IMRN International Mathematics Research Notices
1995, No. 2

The Dimension of the Space of Cusp Forms of Weight One

W. Duke

1 Introduction

It is a basic problem to determine the dimension of the space of cusp forms of a given type. For classical holomorphic forms of integral weight larger than one, the dimension is well understood by means of either the Riemann-Roch theorem or the Selberg trace formula. The case of weight one, however, remains mysterious. From the point of view of spectral theory, this is because these forms belong to an eigenvalue of the Laplacian which is not isolated; the difficulty of estimating nontrivially the multiplicity of such an eigenvalue is well known.

Suppose, for example, that q is a prime and that $S_1(q)$ is the space of holomorphic cusp forms for $\Gamma_0(q)$ of weight one with character $(\cdot/q)$, the Legendre symbol. No nonzero cusp forms may exist unless $(-1/q) = -1$, so assume $q \equiv 3 \pmod 4$. Hecke discovered that the existence of such cusp forms is tied up with the class number h of $\mathbf{Q}(\sqrt{-q})$; if χ is any nontrivial (hence nonreal) class character, then

$$\sum_{\mathfrak{a}} \chi(\mathfrak{a}) e(N(\mathfrak{a})\, z) \in S_1(q), \tag{1}$$

where $\mathfrak{a}$ runs over all nonzero integral ideals of $\mathbf{Q}(\sqrt{-q})$. There are $(h-1)/2$ independent forms of this type, so by Siegel's theorem we have the (ineffective) lower bound

$$\dim S_1(q) \gg_{\varepsilon} q^{1/2-\varepsilon}$$

for all $\varepsilon > 0$.

In general, $S_1(q)$ is not spanned by forms of Hecke's type. An active area of research is the construction of specific examples demonstrating this. (See [F] and its references.)

Received 21 November 1994. Revision received 19 December 1994.
Communicated by Barry Mazur.

Such exotic forms seem quite rare, however, and it appears reasonable to expect that in fact for all $\varepsilon > 0$

$$\dim S_1(q) = \frac{1}{2}(h-1) + O_\varepsilon(q^\varepsilon).$$

In particular, this would imply that $\dim S_1(q) \ll q^{1/2} \log q$. Rather less, however, is actually known. Serre has shown [S] that if $q = 24m-1$ or $24m+7$, then $\dim S_1(q) \leq m-(h-1)/2$, while if $q = 24m+11$ or $24m+19$, then $\dim S_1(q) \leq m-h+1$. Sarnak has informed me that the Selberg trace formula for weight one [Hej, Chapter 9] with a suitably chosen test function yields the bound

$$\dim S_1(q) \ll \frac{q}{\log q},$$

which, for large values of q, is currently the best known.[1] The main object here is to improve this estimate.

Theorem 1. For q prime,

$$\dim S_1(q) \ll q^{11/12} \log^4 q,$$

with an absolute implied constant. □

Roughly speaking, the idea of the proof is to exploit two conflicting properties of the Fourier coefficients of newforms in $S_1(q)$ not of Hecke's type (1): their approximate orthogonality and the finiteness of the number of their possible values at primes. The first property is a consequence of their belonging to automorphic forms, while the second is a consequence of the Deligne-Serre theorem. Taken together, these properties limit the number of possible newforms which may exist.

Following Serre [S], Theorem 1 has an application to the quotient $X_0^*(q)$ of the modular curve $X_0(q)$ by the Fricke involution $z \mapsto -1/qz$ when $q = 24m-1$ is prime. In this case the genus of $X_0^*(q)$ is $m-(h-1)/2$.

Corollary. For $q = 24m-1$ prime, the space of differential forms of the first kind on $X_0^*(q)$, with a zero of order at least m at the cusp, has dimension that is $O(q^{11/12} \log^4 q)$. □

A different kind of application of these ideas is to bound the number $m_4(-q)$ of quartic number fields of discriminant $-q$.

[1] In an unpublished note, "Accumulation of the Eigenvalues of the Hyperbolic Laplacian at 1/4," J.-M. Deshouillers and H. Iwaniec obtained in this way the bound $O(q \log^{-3} q)$ for the multiplicity of the eigenvalue 1/4 in the case of weight-zero Maass cusp forms for $\Gamma_0(q)$ with trivial character.

Theorem 2. For q prime,

$$m_4(-q) \ll q^{7/8} \log^4 q,$$

with an absolute implied constant. □

To put this result in context it may be of interest to show what follows from algebraic number theory and trivial bounds for class numbers. By means of class field theory, Heilbronn [Hei] showed that

$$m_4(-q) = \frac{4}{3} \sum_k h_2(k),$$

where k runs over all cubic number fields of discriminant $-q$. Here, for any ℓ and any number field k, $h_\ell(k)$ denotes the number of ideal classes of k of (exact) order ℓ. Furthermore, by [Ha], the number of cubic fields in the sum is $(3/2)h_3(\mathbf{Q}(\sqrt{-q}))$. For the class number $h(k)$ of any number field k of degree $n > 1$ and discriminant D, we have the bound

$$h(k) \ll |D|^{1/2} \log^{n-1} |D|,$$

where the implied constant depends only on n (see [N, page 153]). Since $h_\ell(k) \le h(k)$, we deduce that

$$m_4(-q) \ll q \log^3 q$$

with an absolute implied constant. The improvement of this given in Theorem 2 requires both the classification of quartic fields of discriminant $-q$ by odd octahedral Galois representations of conductor q given in [S] and the proof in this case of the Artin conjecture given in [T]. If we assume the Artin conjecture for icosahedral representations, then similarly we can deduce that the number of nonreal quintic fields of discriminant q^2 whose normal closure has Galois group A_5 is $O(q^{11/12} \log^4 q)$.

2 Approximate orthogonality of Fourier coefficients

For $q \in \mathbf{Z}^+$ and ε an odd Dirichlet character mod q, let $S_1(q) = S_1(q, \varepsilon)$ be the set of all holomorphic cusp forms for $\Gamma_0(q)$ of weight one with character ε. Thus $f \in S_1(q)$ satisfies, for $\gamma \in \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(q)$,

$$f(\gamma z) = \varepsilon(d)\,(cz + d)\,f(z),$$

and $(\operatorname{Im} z)^{1/2}\,|f(z)|$ is uniformly bounded on the upper half-plane $\mathbf{H}$. The vector space $S_1(q)$ is finite-dimensional and has an inner product

$$\langle f, g \rangle = \int_{\Gamma_0(q)\backslash \mathbf{H}} f(z)\bar{g}(z)\,y^{-1}\,dx\,dy.$$

Each $f \in S_1(q)$ has the Fourier expansion at ∞

$$f(z) = \sum_{n \geq 1} a_f(n)\, e(nz).$$

The object of this section is to establish the following mean value result which expresses the approximate orthogonality of the $a_f(n)$ over any fixed orthonormal basis $\mathcal{B}$ for $S_1(q)$.

Proposition 1. For arbitrary $c_n \in \mathbf{C}$ with $1 \leq n \leq N$, we have

$$\sum_{f \in \mathcal{B}} \left| \sum_{n \leq N} c_n\, a_f(n) \right|^2 \ll \left(1 + \frac{N}{q}\right) \sum_{n \leq N} |c_n|^2$$

with an absolute implied constant. □

The proof we give of this uses the following well-known duality principle.

Lemma 1. Suppose that V is a finite-dimensional inner product space over $\mathbf{C}$ with an orthonormal basis $\mathcal{B}$. Let $\{v_1, \ldots, v_N\} \subset V$ be a finite set of vectors, and let Δ be a positive number. Then the inequality

$$\sum_{n \leq N} |\langle u, v_n \rangle|^2 \leq \Delta \langle u, u \rangle \tag{2}$$

holds for all $u \in V$ if and only if

$$\sum_{f \in \mathcal{B}} \left| \sum_{n \leq N} c_n \langle f, v_n \rangle \right|^2 \leq \Delta \sum_{n \leq N} |c_n|^2 \tag{3}$$

holds for all $c_n \in \mathbf{C}$. □

Proof. Define the operator $A : \mathbf{C}^N \to V = \mathbf{C}^{\mathcal{B}}$ by $(c_n) \mapsto \sum c_n v_n$. Its adjoint $A^* : V \to \mathbf{C}^N$ is given by $u \mapsto (\langle u, v_n \rangle)$. Let $\|A\|$ and $\|A^*\|$ be their norms. Inequality (2) means that $\|A^*\| \leq \Delta$, while (3) can be rewritten

$$\left\langle \sum c_n v_n, \sum c_n v_n \right\rangle \leq \Delta \sum |c_n|^2,$$

which means that $\|A\| \leq \Delta$. Thus the lemma follows from the well-known fact that $\|A\| = \|A^*\|$. ■

Taking $v_n = \sum_{f \in \mathcal{B}} \bar{a}_f(n)\, f$, we see that Proposition 1 is reduced to the following lemma.

Lemma 2. For any $f \in S_1(q)$,

$$\sum_{n \le N} |a_f(n)|^2 \ll \left(1 + \frac{N}{q}\right) \langle f, f \rangle$$

with an absolute implied constant. □

Proof. We employ a technique of Iwaniec (unpublished) to bound such sums uniformly. For any $y > 0$, we have

$$\sum_{n \ge 1} e^{-4\pi n y} |a_f(n)|^2 = \int_0^1 |f(x + iy)|^2 \, dx.$$

Thus, for any $Y > 0$,

$$\sum_{n \le N} F_Y(n) \, |a_f(n)|^2 \le \int_Y^\infty \int_0^1 |f(z)|^2 y^{-1} \, dx \, dy, \tag{4}$$

where

$$F_Y(n) = \int_1^\infty e^{-4\pi n Y y} y^{-1} \, dy \ \ge \int_1^\infty e^{-4\pi N Y y} \, y^{-1} \, dy \tag{5}$$

for $n \le N$. Setting

$$P(Y) = \{z \in \mathbf{H}\,;\ 0 < \operatorname{Re} z \le 1,\ \operatorname{Im} z > Y\},$$

we have

$$\int_Y^\infty \int_0^1 |f(z)|^2 \, y^{-1} \, dx \, dy \le \max_{z \in P(Y)} \#\{\gamma \in \Gamma_0(q)/\{\pm 1\}\,;\ \gamma z \in P(Y)\} \, \langle f, f \rangle. \tag{6}$$

Now for fixed $z = x + iy \in \mathbf{H}$ the condition that $\gamma z = (az + b)/(cqz + d) \in P(Y)$ is imposed by requiring that

$$\operatorname{Im}((az + b)/(cqz + d)) = y|cqz + d|^{-2} > Y \tag{7}$$

and that

$$0 < \operatorname{Re}((az + b)/(cqz + d)) \le 1. \tag{8}$$

For fixed c and d satisfying (7), a and b are determined by (8). It follows that for $z \in \mathbf{H}$

$$\begin{aligned}\#\{\gamma \in \Gamma_0(q)/\{\pm 1\}; \gamma z \in P(Y)\} &\le 1 + \#\{c > 0, d;\ |cqz + d|^2 < yY^{-1}\} \\ &\le 1 + \#\{c, d;\ 0 < c < q^{-1}(yY)^{-1/2} \text{ and } |cqx + d|^2 < yY^{-1}\}\end{aligned}$$

by taking real and imaginary parts. Hence for $z \in P(Y)$

$$\begin{aligned}\#\{\gamma \in \Gamma_0(q)/\{\pm 1\}\}\,;\gamma z \in P(Y)\} &\le 1 + q^{-1}(yY)^{-1/2}\,(1 + 2y^{1/2}Y^{-1/2})\\ &= 1 + q^{-1}(yY)^{-1/2} + 2q^{-1}Y^{-1}\\ &\le 1 + 3q^{-1}Y^{-1}.\end{aligned}$$

Thus from (4) and (6) we get

$$\sum_{n\le N} F_Y(n)\,|a_f(n)|^2 \le (1 + 3q^{-1}Y^{-1})\,\langle f, f\rangle.$$

Choosing $Y = 3N^{-1}$ and using (5), we get Lemma 2 (with 10^{18} for the absolute constant). ■

3 Consequences of the Deligne-Serre theorem

Let $N_1(q,\varepsilon) \subset S_1(q,\varepsilon)$ be the set of normalized newforms. For $f \in N_1(q,\varepsilon)$ the associated L-function is an Euler product

$$L_f(s) = \sum_{n\ge 1} a_f(n)\,n^{-s} = \prod_p (1 - a_f(p)\,p^{-s} + \varepsilon(p)\,p^{-2s})^{-1}. \tag{9}$$

The Deligne-Serre theorem [DS] states that $L_f(s)$ is the Artin L-function of an irreducible two-dimensional Galois representation ρ of conductor q with $\det\rho = \varepsilon$ (via the Artin map). We shall use two consequences of this result. The first is that $a_f(n)$ satisfies the Ramanujan bound

$$|a_f(n)| \le d(n), \tag{10}$$

where $d(n)$ is the divisor function. The second is that f may be classified as being of dihedral, tetrahedral, octahedral, or icosahedral type according to whether the image of ρ in $PGL(2,\mathbf{C})$ is D_h, A_4, S_4, or A_5.

We now restrict attention to the case that q is prime and $\varepsilon(\cdot) = \left(\cdot/q\right)$, and write $N_1(q)$ for $N_1(q,\varepsilon)$. Also, let $N_{dih}, N_{oct}, N_{ico}$ be the forms in $N_1(q)$ of dihedral, octahedral, and icosahedral type, respectively. It is shown in [S, page 343 of *OEuvres*] that f is of dihedral type exactly when it is of Hecke's type (1).

Proposition 2. Suppose that $f \in N_1(q)$ for q prime. Then

(a) $\langle f, f \rangle \ll q \log^3 q$ with an absolute implied constant.

(b) If f is not of dihedral type, then it is either of octahedral or icosahedral type. If $f \in N_{oct}$, then

$$a_f(p^8) - a_f(p^4) - \left(\frac{p}{q}\right) a_f(p^2) = 1, \tag{11}$$

while if $f \in N_{ico}$, then

$$a(p^{12}) - a_f(p^8) - \left(\frac{p}{q}\right) a_f(p^2) = 1 \tag{12}$$

for all primes $p \neq q$. □

Proof. (a) The Rankin-Selberg convolution

$$\varphi(s) = \sum_{n \geq 1} b(n)\, n^{-s} = (1 + q^{-s})\, \zeta(2s) \sum_{n \geq 1} |a_f(n)|^2 n^{-s} \tag{13}$$

is entire of order one except for a simple pole at $s = 1$ with

$$\operatorname*{Res}_{s=1} \varphi(s) = \frac{2\pi^2}{q} \langle f, f \rangle,$$

and satisfies the functional equation

$$\Phi(s) = \left(\frac{q}{4\pi^2}\right)^s \Gamma(s)^2 \varphi(s) = \Phi(1 - s) \tag{14}$$

by [L]. Choose $F \in C_c^\infty(0, \infty)$ such that $\int_0^\infty F(x)\, dx = 1$. Then the Mellin transform

$$\hat{F}(s) = \int_0^\infty F(x)\, x^s \frac{dx}{x}$$

is entire, of rapid decay in vertical strips, and $\hat{F}(1) = 1$. By Mellin inversion,

$$F(x) = \frac{1}{2\pi i} \int_{\operatorname{Re} s = 2} \hat{F}(s) x^{-s}\, ds$$

for $x > 0$. Thus

$$\sum_{n \geq 1} b(n) F(n/q^2) = \frac{1}{2\pi i} \int_{\operatorname{Re} s = 2} \hat{F}(s) \varphi(s) q^{2s}\, ds \tag{15}$$

$$= 2\pi^2 q \langle f, f \rangle + \frac{1}{2\pi i} \int_{\operatorname{Re} s = -1} \hat{F}(s) \varphi(s) q^{2s}\, ds,$$

the shift of the line of integration from $\mathrm{Re}\, s = 2$ to $\mathrm{Re}\, s = -1$ being justified by a standard application of the Phragmén-Lindelöf theorem. It follows from (14), (13), and (10) that

$$|\varphi(-1+it)| \le (4\pi^2)^{-3} q^3 \frac{|\Gamma(2-it)|^2}{|\Gamma(-1+it)|^2}(1+q^{-2})|\zeta(4-2it)| \sum_{n\ge1} d(n)^2 n^{-2},$$

and by Stirling's formula this is $\ll q^3(|t|+1)^6$ with an absolute constant. Hence

$$\frac{1}{2\pi i}\int_{\mathrm{Re}(s)=-1} \hat{F}(s)\varphi(s)q^{2s}\,ds \ll q\int_{-\infty}^{\infty} |\hat{F}(-1+it)|(|t|+1)^6\,dt \ll q \tag{16}$$

since $\hat{F}(-1+it) \ll (1+|t|)^{-8}$. By (15) and (16) we get

$$\langle f, f\rangle = \frac{1}{2\pi^2 q}\sum_{n\ge1} b(n)\, F(n/q^2) + O(1)$$

where the implied constant depends only on F. Now, by (13) and (10)

$$0 \le b(n) \le 2\sum_{n=m\ell^2} d^2(m) = 2d_4(n)$$

where $d_4(n)$ is the number of factorizations of n into four factors. Thus

$$\langle f, f\rangle \ll \frac{1}{q}\sum_{n\ge1} d_4(n)\, F(n/q^2) \ll q\,\log^3 q$$

with an absolute implied constant.

(b) The fact that f cannot be tetrahedral is shown in [S, Theorem 7 (c)]. It is also observed in [S, page 362] that if $p \ne q$ and $f \in N_{\mathrm{oct}}$, then

$$\left(\frac{p}{q}\right) a(p^2) \in \{-1, 0, 1, 3\}, \tag{17}$$

while if $f \in N_{\mathrm{ico}}$, then

$$\left(\frac{p}{q}\right) a(p^2) \in \left\{-1, 0, 3, \frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2}\right\}. \tag{18}$$

For general $f \in N_1(q, \varepsilon)$ it follows from (11) that, if we set $x = \bar{\varepsilon}(p)\, a_f(p^2)$, then for $p \ne q$ and $n \ge 0$,

$$\bar{\varepsilon}^n(p)\, a_f(p^{2n}) = P_n(x), \tag{19}$$

where $P_0(x) = 1$, $P_1(x) = x$, and $P_{n+1}(x) = (x-1)\,P_n(x) - P_{n-1}(x)$ for $n \ge 1$. Thus $P_2(x) = x^2 - x - 1$, $P_3(x) = x^3 - 2x^2 - x + 1$, etc. One may check that

$$x(x+1)(x-1)(x-3) + 1 = P_4(x) - P_2(x) - P_1(x),$$

while

$$x(x+1)(x-2)(x-3)(x^2-x-1)+1 = P_6(x) - P_4(x) - P_1(x).$$

Thus by (17)–(19) we finish the proof of (b). ∎

Remark. It has been checked that no linear form $\sum_{\ell=1}^{7} c_\ell\, a_f(p^\ell)$ takes positive values for all possible values of $a_f(p)$, $p \neq q$, when $f \in N_{oct}$. A similar remark applies to $f \in N_{ico}$.

4 Counting newforms and quartic fields

By combining Propositions 1 and 2 we may now estimate the number of newforms of octahedral or icosahedral type.

Proposition 3. For q prime,

(a) $\#N_{oct} \ll q^{7/8} \log^4 q$

(b) $\#N_{ico} \ll q^{11/12} \log^4 q$

with absolute implied constants. □

Proof. By Propositions 1 and 2 (a) we have for any $c_n \in \mathbf{C}$

$$\sum_{f \in N_1(q)} \left| \sum_{n \le N} c_n\, a_f(n) \right|^2 \ll (q+N) \log^3 q \sum_{n \le N} |c_n|^2,$$

using the well-known fact that distinct newforms are orthogonal. Thus we deduce by positivity the inequality

$$\sum_{f \in N_{oct}} \left| \sum_{n \le N} c_n\, a_f(n) \right|^2 \ll (q+N) \log^3 q \sum_{n \le N} |c_n|^2. \tag{20}$$

Choose $N = q$ and

$$c_n = \begin{cases} 1, & n = p^8 \le q, \\ -1, & n = p^4 \le q^{1/2}, \\ -\left(p/q\right), & n = p^2 \le q^{1/4}, \\ 0, & \text{otherwise}, \end{cases}$$

for p prime. By the prime number theorem and (11),

$$\sum_{n \le N} c_n a_f(n) \sim \frac{8q^{1/8}}{\log q}, \quad \text{for } f \in N_{oct},$$

while

$$\sum_{n \le N} |c_n|^2 \sim \frac{24q^{1/8}}{\log q}.$$

Hence, by (20),

$$\#N_{oct} \ll q^{7/8} \log^4 q.$$

A similar argument works for N_{ico}. ■

Theorem 1 now follows from Propositions 2 (b) and 3, since

$$\#N_{dih} = \frac{h-1}{2} \ll q^{1/2} \log q.$$

For its corollary, we use the fact, shown in [S], that the required space of differentials has dimension

$$\frac{1}{2} \dim S_1(q) - \frac{1}{4}(h-1).$$

Theorem 2 follows from Proposition 3 (a), Theorem 8 of [S], and Tunnell's proof [T] of the Artin conjecture for octahedral Galois representations. Here we also use the fact that the Galois closure of any quartic extension with discriminant $-q$ has Galois group S_4, since otherwise the discriminant would not be square free (e.g., see [B]).

Acknowledgments

I would like to thank J-P. Serre for his many helpful comments. This paper was written while I was participating in the 1994–95 program on automorphic forms at the Mathematical Sciences Research Institute in Berkeley. I thank the organizers and the Institute for providing financial support and a congenial atmosphere in which to work.

Author's research was supported in part by National Science Foundation grants DMS-9202022 and DMS-9022140.

References

[B] A. M. Baily, *On the density of discriminants of quartic fields*, J. Reine Angew. Math. **315** (1980), 190–210.

[DS] P. Deligne and J-P. Serre, *Formes modulaires de poids 1*, Ann. Sci. École Norm. Sup. (4) **7** (1974), 507–530; also in *OEuvres*, *Vol. III*, Springer-Verlag, Berlin, 1986, 193–216.

[F] G. Frey, "Construction and arithmetical applications of modular forms of low weight" in *Elliptic Curves and Related Topics*, CRM Proc. and Lecture Notes **4**, Amer. Math. Soc., Providence, 1994, 1–21.

[Ha] H. Hasse, *Arithmetische Theorie der Kubischen Zahlkörper auf Klassenkörper-theoretischer Grundlage*, Math. Z. **31** (1930), 565–582.

[Hei] H. Heilbronn, "On the 2-classgroup of cubic fields" in *Studies in Pure Mathematics*, ed. by L. Mirsky, Academic Press, London, 1971, 117–119.

[Hej] D. A. Hejhal, *The Selberg Trace Formula for* PSL(2, **R**), *Vol. 2*, Lectures Notes in Math. **1001**, Springer-Verlag, Berlin, 1983.

[N] W. Narkiewicz, *Elementary and Analytic Theory of Algebraic Numbers*, 2nd ed., Springer-Verlag, Berlin, 1990.

[L] W.-C. Li, L-*series of Rankin type and their functional equations*, Math. Ann. **244** (1979), 135–166.

[S] J-P. Serre, "Modular forms of weight one and Galois representations" in *Algebraic Number Fields*, ed. by A. Fröhlich, Academic Press, 1977, 193–268; also in *OEuvres*, *Vol. III*, Springer-Verlag, Berlin, 1986, 292–367.

[T] J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. Amer. Math. Soc. **5** (1981), 173–175.

Current address: Mathematical Sciences Research Institute, 1000 Centennial Drive, Berkeley, California 94720, USA

Permanent address: Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903, USA; duke@math.rutgers.edu