for some $h \in \mathbb{R}$, and it may be seen that $k(\theta')a(x_1') \in n'a'K$ and $k(\theta')ga(x_2') \in n'a'a(h)K$, so that h is the signed distance from $k(\theta')\ell(x_1')$ to $k(\theta')g\ell(x_2')$ along v. □

Given a pair of geodesics ℓ_1 and ℓ_2, we say that a geodesic j is a critical geodesic for (ℓ_1, ℓ_2) if j meets ℓ_1 and ℓ_2 at angles of $\pm\alpha$. We may therefore rephrase Lemma 7.1 as saying that $(x_1, x_2, \theta, g, \rho)$ is a critical point of ϕ exactly when $(\ell, g\ell)$ has a critical geodesic j, $\ell(x_1)$ and $g\ell(x_2)$ both lie on j, and $k(\theta)j$ is vertical. As in Lemma 7.1, we define the aperture of a critical point to be the signed distance from $\ell(x_1)$ to $g\ell(x_2)$ on the geodesic j.

We shall now calculate the Hessian of ϕ at its critical points. Let (x_1', x_2', θ') be a critical point of ϕ, and define functions $\beta_i(\theta)$ by

$$k(\theta)a(x_1') \in NAk\big(\beta_1(\theta)\big), \qquad k(\theta)ga(x_2') \in NAk\big(\beta_2(\theta)\big). \tag{48}$$

We let $\beta_i' = \beta_i(\theta')$. It follows from Lemma 7.1 that $\beta_i' \in \{\pm\alpha\}$. Let h be the aperture of the critical point, so that

$$k(\beta_1')a(x_1')^{-1}ga(x_2') \in \begin{pmatrix} e^h & 0 \\ 0 & 1 \end{pmatrix} K.$$

We define $\kappa = \frac{\partial\beta_1}{\partial\theta}(\theta')$, which is nonzero by Lemma 6.1. The Hessian of ϕ at (x_1', x_2', θ') is given by the following proposition.

PROPOSITION 7.2

The Hessian of ϕ *at* (x_1', x_2', θ') *with respect to the coordinates* (x_1, x_2, θ) *is*

$$D = \begin{pmatrix} \frac{1}{2}\sin^2\alpha & 0 & \kappa\sin\beta_1' \\ 0 & -\frac{1}{2}\sin^2\alpha & -\kappa e^h \sin\beta_2' \\ \kappa\sin\beta_1' & -\kappa e^h\sin\beta_2' & \kappa^2(1-e^{2h})/2 \end{pmatrix}.$$

The determinant of D is

$$|D| = \frac{3}{8}\kappa^2\sin^4\alpha(1-e^{2h}),$$

which is nonzero unless $h = 0$*; that is, the points* $\ell(x_1')$ *and* $g\ell(x_2')$ *coincide in* $\mathbb{H}$.

Proof

It is clear that $\partial^2\phi/\partial x_1\,\partial x_2$ is identically 0. To calculate $\partial^2\phi/\partial x_1^2$, define $\gamma : \mathbb{R} \to \mathbb{R}/2\pi\mathbb{Z}$ by the condition that $k(\theta')a(x_1' + t) \in NAk(\gamma(t))$. Our assumption that we are at a critical point implies that $\gamma(0) = \beta_1' = \pm\alpha$. Lemma 6.3 gives

$$\frac{\partial}{\partial t}\phi(x_1' + t, x_2', \theta') = \rho - \cos\gamma(t)$$

and

$$\frac{\partial^2\phi}{\partial x_1^2}(x_1', x_2', \theta') = \sin\beta_1' \frac{\partial\gamma}{\partial t}(0). \tag{49}$$

We have

$$\begin{aligned} k(\theta')a(x_1' + t) &\in NAk\big(\gamma(t)\big), \\ NAk(\beta_1')a(t) &= NAk\big(\gamma(t)\big), \\ k(\beta_1')a(t) &\in NAk\big(\gamma(t)\big). \end{aligned}$$

Equation (33) then gives $\tan(\gamma(t)/2) = e^t\tan(\beta_1'/2)$, so that

$$\begin{aligned} \frac{\partial\gamma}{\partial t}\sec^2\big(\gamma(t)/2\big) &= e^t\tan(\beta_1'/2), \\ \frac{\partial\gamma}{\partial t}(0) &= \cos^2(\beta_1'/2)\tan(\beta_1'/2) \\ &= \frac{1}{2}\sin\beta_1'. \end{aligned}$$

Substituting this into (49) gives

$$\frac{\partial^2\phi}{\partial x_1^2}(x_1', x_2', \theta') = \frac{1}{2}\sin^2\beta_1' = \frac{1}{2}\sin^2\alpha.$$

The calculation of $\partial^2\phi/\partial x_2^2$ is identical, with the exception of a change in sign.

To calculate $\partial^2\phi/\partial\theta\,\partial x_1$, we again have

$$\frac{\partial\phi}{\partial x_1}(x_1', x_2', \theta) = \rho - \cos\beta_1(\theta)$$

and

$$\frac{\partial^2\phi}{\partial\theta\,\partial x_1}(x_1', x_2', \theta') = \sin\beta_1' \frac{\partial\beta_1}{\partial\theta}(\theta') = \kappa\sin\beta_1'.$$

We likewise have

$$\frac{\partial^2\phi}{\partial\theta\,\partial x_2}\phi(x_1', x_2', \theta') = -\sin\beta_2' \frac{\partial\beta_2}{\partial\theta}(\theta'),$$

and we shall express $\frac{\partial\beta_2}{\partial\theta}(\theta')$ in terms of κ and h. We recall that

$$k(\beta_1')a(x_1')^{-1}ga(x_2') = a(h)k(\theta_0)$$

for some θ_0, and so

$$k(\theta)a(x_1')k(-\beta_1')a(h)k(\theta_0) = k(\theta)ga(x_2').$$

Substituting both parts of (48) into this gives

$$\begin{aligned} NAk\big(\beta_1(\theta)\big)k(-\beta_1')a(h)k(\theta_0) &= NAk\big(\beta_2(\theta)\big), \\ k\big(\beta_1(\theta) - \beta_1'\big)a(h) &\in NAk\big(\beta_2(\theta) - \theta_0\big). \end{aligned}$$

By setting $\theta = \theta'$ we see that $\theta_0 = \beta_1'$. Equation (33) then gives

$$e^h \tan\big(\big(\beta_1(\theta) - \beta_1'\big)/2\big) = \tan\big(\big(\beta_2(\theta) - \beta_2'\big)/2\big),$$

and differentiating both sides with respect to θ and evaluating at $\theta = \theta'$ gives

$$\frac{\partial\beta_2}{\partial\theta}(\theta') = \kappa e^h.$$

It follows that

$$\frac{\partial^2\phi}{\partial\theta\,\partial x_2}\phi(x_1', x_2', \theta') = -\kappa e^h \sin\beta_2'.$$

To calculate $\partial^2\phi/\partial\theta^2$, we have as in (47) that

$$\frac{\partial\phi}{\partial\theta}(x_1', x_2', \theta) = \frac{\partial\beta_1}{\partial\theta}\frac{\partial}{\partial\beta} A\big(k(\beta)a(x_1')^{-1}ga(x_2')\big)\Big|_{\beta=\beta_1(\theta)}.$$

Because

$$\frac{\partial}{\partial\beta}A\big(k(\beta)a(x_1')^{-1}ga(x_2')\big)\Big|_{\beta=\beta_1'}=0,$$

we have

$$\begin{aligned}\frac{\partial^2\phi}{\partial\theta^2}(x_1',x_2',\theta')&=\kappa^2\frac{\partial^2}{\partial\beta^2}A\big(k(\beta)a(x_1')^{-1}ga(x_2')\big)\Big|_{\beta=\beta_1'}\\&=\kappa^2\frac{\partial^2}{\partial\beta^2}A\big(k(\beta-\beta_1')a(h)\big)\Big|_{\beta=\beta_1'}.\end{aligned}$$

It is a standard calculation that

$$\frac{\partial^2}{\partial\beta^2}A\big(k(\beta)a(h)\big)\Big|_{\beta=0}=(1-e^{2h})/2,$$

and this completes the proof. □

7.2. The function ψ

Define $\mathcal{P}=\mathbb{R}/2\pi\mathbb{Z}\times G\times[\delta,1-\delta]$, and define $\mathcal{S}\subset\mathcal{P}$ to be the set where one of the geodesics $k(\theta)\ell$ and $k(\theta)g\ell$ is vertical. Note that $\mathcal{S}$ is closed and contains at most 4 values of θ for each fixed (g,ρ). We may define functions

$$\xi_1,\xi_2:\mathcal{P}\setminus\mathcal{S}\to\mathbb{R}$$

by requiring that $k(\theta)\ell(\xi_1(\theta,g,\rho))$ is the unique point on $k(\theta)\ell$ at which the tangent vector to the geodesic makes an angle of α with the upward-pointing vector, and likewise for $\xi_2(\theta,g,\rho)$ and $k(\theta)g\ell$. As ξ_1 does not depend on g, we will omit this argument of the function. We have

$$k(\theta)a\big(\xi_1(\theta,\rho)\big)\in NAk(\epsilon_1\alpha),\qquad k(\theta)ga\big(\xi_2(\theta,g,\rho)\big)\in NAk(\epsilon_2\alpha)\tag{50}$$

for $\epsilon_i\in\{\pm1\}$, and so equation (33) gives

$$e^{\xi_1(\theta,\rho)}\tan(\theta/2)=\tan(\epsilon_1\alpha/2),\qquad e^{\xi_2(\theta,g,\rho)}\tan(\theta/2)=\tan(\epsilon_2\alpha/2).$$

Moreover, it may be seen that $\epsilon_1=1$ if and only if the geodesic $k(\theta)\ell$ runs from right to left in the upper half-plane model of $\mathbb{H}$, which is equivalent to $\theta\in(0,\pi)$, and likewise for ϵ_2.

It follows from Lemma 6.3 that $\xi_1(\theta,\rho)$ and $\xi_2(\theta,g,\rho)$ may also be characterized as the unique functions such that

$$\frac{\partial\phi}{\partial x_1}\big(\xi_1(\theta,\rho),x_2,\theta\big)=\frac{\partial\phi}{\partial x_2}\big(x_1,\xi_2(\theta,g,\rho),\theta\big)=0.\tag{51}$$

We define

$$\psi : \mathcal{P} \setminus \mathcal{S} \to \mathbb{R},$$
$$\psi(\theta, g, \rho) = \phi\big(\xi_1(\theta, \rho), \xi_2(\theta, g, \rho), \theta, g, \rho\big).$$

LEMMA 7.3
(θ', g', ρ') is a critical point of ψ exactly when $(\xi_1(\theta'), \xi_2(\theta'), \theta', g', \rho')$ is a critical point of ϕ. If (θ', g', ρ') is a critical point of ψ, let κ and h be the values associated to the corresponding critical point of ϕ. We then have

$$\frac{\partial^2 \psi}{\partial \theta^2}(\theta', g', \rho') = -\frac{3}{2}\kappa^2(1 - e^{2h}).$$

Proof
We shall fix g and ρ and omit them from the arguments of ϕ and ψ. Let D be the Hessian of ϕ calculated in Proposition 7.2. If we apply the chain rule to ψ and substitute $\theta = \theta'$, we obtain

$$\frac{\partial^2 \psi}{\partial \theta^2}(\theta') = \Big(\frac{\partial \xi_1}{\partial \theta}(\theta'), \frac{\partial \xi_2}{\partial \theta}(\theta'), 1\Big) D \Big(\frac{\partial \xi_1}{\partial \theta}(\theta'), \frac{\partial \xi_2}{\partial \theta}(\theta'), 1\Big)^t.$$

To calculate $\frac{\partial \xi_1}{\partial \theta}(\theta')$ and $\frac{\partial \xi_2}{\partial \theta}(\theta')$, we differentiate (51) with respect to θ and set $\theta = \theta'$ to obtain

$$\frac{\partial^2 \phi}{\partial \theta\, \partial x_1}\big(\xi_1(\theta'), \xi_2(\theta'), \theta'\big) + \frac{\partial \xi_1}{\partial \theta}(\theta') \frac{\partial^2 \phi}{\partial x_1^2}\big(\xi_1(\theta'), \xi_2(\theta'), \theta'\big) = 0. \tag{52}$$

Substituting the second partial derivatives of ϕ calculated in Proposition 7.2 gives

$$\frac{\partial \xi_1}{\partial \theta}(\theta') = \frac{-2\kappa}{\sin \beta_1'},$$

and likewise

$$\frac{\partial \xi_2}{\partial \theta}(\theta') = \frac{-2\kappa e^h}{\sin \beta_2'}.$$

The lemma follows on substituting these into (52). ☐

It follows that the set of (g, ρ) for which the function $\psi(\theta, g, \rho)$ has a degenerate critical point is exactly those (g, ρ) for which either $\ell = g\ell$ as oriented geodesics, so that $g \in A$, or $\angle(\ell, g\ell) = \pm 2\alpha$. Note that these two cases are distinct, as $\alpha \in (0, \pi/2)$. In the first case the function $\psi(\theta, g, \rho)$ vanishes identically. In the second case, $\psi(\theta, g, \rho)$ has only a single degenerate critical point, as no oriented geodesic can cross ℓ and $g\ell$ making an angle of α with both except at their point of intersection. We will show in Section 7.4 that this degeneracy is cubic. To determine the location of this critical point, the condition that $\angle(\ell, g\ell) = \pm 2\alpha$ implies that $g \in a(y)k(\pm 2\alpha)A$ for some $y \in \mathbb{R}$, so that $\ell \cap g\ell = \ell(y)$. The angle bisector of the two geodesics at the point $\ell(y)$ is $a(y)k(\pm\alpha)\ell$, and the critical point of $\psi(\theta, g, \rho)$ is the θ such that the positive endpoint of $k(\theta)a(y)k(\pm\alpha)\ell$ is $i\infty$. This is equivalent to the condition $k(\theta)a(y)k(\pm\alpha) \in NA$, and (33) then gives $\cot\theta/2 = \mp e^y \cot\alpha/2$.

We define

$$\mathcal{D}_1 = \{(\theta, g, \rho) \in \mathcal{P} \setminus \mathcal{S} \mid g \in A\},$$
$$\mathcal{D}_2^\pm = \{(\theta, g, \rho) \in \mathcal{P} \setminus \mathcal{S} \mid g \in a(y)k(\pm 2\alpha)A, \cot\theta/2 = \mp e^y \cot\alpha/2\}$$

to be the three sets on which ψ has a degenerate critical point. We also define $\overline{\mathcal{P}} = G \times [\delta, 1-\delta]$ and define

$$\overline{\mathcal{D}}_1 = A \times [\delta, 1-\delta],$$
$$\overline{\mathcal{D}}_2^\pm = \{(g, \rho) \in \overline{\mathcal{P}} \mid g \in Ak(\pm 2\alpha)A\}$$

to be the projections of $\mathcal{D}_1$ and $\mathcal{D}_2^\pm$ to $\overline{\mathcal{P}}$.

7.3. *The degenerate set* $\mathcal{D}_1$

As $\psi(\theta, g, \rho) = \psi(\theta, ga, \rho)$ for $a \in A$, we may study the degeneracy of ψ near $\mathcal{D}_1$ by differentiating $\psi(\theta, \exp(X), \rho)$ at $X = 0$ as in the following proposition.

PROPOSITION 7.4
If $X = \begin{pmatrix} 0 & X_1 \\ X_2 & 0 \end{pmatrix} \in \mathfrak{g}$, *then*

$$\frac{\partial}{\partial t}\psi\big(\theta, \exp(tX), \rho\big)\Big|_{t=0} = \epsilon \sin\alpha (e^{-\xi_2(\theta,e,\rho)} X_1 + e^{\xi_2(\theta,e,\rho)} X_2), \tag{53}$$

where ϵ *is* 1 *if* $\theta \in (0, \pi)$ *and* -1 *otherwise. In particular,* $\partial\psi/\partial t(\theta, \exp(tX), \rho)|_{t=0}$ *has no degenerate critical points unless* $X = 0$.

Proof
We have

$$\begin{aligned}\frac{\partial}{\partial t}\psi\big(\theta, \exp(tX), \rho\big)\Big|_{t=0} &= \frac{\partial}{\partial t}\phi\big(\xi_1(\theta,\rho), \xi_2\big(\theta, \exp(tX), \rho\big), \theta, \exp(tX), \rho\big)\Big|_{t=0} \\ &= \frac{\partial \phi}{\partial x_2}\big(\xi_1(\theta,\rho), \xi_2(\theta, e, \rho), \theta, e, \rho\big)\frac{\partial}{\partial t}\xi_2\big(\theta, \exp(tX), \rho\big)\Big|_{t=0} \\ &\quad + \frac{\partial}{\partial t}\phi\big(\xi_1(\theta,\rho), \xi_2(\theta, e, \rho), \theta, \exp(tX), \rho\big)\Big|_{t=0}.\end{aligned}$$

The first term vanishes by (51), so we are left with

$$\begin{aligned}\frac{\partial}{\partial t}\psi\big(\theta,\exp(tX),\rho\big)\Big|_{t=0} &= \frac{\partial}{\partial t}\phi\big(\xi_1(\theta,\rho),\xi_2(\theta,e,\rho),\theta,\exp(tX),\rho\big)\Big|_{t=0}\\ &= \frac{\partial}{\partial t}A\big(k(\theta)\exp(tX)a\big(\xi_2(\theta,e,\rho)\big)\big)\Big|_{t=0}.\end{aligned}$$

We shall abbreviate $\xi_2(\theta,e,\rho)$ to $\xi_2(\theta)$ for the remainder of the proof. Write the first-order approximation to the Iwasawa decomposition of $k(\theta)\exp(tX)a(\xi_2(\theta))$ as

$$\begin{aligned}&k(\theta)\exp(tX)a\big(\xi_2(\theta)\big)\\ &\quad= n\exp\big(tX_N+O(t^2)\big)a\exp\big(tX_A+O(t^2)\big)k\exp\big(tX_K+O(t^2)\big),\end{aligned}$$

where $X_N\in\mathfrak{n}$, $X_A\in\mathfrak{a}$, and $X_K\in\mathfrak{k}$. As in equation (50), we have $k=k(\alpha)$ if $\theta\in(0,\pi)$ and $k=k(-\alpha)$ if $\theta\in(-\pi,0)$. We first assume that $\theta\in(0,\pi)$. Rearranging and equating first-order terms gives

$$\begin{aligned}X &= \mathrm{Ad}\big(a\big(\xi_2(\theta)\big)k(\alpha)^{-1}a^{-1}\big)X_N+\mathrm{Ad}\big(a\big(\xi_2(\theta)\big)k(\alpha)^{-1}\big)X_A\\ &\qquad+\mathrm{Ad}\big(a\big(\xi_2(\theta)\big)\big)X_K,\\ \mathrm{Ad}\big(k(\alpha)a\big(\xi_2(\theta)\big)^{-1}\big)X &= \mathrm{Ad}(a^{-1})X_N+X_A+\mathrm{Ad}\big(k(\alpha)\big)X_K.\end{aligned}$$

As $\mathrm{Ad}(a^{-1})X_N$ and $\mathrm{Ad}(k(\alpha))X_K$ lie in $\mathfrak{a}^\perp\subset\mathfrak{g}$, we see that X_A is the orthogonal projection of $\mathrm{Ad}(k(\alpha)a(\xi_2(\theta))^{-1})X$ to $\mathfrak{a}$. A calculation gives

$$X_A=\sin\alpha(e^{-\xi_2(\theta)}X_1+e^{\xi_2(\theta)}X_2)\begin{pmatrix}1/2 & 0\\ 0 & -1/2\end{pmatrix},$$

so that

$$\frac{\partial}{\partial t}A\big(k(\theta)\exp(tX)a\big(\xi_2(\theta)\big)\big)\Big|_{t=0}=\sin\alpha(e^{-\xi_2(\theta)}X_1+e^{\xi_2(\theta)}X_2).$$

This proves (53) when $\theta\in(0,\pi)$, and the other case is identical.

We now prove that $\partial\psi/\partial t(\theta,\exp(tX),\rho)|_{t=0}$ has no degenerate critical points if $X\neq 0$ and $\theta\in(0,\pi)$. We define $f(x)=\sin\alpha(X_1e^{-x}+X_2e^x)$, so that

$$\frac{\partial}{\partial t}A\big(k(\theta)\exp(tX)a\big(\xi_2(\theta)\big)\big)\Big|_{t=0}=f\big(\xi_2(\theta)\big).$$

Differentiating equation (33) gives

$$\frac{\partial\xi_2}{\partial\theta}=-\frac{1}{2}e^{-\xi_2(\theta)}\tan(\alpha/2)\csc^2(\theta/2),$$

so that $\partial\xi_2/\partial\theta$ is always nonzero. Suppose that $X\neq 0$ and that θ is a degenerate critical point of $\partial\psi/\partial t(\theta,\exp(tX),\rho)|_{t=0}$. We then have

$$\begin{aligned} 0 &= \frac{\partial^2}{\partial\theta\,\partial t} A\big(k(\theta)\exp(tX)a\big(\xi_2(\theta)\big)\big)\Big|_{t=0} \\ &= f'\big(\xi_2(\theta)\big)\frac{\partial \xi_2}{\partial\theta} \\ &= f'\big(\xi_2(\theta)\big). \end{aligned}$$

Differentiating again with respect to θ gives

$$\begin{aligned} 0 &= \frac{\partial^3}{\partial^2\theta\,\partial t} A\big(k(\theta)\exp(tX)a\big(\xi_2(\theta)\big)\big)\Big|_{t=0} \\ &= f''\big(\xi_2(\theta)\big)\Big(\frac{\partial \xi_2}{\partial\theta}\Big)^2 \\ &= f''\big(\xi_2(\theta)\big), \end{aligned}$$

but this is a contradiction as it may be easily checked that f has no degenerate critical points unless $X = 0$. The case of $\theta \in (-\pi, 0)$ is identical. □

Define $P = \mathbb{R}/2\pi\mathbb{Z} \times \mathfrak{a}^{\perp} \times [\delta, 1-\delta]$, and define $S = \{(\theta, X, \rho) \in P \mid (\theta, \exp(X), \rho) \in \mathcal{S}\}$. S is again closed and contains at most 4 values of θ for each fixed (X, ρ).

LEMMA 7.5
There is an open neighborhood $0 \in U \subset \mathfrak{a}^{\perp}$ such that for all $X \in U$, all $\rho \in [\delta, 1-\delta]$, and all $b \in C_0^{\infty}(P \setminus S)$, we have

$$\int b(\theta, X, \rho) e^{is\psi(\theta, \exp(X), \rho)}\, d\theta \ll \big(1 + s\|X\|\big)^{-1/2},$$

where $\|X\|$ is as in (8). The implied constant depends only on δ, $\|b\|_\infty$, and $\|\partial b/\partial\theta\|_\infty$.

Proof
Define the map $X : \mathbb{R} \times \mathbb{R}/2\pi\mathbb{Z} \to \mathfrak{a}^{\perp}$ by

$$X(r, \gamma) = \begin{pmatrix} 0 & r\sin\gamma \\ r\cos\gamma & 0 \end{pmatrix}.$$

We define

$$\widetilde{P} = \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R} \times \mathbb{R}/2\pi\mathbb{Z} \times [\delta, 1-\delta]$$

and

$$\widetilde{S} = \big\{(\theta, r, \gamma, \rho) \subset \widetilde{P} \;\big|\; \big(\theta, X(r, \gamma), \rho\big) \in S\big\}.$$

We define $\widetilde{b}(\theta, r, \gamma, \rho) \in C_0^\infty(\widetilde{P} \setminus \widetilde{S})$ and $\widetilde{\psi}(\theta, r, \gamma, \rho) \in C^\infty(\widetilde{P} \setminus \widetilde{S})$ to be the pullbacks of b and ψ under X. We know that $\widetilde{\psi}$ vanishes when $r = 0$, and as $\widetilde{\psi}$ is smooth (in fact, real analytic) we have that $\widetilde{\psi}/r$ is again a smooth function. Proposition 7.4 implies that $\widetilde{\psi}/r$ has no degenerate critical points when $r = 0$, and so there is some $\epsilon > 0$ such that it also has no degenerate critical points on the set $\mathrm{supp}(\widetilde{b}) \cap (\mathbb{R}/2\pi\mathbb{Z} \times [-\epsilon, \epsilon] \times \mathbb{R}/2\pi\mathbb{Z} \times [\delta, 1-\delta])$. If we define $U = X((-\epsilon, \epsilon) \times \mathbb{R}/2\pi\mathbb{Z})$, the result now follows from the van der Corput lemma (see, e.g., [26, Proposition 2, Section 1.2, Chapter VIII] and the corollary there). $\square$

COROLLARY 7.6

If $(a', \rho') \in \overline{\mathcal{D}}_1$, there is an open neighborhood $(a', \rho') \in U \subset \overline{\mathcal{P}}$ such that for all $b \in C_0^\infty(\mathcal{P} \setminus \mathcal{S})$ and all $(g, \rho) \in U$, we have

$$\int_0^{2\pi} b(\theta, g, \rho) e^{is\psi(\theta, g, \rho)}\, d\theta \ll \big(1 + sd(g, A)\big)^{-1/2}.$$

The implied constant depends only on δ, (a', ρ'), $\|b\|_\infty$, and $\|\partial b/\partial\theta\|_\infty$.

Proof

Let $U_X \subset \mathfrak{a}^\perp$ be as in Lemma 7.5. If $g = \exp(X)a(x)$ for $X \in U_X$ and $|x| \le B$, we have $d(g, A) \sim \|X\|$, where the implied constants depend only on B. As $\psi(\theta, ga, \rho) = \psi(\theta, g, \rho)$ for $a \in A$, the result follows from Lemma 7.5. $\square$

7.4. The degenerate set $\mathcal{D}_2^\pm$

The next proposition proves that ψ has a cubic degeneracy on $\mathcal{D}_2^\pm$.

PROPOSITION 7.7

If $(\theta', g', \rho') \in \mathcal{D}_2^\pm$, we have $\partial^3\psi/\partial\theta^3(\theta', g', \rho') \ne 0$.

Proof

Suppose that $g' = a(y)k(2\alpha)a_2$. Let $\epsilon > 0$, and define $g = a(y)k(2\alpha + \epsilon)a_2$. If ϵ is chosen sufficiently small, the pair $(\ell, g\ell)$ will have exactly two critical geodesics ℓ_1 and ℓ_2 as shown in Figure 1. The triangles AB_1C_1 and AB_2C_2 both have angular defect, and hence area, ϵ. Our assumption that α was bounded away from 0 and $\pi/2$ then implies that $AB_1 = AB_2 \sim \epsilon^{1/2}$ and $B_1C_1 = B_2C_2 \sim \epsilon^{1/2}$, where the implied constants depend only on δ. The critical points θ_i corresponding to ℓ_i are the solutions to

$$\cot\theta_1/2 = -e^{y+AB_1}\cot\alpha/2, \qquad \cot\theta_2/2 = -e^{y-AB_1}\cot\alpha/2.$$

It follows that $0 > \theta_1 > -\alpha > \theta_2 > -\pi$ and also that $\theta_1 - \theta_2 \sim \epsilon^{1/2}$. The apertures h_i of the critical points θ_i are given by $h_1 = -B_1C_1 \sim -\epsilon^{1/2}$ and $h_2 = B_2C_2 \sim \epsilon^{1/2}$,

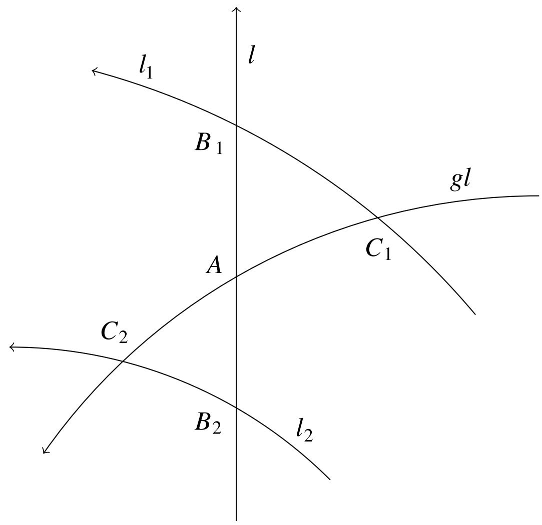

Figure 1. Two degenerating critical geodesics.

so that Lemma 7.3 gives

$$\frac{\partial^2\psi}{\partial\theta^2}(\theta_1, g, \rho') \sim -\epsilon^{1/2}, \qquad \frac{\partial^2\psi}{\partial\theta^2}(\theta_2, g, \rho') \sim \epsilon^{1/2}.$$

It follows that there is $\theta_0 \in [\theta_2, \theta_1]$ at which

$$\frac{\partial^3\psi}{\partial\theta^3}(\theta_0, g, \rho') = \frac{\frac{\partial^2\psi}{\partial\theta^2}(\theta_2, g, \rho') - \frac{\partial^2\psi}{\partial\theta^2}(\theta_1, g, \rho')}{\theta_2 - \theta_1} \sim -1,$$

and shrinking ϵ to 0 gives the result. The case $g' \in Ak(-2\alpha)A$ is identical. □

COROLLARY 7.8

If $(g', \rho') \in \overline{\mathcal{D}}_2^{\pm}$, there is an open neighborhood $(g', \rho') \in U \subset \overline{\mathcal{P}}$ such that for all $b \in C_0^\infty(\mathcal{P} \setminus \mathcal{S})$ and all $(g, \rho) \in U$, we have

$$\int_0^{2\pi} b(\theta, g, \rho) e^{is\psi(\theta, g, \rho)}\, d\theta \ll s^{-1/3}.$$

The implied constant depends only on δ, (g', ρ'), $\|b\|_\infty$, and $\|\partial b/\partial\theta\|_\infty$.

Proof

By Proposition 7.7, there exist $\sigma > 0$ and neighborhoods U_θ of θ' and U of (g', ρ'), all depending on δ and (g', ρ'), such that $(U_\theta \times U) \cap \mathcal{S} = \emptyset$ and $|\partial^3\psi/\partial\theta^3| \geq \sigma > 0$ on $U_\theta \times U$. As $\psi(\theta, g', \rho')$ only has a critical point at θ', by shrinking U we may also assume that ψ has no critical points on $(\mathbb{R}/2\pi\mathbb{Z} \setminus U_\theta) \times U \setminus \mathcal{S}$. The result then follows from the van der Corput lemma. □

7.5. Bounds for $J(s,t,g)$

We shall use the results of the previous sections to prove the following proposition, which implies Proposition 3.5 in the case $t/\lambda \in [\delta, 1-\delta]$ after inverting the Harish-Chandra transform.

PROPOSITION 7.9

If $g \in G$ satisfies $d(g,e) \le 1$ and $t/s \in [\delta, 1-\delta]$, we have

$$J(s,t,g) \ll_\delta \begin{cases} s^{-1}(1+sd(g,A))^{-1/2} & \text{when } d(g,A) \le s^{-1/3}, \\ s^{-4/3} & \text{when } d(g,A) \ge s^{-1/3}. \end{cases}$$

Proof

If we substitute the expression (43) into the definition of $J(s,t,g)$, we obtain

$$\int_0^{2\pi} \iint_{-\infty}^{\infty} b_0(x_1) b_0(x_2) e^{-it(x_1-x_2)} \times \exp\Big((1/2-is)\Big(A\big(k(\theta)ga(x_2)\big) - A\big(k(\theta)a(x_1)\big)\Big)\Big) \frac{d\sigma}{d\theta}\, dx_1\, dx_2\, d\theta.$$

We let $b \in C_0^\infty(G)$ be a function that is equal to 1 when $d(g,e) \le 1$ and introduce a factor of $b(g)$ into the integral. When $d(g,e) \le 1$ we then have

$$J(s,t,g) = \int_0^{2\pi} \iint_{-\infty}^{\infty} e^{-is\phi(x_1,x_2,\theta,g,\rho)} c(x_1,x_2,\theta,g,\rho)\, dx_1\, dx_2\, d\theta,$$

where $c \in C_0^\infty(\mathbb{R}^2 \times \mathcal{P})$ is the combination of all of the amplitude factors. The following lemma eliminates the variables x_1 and x_2.

LEMMA 7.10

There is a function $c_1 \in C_0^\infty(\mathcal{P} \setminus \mathcal{S})$ such that for all $(g,\rho) \in G \times [\delta, 1-\delta]$ we have

$$\int_0^{2\pi} \iint_{-\infty}^{\infty} e^{-is\phi(x_1,x_2,\theta,g,\rho)} c(x_1,x_2,\theta,g,\rho)\, dx_1\, dx_2\, d\theta = s^{-1} \int_0^{2\pi} e^{-is\psi(\theta,g,\rho)} c_1(\theta,g,\rho)\, d\theta + O(s^{-2}),$$

where the implied constant depends on δ and on (g,ρ) in a locally uniform manner.

Proof
We shall apply stationary phase in the x_i variables. For fixed (θ, g, ρ), the function $\phi(x_1, x_2)$ has one critical point at $(\xi_1(\theta,\rho), \xi_2(\theta, g, \rho))$ if $(\theta, g, \rho) \notin \mathcal{S}$, and none otherwise. Moreover, it may be shown in the same way as the proof of Proposition 7.2 that the Hessian at this critical point is

$$D = \begin{pmatrix} \frac{1}{2}\sin^2\alpha & 0 \\ 0 & -\frac{1}{2}\sin^2\alpha \end{pmatrix},$$

so that the critical point is uniformly nondegenerate.

Define

$$\mathcal{P}_0 = \big\{(\theta, g, \rho) \in \mathcal{P} \setminus \mathcal{S} \mid \big(\xi_1(\theta,\rho), \xi_2(\theta, g, \rho), \theta, g, \rho\big) \in \operatorname{supp}(c)\big\},$$

so that $\mathcal{P}_0$ is compact and $\mathcal{P}_0 \cap \mathcal{S} = \emptyset$. If we define $c_1 \in C_0^\infty(\mathcal{P} \setminus \mathcal{S})$ by

$$c_1(\theta, g, \rho) = \frac{2\pi}{\sin^2\alpha} c\big(\xi_1(\theta,\rho), \xi_2(\theta, g, \rho), \theta, g, \rho\big),$$

then we have $\operatorname{supp}(c_1) \subseteq \mathcal{P}_0$, and stationary phase gives

$$\begin{aligned} &\iint_{-\infty}^{\infty} e^{-is\phi(x_1,x_2,\theta)} c(x_1, x_2, \theta, g, \rho)\, dx_1\, dx_2 \\ &\quad = e^{-is\psi(\theta,g,\rho)} s^{-1} c_1(\theta, g, \rho) + O_\delta(s^{-2}) \end{aligned} \tag{54}$$

locally uniformly on $\mathcal{P} \setminus \mathcal{S}$. We also have

$$\iint_{-\infty}^{\infty} e^{-is\phi(x_1,x_2,\theta)} c(x_1, x_2, \theta, g, \rho)\, dx_1\, dx_2 \ll_{A,\delta} s^{-A}$$

locally uniformly on $\mathcal{P} \setminus \mathcal{P}_0$. Therefore, if we extend c_1 to a function in $C^\infty(\mathcal{P})$ by making it 0 on $\mathcal{S}$, then (54) holds locally uniformly on $\mathcal{P}$ and the lemma follows. □

We now apply Corollaries 7.6 and 7.8. Let $B = \{g \in G : d(g, e) \le 1\}$. Corollary 7.6 implies that there is an open neighborhood U_1 of $\overline{\mathcal{D}}_1 \cap (B \times [\delta, 1-\delta])$ in $\overline{\mathcal{P}}$ such that

$$J(s, t, g) \ll_\delta s^{-1}\big(1 + sd(g, A)\big)^{-1/2}$$

when $(g, \rho) \in U_1 \cap (B \times [\delta, 1-\delta])$, and Corollary 7.8 implies that there is a neighborhood U_2 of $\overline{\mathcal{D}}_2^{\pm} \cap (B \times [\delta, 1-\delta])$ such that $J(s, t, g) \ll_\delta s^{-4/3}$ when $(g, \rho) \in U_2 \cap (B \times [\delta, 1-\delta])$. As ψ has no degenerate critical points outside $\overline{\mathcal{D}}_1 \cup \overline{\mathcal{D}}_2^{\pm}$, we also have $J(s, t, g) \ll_\delta s^{-3/2}$ when $(g, \rho) \in (B \times [\delta, 1-\delta]) \setminus (U_1 \cup U_2)$. As the bound in Proposition 7.9 is the maximum of these three bounds, this completes the proof. □

It remains to discuss the case when $t = 0$, so that $\alpha = \pi/2$. The proof proceeds as before, until the analysis of the degenerate critical points of ψ. These degeneracies now occur when

$$g \in A \cup \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} A,$$

and the function ψ vanishes identically at these points. These degeneracies may be treated in exactly the same way as $\mathcal{D}_1$ in Section 7.3, which gives the bound

$$J(s,0,g) \ll s^{-1}\big(1 + sd(g,A)\big)^{-1/2}$$

when $d(g,e) \leq 1$. Inverting the Harish-Chandra transform completes the proof.

Acknowledgments. We would like to thank Xiaoqing Li, Peter Sarnak, Christopher Sogge, Nicolas Templier, Akshay Venkatesh, and Steve Zelditch for many helpful discussions. We would also like to thank the referees for many helpful comments.

The author's work was supported by National Science Foundation grant DMS-1501230.

References

[1] G. V. AVAKUMOVIĆ, *Über die Eigenfunktionen auf geschlossenen Riemannschen Mannigfaltigkeiten*, Math. Z. **65** (1956), 327–344. MR 0080862. *(463)*

[2] M. D. BLAIR and C. D. SOGGE, *Refined and microlocal Kakeya–Nikodym bounds for eigenfunctions in two dimensions*, Anal. PDE **8** (2015), 747–764. MR 3353830. DOI 10.2140/apde.2015.8.747. *(464)*

[3] V. BLOMER and F. BRUMLEY, *On the Ramanujan conjecture over number fields*, Ann. of Math. (2) **174** (2011), 581–605. MR 2811610. DOI 10.4007/annals.2011.174.1.18. *(466)*

[4] V. BLOMER and R. HOLOWINSKY, *Bounding sup-norms of cusp forms of large level*, Invent. Math. **179** (2010), 645–681. MR 2587342. DOI 10.1007/s00222-009-0228-0. *(464)*

[5] V. BLOMER, R. KHAN, and M. YOUNG, *Distribution of mass of holomorphic cusp forms*, Duke Math. J. **162** (2013), 2609–2644. MR 3127809. DOI 10.1215/00127094-2380967. *(464)*

[6] V. BLOMER and P. MÁGA, *The sup-norm problem for $PGL(4)$*, Int. Math. Res. Not. IMRN **2015**, no. 14, 5311–5332. MR 3384442. DOI 10.1093/imrn/rnu100. *(464)*

[7] ———, *Subconvexity for sup-norms of automorphic forms on $PGL(n)$*, preprint, arXiv:1405.6691v1 [math.NT]. *(464)*

[8] V. BLOMER and PH. MICHEL, *Sup-norms of eigenfunctions on arithmetic ellipsoids*, Int. Math. Res. Not. IMRN **2011**, no. 21, 4934–4966. MR 2852302. DOI 10.1093/imrn/rnq280. *(464)*

[9] V. BLOMER and A. POHL, *The sup-norm problem on the Siegel modular space of rank two*, preprint, arXiv:1402.4635v2 [math.NT]. *(464)*

[10] J. BOURGAIN, "Geodesic restrictions and L^p estimates for eigenfunctions of Riemannian surfaces" in *Linear and Complex Analysis*, Amer. Math. Soc. Transl., Adv. Math. Sci., 2009, 27–35. MR 2500507. *(465)*

[11] N. BURQ, P. GÉRARD, and N. TZVETKOV, *Restrictions of the Laplace-Beltrami eigenfunctions to submanifolds*, Duke Math. J. **138** (2007), 445–486. MR 2322684. DOI 10.1215/S0012-7094-07-13834-1. *(463, 483)*

[12] X. CHEN, *An improvement on eigenfunction restriction estimates for compact boundaryless Riemannian manifolds with nonpositive sectional curvature*, Trans. Amer. Math. Soc. **367** (2015), 4019–4039. MR 3324918. DOI 10.1090/S0002-9947-2014-06158-8. *(464)*

[13] X. CHEN and C. SOGGE, *A few endpoint geodesic restriction estimates for eigenfunctions*, Comm. Math. Phys. **329** (2014), 435–459. MR 3210140. DOI 10.1007/s00220-014-1959-3. *(464)*

[14] ———, *On integrals of eigenfunctions over geodesics*, Proc. Amer. Math. Soc. **143** (2015), 151–161. MR 3272740. DOI 10.1090/S0002-9939-2014-12233-7. *(465)*

[15] M. EICHLER, *Lectures on Modular Correspondences*, Tata Inst. Fund. Res. Lect. Math. **9**, Tata Inst. Fund. Res., Mumbai, 1955. *(468)*

[16] R. GANGOLLI, *On the plancherel formula and the Paley–Wiener theorem for spherical functions on semisimple Lie groups*, Ann. of Math. (2) **93** (1971), 105–165. MR 0289724. *(470)*

[17] R. GANGOLLI and V. S. VARADARAJAN, *Harmonic analysis of spherical functions on real reductive groups*, Ergeb. Math. Grenzgeb. **101**, Springer, Berlin, 1988. MR 0954385. DOI 10.1007/978-3-642-72956-0. *(470)*

[18] A. GHOSH, A. REZNIKOV, and P. SARNAK, *Nodal domains of Maass forms, I*, Geom. Funct. Anal. **23** (2013), 1515–1568. MR 3102912. DOI 10.1007/s00039-013-0237-4. *(467)*

[19] G. HARCOS and N. TEMPLIER, *On the sup-norm of Maass cusp forms of large level, II*, Int. Math. Res. Not. IMRN **2011**, no. 20, 4764–4774. MR 2989618. *(464)*

[20] H. IWANIEC and P. SARNAK, *L^∞ norms of eigenfunctions of arithmetic surfaces*, Ann. of Math. (2) **141** (1995), 301–320. MR 1324136. DOI 10.2307/2118522. *(464, 466, 467, 473, 476, 483)*

[21] B. M. LEVITAN, *On the asymptoptic behavior of the spectral function of a self-adjoint differential equation of second order*, Isv. Akad. Nauk SSSR Ser. Mat. **16** (1952), 325–352. MR 0058067. *(463)*

[22] S. MARSHALL, *Restrictions of SL_3 Maass forms to maximal flat subspaces*, Int. Math. Res. Not. IMRN **2015**, no. 16, 6988–7015. DOI 10.1093/imrn/rnu155. *(465)*

[23] ———, *L^p bounds for higher rank eigenfunctions and asymptotics of spherical functions*, to appear in J. Eur. Math. Soc., preprint, arXiv:1106.0534. *(465, 483, 484)*

[24] K. MARTIN and D. WHITEHOUSE, *Central L-values and toric periods for GL_2*, Int. Math. Res. Not. IMRN **2009**, no. 1, 141–191. MR 2471298. DOI 10.1093/imrn/rnn127. *(467)*

[25] P. MICHEL and A. VENKATESH, *The subconvexity problem for GL_2*, Publ. Math. Inst. Hautes Études Sci. **111** (2010), 171–271. MR 2653249. DOI 10.1007/s10240-010-0025-8. *(467)*

[26] T. MURPHY and E. STEIN, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*. Princeton Univ. Press, Princeton, 1993. MR 1232192. *(502)*

[27] A. REZNIKOV, *A uniform bound for geodesic periods of eigenfunctions on hyperbolic surfaces*, Forum Math. **27** (2015), 1569–1590. MR 3341487. DOI 10.1515/forum-2012-0185. *(465)*

[28] P. SARNAK, *Spectra of hyperbolic surfaces*, Bull. Amer. Math. Soc. **40** (2003), 441–478. MR 1997348. DOI 10.1090/S0273-0979-03-00991-1. *(464)*

[29] A. SELBERG, "Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces, with applications to Dirichlet series" in *Collected Papers, Vol. 1*, Springer, New York, 1989. *(470)*

[30] C. SOGGE, *Concerning the L^p norm of spectral clusters of second order elliptic operators on a compact manifold*, J. Funct. Anal. **77** (1988), 123–138. MR 0930395. DOI 10.1016/0022-1236(88)90081-X. *(463)*

[31] C. SOGGE, J. TOTH, and S. ZELDITCH, *About the blowup of quasimodes on Riemannian manifolds*, J. Geom. Anal. **21** (2009), 150–173. MR 2755680. DOI 10.1007/s12220-010-9168-6. *(464)*

[32] C. SOGGE and S. ZELDITCH, *Riemannian manifolds with maximal eigenfunction growth*, Duke Math. J. **114** (2002), 387–437. MR 1924569. DOI 10.1215/S0012-7094-02-11431-8. *(464)*

[33] ———, "On eigenfunction restriction estimates and L^4-bounds for compact surfaces with nonpositive curvature" in *Advances in Analysis: The Legacy of Elias M. Stein*, Princeton Univ. Press, Princeton, 2014, 447–462. *(464)*

[34] N. TEMPLIER, *On the sup-norm of Maass cusp forms of large level*, Selecta Math. **16** (2010), 501–531. MR 2734340. DOI 10.1007/s00029-010-0026-y. *(464)*

[35] J.-L. WALDSPURGER, *Sur les valeurs de fonctions L-automorphes en leur centre de symétrie*, Comp. Math. **54** (1985), 173–242. MR 0783511. *(466)*

[36] S. ZELDITCH, *Kuznecov sum formulae and Szegö limit formulae on manifolds*, Comm. Partial Differential Equations **17** (1992), 221–260. MR 1151262. DOI 10.1080/03605309208820840. *(465)*

Department of Mathematics, University of Wisconsin–Madison, Madison, Wisconsin, USA;
marshall@math.wisc.edu