

# The Behaviour of Eigenstates of Arithmetic Hyperbolic Manifolds

**Zeév Rudnick, Peter Sarnak**

Department of Mathematics, Princeton University, Princeton, NJ 08544, USA

Received: 28 April 1993

**Abstract.** In this paper we study some problems arising from the theory of Quantum Chaos, in the context of arithmetic hyperbolic manifolds. We show that there is no strong localization (“scarring”) onto totally geodesic submanifolds. Arithmetic examples are given, which show that the random wave model for eigenstates does not apply universally in 3 degrees of freedom.

## 1. Introduction

Let  $X = \Gamma \backslash \mathbf{H}^2$  be a compact hyperbolic surface, with  $\Gamma \subset PSL(2, \mathbf{R})$  a discrete co-compact subgroup, and  $\mathbf{H}^2$  the hyperbolic plane. As is well known, the geodesic flow on the unit cotangent bundle  $S^*X$  is ergodic, Anosov and displays chaotic features [1]. Denote by  $\Delta$  the Laplace-Beltrami operator for  $X$ , and by  $\lambda_j$  (respectively  $\phi_j$ ) its eigenvalues (resp. an orthonormal basis of eigenstates).  $\Delta$  is the quantization of the Hamiltonian generating the geodesic flow. For this reason the behaviour of  $\lambda_j$  and  $\phi_j$  has been studied extensively in the context of Quantum Chaos. One of the central questions is whether the  $\phi_j$ ’s behave like random waves, or if they display some localization or other structure related to the classical trajectories. In the case of the billiards in the Bunimovich stadium (which is a somewhat more complicated chaotic system), Heller [11] unexpectedly found that certain states are enhanced on a finite union of periodic (unstable) orbits. He called this phenomenon “scarring.” In our case, the numerical evidence [9, 2, 3] points to the eigenstates behaving like random waves. We refer to Heller [11] for a discussion of random waves.

In this paper we examine these issues for  $\Gamma$  arithmetic, in fact for certain congruence subgroups derived from quaternion algebras. These are of course special, but in connection with the questions at hand we expect that our results are typical of

the truth in the general case. Note however that in connection with the statistics of the spectrum, arithmetic and non-arithmetic groups display quite different behaviours [2, 3, 4, 16].

Define the probability measures  $\mu_j$  on  $X$  by

$$d\mu_j = |\phi_j(z)|^2 d\text{vol}(z). \tag{1.1}$$

The quantum mechanical interpretation of these probability measures, as the probability density for finding a particle in the state  $\phi_j$  at the point  $z$ , is well known. A (weak-\*) limit  $\nu$  of the sequence  $\mu_j$  will be called a quantum limit. A general result due to Schnirelman [19], Zelditch [23], Colin de Verdiere [6] asserts that if the geodesic flow on  $S^*(X)$  is ergodic then for almost all  $j$ ,  $\mu_j \rightarrow d\text{vol}$ . In fact one can define an appropriate extension  $\tilde{\mu}_j$  of  $\mu_j$  to phase space  $T^*(X)$  and they show that any limit  $\tilde{\nu}$  of  $\tilde{\mu}_j$  is invariant under the geodesic flow. While this restricts the set of possible  $\tilde{\nu}$ 's, it is a basic feature of chaotic systems that the set of such invariant measures is very large and complicated, as are its typical members. The simplest and most localized such measure is the arc-length measure supported on a union of periodic geodesics. Whether these could occur as quantum limits has been raised by Colin de Verdiere on a number of occasions [6]. It is related to Heller's scarring in its strongest possible form. Since the latter concerns enhancement of probabilities, we define

**Definition 1.1.** A subsequence  $\mu_{j_k}$  is said to scar strongly to a closed subset  $S \subset X$  if  $\mu_{j_k} \rightarrow \nu$  and  $\emptyset \neq \text{singsupp } \nu \subset S$ .

For  $\Gamma \subset SL(2, \mathbf{R})$  derived from an Eichler order in a quaternion algebra over  $\mathbf{Q}$  (see Sect. 2 for definitions), there is a commutative self-adjoint Hecke algebra  $\mathcal{H}$  of arithmetically defined operators on  $L^2(X)$ . Moreover  $\mathcal{H}$  commutes with  $\Delta$ . Hence we may and will assume that  $\phi_j$  are also eigenfunctions of  $\mathcal{H}$ . This is probably automatic, since all evidence points to the spectrum of  $\Delta$  being simple (except for multiplicities imposed by finite symmetry conditions). In any event, such eigenfunctions are the arithmetically interesting ones. The following asserts that strong scarring on closed geodesics for such  $X$  is impossible and in particular answers the above question of Colin de Verdiere for these surfaces.

**Theorem 1.1.** *Let  $X = \Gamma \backslash \mathbf{H}^2$  be an arithmetic surface derived from a quaternion algebra as above. Let  $\nu$  be a quantum limit and  $\sigma$  the support of its singular part  $\nu^s$ . Then if  $\sigma$  is contained in the union of a finite number of points and closed geodesics then  $\sigma = \emptyset$ .*

The theorem asserts that if  $\nu^s$  has support in a finite union of closed geodesics then  $\nu$  is absolutely continuous with respect to  $d\text{vol}$ . Theorem 1.1 is a first step towards proving that the  $\mu_j$  are individually becoming equidistributed.

**Conjecture (Quantum unique ergodicity).** *Let  $X$  be a compact manifold of negative curvature. Then the measures  $\mu_j$  converge to  $d\text{vol}$ .*

If this is true it is remarkable, since it asserts that at the quantum level and its semi-classical limit, there is little manifestation of chaos from this point of view. In particular, one would have quantum unique ergodicity (i.e. only one possible quantum limit) while classical unique ergodicity (i.e. uniqueness of the invariant measure for the Hamiltonian flow) is never satisfied for chaotic systems.

Our methods can be extended to hyperbolic 3-manifolds  $\Gamma \backslash \mathbf{H}^3$ , where  $\Gamma \subset SL(2, \mathbf{C})$  is derived from a quaternion algebra over an imaginary quadratic number field of class number 1. In this case the result asserts that there is no strong scarring on proper closed totally geodesic subspaces, i.e. finite sets of points, closed geodesics or closed immersed hyperbolic surfaces.

All of the above is consistent with the random wave theory and indeed for  $n = 2$  this seems to be the true picture. However for  $n = 3$  and  $\Gamma$  arithmetic we now show that the random wave picture does not apply universally. The model of random waves requires that the extrema of the eigenstates do not grow too rapidly with the energy  $\lambda$ . In particular [9] that

$$\|\phi_j\|_\infty \approx \sqrt{\log \lambda_j}. \tag{1.2}$$

This is to be compared with what is known for a general compact manifold [22],

$$\|\phi_j\|_\infty = O(\lambda_j^{(n-1)/4}). \tag{1.3}$$

For  $n = 2$ , (1.2) is consistent with what can be proven about the  $L^\infty$  norms of the eigenstates [14]. The following concerns  $n = 3$ . Let  $F(x)$ ,  $x = (x_1, x_2, x_3, x_4)$  be an integral quadratic form of signature  $(3, 1)$ .  $V = \{x : F(x) = -1\}$  is a two sheeted hyperboloid giving a model for  $\mathbf{H}^3$ . Let  $\Gamma = O(F, \mathbf{Z})$  be the group of integral  $4 \times 4$  matrices preserving  $F$ . We assume that  $F$  is anisotropic over the rationals, that is  $F(x) \neq 0$  for  $0 \neq x \in \mathbf{Q}^4$ , then it is known [5] that  $X_F = \Gamma \backslash \mathbf{H}^3$  is a compact hyperbolic 3-manifold (arithmetic).

**Theorem 1.2.** *Let  $X_F$  be as above Then there is a sequence  $\phi_{j_k}$  of eigenstates satisfying*

$$\|\phi_{j_k}\|_\infty \gg \lambda_{j_k}^{1/4}. \tag{1.4}$$

This is proven using Siegel’s theory of theta functions. Indeed the eigenstates above with this singular behaviour are “theta lifts.” That theta lifts are singular in other aspects and contexts is not new – for example the failure of the “Ramanujan conjectures” [13] or the Sato-Tate conjectures. The reason for this behaviour in Theorem 1.2 is beautifully illustrated in the following special example: Let

$$F(x) = x_1^2 + x_2^2 + x_3^2 - 7x_4^2, \tag{1.5}$$

and as before  $V = \{x : F(x) = -1\}$ ,  $\Gamma = O(F, \mathbf{Z})$ ,  $Y = \Gamma \backslash V$ .

**Theorem 1.3.** *Let  $P = (2, 1, 1, 1) \in Y$ , then for all the  $cR^3$  eigenstates of  $Y$  with  $\sqrt{\lambda_j} \leq R$  ( $R \rightarrow \infty$ ) at most  $O(R^2)$  do not vanish at  $P$ .*

The fact that there so many eigenstates that vanish at  $P$  is what forces those that don’t to be very large at  $P$ . It is an interesting problem to explain the geometric source (rather than arithmetic) of these singular eigenstates and the above vanishing at  $P$ . In view of (1.2) and Theorem 1.2, these  $O(R^2)$  singular eigenstates do not behave like random waves.

The results of this paper were announced in the Schur lectures [18]. In those lectures further background on quantum chaos and related results may be found.

## 2. Non-Scarring

### 2.1 Correspondences

We begin with a general observation related to localization of eigenfunctions of Hecke operators. Let  $X$  be a Riemannian manifold. A correspondence  $C$  on  $X$  is a multivalued map given locally by

$$C : x \mapsto \{S_1(x), S_2(x), \dots, S_r(x)\}, \tag{2.1}$$

where each branch  $S_j(x)$  is locally an isometry and  $C$  is globally well defined. Associated to such a correspondence we have a Hecke operator  $T_C : L^2(X) \rightarrow L^2(X)$  defined by

$$T_C f(x) = \sum_{j=1}^r f(S_j x). \tag{2.2}$$

**Definition 2.1.** Let  $\Lambda \subset X$  be a closed set. We say that a correspondence  $C$  separates  $\Lambda$  if there is a point  $z \in X - \Lambda$  such that  $Cz \cap \Lambda$  consists of exactly one point. That is to say,  $S_1 z = w \in \Lambda$  and  $S_j z \notin \Lambda$  for  $j = 2, \dots, r$ .

**Lemma 2.1.** Let  $\Lambda \subset X$  be closed and of zero volume, and let  $C$  be a correspondence which separates  $\Lambda$ . If  $\phi_j$  is a sequence of eigenfunctions of  $T_C$  with  $\|\phi_j\| = 1$  and  $\nu = \lim_{j \rightarrow \infty} |\phi_j(z)|^2 d \text{vol}(z)$  exists, then the singular support of  $\nu$  cannot equal  $\Lambda$ .

*Proof.* By assumption,

$$\lambda_j(C)\phi_j(x) = \phi_j(S_1 x) + \sum_{k=2}^r \phi_j(S_k x). \tag{2.3}$$

Moreover, for all  $j$

$$|\lambda_j(C)| \leq r. \tag{2.4}$$

Let  $z, w$  be the points whose existence is ensured by Definition 2.1. Let  $U$  be a fixed small neighborhood of  $w$  in  $\Lambda$ . Now if  $\text{singsupp}(\nu) = \Lambda$ , we have

$$\nu(U) \neq 0. \tag{2.5}$$

For a decreasing sequence  $\varepsilon_n \searrow 0$ , set

$$U_n = \{x \in X : \text{dist}(x, U) < \varepsilon_n\}. \tag{2.6}$$

Then  $U_n \supseteq U_{n+1}$  and  $\bigcap_{n=1}^\infty U_n = U$ . Hence  $\nu(U_n) \searrow \nu(U)$  as  $n \rightarrow \infty$ .

For  $n$  large enough (and  $U$  small enough but fixed), let  $V_n$  be the corresponding neighborhood of  $z$  so that

$$S_1(V_n) = U_n$$

and

$$S_j(V_n) \cap \Lambda = \emptyset, \quad j = 2, \dots, r$$

From (2.3), (2.4) and the Cauchy-Schwartz inequality we have for  $n$  sufficiently large,

$$\int_{V_n} |\phi_j(S_1 x)|^2 dx = \int_{U_n} |\phi_j(x)|^2 dx \leq r \left( r \int_{V_n} |\phi_j(x)|^2 dx + \sum_{k=2}^r \int_{S_k(V_n)} |\phi_j(x)|^2 dx \right).$$

Letting  $j \rightarrow \infty$ , we conclude that

$$\nu(U_n) \leq r \left( r\nu(V_n) + \sum_{k=2}^r \nu(S_k(V_n)) \right).$$

Finally, letting  $n \rightarrow \infty$  we have

$$\nu(U) \leq r \left( r\nu(V_\infty) + \sum_{k=2}^r \nu(S_k(V_\infty)) \right), \tag{2.7}$$

where  $V_\infty = S_1^{-1}(U)$ .

Now  $V_\infty$  and  $S_k V_\infty$  all have zero volume and since we are assuming  $\nu$  is absolutely continuous with respect to  $d \text{ vol}$  outside  $A$ , it follows that the R.H.S. of (2.7) is 0. Hence  $\nu(U) = 0$  which contradicts (2.5). This proves the lemma.  $\square$

### 2.2. Quaternion Algebras

We next show how Lemma 2.1 may be used when  $X$  is a hyperbolic manifold derived from a quaternion algebra, and  $A$  is contained in a finite union of closed totally geodesic subspaces. We begin with a quick review of quaternion algebras over  $\mathbf{Q}$ . For details (and proofs) of what follows see Eichler [8]. Let  $a, b \in \mathbf{Z}$  be square-free integers with (say)  $a > 0$ . Let  $A = \left( \frac{a, b}{\mathbf{Q}} \right)$  be the corresponding indefinite quaternion algebra, which is defined as follows: Elements of  $A$  are

$$\alpha = x_0 + x_1\omega + x_2\Omega + x_3\omega\Omega, \tag{2.8}$$

where

$$\omega^2 = a, \quad \Omega^2 = b, \quad \omega\Omega + \Omega\omega = 0 \tag{2.9}$$

and  $x_j \in \mathbf{Q}$ .

The trace and norm maps are defined for  $\alpha$  as in (2.8) by

$$\text{tr}(\alpha) = \alpha + \bar{\alpha}, \quad N(\alpha) = \alpha\bar{\alpha}, \tag{2.10}$$

where  $\bar{\alpha} = x_0 - x_1\omega - x_2\Omega - x_3\omega\Omega$ .

We assume that  $A$  is a division algebra, which is equivalent to  $N(\alpha) \neq 0$  if  $\alpha \neq 0$ , i.e. that the rational quadratic form  $N(\alpha) = x_0^2 - ax_1^2 + bx_2^2 - abx_3^2$  is anisotropic over  $\mathbf{Q}$ . Fix once and for all an embedding of  $A$  into the  $2 \times 2$  real matrices by setting

$$\phi(\alpha) = \begin{pmatrix} x_0 + x_1\sqrt{a} & x_2 + x_3\sqrt{a} \\ b(x_2 - x_3\sqrt{a}) & x_0 - x_1\sqrt{a} \end{pmatrix}. \tag{2.11}$$

$\phi$  is an algebra homomorphism, and in fact extends to an isomorphism  $\phi: A \otimes_{\mathbf{R}} \mathbf{R} \rightarrow M_2(\mathbf{R})$ .

Note that for  $\alpha \in A$ , we have  $\phi(\alpha) \in M_2(F)$ , where  $F = \mathbf{Q}(\sqrt{a})$ . Indeed, if we set  $\xi = x_0 + x_1\sqrt{a}$ , and  $\eta = x_2 + x_3\sqrt{a}$ , then  $\alpha = \xi + \eta\Omega$  and

$$\phi(\alpha) = \begin{pmatrix} \xi & \eta \\ b\bar{\eta} & \bar{\xi} \end{pmatrix}. \tag{2.12}$$

We also observe that  $\det \phi(\alpha) = N(\alpha)$ ,  $\text{tr} \phi(\alpha) = \text{tr} \alpha$ . In what follows, we omit explicit mention of the embedding  $\phi: A \hookrightarrow M_2(F)$ . It is understood that writing  $\alpha z$  for  $z \in \mathbf{H}^2$  means the linear fractional transformation corresponding to the matrix  $\phi(\alpha)$ . Likewise for the action on binary quadratic forms below.

An order  $R$  in  $A$  is a subring containing 1 and for which  $\text{tr}(\alpha)$ ,  $N(\alpha) \in \mathbf{Z}$  are integers for all  $\alpha \in R$ , and such that  $R$  contains 4 linearly independent vectors over  $\mathbf{Q}$ . For example,  $R_0 = \{\alpha = \xi + \eta\Omega : \xi, \eta \in \mathcal{O}_F\}$  is an order, where  $\mathcal{O}_F$  are the integers of  $F$ . We consider maximal orders (or more generally the orders  $(p, q)$  defined in Eichler [8]). The reason we do so is technical – such orders have class number 1, which makes calculations much simpler. It is quite likely that our methods extend to general orders and quaternion algebras defined over general number fields. Let  $R$  be a maximal order containing  $R_0$ ; since  $A$  is an indefinite quaternion algebra over  $\mathbf{Q}$ , it is unique up to conjugation in  $A$ . For a suitable integer  $D$ , we have  $DR \subseteq R_0$ .

For  $m \in \mathbf{Z}$  let  $R(m)$  be defined by

$$R(m) = \{\alpha \in R : N(\alpha) = m\}. \tag{2.13}$$

$R(1)$  is the group of units (of norm 1) in  $R$  and it acts on  $R(m)$  by multiplication on the left. Under this action,  $R(m)$  decomposes into a finite number of orbits. As long as  $(m, q) = 1$ , where  $q = q(R)$  is a fixed integer, the behaviour of the orbits  $R(1) \backslash R(m)$  is simple. For example [8], for  $p$  a prime

$$|R(1) \backslash R(p)| = p + 1. \tag{2.14}$$

In what follows, we restrict ourselves to the case  $(m, q) = 1$ .

Let  $\Gamma_R$  be the image of  $R(1)$  under the embedding  $\phi$  in (2.11). It is a lattice in  $SL_2(\mathbf{R})$ , and the quotient

$$X_R = \Gamma_R \backslash \mathbf{H}^2 \tag{2.15}$$

is a compact hyperbolic surface (since  $A$  is a division algebra). For  $m$  as above, we have modular correspondences  $\tilde{T}_m : X_R \rightarrow X_R$  (which of course fit into our general definition given earlier), defined by

$$\tilde{T}_m : z \mapsto \phi(R(1) \backslash R(m))z. \tag{2.16}$$

The corresponding Hecke operators on  $L^2(X_R)$  are denoted by  $T_m$ , and are given by

$$T_m f(z) = \sum_{\alpha \in R(1) \backslash R(m)} f(\phi(\alpha)z). \tag{2.17}$$

An element  $\alpha \in R$  is primitive if there is no rational integer  $t > 1$  such that  $\alpha/t \in R$ . The set  $R^{pr}(m) = \{\alpha \in R(m) : \alpha \text{ primitive}\}$  is invariant under multiplication by  $R(1)$  and so we can define a modular correspondence  $C_m$  by

$$C_m : z \mapsto \bigcup_{\alpha \in R(1) \backslash R^{pr}(m)} \phi(\alpha)z. \tag{2.18}$$

The corresponding Hecke operators are also denoted by  $C_m$ , and they satisfy the following properties:

- (1)  $T_m = \sum_{t^2|m} C_{m/t^2}$ ;
- (2)  $T_m^* = T_m$  (self-adjointness);
- (3)  $T_n T_m = \sum_{d|(n,m)} d T_{mn/d^2}$ ;
- (4)  $T_m$  commutes with  $\Delta$ .

In view of (2), (3), (4) above we may simultaneously diagonalize  $\Delta$  and  $T_m$ , for  $(m, q) = 1$ . Let  $\phi_j$  be an orthonormal basis (o.n.b.) of simultaneous eigenfunctions. Our results about non-scarring are (proven) only for such an o.n.b. In view of Lemma 2.1, the main theorem will be established once we show that if  $\Lambda \subset X_R$  is contained in a finite union of closed geodesics (or points), then there is a correspondence  $C_m$ ,  $(m, q) = 1$ , which separates  $\Lambda$ . The next two subsections are devoted to proving this.

### 2.3. Binary Forms

First we recall the relation between binary quadratic forms, points in  $\mathbf{H}^2$  and closed geodesics. Let  $M = [a, b, c]$ ,  $a, b, c \in \mathbf{R}$  be the binary quadratic form

$$M(x, y) = ax^2 + bxy + cy^2.$$

We will also represent  $M$  by the symmetric matrix  $M = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$ . To such a form  $M$  whose discriminant  $d < 0$ , we associate the root  $z(M)$  of  $M(z, 1) = 0$  with  $\text{Im}(z) > 0$ . In this way the action of  $g \in SL_2(\mathbf{R})$  by a linear fractional map corresponds to the map on symmetric matrices  $M \mapsto M[g] = g^t M g$ . Specifically, we have  $z(M[g]) = g^{-1} z(M)$ . Of course, all forms proportional to  $M$  give rise to the same point in  $\mathbf{H}^2$  so that we think of  $\mathbf{H}^2$  as proportionality classes of binary quadratic forms with negative discriminant. If  $d > 0$  the roots of  $M(z, 1) = 0$  are both real and we associate to the class of  $M$  the geodesic in  $\mathbf{H}^2$  which connects these two roots. Again the linear fractional action of  $g \in SL(\mathbf{R})$  on the geodesic corresponds to  $M \mapsto M' = M[g]$  as above. The following is the key lemma:

**Lemma 2.2.** *Let  $M, M_1, \dots, M_r$  be a finite set of binary quadratic forms. Then there are infinitely many primes  $p$  such that*

$$M[\alpha] \neq \lambda_j(\alpha) M_j, \quad j = 1, \dots, r$$

for any primitive  $\alpha \in R$  of norm  $p$  or  $p^2$ .

This lemma says that one can separate one (class of) forms from any finite set of forms by a modular correspondence.

We start by making a reduction to the case that all the forms  $M, M_1, \dots, M_r$  in Lemma 2.2 are proportional to  $F$ -rational forms, that is to say, if  $M = [A, B, C]$  then after multiplying  $M$  by a constant, we have  $A, B, C \in F$ . Recall that there is an integer  $D$  such that  $DR \subset R_0$ .

**Lemma 2.3.** *Let  $M = [A, B, C]$ ,  $M' = [A', B', C']$  be binary quadratic forms for which there are 3 distinct primes  $p_1, p_2, p_3, p_i \nmid abD$ , for which either*

- (1) *There are elements  $\alpha_i \in R(p_i)$  such that  $M[\alpha_i] = \lambda_i M'$ ,  $i = 1, 2, 3$ , or*

(2) *There are primitive elements  $\alpha_i \in R^{pr}(p_i^2)$  such that  $M[\alpha_i] = \lambda_i M'$ ,  $i = 1, 2, 3$ . Then  $M, M'$  are proportional to  $F$ -rational forms.*

*Proof.* We begin by noting that if  $M'$  is split over  $F$ , then so is  $M$  and so both are proportional to a form  $M_0[\gamma]$ ,  $M_0(x, y) = 2xy$ ,  $\gamma \in GL_2(F)$ , which is  $F$ -rational. So we may assume that both  $M, M'$  are anisotropic over  $F$ , and in particular that  $A'C' \neq 0$ .

Taking determinants in the equation  $M[\alpha_i] = \lambda_i M'$ ,  $\alpha_i \in R^{pr}(p_i^n)$ , we find that  $N(\alpha_i)^2 \det(M) = \lambda_i^2 \det(M')$ , or that

$$\lambda_i = \kappa e_i p_i^n, \quad \kappa = \sqrt{\det(M)/\det(M')}, \quad e_i = \pm 1,$$

where  $n = 1$  or  $2$ . The equation  $M[\alpha_i] = \lambda_i M'$  implies in particular that

$$A\eta_i^2 + B\eta_i\bar{\xi}_i + C\bar{\xi}_i^2 = C'\kappa e_i p_i^n, \quad i = 1, 2, 3. \tag{2.19}$$

We may think of this as a system of 3 equations for  $A, B, C$ :

$$\Psi \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \kappa C' \begin{pmatrix} e_1 p_1^n \\ e_2 p_2^n \\ e_3 p_3^n \end{pmatrix},$$

where

$$\Psi = \begin{pmatrix} \eta_1^2 & \eta_1\bar{\xi}_1 & \bar{\xi}_1^2 \\ \eta_2^2 & \eta_2\bar{\xi}_2 & \bar{\xi}_2^2 \\ \eta_3^2 & \eta_3\bar{\xi}_3 & \bar{\xi}_3^2 \end{pmatrix}.$$

If  $\Psi$  is invertible, we will find that

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \kappa C' \Psi^{-1} \begin{pmatrix} e_1 p_1^n \\ e_2 p_2^n \\ e_3 p_3^n \end{pmatrix},$$

and since  $\Psi \in M_2(F)$ , we see that  $M$  is proportional to an  $F$ -form. Thus we need to consider the possibility that  $\det \Psi = 0$ . We have

$$\det \Psi = \prod_{i < j} (\eta_i \bar{\xi}_j - \eta_j \bar{\xi}_i),$$

and so if  $\det \Psi = 0$ , one of the factors must vanish – say  $\eta_1 \bar{\xi}_2 - \eta_2 \bar{\xi}_1 = 0$ . This means there is a linear dependence relation, say

$$(\eta_1, \bar{\xi}_1) = \mu(\eta_2, \bar{\xi}_2)$$

with  $\mu = \mu(n) \in F$ . Substituting into (2.19), we find that

$$\begin{aligned} \kappa C' e_1 p_1^n &= A\eta_1^2 + B\eta_1\bar{\xi}_1 + C\bar{\xi}_1^2 \\ &= \mu^2(A\eta_2^2 + B\eta_2\bar{\xi}_2 + C\bar{\xi}_2^2) \\ &= \mu^2 \kappa C' e_2 p_2^n, \end{aligned} \tag{2.20}$$

and since  $C' \neq 0$ , we find

$$p_1^n = \mu^2 p_2^n. \tag{2.21}$$



If  $n = 1$ , (2.21) implies that  $\mu^2 = p_1/p_2$ , where  $\mu \in F$ . This implies that  $p_1/p_2$  is a square in  $F = \mathbf{Q}(\sqrt{a})$ , which we now show cannot happen. Indeed, if we write  $\mu = u + v\sqrt{a}$ ,  $u, v \in \mathbf{Q}$ , then on squaring we find

$$p_1/p_2 = u^2 + av^2 + 2uv\sqrt{a},$$

and so either  $u = 0$  or  $v = 0$ . If  $v = 0$  we find  $p_1/p_2 = u^2$  is a rational square, which is clearly impossible if  $p_1 \neq p_2$ . If  $u = 0$  then  $p_1/p_2 = av^2$ , and since we assume that  $p_1, p_2 \nmid a$ , this is again impossible. This finishes the proof for the case  $n = 1$ .

In the case  $n = 2$ , we find from (2.21) that  $\mu = \pm p_1/p_2$ ; we will for simplicity assume the sign is  $+$ . Then we find

$$\xi_1 = \frac{p_1}{p_2} \xi_2, \quad \eta_1 = \frac{p_1}{p_2} \eta_2. \tag{2.22}$$

Multiplying (2.22) by  $D$  so as to make all quantities integers in  $\mathcal{O}_F$ , and setting  $\xi'_i = D\xi_i, \eta'_i = D\eta_i$ , we have from (2.22)

$$p_2\xi'_1 = p_1\xi'_2, \quad p_2\eta'_1 = p_1\eta'_2.$$

Taking  $x, y \in \mathbf{Z}$  such that  $xp_1 + yp_2 = 1$ , we find that

$$\xi'_1 = (xp_1 + yp_2)\xi'_1 = p_1(x\xi'_1 + y\xi'_2).$$

This means that  $\xi'_1 = p_1\theta, \theta \in \mathcal{O}_F$ . Likewise we see that  $\eta'_1 = p_1\zeta, \zeta \in \mathcal{O}_F$ . This implies that  $D\alpha_1 = p_1\beta$ , with

$$\beta = \begin{pmatrix} \theta & \zeta \\ b\bar{\zeta} & \bar{\theta} \end{pmatrix} \in R_0 \subseteq R.$$

Now taking  $u, v \in \mathbf{Z}$  such that  $uD + vp_1 = 1$  we find

$$\alpha_1 = (uD + vp_1)\alpha_1 = p_1(u\beta + v\alpha_1) = p_1\gamma$$

with  $\gamma = u\beta + v\alpha_1 \in R$ , and so  $\alpha_1$  cannot be a primitive element.  $\square$

We can now proceed with the proof of Lemma 2.2, assuming that the forms  $M, M_1, \dots, M_r$  are  $F$ -rational. By clearing denominators, we can assume that  $M = [A, B, C], M_i = [A_i, B_i, C_i]$  with all elements integers:  $A, B, \dots, C_i \in \mathcal{O}_F$ . On taking determinants in the relation  $M[\alpha] = \lambda_i M_i$ , we find that  $\lambda_i = \kappa_i p^n, \kappa_i = \pm \sqrt{\det(M)/\det(M_i)}$ . From the original equation  $M[\alpha] = \lambda_i M_i$ , we find that  $\lambda_i \in F$  since the forms are  $F$ -rational. From now on, we omit from consideration those finitely many primes  $p$  for which  $\kappa_i$  are not units, or dividing  $abD$ . Thus  $M[\alpha] = p^n \kappa_i M_i, \kappa_i \in F$  a unit for  $p$ .

Writing  $\alpha_1 = D\alpha = \begin{pmatrix} \xi & \eta \\ b\bar{\eta} & \bar{\xi} \end{pmatrix} \in R_0, \xi, \eta \in \mathcal{O}_F$ , we find that

$$A\eta^2 + B\eta\bar{\xi} + C\bar{\xi}^2 = p^n \kappa_i C_i D^2, \tag{2.23}$$

$$A\xi^2 + Bb\xi\bar{\eta} + Cb^2\bar{\eta}^2 = p^n \kappa_i A_i D^2. \tag{2.24}$$

We now separate the argument into two cases. Clearly, if  $M[\alpha_i] = \lambda_i M_i$ , then both  $M, M_i$  are split over  $F$  or both are anisotropic over  $F$ . We then treat separately the case when all  $M, M_1, \dots, M_r$  are anisotropic over  $F$  or are all split over  $F$ .

*The Anisotropic Case.* Let  $d = B^2 - 4AC$  be the discriminant of  $M$ .  $d \in \mathcal{O}_F$  is not a square in  $F$  iff  $M$  is anisotropic over  $F$ . Thus there are infinitely many prime

ideals  $P \subset \mathcal{O}_F$  such that  $M$  is still anisotropic on reducing modulo  $P$ . This can be seen without invoking Dirichlet's theorem on primes in a progression, as follows: Consider the quadratic extension  $K = F(\sqrt{d})$ . For unramified primes  $P$  of  $F$ , we have  $P$  splits in  $K$  iff  $d$  is a square modulo  $P$ . Thus  $d$  is not a square mod  $P$  for only finitely many primes  $P$  is equivalent to saying that only finitely many primes of  $F$  split in  $K = F(\sqrt{d})$ . Now consider the Dedekind zeta-function of  $K$ : It has a factorization  $\zeta_K(s) = \zeta_F(s)L(s, \chi_{K/F})$ , where  $L(s, \chi_{K/F}) = \prod_P (1 - \chi_{K/F}(P)N(P)^{-s})^{-1}$ , with  $\chi_{K/F}(P) = 0, 1, -1$  depending on whether  $P$  is ramified, split or inert in  $K$ . If we assume that only finitely primes of  $F$  split in  $K$ , then we will get that  $\zeta_K(s) = \beta(s)\zeta_F(2s)$ , where  $\beta(s)$  is a finite Euler product. However  $\zeta_K(s)$  has a (simple) pole at  $s = 1$ , while  $\zeta_F(2s)$  is holomorphic at  $s = 1$ , which is the required contradiction. For such a prime  $P$ , we assume either  $P = (p)$  is inert ( $p$  a rational prime), or  $P\bar{P} = (p)$  ( $P$  split).

In case  $P = (p)$  is inert, in (2.23) and (2.24) we reduce modulo  $p$  to find that  $(\eta, \bar{\xi})$  and  $(\xi, b\bar{\eta})$  are isotropic vectors for  $M \bmod P$ , and since we arranged that  $M$  is anisotropic modulo  $P$ , this implies that  $\eta \equiv \xi \equiv \bar{\xi} \equiv b\bar{\eta} \equiv 0 \pmod p$  (recall that  $\xi, \eta \in \mathcal{O}_F$ ). Thus  $\xi = p\xi', \eta = p\eta'$ , and  $\alpha_1 = p\alpha'$ , with  $\alpha' = \begin{pmatrix} \xi' & \eta' \\ b\bar{\xi}' & \bar{\eta}' \end{pmatrix} \in R_0 \subset R$ .

Now pick  $x, y \in \mathbf{Z}$  such that  $xD + yp = 1$ . Then  $\alpha = (xD + yp)\alpha = xp\alpha' + py\alpha = p\beta$ , where  $\beta = x\alpha' + y\alpha \in R$  since  $\alpha' \in R_0 \subset R$ . This means that  $\alpha$  is not primitive.

In the case  $P$  is split, so that  $P\bar{P} = (p)$ , we reduce (2.23), (2.24) modulo  $P$  to find as before that  $\xi, \bar{\xi}, \eta, b\bar{\eta} \in P$ . However, if both  $\xi, \bar{\xi} \in P$  then since  $(P, \bar{P}) = (1)$ , we find  $\xi \in P \cap \bar{P} = P\bar{P} = (p)$ , and so  $\bar{\xi} \in (p)$  as well. Likewise  $\eta, \bar{\eta} \in (p)$  (we assume  $p \nmid b$ ). Thus as in the inert case,  $N(\alpha) = p$  is impossible and  $N(\alpha) = p^2$  contradicts  $\alpha$  being primitive.

*The Split Case.* This means that we can write

$$M(x, y) = (ux + vy)(u'x + v'y)$$

or  $M = l^u l'$ , where  $l = (u, v)$ ,  $l' = (u', v') \in \mathcal{O}_F^2$ . Likewise  $M_i = l_i^u l_i'$ . Now if we assume  $M[\alpha] = \lambda_i M_i$ , then  $\alpha$  takes the linear forms  $l, l'$  to multiples of  $l_i, l_i'$ :

$$l\alpha = \mu_i l_i, \quad l'\alpha = \nu_i l_i', \quad \mu_i \nu_i = \lambda_i = \kappa_i p^n. \tag{2.25}$$

We now assume that  $p$  is inert in  $F$ , i.e. that  $\left(\frac{a}{p}\right) = -1$ . Then  $p \mid \mu_i \nu_i$  implies  $p \mid \mu_i$  or  $p \mid \nu_i$  – say  $p \mid \nu_i$ . Thus

$$l\alpha = \mu_i l_i \equiv 0 \pmod p \tag{2.26}$$

and in particular we have

$$u\eta + v\bar{\xi} \equiv 0 \pmod p. \tag{2.21}$$

Since  $M = l^u l'$  is nondegenerate, either  $u \neq 0$  or  $v \neq 0$ , say the latter. By omitting finitely many primes, we may assume that  $v \not\equiv 0 \pmod p$ . Then we have

$$\bar{\xi} \equiv -\frac{u}{v} \eta \pmod p,$$

and substituting into  $N(\alpha) = \xi\bar{\xi} - b\eta\bar{\eta} = p^n$  we find:

$$\left(\frac{u\bar{u}}{v\bar{v}} - b\right)\eta\bar{\eta} \equiv 0 \pmod p. \tag{2.28}$$

We now make a further restriction on  $p$ , namely  $p \nmid (u\bar{u} - bv\bar{v})$ . Since  $A = \left(\frac{a, b}{\mathbf{Q}}\right)$  is a division algebra,  $u\bar{u} - bv\bar{v} \neq 0$  and so this omits finitely many primes from consideration (recall  $u, v, b$  are fixed).

Then for  $p$  outside this finite set and satisfying  $\left(\frac{a}{p}\right) = -1$ , (2.28) implies  $\eta\bar{\eta} \equiv 0 \pmod p$  and so  $p \mid \eta$  or  $p \mid \bar{\eta}$ . Since  $p$  is a rational prime, either condition implies both, so  $\eta \equiv \bar{\eta} \equiv 0 \pmod p$ . From (2.27) we then see that  $\bar{\xi} \equiv 0 \pmod p$ , and so  $\xi \equiv 0 \pmod p$ . Thus again we find that  $N(\alpha) = p$  is impossible, and if  $N(\alpha) = p^2$  then  $\alpha$  is not primitive. This finishes the proof of Lemma 2.2.

### 2.4. Separation

We now use Lemma 2.2 to prove that if  $\Lambda \subset F_1 \cup F_2 \cup \dots \cup F_r$  is contained in a finite union of closed geodesics  $F_i$  in  $X_R$ , then there is a prime  $p$  with  $(p, q(R)) = 1$  such that  $C_p$  separates  $\Lambda$  in the sense of Definition 2.1, and together with case (i) below we thus prove Theorem 1.1. There are two cases:

(i)  $\Lambda$  is finite: We may write  $\Lambda = \{z_1, \dots, z_l\}$ ,  $z_j \in X_R$ . Let  $\hat{z}_1, \dots, \hat{z}_l \in \mathbf{H}^2$  be representatives for these points, and let  $\hat{M}_1, \dots, \hat{M}_l$  be the corresponding quadratic forms. According to Lemma 2.2, we can separate  $\hat{M}_1$  from  $\hat{M}_1, \dots, \hat{M}_l$  for suitable  $p$ . That is for all primitive  $\alpha$  in  $R(p)$  or  $R(p^2)$ ,

$$\Gamma_R \alpha \hat{z}_1 \neq \hat{z}_j, \quad j = 1, 2, \dots, l. \tag{2.29}$$

Now choose representatives  $\sigma_1, \dots, \sigma_{p+1} \in R(1) \setminus R(p)$  [see (2.14)]. Properties (1), (2), (3) imply that to each  $\sigma_j$  there is a unique index  $j'$  such that  $\sigma_j \sigma_{j'} \in p^2 R(1)$ . In fact, conjugation acts on  $R(p)$  and  $\sigma_j \bar{\sigma}_j = p^2$  so that we may take  $\sigma_{j'} \in \Gamma_R \bar{\sigma}_j$ . We denote this unique representative by  $\bar{\sigma}_j$ . Moreover  $\sigma_j \sigma_i$  is primitive unless  $\sigma_i = \bar{\sigma}_j$ . Let

$$w = \Gamma_R \sigma_1 \hat{z}_1 \in X_R. \tag{2.30}$$

From (2.29) we see that

$$w \neq \Gamma_R \hat{z}_j = z_j, \quad j = 1, \dots, l,$$

that is  $w \notin \Lambda$ . Consider

$$\begin{aligned} C_p w &= \{\sigma_1 w, \sigma_2 w, \dots, \sigma_{p+1} w\} \\ &= \{\Gamma_R \sigma_1 \sigma_1 \hat{z}_1, \Gamma_R \sigma_2 \sigma_1 \hat{z}_1, \dots, \Gamma_R \sigma_{p+1} \sigma_1 \hat{z}_1\}. \end{aligned}$$

Now  $\bar{\sigma}_1 \sigma_1 \hat{z}_1 = \Gamma_R \hat{z}_1$  while all other products  $\sigma_j \sigma_1$  are primitive, and so

$$C_p w = \{z_1\} \cup B, \tag{2.31}$$

where  $B \subset \{\sigma \hat{z}_1 : N(\sigma) = p^2, \sigma \text{ primitive}\}$ . According to (2.29),  $B \cap \Lambda = \emptyset$ . We have shown that with the above choice of  $w$  and correspondence  $C = C_p$ ,  $C$  separates  $\Lambda$  as needed.

(ii)  $\Lambda$  infinite: Let  $\hat{F}_1, \dots, \hat{F}_l$  be geodesics on  $\mathbf{H}^2$  representing  $F_1, \dots, F_l$  and let  $\hat{M}_1, \dots, \hat{M}_l$  be the corresponding quadratic forms. Let  $p$  be chosen as in Lemma 2.2, separating  $\hat{M}_1$  from  $\hat{M}_1, \dots, \hat{M}_l$ . So for  $\sigma$  primitive of norm  $p^n$ ,  $n = 1, 2$ ,

$$\hat{M}_1[\sigma] \neq \lambda_j \hat{M}_j. \tag{2.32}$$

If we have forms  $M, N$  corresponding to closed geodesics  $F, G$  on  $X_R$ , then either  $F \cap G$  is finite on  $X_R$  or for some  $\sigma \in \Gamma_R$ ,

$$M[\sigma] = \lambda N \tag{2.33}$$

(this is the only place where we use the fact that the geodesics are closed). Hence from (2.32) we see that for  $p$  as above, the intersections

$$\mu_1 := C_p(F_1) \cap (F_1 \cup F_2 \cup \dots \cup F_r), \quad \mu_2 := C_{p^2}(F_1) \cap (F_1 \cup F_2 \cup \dots \cup F_r) \tag{2.34}$$

are finite subsets of  $X_R$ . Let

$$\nu_1 = \{z \in X_R : C_p(z) \cap \mu_1 \neq \emptyset\}, \quad \nu_2 = \{z \in X_R : C_{p^2}(z) \cap \mu_1 \neq \emptyset\}. \tag{2.35}$$

$\nu_1$  and  $\nu_2$  are also finite sets. We have tacitly assumed that  $\Lambda \cap F_1$  is infinite; since  $\Lambda$  is infinite this is no loss of generality. Hence we may find some  $z \in \Lambda \cap F_1$  such that  $z \notin \nu_1 \cup \nu_2$ .

As before, let  $\hat{z} \in \mathbf{H}^2$  with  $\Gamma_R \hat{z} = z$ , and let  $w = \Gamma_R \sigma_1 \hat{z}$ . Firstly  $w \notin \Lambda$ , since otherwise

$$w \in \Lambda \cap C_p(F_1) \subset C_p(F_1) \cap (F_1 \cup \dots \cup F_l) = \mu_1,$$

that is  $\Gamma_R z \in \nu_1$  which contradicts  $z \notin \nu_1 \cup \nu_2$ . Secondly, as before

$$C_p w = \{\Gamma_R \sigma_1 \sigma_1 z, \Gamma_R \sigma_2 \sigma_1 z, \dots, \Gamma_R \sigma_{p+1} \sigma_1 z\} = \Gamma_R z \cup B,$$

where  $B \subset C_{p^2}(z)$ . Thus if  $B \cap \Lambda \neq \emptyset$  then  $C_{p^2}(z) \cap \Lambda \neq \emptyset$  and since  $z \in F_1$  we must have from (2.34) that  $C_{p^2}(z) \cap \nu_2 \neq \emptyset$ . This contradicts our choice  $z \notin \nu_1 \cup \nu_2$ . So in this case too we have shown that  $C_p$  separates  $\Lambda$ , which completes the proof of Theorem 1.1.

### 3. Hyperbolic 3-Manifolds

#### 3.1 Theta Functions

We begin with a general construction using theta series of Siegel. We borrow from Shintani’s treatment of these functions [21]. Let  $Q(x_1, \dots, x_n)$  be an integral quadratic form of signature  $(n - 1, 1)$ . We assume that  $Q$  is anisotropic over  $\mathbf{Q}$ , which means that  $Q(x) \neq 0$  if  $0 \neq x \in \mathbf{Q}^n$ . It is a well known theorem of Meyer [5] that this can only happen if  $n = 2, 3, 4$ . Let  $G(\mathbf{R})$  be the connected component of the identity in the orthogonal group of  $Q$ . It is a subgroup of index 4 in the real orthogonal group of  $Q$ , and denote by  $\Gamma = G(\mathbf{Z})$  the elements of  $G$  with integer entries. It is well known [5] that under the assumption  $Q$  anisotropic, the quotient  $\Gamma \backslash G(\mathbf{R})$  is compact. Let  $Q_1 = 2Q$  and let  $R$  be a majorant of  $Q_1$ , that is  $R$  is a positive definite symmetric  $n \times n$  matrix satisfying

$$RQ_1^{-1}R = Q_1. \tag{3.1}$$

Denote by  $f(x)$  the Gaussian on  $\mathbf{R}^n$  given by

$$f(x) = e^{-\pi x^t R x}. \tag{3.2}$$

Define the  $\theta$ -function on  $\mathbf{H}^2 \times G(\mathbf{R})$  by

$$\theta(z, g) = \sum_{x \in \mathbf{Z}^n} v^{n/4} e\left(\frac{uQ_1(x)}{2}\right) f(\sqrt{v}g^{-1}x), \tag{3.3}$$

where  $z = u + iv$ , and  $e(z) = \exp(2\pi iz)$ .

The series (3.3) clearly converges absolutely and uniformly for  $(z, g)$  on compacta. Since  $Q_1(x)/2 = Q(x) \in \mathbf{Z}$  for  $x \in \mathbf{Z}^n$  and  $\gamma\mathbf{Z}^n = \mathbf{Z}^n$  for  $\gamma \in \Gamma$ , from (3.3) it is clear that

$$\theta(z, \gamma g) = \theta(z, g), \quad \gamma \in G(\mathbf{Z}), \tag{3.4}$$

$$\theta(z + 1, g) = \theta(z, g). \tag{3.5}$$

$\theta$  also has extra transformation properties in  $z$  which come from Poisson summation. Indeed using Proposition 1.6 and 1.7 of Shintani [21], we get

$$\begin{aligned} \theta(\gamma z, g) &= \chi(\gamma) \left( \frac{cz + d}{|cz + d|} \right)^{(n-2)/2} \theta(z, g), \\ \gamma \in \Gamma_0(4D) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}) : c \equiv 0 \pmod{4D} \right\}, \end{aligned} \tag{3.6}$$

where  $D = \text{disc } Q$ , and in the case that  $n$  is even, which is the interest here, the character  $\chi$  of  $\Gamma_0(4D)$  is given by

$$\chi \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) = \begin{cases} \left( \frac{D}{-d} \right) \varepsilon_d^{-n}, & d < 0 \\ \left( \frac{D}{d} \right) \varepsilon_d^{-n}, & d > 0 \end{cases} \tag{3.1}$$

with  $\varepsilon_d = 1$  or  $i$  according as  $d \equiv 1$  or  $3 \pmod{4}$  and  $\left( \frac{D}{d} \right)$  is the Kronecker symbol [15].

A final transformation property of  $\theta$  that is clear from the series (3.3) is the transformation under the maximal compact subgroup  $K = K_R = \{g \in G(\mathbf{R}) : g^t R g = R\}$ :

$$\theta(z, gk) = \theta(z, g), \quad k \in K. \tag{3.8}$$

Besides all these transformation properties  $\theta(z, g)$  satisfies a differential equation, which is the reason it is important for our eigenfunction analysis. Let  $L$  be the Casimir operator on  $G(\mathbf{H})$ , normalized as in Shintani, and let

$$\Delta_k = 4v^2 \left( \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2} \right) - 4ikv \frac{\partial}{\partial u}. \tag{3.9}$$

Then we have

$$\Delta_{(n-2)/2} \theta(z, g) = L_g \theta(z, g) + \nu_n \theta(z, g), \tag{3.10}$$

where

$$\nu_n = n \left( \frac{n}{4} - 1 \right). \tag{3.11}$$

Now  $V = G(\mathbf{R})/K$  gives a model for hyperbolic  $(n - 1)$ -space, and the Casimir operator  $L$  acting on functions on  $G(\mathbf{R})$  which are right  $K$ -invariant is simply a multiple of the Laplacian. Thus  $L$  is self-adjoint on  $L^2(X) = L^2(\Gamma \backslash G(\mathbf{R})/K)$  while

$\Delta_{(n-2)/2}$  is self-adjoint on  $L^2\left(\Gamma_0(4D)\backslash\mathbf{H}^2, \chi, \frac{n-2}{2}\right)$ , the functions  $f:\mathbf{H}^2 \rightarrow \mathbf{C}$  satisfying

$$f(\gamma z) = \chi(\gamma) \left(\frac{cz+d}{|cz+d|}\right)^{(n-2)/2} f(z), \quad \gamma \in \Gamma_0(4D), \tag{3.12}$$

$$\int_{\Gamma_0(4D)\backslash\mathbf{H}^2} |f(z)|^2 \frac{du dv}{v^2} < \infty. \tag{3.13}$$

In view of the above, it follows from (3.10) and an integration by parts that if  $\Phi \in L^2(X)$  is an eigenfunction of  $L$  satisfying  $L\Phi = \lambda\Phi$ , then

$$F(z) = \int_{\Gamma\backslash G(\mathbf{R})} \theta(z, g)\Phi(g)dg \tag{3.14}$$

satisfies

$$\Delta_{(n-2)/2}F = (\lambda + \nu_n)F, \tag{3.15}$$

that is  $F$  is an eigenfunction of the Laplacian and we will see now that if  $\lambda \neq 0$  then  $F$  is in  $L^2(\Gamma_0(4D)\backslash\mathbf{H}^2)$ , in fact is a cusp form.

In the case  $n = 2$ ,  $G(\mathbf{Z})\backslash G(\mathbf{R})/K$  is a circle and its eigenfunctions  $\Phi_m$ ,  $m \in \mathbf{Z}$ , are just exponentials. In this way we obtain an explicit arithmetic progression of eigenvalues of cusp forms of  $(\Gamma_0(4D)\backslash\mathbf{H}^2, \chi)$ . These ‘‘lifted’’ eigenforms  $F_m$  were first constructed by Maass [17]. Our main concern is the case  $n = 4$ .

If  $F(z)$  is as in (3.14), then since  $F(z+1) = F(z)$ , it has a Fourier expansion

$$F(z) = \sum_{m=-\infty}^{\infty} a_m(v)e(mu). \tag{3.16}$$

From (3.15), the coefficients  $a_m(v)$  satisfy an ODE. We now compute  $a_m(v)$ :

$$\begin{aligned} a_m(v) &= \int_0^1 \int_{\Gamma\backslash G(\mathbf{R})} \theta(u+iv, g)\Phi(g)dg e(-mu)du \\ &= v^{n/4} \int_{\Gamma\backslash G(\mathbf{R})} \sum_{Q(x)=m} f(\sqrt{v}g^{-1}x)\Phi(g)dg. \end{aligned} \tag{3.17}$$

Since  $Q$  is anisotropic, when  $m = 0$  the only solution  $x \in \mathbf{Z}^n$  to  $Q(x) = 0$  is  $x = 0$ . Hence the constant term  $a_0(v)$  is given by

$$a_0(v) = v^{n/4} \int_{\Gamma\backslash G(\mathbf{R})} \Phi(g)dg. \tag{3.18}$$

In particular if  $\Phi$  is the not constant eigenfunction then  $a_0(v) = 0$ .

Next assume  $m \neq 0$ ; as is well known [5], the integer points on the quadric  $V_m = \{x: Q(x) = m\}$  decompose into a finite number of  $\Gamma$  orbits. Let  $x_1, \dots, x_h \in V_m(\mathbf{Z})$

be inequivalent representatives for these orbits, and let  $\Gamma_j \subset \Gamma$  be the stabilizer of  $x_j$  in  $\Gamma$ . Then

$$\begin{aligned} a_m(v) &= v^{n/4} \int_{\Gamma \backslash G(\mathbf{R})} \sum_{j=1}^h \sum_{\gamma \in \Gamma_j \backslash \Gamma} f(\sqrt{v}g^{-1}\gamma^{-1}x_j)\Phi(g)dg \\ &= \sum_{j=1}^h I_j^m, \end{aligned} \tag{3.19}$$

where

$$\begin{aligned} I_j^m &= v^{n/4} \int_{\Gamma \backslash G(\mathbf{R})} \sum_{\gamma \in \Gamma_j \backslash \Gamma} f(\sqrt{v}g^{-1}\gamma^{-1}x_j)\Phi(g)dg \\ &= v^{n/4} \int_{\Gamma_j \backslash G(\mathbf{R})} f(\sqrt{v}g^{-1}x_j)\Phi(g)dg. \end{aligned} \tag{3.20}$$

We now assume that the sign of  $m$  is such that  $\text{Stab}(x_j) \subset G(\mathbf{R})$  is compact, i.e.  $V_m$  is a two sheeted hyperboloid. If the signature of the form  $Q$  is  $(n - 1, 1)$ , then  $m < 0$ . The group  $G(\mathbf{R})$  has two orbits on  $V_m$ , which are the connected components  $V_m^\pm$  of this two-sheeted hyperboloid. In this setting, the stabilizer  $\Gamma_j$  is finite; let  $w_j$  be its order. We then have

$$I_j^m = \frac{v^{n/4}}{w_j} \int_{G(\mathbf{R})} f(\sqrt{v}g^{-1}x_j)\Phi(g)dg. \tag{3.21}$$

Choose  $x_0 \in V_m^+$  such that  $\text{Stab}_G(x_0) = K$ , where  $K$  is the fixed maximal compact subgroup in (3.8). Let  $g_j \in G(\mathbf{R})$  be such that  $g_j x_j = \pm x_0$ , the sign depending on which sheet of the hyperboloid  $x_j$  is found. Then we have, using the fact that  $f(x)$  is even and so we need not worry about the sign,

$$I_g^m = \frac{v^{n/4}}{w_j} \int_{G(\mathbf{R})} f(\pm\sqrt{v}g^{-1}x_0)\Phi(g_j g)dg = \frac{v^{n/4}}{w_j} \int_{G(\mathbf{R})} f(\sqrt{v}g^{-1}x_0)\Phi(g_j g)dg. \tag{3.22}$$

So that if  $\Psi(g) = \Phi(g_j g)$  then  $L\Psi = \lambda\Psi$ . By the theory of spherical functions [10] (uniqueness of bi- $K$ -invariant eigenfunctions of  $L$ ), we have

$$I_j^m = \frac{v^{n/4}}{w_j} \Phi(g_j) \int_{G(\mathbf{R})} f(\sqrt{v}g^{-1}x_0)\omega_\lambda(g)dg, \tag{3.23}$$

where  $\omega_\lambda(g)$  is the unique spherical function with eigenvalue  $\lambda$  and  $\omega_\lambda(e) = 1$ .

From (3.19) and (3.23), we obtain

$$a_m(v) = \tilde{f}(\lambda; v) \sum_{j=1}^h \frac{1}{w_j} \Phi(g_j), \tag{3.24}$$

where  $\tilde{f}(\lambda; v)$  is the special function

$$\tilde{f}(\lambda; v) = v^{n/4} \int_{G(\mathbf{R})} f(\sqrt{v}g^{-1}x_0)\omega_\lambda(g)dg. \tag{3.25}$$

One can easily check by letting  $v \rightarrow \infty$  that  $\tilde{f}(\lambda; v) \neq 0$ . Calculations along the above lines of Fourier coefficients of lifted forms  $F_\Phi(z)$  were done by Maass in [17].

From (3.24) we may draw the important conclusion that if  $F_\Phi(z) \equiv 0$  then

$$\sum_{j=1}^h \frac{1}{w_j} \Phi(g_j) = 0. \tag{3.26}$$

### 3.2. Proof of Theorem 1.1

We next show that when  $n = 4$  almost all the lifted forms  $F_\Phi(z) \equiv 0$ . If  $\langle \Phi, 1 \rangle = 0$  then according to (3.18),  $a_0(v) = 0$ . One can show that the same is true for the constant terms at the other cusps, so that  $F_\Phi$  is a cusp form on  $\Gamma_0(4D)\backslash\mathbf{H}^2$ . Let  $\mathcal{E} = \mathcal{E}\left(\Gamma_0(4D)\backslash\mathbf{H}^2, \chi, \frac{n-2}{2}\right)$  denote this space of cusp forms.  $\Delta_{(n-2)/2}$  has discrete spectrum on  $\mathcal{E}$ , and we let  $\{F_j\}$  be an orthonormal basis of eigenfunctions with eigenvalues  $\mu_j$ . One can show that Weyl’s law holds for  $\mathcal{E}$  [20], that is:

$$\sum_{\sqrt{\mu_j} \leq R} 1 \sim \frac{\text{vol}(\Gamma_0(4D)\backslash\mathbf{H}^2)}{4\pi} R^2, \quad \text{as } R \rightarrow \infty. \tag{3.21}$$

Integration against  $\theta$  gives a linear map  $\Theta: \mathcal{E} \rightarrow L^2(X)$ :

$$\Theta(F)(g) = \int_{\Gamma_0(4D)\backslash\mathbf{H}^2} \overline{\theta(z, g)} F(z) \frac{du dv}{v^2}. \tag{3.28}$$

In view of (3.10),  $\Theta$  takes cusp forms to eigenfunctions of  $L$  on  $X$ . Let  $\mathcal{V}$  be the orthogonal complement in  $L^2(X)$  to the space spanned by the image of  $\Theta$  and the constant function. If  $\Phi \in \mathcal{V}$  then  $\langle \Phi, \Theta(F) \rangle = 0$  for all  $F \in \mathcal{E}$ , that is

$$\begin{aligned} 0 &= \int_{\Gamma \backslash G(\mathbf{R})} \Phi(g) \int_{\Gamma_0(4D)\backslash\mathbf{H}^2} \theta(z, g) \overline{F(z)} \frac{du dv}{v^2} dg \\ &= \int_{\Gamma_0(4D)\backslash\mathbf{H}^2} F_\Phi(z) \overline{F(z)} \frac{du dv}{v^2}. \end{aligned} \tag{3.29}$$

Thus if  $\Phi \in \mathcal{V}$ , since  $F_\Phi(z)$  is cuspidal and  $\langle F_\Phi, F \rangle = 0$  for all  $F \in \mathcal{E}$ , we conclude that  $F_\Phi \equiv 0$ . Hence from (3.26) we conclude that for  $\Phi \in \mathcal{V}$ ,

$$\sum_{j=1}^h \frac{1}{w_j} \Phi(g_j) = 0 \tag{3.30}$$

(both  $h$  and the elements  $g_j$  depend on  $m$ ).

We now show that for some  $1 \leq i \leq h$ , we have an infinite sequence of eigenvalues  $\lambda_k$  (which correspond to theta-lifts), such that

$$\Phi_k(g_i) \gg \lambda_k^{1/4}. \tag{3.31}$$

This of course implies the lower bound of Theorem 1.1.



To see this, we first note that

**Lemma 3.1.** *If  $\Phi_k$  is an orthonormal basis of eigenfunctions in  $L^2(X)$ , then*

$$\sum_{\sqrt{\lambda_k} \leq R} \left| \sum_{j=1}^h \frac{1}{w_j} \Phi_k(g_j) \right|^2 \sim (4\pi)^{-3/2} \sum_{j=1}^h \frac{1}{w_j^2} R^3, \quad \text{as } R \rightarrow \infty. \tag{3.32}$$

*Proof.* For a general compact Riemannian manifold of dimension 3, one has [12]:

$$\sum_{\sqrt{\lambda_k} \leq R} \Phi_k(x) \overline{\Phi_k(y)} = \begin{cases} (4\pi)^{-3/2} R^3 + o(R^3), & x = y \\ o(R^3), & x \neq y \end{cases}, \tag{3.35}$$

where the sum is over an orthonormal basis of eigenfunctions in  $L^2(X)$ . Squaring the inner sum in (3.32) and switching the order of summation, we get the required result upon using (3.33).  $\square$

Next we note that of the summands in (3.32), all but  $O(R^2)$  vanish: Indeed, if  $\Phi_k \in \mathcal{F}$  then the inner sum vanishes by (3.30). The remaining part of the spectrum is the image of  $\Theta: \mathcal{E} \rightarrow L^2(X)$ , which by Weyl’s law for  $\mathcal{E}$  (3.27) has only  $O(R^2)$  elements in the above range. This shows that

$$\sum_{j=1}^h \frac{1}{w_j} \Phi_k(g_j) = \Omega(\lambda_k^{1/4}), \quad \text{as } k \rightarrow \infty. \tag{3.31}$$

From this it follows that at least at one of the points  $g_j$ , an infinite sequence of the eigenfunctions must be as large as  $\lambda_k^{1/4}$ , hence we get Theorem 1.2.

### 3.3. An Example

We conclude this section with an example of a quadratic form for which there is a single point  $g_1 \in G(\mathbf{R})$  for which every eigenfunction  $\Phi \in \mathcal{F}$  vanishes. To this end, let

$$Q(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 - 7x_4^2. \tag{3.35}$$

This form is anisotropic, since one checks that  $Q(x) \equiv 0 \pmod{8}$  only if  $x \equiv 0 \pmod{2}$ .

**Proposition 3.1.**  *$O(Q, \mathbf{Z})$  has a single orbit on  $V_{-1}(\mathbf{Z}) = \{x \in \mathbf{Z}^4 : Q(x) = -1\}$ .*

*Proof.* Clearly the point  $P = (2, 1, 1, 1)$  is in  $V_{-1}(\mathbf{Z})$ . We now make the crucial observation that if there is an integer vector  $x$  such that  $Q(x) = \pm 1$ , then there is an orthogonal decomposition over  $\mathbf{Z}: \mathbf{Z}^4 = \mathbf{Z}P \oplus \Lambda$ , for a 3-dimensional lattice  $\Lambda \subset \mathbf{Z}^4$ . This is because the orthogonal projection onto  $\mathbf{R}P$  is given by

$$Tx = \frac{B(x, P)}{2Q(P)} P = \pm \langle x, P \rangle P \tag{3.36}$$

with  $B(x, y) = 2\langle x, y \rangle = Q(x + y) - Q(x) - Q(y)$  being the associated bilinear form. Equation (3.36) clearly takes integer vectors to integer multiples of  $P$  [this is where  $Q(P) = \pm 1$  is used!], and so any integer vector  $x \in \mathbf{Z}^4$  may be written uniquely as

$$x = Tx + (I - T)x \tag{3.37}$$

which is an orthogonal sum of integers vectors. The lattice  $\Lambda$  is just  $(I - T)\mathbf{Z}^4$ .

Pick an integral basis  $l_1, l_2, l_3$  of  $\Lambda$ . Then the matrix of the form  $Q$  with respect to the basis  $P, l_1, l_2, l_3$  of  $\mathbf{Z}^4$  is of the form

$$\begin{pmatrix} -1 & & & \\ & & & \\ & & & \\ & & & q \end{pmatrix}. \tag{3.38}$$

Since the discriminant of  $Q$  is  $-7$ , the discriminant of  $\Lambda$  is  $7$ , and the form  $q = Q|_{\Lambda}$  is positive definite.

Now suppose  $P' \in \mathbf{Z}^4$  is another vector such that  $Q(P') = -1$ ; we want to find  $\beta \in O(Q, \mathbf{Z})$  such that  $\beta(P) = P'$ . Repeating the above process, we get another orthogonal decomposition  $\mathbf{Z}^4 = \mathbf{Z}P' \oplus \Lambda'$  with  $q' = Q|_{\Lambda'}$  is a positive definite integral ternary form of discriminant  $7$ .

Now it is known [7] that there is only one equivalence class of positive definite ternary integral forms of discriminant  $7$ . Thus there is an integer basis  $\{l'_1, l'_2, l'_3\}$  of the lattice  $\Lambda'$  such that  $Q(l'_i, l'_j) = Q(l_i, l_j)$ , for all  $i, j = 1, 2, 3$ . Now let  $\beta: \mathbf{Z}^4 \rightarrow \mathbf{Z}^4$  be given by

$$\beta = \begin{cases} P \mapsto P' \\ l_i \mapsto l'_i, \quad i = 1, 2, 3 \end{cases}. \tag{3.39}$$

Then  $\beta$  is an isometry and lies in  $GL(4, \mathbf{Z})$ , so that  $\beta \in O(Q, \mathbf{Z})$  and  $\beta(P) = P'$  as required.  $\square$

From this, we deduce that  $\Gamma$  has exactly one orbit on the integer points of each sheet  $V_{-1}^{\pm}(\mathbf{Z})$  of the hyperboloid. Indeed, the special orthogonal group  $SO(Q, \mathbf{Z})$  still has only one orbit on  $V_{-1}(\mathbf{Z})$ , since  $P = (2, 1, 1, 1)$  is stabilized by the improper rotation

$$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{3.40}$$

The number of  $\Gamma$  orbits on  $V_{-1}(\mathbf{Z})$  is thus at most two, since  $\Gamma$  has index  $2$  in  $SO(Q, \mathbf{Z})$  [the rotation  $-I \in SO(Q, \mathbf{Z})$  is not in  $\Gamma$ ]. It is then exactly two, since  $\Gamma$  cannot interchange the two sheets of  $V_{-1}$ . As representative we may then take  $x_1 = P, x_2 = -P$ .

According to the recipe in the previous section, we take  $g_1 \in G(\mathbf{R})$  so that  $g_1 x_1 = x_0$ ; then we can take  $g_2 = g_1$ , since  $g_1 x_2 = -g_1 P = -x_0$ . The formula (3.24) for the Fourier coefficient  $a_{-1}(v)$  collapses to a single term

$$a_{-1}(v) = v \tilde{f}(\lambda; v) \frac{2}{w_1} \Phi(g_1), \tag{3.41}$$

and so if  $\Phi \in \mathcal{S}'$ , we find that  $\Phi(g_1) = 0!$

*Acknowledgement* We wish to thank J. Conway for showing us this simple method of proving Proposition 3.1.

## References

1. Arnold, V., Avez, A.: Ergodic problems of classical mechanics. New York: Benjamin 1968
2. Aurich, R., Steiner, F.: Energy level statistics of the Hadamard-Gutzwiller ensemble *Physica D* **43**, 155–180 (1990)
3. Aurich, R., Steiner, F.: Statistical properties of highly excited quantum eigenstates of a strongly chaotic system Preprint DESY 92-091, June 1992
4. Bogomolny, E.B., Georgeot, B., Giannoni, M., Schmidt, C.: Chaotic billiards generated by arithmetic groups. *Phys. Rev. Lett.* **69**, 1477–1480 (1992)
5. Cassels, J.W.: Rational quadratic forms. New York: Academic Press (1978)
6. Colin de Verdiere, Y.: Ergodicité et fonctions propre du laplacien *Commun. Math. Phys.* **102**, 497–502 (1985)
7. Conway, J., Sloane, N.: Sphere packings, lattices and groups Berlin, Heidelberg, New York: Springer 1988
8. Eichler, M.: Lectures on modular correspondences *Tata Institute* **9**, 1955
9. Hejhal, D., Rackner, B.: On the topography of Maass waveforms for  $PSL(2, \mathbb{Z})$ : Experiments and heuristics. *Experimental Math* **1**, 275–306 (1992)
10. Helgason, S.: Groups and Geometric Analysis. New York: Academic Press (1984)
11. Heller, E.J.: In: Chaos and Quantum Physics, Les Houches 1989 (ed. by M.J. Giannoni, A. Voros, and J. Zinn-Justin), Amsterdam: North-Holland, 1991, pp. 549–661
12. Hormander, L.: The Analysis of Linear Partial Differential Operators, Vol. I–IV, Berlin, Heidelberg, New York: Springer-Verlag, 1985
13. Howe, R., Piatetski-Shapiro, I.: A counter-example to the “generalized Ramanujan conjecture” for (quasi-) split groups. *Proc. Symp. in Pure Math* vol. **33**, Amer. Math. Soc. 315–322 (1979)
14. Iwaniec, H., Sarnak, P.:  $L^\infty$  norms of eigenfunctions of arithmetic surfaces. Preprint
15. Landau, E.: Elementary Number Theory. New York: Chelsea Pub. Co., 1958
16. Luo, W, Sarnak, P.: Number variance for arithmetic hyperbolic surfaces. To appear in *Commun. Math. Phys*
17. Maass, H.: Über die räumliche Verteilung der Punkte in Gittern mit indefiniter Metrik. *Math. Annalen* **138**, 287–315 (1959)
18. Sarnak, P.: Arithmetic Quantum Chaos. Schur Lectures, Tel Aviv 1992, preprint
19. Schnirelman, A.: *Usp. Mat. Nauk* **29**, 181–182 (1974)
20. Selberg, A.: Gottingen Lectures. In: *Collected Works*, vol. **1**, Berlin, Heidelberg, New York: Springer-Verlag
21. Shintani, T.: On construction of holomorphic forms of half integral weight. *Nagoya Math. J.* **58**, 83–126 (1975)
22. Seeger, A., Sogge, C.D.: Bounds for eigenfunctions of differential operators. *Indiana University Math. J.* **38**, 669–682 (1989)
23. Zelditch, S.: Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.* **55**, 919–941 (1987)

Communicated by Ya. G. Sinai

