ACTA ARITHMETICA
LVI (1990)

Small eigenvalues of Laplacian for $\Gamma_0(N)$

by

HENRYK IWANIEC (New Brunswick, N.J.)

1. Introduction. The Hecke congruence group of level N,

$$\Gamma = \Gamma_0(N) = \left\{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbf{Z}):\ c \equiv 0 \pmod N\right\}$$

acts on the upper half-plane H (equipped with the Poincaré metric) discontinuously by

$$\gamma z = \frac{az+b}{cz+d} \quad \text{if} \quad \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

giving the quotient space $\Gamma \backslash H$ of finite volume

$$V = \mathrm{vol}(\Gamma \backslash H) = \int_{\Gamma\backslash H} d\mu z$$

with respect to the invariant measure $d\mu z = y^{-2}\,dx\,dy$. Let χ be a Dirichlet character to the modulus N; it induces a character of Γ by $\chi(\gamma) = \chi(d)$. Let $\mathscr{A}(\Gamma \backslash H, \chi)$ be the space of functions $f\colon H \to \mathbf{C}$ which satisfy the automorphy equation

$$f(\gamma z) = \chi(\gamma) f(z).$$

Thus for any $f, g \in \mathscr{A}(\Gamma \backslash H, \chi)$ the product $f(z)\bar{g}(z)$ is Γ-invariant. Let $\mathscr{L}(\Gamma \backslash H, \chi)$ be the subspace of $\mathscr{A}(\Gamma \backslash H, \chi)$ of functions $f(z)$ such that $\langle f, f\rangle$ is finite, where

$$\langle f, g\rangle = \int_{\Gamma\backslash H} f(z)\bar{g}(z)\,d\mu z$$

is the inner product in $\mathscr{L}(\Gamma \backslash H, \chi)$. The Laplace–Beltrami operator

$$\Delta = y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

acts on a dense subspace of $\mathscr{L}(\Gamma \backslash H, \chi)$, and it is essentially a self-adjoint, unbounded operator. A. Selberg [26] showed that Δ has a point spectrum $\lambda_0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \ldots$, say, with $\lambda_j \to \infty$ as $j \to \infty$, and it has a continuous spectrum covering the interval $(\frac{1}{4}, \infty)$ with multiplicity equal to the number of singular cusps of Γ. The subspace $\mathcal{L}_c(\Gamma\backslash H, \chi)$ of the continuous spectrum is spanned by Eisenstein series and the subspace $\mathcal{L}_0(\Gamma\backslash H, \chi)$ of the point spectrum is spanned by Maass cusp forms together with a constant function of eigenvalue $\lambda_0 = 0$ in the case of the trivial character. If, however, χ is a non-trivial character then the entire point spectrum is cuspidal. In the latter case we formally set $\lambda_0 = 0$, so $\lambda_1 \leqslant \lambda_2 \leqslant \ldots$ always accounts for the cuspidal spectrum. As usual we set

$$\lambda_j = s_j(1-s_j) \quad \text{with} \quad s_j = \tfrac{1}{2}+it_j,$$

so t_j is real if $\lambda_j \geqslant \frac{1}{4}$ and it_j is real otherwise.

A. Selberg [27] conjectured that for congruence groups we have

$$\lambda_1 \geqslant \tfrac{1}{4}. \tag{1}$$

In other words, the cuspidal spectrum lies on the continuous one. The conjecture is known to be true for a few groups of small level (see [24] and [14]). In fact, for the modular group $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ the lowest eigenvalue is quite large, $\lambda_1 = 91.14\ldots$ (see the numerical computation of D. Hejhal [12]). But as the level tends to infinity one can find many eigenvalues near $\frac{1}{4}$ (see [8]). Moreover, the point $\lambda = \frac{1}{4}$ belongs to the spectrum of Δ on $\mathcal{L}_0(\Gamma_0(p)\backslash H, \chi_p)$ with $p \equiv 1 \pmod 4$ provided the class number of $\mathbf{Q}(\sqrt{p})$ exceeds 1 (see [31]). Here χ_p is the quadratic character

$$\chi_p(d) = \left(\frac{d}{p}\right).$$

The assumption that Γ is a congruence group cannot be dropped from the Selberg conjecture, for one can find a compact smooth surface $\Gamma\backslash H$ of genus $g \geqslant 2$ with λ_{2g-3} as small as desired (see [5]).

J.-P. Serre [28] interpreted Selberg's eigenvalue conjecture as an analogue of the Ramanujan conjecture for the "infinite place", so today the conjecture lies within the scope of the program of R. Langlands outlined in [20]. Although Langlands' program happened to be partially successful for "finite places" (see [22], [23], [29]) it fails so far to work for the "infinite place".

Yet, A. Selberg [27] was able to show that

$$\lambda_1 \geqslant \tfrac{3}{16} \tag{2}$$

by applying A. Weil's bound for the Kloosterman sums

$$S_\chi(m, n; c) = \sum_{d \,(\mathrm{mod}\, c)} \chi(d)\, e_c(m\bar{d}+nd),$$

namely that

$$|S_\chi(m, n; c)| \leqslant (m, n, c)^{1/2} c^{1/2} \tau(c). \tag{3}$$

Thus Selberg appealed (indirectly) to the Riemann hypothesis for curves over finite fields. Later S. Gelbart and H. Jacquet [10] established the strict inequality $\lambda_1 > \frac{3}{16}$ without using bounds for Kloosterman sums by means of a lift from GL_2 to GL_3. Recently we [18] gave an estimate for special sums of Kloosterman sums from which (2) follows with minimum arithmetic being involved.

A relation between the spectrum of Δ and the Kloosterman sums $S_\chi(m, n; c)$ emerges in the spectral representation for the zeta-function

$$Z(s) = \sum_{c \equiv 0 \,(\mathrm{mod}\, N)} c^{-2s} S_\chi(m, n; c). \tag{4}$$

Selberg [27] has shown that $Z(s)$ has meromorphic continuation to the whole complex s-plane. For suitable m, n the poles of $Z(s)$ in the half-plane $\mathrm{Re}\, s > \frac{1}{2}$ are at the points s_j of the segment $(\frac{1}{2}, 1)$. These points will be called *exceptional*; they correspond to the exceptional eigenvalues

$$0 < \lambda_j = s_j(1-s_j) < \tfrac{1}{4}. \tag{5}$$

Since (3) implies that the series (4) converges absolutely in $\mathrm{Re}\, s > \frac{3}{4}$ it follows that $Z(s)$ is holomorphic in $\mathrm{Re}\, s > \frac{3}{4}$, whence $\lambda_j \geqslant \frac{3}{16}$. For more information about the zeta-function $Z(s)$ see [11].

In this paper we shall utilize a somewhat stronger connection (a quantitative one) between the spectrum and the Kloosterman sums. Let us choose a complete orthonormal system $\{u_j(z)\}_{j=1}^{\infty}$ of Maass cusp forms in $\mathcal{L}_0(\Gamma\backslash H, \chi)$. Each form has the Fourier expansion

$$u_j(z) = y^{1/2} \sum_{n \neq 0} \varrho_j(n) K_{s_j-1/2}(2\pi |n| y)\, e(nx), \tag{6}$$

where $K_s(y)$ is the Macdonald–Bessel function and the $\varrho_j(n)$ are complex numbers called the *Fourier coefficients* of $u_j(z)$. Let $f(v)$ be a test function, smooth and compactly supported in $(0, \infty)$. For $N \geqslant 1$ and $X \geqslant 1$ we set

$$\mathcal{F}_N(X) = \sum_{c \equiv 0 \,(\mathrm{mod}\, N)} c^{-1} f(2\pi X/c)\, S_\chi(1, 1; c),$$

$$\mathcal{G}_N(X) = \sum_{1/2 < s_j < 1} \sin(\pi s_j)\, \hat{f}(1-2s_j)\, \Gamma(2s_j-1)(X^{2s_j-1} - X^{1-2s_j})\, |\varrho_j(1)|^2,$$

where $\hat{f}$ stands for the Mellin transform of f. Then we have

$$\mathcal{F}_N(X) = \mathcal{G}_N(X) + O_f(1) \tag{7}$$

up to the error term $O_f(1)$ which is bounded by a constant depending on the test function f alone. This is the Kuznetsov formula for $\Gamma = \Gamma_0(N)$ with some undisplayed spectral terms being estimated by $O_f(1)$. The main ingredients for a proof of (7) have been manufactured in [7]. Since the arguments are almost identical to those used in [19] we omit them here.

Notice that the terms of $\mathcal{G}_N(X)$ are non-negative provided f is non-negative, which we henceforth assume. Therefore it is again evident that (3) implies (2) by letting X in (7) tend to ∞. Moreover, it is clear that in order to improve (2) along these lines one must inquire into the oscillatory behavior of the Kloosterman sums $S_\chi(1, 1; c)$ with respect to the modulus c. One expects that the variation in sign of the Kloosterman sums may result in a considerable cancellation of terms in $\mathcal{F}_N(X)$ to the effect that

$$\mathcal{F}_N(X) \ll X^\varepsilon$$

for any $\varepsilon > 0$. This estimate, of course, is equivalent to Selberg's eigenvalue conjecture.

Notation. Throughout this paper we use the notation accepted in number theory. In particular, we set $e(z) = e^{2\pi i z}$, $e_c(z) = e(z/c)$, $\tau(c)$ is the divisor function, $(a, b, \ldots)$ is the greatest common divisor of $a, b, \ldots$ and $\bar{d} \pmod c$ is the multiplicative inverse of d modulo c. If χ is the trivial character to the modulus c then $S_\chi(m, n; c)$ is the classical Kloosterman sum which we denote by $S(m, n; c)$. If $m = n$ we abbreviate $S_\chi(n, n; c) = S_\chi(n; c)$ and $S(n, n; c) = S(n; c)$.

Either expression $f = O(g)$ or $f \ll g$ means that $|f| \leqslant \alpha g$, where α is a suitable positive constant which may depend on the relevant parameters to be specified occasionally. For instance, we have $\tau(c) \ll c^\varepsilon$ for any $\varepsilon > 0$, the constant implied in the symbol $\ll$ depending on ε only.

2. Statement of results. In practice a few exceptional eigenvalues do not matter, but a large number of them may cause a problem. Therefore it is important to investigate the distribution of the points s_j in the segment $(\frac{1}{2}, 1)$ from a statistical point of view. A natural may to pursue this program is to count all points s_j with certain weights such that the larger s_j is the heavier the weight attached to s_j. The following inequality, called the "density theorem", is suitable:

$$\sum_{1/2 < s_j < 1} V^{c(s_j - 1/2)} \ll V^{1+\varepsilon}, \tag{8}$$

where ε is any positive number and the constant implied in $\ll$ depends on ε alone. Here the points s_j are counted with the multiplicity of the eigenvalues $\lambda_j = s_j(1-s_j)$.

It has been shown in [19] and [16] that the density theorem holds true in $\mathcal{L}(\Gamma_0(p)\backslash H, 1)$ with exponents $c = \frac{24}{11}$ and $c = \frac{12}{5}$ respectively. In this paper we improve these results substantially.

THEOREM 1. *The density theorem holds true in $\mathcal{L}(\Gamma_0(N)\backslash H, 1)$ with exponent $c = 4$.*

Let us recall that the volume of $\Gamma_0(N)\backslash H$ is equal to

$$V = \tfrac{1}{3}\pi N \prod_{p|N} (1 + 1/p), \tag{9}$$

thus the density theorem can be expressed in terms of the level, namely we have

$$\sum_{1/2 < s_j < 1} N^{4(s_j - 1/2)} \ll N^{1+\varepsilon}, \tag{10}$$

or equivalently, we have

$$\#\{j:\ s_j \geqslant \alpha\} \ll N^{3-4\alpha+\varepsilon} \tag{11}$$

for any $\alpha > \frac{1}{2}$ and any $\varepsilon > 0$, the constant implied in $\ll$ depending on ε alone.

As pointed out in [15] the density theorem with exponent $c = 4$ would follow from a suitable lower bound for the Fourier coefficients of the Maass cusp form. Until now the latter objective seemed to be more difficult than the former, yet, it could have been accomplished by assuming the Ramanujan conjecture for finite places. In this paper we establish the desired bound unconditionally.

THEOREM 2. *Let $u(z)$ be a Maass cusp form for the group $\Gamma = \Gamma_0(N)$ with a multiplier given by Dirichlet's character $\chi \pmod N$ and with eigenvalue $\lambda = s(1-s)$. Suppose $u(z)$ is a newform. We then have*

$$\frac{V}{\sin \pi s} \frac{|\varrho(1)|^2}{\langle u, u\rangle} \gg (\lambda N)^{-\varepsilon} \tag{12}$$

for any $\varepsilon > 0$, the constant implied in $\gg$ depending on ε only.

This result is the best possible in either parameter λ or N. We expect, but have not been able to prove, that the reverse inequality holds true with any $\varepsilon < 0$. Assuming the Ramanujan conjecture L. A. Takhtadzhyan and A. I. Vinogradov [32] showed for $\Gamma = SL_2(\mathbf{Z})$ that

$$(\log \lambda)^{-3}(\log\log \lambda)^{-3} \ll \frac{V}{\sin \pi s} \frac{|\varrho(1)|^2}{\langle u, u\rangle} \ll (\log \lambda)(\log\log \lambda).$$

Weaker but unconditional results can be found in [30].

Now the density theorem with exponent $c = 4$ follows directly from (12) through (3) and (7) (see Section 4). No cancellation between Kloosterman sums is required to this end.

Our main goal in this paper is to get further improvements for almost all groups $\Gamma_0(p)$. The following estimate for bilinear forms in Kloosterman sums $S_{\chi_p}(1; pqr)$ is basic in this work.

THEOREM 3. *For any complex numbers α_p, β_q we have*

$$\sum_{p \leqslant P} \sum_{q \leqslant Q} \alpha_p \beta_q S_{\chi_p}(1; pqr) \ll \|\alpha\|\, \|\beta\| \sigma_r(P, Q) \tag{13}$$

where p, q range over primes, $\|\alpha\|$, $\|\beta\|$ stand for the l^2-norms of the sequences (α), (β) respectively and

$$\sigma_r(P, Q) = (rPQ)^{1/2}(P^{1/2} + r^{1/4} Q \log 3rQ)(\tau(r))^{28}, \tag{14}$$

the implied constant being absolute.

When compared with $\sigma_r(P, Q) = PQr^{1/2}\tau(r)$ obtainable directly by (3) our result shows that a cancellation between Kloosterman sums $S_{\chi_p}(1; pqr)$ exists if Q is large and P is a bit larger than $Qr^{1/2}$. This is a good range for applications. In the proof of Theorem 3 an exponential sum in three variables over a finite field arises, it is estimated by means of the Riemann hypothesis for varieties (Deligne's theory). We should like to emphasize that building up the complete exponential sum in three variables rests on the approximate reciprocity formula

$$\frac{1}{2\sqrt{p}} S_{\chi_p}(\bar{q}; p) = \frac{\varepsilon(p, q)}{2\sqrt{q}} S_{\chi_q}(\bar{p}; q) + O\left(\frac{1}{pq}\right), \tag{15}$$

where $\varepsilon(p, q) = 1, i, i^2, i^3$ for $(p, q) \equiv (1, 1), (-1, 1), (-1, -1), (1, -1) \pmod 4$ respectively, and $\bar{q}$, $\bar{p}$ are the multiplicative inverses of q, p modulo p, q respectively.

Theorem 3 will be applied to explore the exceptional spectrum of Δ in $\mathscr{L}(\Gamma_0(p)\backslash H; \chi_p)$ through the connection (7). This is by no means immediate. We shall apply (7) for the subgroups $\Gamma_0(pq)$ rather than for $\Gamma_0(p)$ and we shall average over q to gain a flexibility on the Kloosterman sums side. On the spectral side of (7) the positivity of terms becomes vital as well as the lower bound (12). Combining all the described ideas we shall prove

THEOREM 4. *We have*

$$\sum_{p \leqslant P} \sum_{1/2 < s_j < 1} p^{5(s_j - 1/2)} \ll P^{13/6+\varepsilon}, \tag{16}$$

where the innermost sum ranges over the exceptional eigenvalues of Δ in $\mathscr{L}(\Gamma_0(p)\backslash H; \chi_p)$ and ε is any positive number, the constant implied in $\ll$ depending on ε alone.

Theorem 4 easily implies

COROLLARY. *Let $0 < \delta < \frac{1}{12}$. We have*

$$\operatorname{Spec} \mathscr{L}(\Gamma_0(p)\backslash H, \chi_p) \subset [\tfrac{44}{225} - (7+3\delta)\delta/75, \infty)$$

except for at most $O(P^{1-\delta+\varepsilon})$ primes $p \leqslant P$.

Observe that $\frac{44}{225} > \frac{3}{16}$, so we have improved Selberg's bound (2) for almost all groups $\Gamma_0(p)$.

The author acknowledges numerous discussions with E. Bombieri and S. Sperber on exponential sums over varieties.

3. An upper bound for the L^2-norm of newforms. The newforms for the group $\Gamma_0(N)$ were introduced by A. O. Atkin and J. Lehner [3]. Originally only holomorphic modular forms of weight k, level N and trivial character $\chi = 1$ were considered. Later W. Li [21] extended the theory to any character $\chi \pmod N$, but still restricted herself to holomorphic forms. This restriction is not important and her theory applies to Maass cusp forms as well. Since the transfer has not been presented in the literature we shall give in this section a brief survey of main concepts and results.

Let T_n be the Hecke operator defined by

$$T_n = \frac{1}{\sqrt{n}} \sum_{ad=n} \chi(d) \sum_{b \pmod d} \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}.$$

Only those T_n with $(n, N) = 1$ are of interest for the group $\Gamma_0(N)$.

PROPOSITION 1. *If $(n, N) = 1$ then T_n acts in $\mathscr{L}_0(\Gamma_0(N)\backslash H, \chi)$, it is bounded and $\chi(n)$-hermitian, i.e. $\langle T_n f, g\rangle = \chi(n)\langle f, T_n g\rangle$. For $(mn, N) = 1$ we have*

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

In particular, the operators T_m, T_n commute. Clearly, they also commute with the Laplace–Beltrami operator Δ. Consequently, there is a basis in $\mathscr{L}_0(\Gamma_0(N)\backslash H, \chi)$ of Maass cusp forms which are common eigenfunctions of all T_n with $(n, N) = 1$.

For $p|N$ the Hecke operator is defined by

$$U_p = \frac{1}{\sqrt{p}} \sum_{b \pmod p} \begin{bmatrix} 1 & b \\ 0 & p \end{bmatrix}.$$

Clearly, U_p acts on $\mathscr{L}_0(\Gamma_0(N)\backslash H, \chi)$ and commutes with T_n for $(n, N) = 1$. But U_p is not hermitian, nor even a normal operator. Therefore, in order to complete the diagonalization through U_p one must split up the space $\mathscr{L}_0(\Gamma_0(N)\backslash H, \chi)$ properly. First observe that if χ is a character to the modulus M and $v(z) \in \mathscr{L}_0(\Gamma_0(M)\backslash H, \chi)$ then $v(dz) \in \mathscr{L}_0(\Gamma_0(N)\backslash H, \chi)$ whenever $dM|N$. This follows from

$$\begin{bmatrix} d & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} \alpha & \beta d \\ \gamma/d & \delta \end{bmatrix}\begin{bmatrix} d & 0 \\ 0 & 1 \end{bmatrix}.$$

Now let $\mathscr{L}_{\text{old}}(\Gamma_0(N)\backslash H, \chi)$ be the subspace of $\mathscr{L}_0(\Gamma_0(N)\backslash H, \chi)$ spanned by forms of type $v(dz)$, where $v(z)$ is a Maass cusp form on $\Gamma_0(M)$ of character $\chi \pmod M$ with $M < N$, $dM|N$ and $v(z)$ is a common eigenfunction of all the Hecke operators T_m with $(m, M) = 1$. Let $\mathscr{L}_{\text{new}}(\Gamma_0(N)\backslash H, \chi)$ be the orthogonal complement to $\mathscr{L}_{\text{old}}(\Gamma_0(N)\backslash H, \chi)$, so

$$\mathscr{L}_0(\Gamma_0(N)\backslash H, \chi) = \mathscr{L}_{\text{old}}(\Gamma_0(N)\backslash H, \chi) \oplus \mathscr{L}_{\text{new}}(\Gamma_0(N)\backslash H, \chi).$$

Clearly, the operator T_n with $(n, N) = 1$ maps $\mathscr{L}_{\text{old}}(\Gamma_0(N)\backslash H, \chi)$ into itself because it commutes with the operator

$$B_d = \begin{bmatrix} d & 0 \\ 0 & 1 \end{bmatrix} \quad \text{for } d|N.$$

Consequently, T_n maps $\mathscr{L}_{\text{new}}(\Gamma_0(N)\backslash H, \chi)$ into itself because it is $\chi(n)$-hermitian. Therefore, there exists a basis in $\mathscr{L}_{\text{new}}(\Gamma_0(N)\backslash H, \chi)$ of Maass cusp forms which are common eigenfunction of all T_n with $(n, N) = 1$. These cusp forms are called *newforms of level N*.

In the Fourier expansion (6) of a newform $u(z)$ the first coefficient does not vanish, so it is customary to normalize newforms by setting $\varrho(1) = 1$. Then the following happens:

PROPOSITION 2. *A normalized newform $u(z)$ on $\Gamma_0(N)$ is an eigenfunction of all U_p with $p|N$ of eigenvalue $\varrho(p)$.*

Finally, let us look back to the space $\mathscr{L}_{\text{old}}(\Gamma_0(N)\backslash H, \chi)$. A Maass cusp form $u(z)$ on $\Gamma_0(N)$ of character $\chi \pmod N$ is called on *oldform* if $u(z) = v(dz)$, where $v(z)$ is a newform on some proper overgroup $\Gamma_0(M)$ of a character $\chi \pmod M$ with $M < N$ and $dM|N$. In this case we say that $u(z)$ is *of level M* and *divisor d*. The following result is true, though it is not obvious:

PROPOSITION 3. *$\mathscr{L}_{\text{old}}(\Gamma_0(N)\backslash H, \chi)$ is spanned by oldforms.*

One of many profits from splitting the space of cusp forms into newforms is the multiplicativity of the Fourier coefficients.

PROPOSITION 4. *Let $u(z)$ be a normalized newform of level N and character χ. Then the Fourier coefficients satisfy*

(17) $$\varrho(m)\varrho(n) = \sum_{d|(m,n)} \chi(d)\varrho(mn/d^2) \quad \textit{for } (mn, N) = 1,$$

(18) $$\varrho(p)\varrho(n) = \varrho(pn) \quad \textit{for } p|N \textit{ and all } n.$$

Now we are ready to prove Theorem 2. We assume that $u(z)$ is a normalized newform and will estimate its L^2-norm. By Proposition 4 we have

$$|\varrho(m)\varrho(n)| \leqslant \sum_{d|(m,n)} |\varrho(mn/d^2)|$$

for all m, n. Hence, letting

$$L(x) = \sum_{1\leqslant n\leqslant x} n^{-1/2}|\varrho(n)|^2$$

we deduce that

$$\begin{aligned} L^2(x) &\leqslant \sum\sum_{m,n\leqslant x} (mn)^{-1/2}\Big(\sum_{d|(m,n)} |\varrho(mn/d^2)|\Big)^2 \\ &\leqslant \sum\sum_{m,n\leqslant x} \tau((m,n))(mn)^{-1/2} \sum_{d|(m,n)} |\varrho(mn/d^2)|^2 \\ &\leqslant \sum_{d\leqslant x} \tau(d)d^{-1} \sum\sum_{k,l\leqslant x/d} \tau((k,l))(kl)^{-1/2}|\varrho(kl)|^2 \\ &\leqslant T(x)\sum_{n\leqslant x^2} t(n)n^{-1/2}|\varrho(n)|^2, \end{aligned}$$

where

$$T(x) = \sum_{d\leqslant x} \tau(d)d^{-1} \leqslant \Big(\sum_{d\leqslant x} d^{-1}\Big)^2 < (\log 3x)^2 \quad \text{and}$$

$$t(n) = \sum_{kl=n} \tau((k,l)) \leqslant c(\varepsilon)n^{\varepsilon}$$

for any $\varepsilon > 0$ and some $c(\varepsilon) > 0$. In this way we obtain the following inhomogeneous inequality for $L(x)$:

(19) $$L^2(x) < c(\varepsilon)(x^{\varepsilon}\log 3x)^2 L(x^2).$$

On the other hand, we shall evaluate $L(x)$ by means of the Rankin–Selberg zeta-function

$$L(u\otimes u;\, w) = \sum_{n=1}^{\infty} |\varrho(n)|^2 n^{-w}.$$

Here the series converges absolutely in $\operatorname{Re} w > 1$. It is known that $L(u\otimes u, w)$ has meromorphic continuation to the whole complex w-plane and it satisfies a certain functional equation (inherited from that for the Eisenstein series). The point $w = 1$ is a simple pole with residue

$$R = \frac{4}{\pi}\frac{\langle u, u\rangle}{\operatorname{vol}(\Gamma\backslash H)}\sin \pi s.$$

In the half-plane $\operatorname{Re} w \geqslant \frac{1}{2}$ there are no other poles. By a standard application of the convexity principle of Phragmén–Lindelöf it follows from the functional equation that

$$(w-1)L(u\otimes u;\, w) \ll (\lambda|w|N)^A R$$

in $\operatorname{Re} w \geqslant \frac{1}{2}$, where A and the implied constant are absolute. Hence one infers that

(20) $$Rx^{1/2} \ll L(x) \ll Rx^{1/2}$$

for $x \gg (\lambda N)^B$, where B and the implied constants are absolute. Inserting (20) into (19) we obtain $R \ll (\lambda N)^{\varepsilon}$ and conclude that

(21) $$\langle u, u\rangle \ll \frac{\operatorname{vol}(\Gamma\backslash H)}{\sin \pi s}(\lambda N)^{\varepsilon}$$

for any $\varepsilon > 0$, the implied constant depending on ε alone. This completes the proof of Theorem 2.

4. Proof of Theorem 1. Define

$$\Omega(N) = \sum_{1/2<s_j<1} N^{4(s_j-1/2)}, \quad \Omega^*(N) = \sum_{1/2<s_j<1}^{*} N^{4(s_j-1/2)},$$

where s_j ranges over the exceptional points for $\Gamma_0(N)$ (counted with multiplicity) and $\sum^*$ means that the sum is restricted to those j for which $u_j(z)$ is a newform of level N. By (3), (7) and (12) we deduce that

$$\Omega^*(N) \ll N^{1+\varepsilon}\mathscr{G}_N(N^2) = N^{1+\varepsilon}(\mathscr{F}_N(N^2)+O(1)) \ll N^{1+\varepsilon}\log 2N.$$

Hence by (2) and Proposition 3 we obtain

$$\Omega(N) \leqslant \sum_{dM|N} \sum_{\substack{u_j \text{ of level } M\\ \text{and divisor } d}} N^{4(s_j-1/2)}$$
$$\leqslant \sum_{dM|N} M^{-1} N\Omega^*(M) \ll N \sum_{dM|N} M^{\varepsilon}\log 2M \ll N^{1+2\varepsilon},$$

completing the proof of Theorem 1.

5. Sums of Kloosterman sums. In this section we shall estimate the bilinear forms

$$\mathscr{A}_r(P, Q) = \sum_{p\leqslant P}\sum_{q\leqslant Q} \alpha_p \beta_q S_{\chi_p}(1; pqr).$$

It will be clear that the hypothesis that p ranges over primes is important and the same hypothesis about q is imposed for convenience. First we reduce $\mathscr{A}_r(P, Q)$ to

$$\mathscr{A}_r^*(P, Q) = \sum_{\substack{p\leqslant P\\ (p,q)=(pq,2r)=1}}\sum_{q\leqslant Q} \alpha_p \beta_q S_{\chi_p}(1; pqr)$$

up to the error term $O(\|\alpha\|\,\|\beta\|(P+Q)\tau(r))$ by means of Weil's estimate (3). Here the Kloosterman sum factors as follows:

$$S_{\chi_p}(1; pqr) = S_{\chi_p}(\overline{qr}; p)\, S(\bar{p}; qr),$$

where $\overline{qr}$, $\bar{p}$ stand for the multiplicative inverses of qr, p modulo p, qr respectively. The Kloosterman sum $S_{\chi_p}(m, n; p)$ can be evaluated in terms of Gauss sums (see [25]).

LEMMA 1. *We have* $S_{\chi_p}(m, n; p) = 0$ *unless* $mn \equiv l^2 \pmod p$, *in which case*

(22) $$S_{\chi_p}(m, n; p) = \varepsilon_p\left(\left(\frac{m}{p}\right)+\left(\frac{n}{p}\right)\right)\cos\left(\frac{4\pi l}{p}\right)p^{1/2},$$

where $\varepsilon_p = i^{((p-1)/2)^2}$ *is the sign of the Gauss sum.*

In particular, we obtain

(23) $$S_{\chi_p}(\overline{qr}; p) = 2\varepsilon_p\left(\frac{qr}{p}\right)\cos\left(\frac{4\pi\overline{qr}}{p}\right)p^{1/2}.$$

Next by the quadratic reciprocity law

$$\left(\frac{q}{p}\right) = \varepsilon_p^{q-1}\left(\frac{p}{q}\right)$$

and by the "reciprocity" formula

$$\frac{\bar{a}}{b}+\frac{\bar{b}}{a} \equiv \frac{1}{ab} \pmod 1$$

we infer that

$$S_{\chi_p}(\overline{qr}; p) = 2\varepsilon_p^q\left(\frac{r}{p}\right)\left(\frac{p}{q}\right)\cos\left(\frac{4\pi\bar{p}}{qr}\right)p^{1/2}+O(p^{1/2}(pqr)^{-1}),$$

$$S_{\chi_p}(1; pqr) = 2\varepsilon_p^q\left(\frac{r}{p}\right)\left(\frac{p}{q}\right)\cos\left(\frac{4\pi\bar{p}}{qr}\right)p^{1/2}S(\bar{p}; qr)+O((pqr)^{-1/2}\tau(r)),$$

$$\mathscr{A}_r(P, Q) = 2\sum_{\substack{p\leqslant P\\ (p,q)=(pq,2r)=1}}\alpha_p\left(\frac{r}{p}\right)p^{1/2}\sum_{q\leqslant Q}\beta_q\varepsilon_p^q\left(\frac{p}{q}\right)\cos\left(\frac{4\pi\bar{p}}{qr}\right)S(\bar{p}; qr)$$
$$+O(\|\alpha\|\,\|\beta\|(P+Q)\tau(r)).$$

For notational simplicity we introduce

$$K(n; c) = e_c(2n)S(n; c).$$

Then by Cauchy's inequality we obtain

$$|\mathscr{A}_r(P, Q)| \leqslant 4\|\alpha\|P^{1/2}\mathscr{B}_r^{1/2}(P, Q)+O(\|\alpha\|\,\|\beta\|(P+Q)\tau(r)), \quad \text{where}$$

$$\mathscr{B}_r(P, Q) = \sum_{\substack{a\leqslant P\\ (a,r)=1}}\left|\sum_{\substack{q\leqslant Q\\ (q,2ar)=1}}\beta_q\left(\frac{a}{q}\right)K(\bar{a}; qr)\right|^2,$$

a ranges over all integers and q over primes. Notice that we have dropped the factor ε_a^q; this can be justified by splitting the summation over q into classes modulo 4 and changing β_q accordingly. We shall prove

LEMMA 2. *For any complex numbers* β_q *we have*

$$|\mathscr{B}_r(P, Q)| \leqslant 8\|\beta\|^2(PQr+Q^3r^{3/2}\log 3Qr)(\tau(r))^{55}.$$

We have

$$\mathscr{B}_r(P, Q) = \sum_{\substack{q_1,q_2\leqslant Q\\ (q_1q_2,2r)=1}}\sum \beta_{q_1}\bar{\beta}_{q_2}\mathscr{B}_{q_1q_2r}(P), \quad \text{where}$$

$$\mathscr{B}_{q_1q_2r}(P) = \sum_{\substack{a\leqslant P\\ (a,2rq_1q_2)=1}}\left(\frac{a}{q_1q_2}\right)K(\bar{a}; q_1r)\bar{K}(\bar{a}; q_2r).$$

If $q_1 = q_2 = q$ we apply Weil's estimate giving

(24) $$|\mathscr{B}_{qqr}(P)| \leqslant 4Pqr\tau^2(r).$$

Now consider $q_1 \neq q_2$, so $(q_1, q_2) = 1$ because q_1, q_2 are primes. In this case we complete the sum by a Fourier technique (see (3.14) of [17] for example) getting

$$\mathscr{B}_{q_1 q_2 r}(P) = \sum_{|h| < q_1 q_2 r} b_h \mathscr{C}_{q_1 q_2 r}(h), \quad \text{where}$$

$$\mathscr{C}_{q_1 q_2 r}(h) = \sum_{a \,(\mathrm{mod}\, q_1 q_2 r)} K(\bar{a};\, q_1 r) K(\bar{a};\, q_2 r)\left(\frac{a}{q_1 q_2}\right) e\left(\frac{ah}{q_1 q_2 r}\right),$$

$b_0 = P/q_1 q_2 r$ and $2|hb_h| \leqslant 1$ for $h \neq 0$. For the complete sum we shall prove

LEMMA 3. *Let* $(q_1, q_2) = (q_1 q_2, 2r) = 1$. *We have*

$$|\mathscr{C}_{q_1 q_2 r}(h)| \leqslant 4 q_1 q_2 (h, r)^{1/2} r^{3/2} (\tau(r))^{54}.$$

Since for $(c_1, c_2) = 1$ we have $K(n;\, c_1 c_2) = K(n\bar{c}_1;\, c_2) K(n\bar{c}_2;\, c_1)$ we find that $\mathscr{C}_{q_1 q_2 r}(h)$ factors as follows:

$$\mathscr{C}_{q_1 q_2 r}(h) = \left(\frac{r}{q_1 q_2}\right) \mathscr{D}_{q_1}(h\bar{q}_2 \bar{r}^2) \bar{\mathscr{D}}_{q_2}(-h\bar{q}_1 \bar{r}^2) \mathscr{E}_r(\bar{q}_1^2, \bar{q}_2^2, h), \quad \text{where}$$

$$\mathscr{D}_q(c) = \sum_{x \,(\mathrm{mod}\, q)}^{*} K(x;\, q)\left(\frac{x}{q}\right) e\left(\frac{c\bar{x}}{q}\right),$$

$$\mathscr{E}_r(a, b, c) = \sum_{x \,(\mathrm{mod}\, r)}^{*} K(ax;\, r) \bar{K}(bx;\, r) e\left(\frac{c\bar{x}}{r}\right).$$

LEMMA 4. *If q is odd prime we have*

$$(25) \qquad |\mathscr{D}_q(c)| \leqslant 2q.$$

Proof. By Lemma 1 we obtain

$$\mathscr{D}_q(c) = \sum_{y \,(\mathrm{mod}\, q)}^{*} S_{\chi_q}(y+2+\bar{y}, c;\, q)$$

$$= \varepsilon_q q^{1/2} \sum_{y \,(\mathrm{mod}\, q)}^{**} \left(\left(\frac{y+2+\bar{y}}{q}\right) + \left(\frac{c}{q}\right)\right) \cos\left(\frac{4\pi z}{q}\right),$$

where $\sum^{**}$ means that y is restricted to the primitive residue classes modulo q such that the congruence $(y+2+\bar{y})c \equiv z^2 \pmod{q}$ is soluble in z.

If $c \equiv 0 \pmod{q}$ then

$$\sum^{**} = \sum_{y \,(\mathrm{mod}\, q)}^{*} \left(\frac{y+2+\bar{y}}{q}\right) = \sum_{y \not\equiv -1 \,(\mathrm{mod}\, q)} \left(\frac{y}{q}\right) = -\left(\frac{-1}{q}\right),$$

whence $\mathscr{D}_q(c) = -\bar{\varepsilon}_q q^{1/2}$. If $c \not\equiv 0 \pmod{q}$ then

$$\sum^{**} = \left(\frac{c}{q}\right) + \left(\frac{c}{q}\right) \sum_{(y+2+\bar{y})c \equiv z^2 \,(\mathrm{mod}\, q)}^{*}\sum^{*} \cos\left(\frac{4\pi z}{q}\right).$$

Here $y \not\equiv -1 \pmod{q}$, so we may change $z = (y+1)t$ getting

$$2\sum^{*}\sum^{*} = \sum_{\substack{yt^2 \equiv c \,(\mathrm{mod}\, q) \\ y \not\equiv -1 \,(\mathrm{mod}\, q)}}\sum 2\cos\left(\frac{4\pi(y+1)t}{q}\right) = S(2c, 2;\, q) - 1 - \left(\frac{-c}{q}\right),$$

whence

$$(26) \qquad \mathscr{D}_q(c) = \frac{1}{2} \varepsilon_q q^{1/2} \left[\left(\frac{c}{q}\right) S(2c, 2;\, q) + \left(\frac{c}{q}\right) - \left(\frac{-1}{q}\right)\right].$$

Now the result follows from Weil's estimate for the classical Kloosterman sums.

We proceed to estimating the sum $\mathscr{E}_r(a, b, c)$ of modulus r. This can be reduced to sums of prime power moduli by means of the factorization

$$\mathscr{E}_r(a, b, c) = \mathscr{E}_{r_1}(a, b, c\bar{r}_2^2) \mathscr{E}_{r_2}(a, b, c\bar{r}_1^2)$$

valid for $r = r_1 r_2$ with $(r_1, r_2) = 1$.

LEMMA 5. *If $r = l^\alpha$ with $\alpha \geqslant 2$, l prime, $l \nmid 2ab$ then*

$$(27) \qquad |\mathscr{E}_r(a, b, c)| \leqslant 4(c, r)^{1/2} r^{3/2} \tau(r).$$

Proof. If $l \nmid 2n$ we have (see [33])

$$K(n;\, r) = \left(\frac{n}{r}\right) r^{1/2} \left[\chi_4(r) e\left(\frac{4n}{r}\right) + 1\right].$$

Hence

$$\mathscr{E}_r(a, b, c) = \left(\frac{ab}{r}\right) r\, [S(2a-2b, c;\, r) + \chi_4(r) S(2a, c;\, r)$$

$$+ \chi_4(r) S(-2b, c;\, r) + S(0, c;\, r)]$$

and (27) follows by (3) (or directly by explicit computation of the Kloosterman sums).

LEMMA 6. *If $r = 2^\alpha$ with $\alpha \geqslant 1$ and $2 \nmid ab$ then*

$$(28) \qquad |\mathscr{E}_r(a, b, c)| \leqslant 4(c, r)^{1/2} r^{3/2} \tau(r).$$

Proof. Similar to that of Lemma 5.

LEMMA 7. *If r is prime, $r \nmid 2ab$ and $r | c$ then*

$$(29) \qquad \mathscr{E}_r(a, b, c) = \delta_{ab} r^2 - \left(\frac{ab}{r}\right) r - r - 1,$$

where $\delta_{ab} = 1$ if $a \equiv b \pmod{r}$ and $\delta_{ab} = 0$ otherwise.

Proof. We have

$$\mathscr{E}_r(a, b, c) = \sum_{x \,(\mathrm{mod}\, r)}^{*} K(ax; r)\bar{K}(bx; r) = r\mathscr{N}(a, b)-(r-1)^2,$$

where $\mathscr{N}(a, b)$ is the number of points on the curve

$$az(y+1)^2 = by(z+1)^2$$

over F_r^*. If $a \equiv b \pmod r$ the curve consists of the line $y = z$ and the hyperbola $yz = 1$. There are $r-1$ points on either component and 2 points on the intersection. Thus $\mathscr{N}(a, b) = 2r-4$ and (29) follows. If $a \not\equiv b \pmod r$ then all points except $(-1, -1)$ are parametrized by $by = azu^2$, $y+1 = (z+1)u$, where $u \neq 0, 1, b/a$ and $u^2 \neq b/a$. Thus $\mathscr{N}(a, b) = r-3-\left(\frac{ab}{r}\right)$ and (29) follows.

Now it remains to estimate $\mathscr{E}_r(a, b, c)$ for r prime with $r \nmid 2abc$. To this end we shall appeal to the Riemann hypothesis for varieties. Deligne's celebrated work [6] does not automatically supply the estimate. It reduces the problem to computing certain l-adic cohomology groups. The complexity of the latter in the general case is formidable. Recently A. Adolphson and S. Sperber [1], [2] succeeded in carrying out the computation of cohomology associated with exponential sums for Laurent polynomials. Their results are quite explicit (expressed in terms of the Newton polyhedron) but unfortunately some hypotheses about the polynomial fail to hold in the case of our sum.

E. Bombieri [4] found ways to go around cohomology by exploiting a bit more than the Riemann hypothesis from Deligne's profound theory, namely the invariance of valuations of roots under the Galois conjugation. In many cases some *ad hoc* estimates for averaged sums may provide an adequate information about the number and magnitude of the roots (see the survey article of C. Hooley [13] in which special sums in two variables are considered). We shall apply similar ideas for sums $\mathscr{E}_r(a, b, c)$ in three variables. We take advantage of Kloosterman sums coupling in $\mathscr{E}_r(a, b, c)$ when estimating the second moment by duality principles.

We set

$$f(x) = x+2+x^{-1}, \quad f(x, y) = af(x)+bf(y), \quad F(x, y, z) = f(x, y)z+cz^{-1}$$

and consider the sum

$$S_q(F) = \sum\sum\sum_{x,y,z \in F_q^*} \psi(F(x, y, z)),$$

where F_q is the field of $q = p^n$ elements and ψ is a non-trivial additive character of F_q. Thus for $q = p$ we have $S_p(F) = \mathscr{E}_p(a, b, c)$.

LEMMA 8. *If $p \nmid 2abc$ we have*

$$\sum_{u \in F_q^*} |S_q(uF)|^2 \leqslant 32q^4. \tag{30}$$

Proof. For $u \in F_q^*$ we have

$$S_q(uF) = \sum_{t \in F_q} \sum_{z \in F_q^*} N(t)\psi(u(tz+cz^{-1})) = \sum_{t \in F_q} \sum_{z \in F_q^*} n(t)\psi(u(tz+cz^{-1})),$$

where $N(t)$ is the number of points on the curve $f(x, y) = t$ over F_q^* and

$$n(t) = N(t)-q^{-1}(q-1)^2 = q^{-1} \sum_{u \in F_q^*} \bar{\psi}(ut) \sum\sum_{x,y \in F_q^*} \psi(uf(x, y)). \tag{31}$$

Thus for any complex numbers λ_u we have

$$S := \sum_{u \in F_q^*} \lambda_u S_q(uF) = \sum_{t \in F_q} n(t) \sum\sum_{u,z \in F_q^*} \lambda_u \psi(u(tz+cz^{-1})).$$

By Cauchy's inequality we obtain $|S|^2 \leqslant AB$, where

$$A = \sum_{t \in F_q} n^2(t), \quad B = \sum_{t \in F_q} \Big|\sum\sum_{u,z \in F_q^*} \lambda_u \psi(u(tz+cz^{-1}))\Big|^2.$$

By (31), Plancherel's theorem and Weil's estimate for Kloosterman sums we obtain

$$A = q^{-1} \sum_{u \in F_q^*} \Big|\sum\sum_{x,y \in F_q^*} \psi(uf(x, y))\Big|^2 \leqslant q^{-1} \sum_{u \in F_q^*} |S(u; q)|^4 \leqslant 16(q-1)q.$$

Next we have

$$\begin{aligned} B &= q \sum\sum\sum\sum_{uz = u_1 z_1 \neq 0} \lambda_u \bar{\lambda}_{u_1} \psi(cuz^{-1}-cu_1 z_1^{-1}) \\ &= q \sum\sum_{u,u_1 \in F_q^*} \lambda_u \bar{\lambda}_{u_1} \sum_{z \in F_q^*} \psi((u^2-u_1^2)z) \\ &= q^2 \sum\sum_{\substack{u,u_1 \in F_q^* \\ u = \pm u_1}} \lambda_u \bar{\lambda}_{u_1} - q\Big|\sum_{u \in F_q^*} \lambda_u\Big|^2 \leqslant 2q^2 \sum_{u \in F_q^*} |\lambda_u|^2. \end{aligned}$$

Setting $\lambda_u = \bar{S}_q(uF)$ we deduce (30) by combining the above inequalities.

Now we appeal to the theory of the zeta-function

$$L^*(F, t) = \exp\Big(\sum_{n=1}^{\infty} S_{p^n}(F)t^n/n\Big).$$

B. Dwork [9] proved that $L^*(F, t)$ is a rational function,

$$L^*(F, t) = \prod_j (1-\omega_j t)^{d_j},$$

say, where the ω_j are distinct complex numbers from the cyclotomic field $\mathbf{Q}(\zeta_p)$. P. Deligne [6] proved the Riemann hypothesis asserting that

$$|\omega_j| = p^{m_j/2},$$

where the m_j are non-negative integers, and E. Bombieri [4] proved that the number of roots ω_j^{-1} counted with the multiplicities $|d_j|$ is bounded by a constant independent of p.

In our case Theorem 2 of [4] yields

$$\sum_j |d_j| \leqslant 17^{13}.$$

It remains to estimate the weights m_j. Evaluating $(d^n/dt^n)\log L^*(F, t)$ at $t = 0$ we get

$$S_q(F) = -\sum_j d_j \omega_j^n.$$

Besides the Riemann hypothesis Deligne proved that under Galois conjugation the roots do not change weight. Let $0 < u < p$ and let ω_{ju} denote the conjugation of ω_j under the automorphism of $Q(\zeta_p)$ over Q which raises ζ_p to the power ζ_p^u. Then the ω_{ju} are roots for $S_q(uF)$ with $|\omega_{ju}| = p^{m_j/2}$ and

$$\sum_{0<u<p} \Big|\sum_j d_j \omega_{ju}^n\Big|^2 = \sum_{0<u<p} |S_q(uF)|^2 \leqslant \sum_{u\in F_q^*} |S_q(uF)|^2 \leqslant 32 p^{4n}$$

by Lemma 8. Suppose there are roots of weight $\geqslant 4$. Let $d_1, \ldots, d_l$ be the multiplicities of roots of the highest weight. As in (5.1) of [4] we infer from the above inequality that

$$(p-1)(d_1^2 + \ldots + d_l^2) \leqslant 32,$$

so $p \leqslant 33$. We conclude that for $p > 33$ all roots have weight $\leqslant 3$ and that

$$|S_q(F)| \leqslant 17^{13} q^{3/2}.$$

In particular, we have

LEMMA 9. *If $p \nmid 2abc$ then*

$$|\mathscr{E}_p(a, b, c)| \leqslant 17^{13} p^{3/2}. \tag{32}$$

Here the restriction $p > 33$ is not necessary because the result is trivial for small p. The constant 17^{13} can be reduced but this is not worth the effort. It is easy to show that

$$|S_q(F)| \leqslant \sum_{z\in F_q^*} |S(z; q)|^2 = q^2 - 2q - 1.$$

Collecting together Lemmas 5–7 and 9 we get

LEMMA 10. *If $(r, ab) = 1$ we have*

$$|\mathscr{E}_r(a, b, c)| \leqslant (c, r)^{1/2} r^{3/2} (\tau(r))^{54}. \tag{33}$$

Next we obtain Lemma 3 by Lemmas 4 and 10. Then by Lemma 3 we infer that

$$|\mathscr{B}_{q_1 q_2 r}(P)| \leqslant 4r (P + q_1 q_2 r^{1/2} \log 3 q_1 q_2 r)(\tau(r))^{55} \tag{34}$$

as long as $(q_1, q_2) = (q_1 q_2, r) = 1$. From (24) and (34) we conclude the assertion of Lemma 2. Finally, by Lemma 2 we complete the proof of Theorem 3.

6. Proof of Theorem 4. If $\{u_j(z)\}$ is an orthonormal basis of Maass cusp forms in $\mathscr{L}(\Gamma_0(p)\backslash H, \chi_p)$ then the system $\{\mu^{-1/2} u_j(z)\}$ with $\mu = [\Gamma_0(p) : \Gamma_0(pq)] \leqslant q+1$ can be completed to an orthonormal basis of Maass cusp forms in $\mathscr{L}(\Gamma_0(pq)\backslash H, \chi_p)$. From this observation we infer that

$$(q+1)^{-1} \mathscr{G}_p(X) \leqslant \mathscr{G}_{pq}(X) = \mathscr{F}_{pq}(X) + O(1)$$

because the terms in $\mathscr{G}_N(X)$ are non-negative, see the formulas prior to (7). We sum over p, q with $P < p \leqslant 2P$, $Q < q \leqslant 2Q$ and get

$$\begin{aligned}(\log Q)^{-1} \sum_p \mathscr{G}_p(X) &\ll \sum_p \sum_q \sum_r (pqr)^{-1} f\left(\frac{4\pi X}{pqr}\right) S_{\chi_p}(1; pqr) + PQ \\ &\ll \sum_r r^{-1/2} (P^{1/2} + r^{1/4} Q \log 3rQ)(\tau(r))^{28} + PQ \\ &\ll (P^{1/2} R^{1/2} + QR^{3/4})(\log X)^{2^{28}} + PQ\end{aligned}$$

by Theorem 3. In the above r ranges over integers with $R \ll r \ll R$ and R is given by $PQR = X$. We choose $Q = P^{1/6}$ and $X = P^{5/2}$ giving

$$\sum_{P<p\leqslant 2P} \sum_{1/2<s_j<1} |\varrho_j(1)|^2 P^{5(s_j - 1/2)} \ll P^{7/6} (\log P)^{2^{28}+1}. \tag{35}$$

Here the $\varrho_j(1)$ are the first Fourier coefficients of $u_j(z)$. We can assume that the $u_j(z)$ are newforms because the old forms on $\Gamma_0(p)$ have level 1, so they are not exceptional. Then Theorem 2 is applicable giving $|\varrho_j(1)|^2 \gg P^{-1-\varepsilon}$. This and (35) yield the assertion of Theorem 4.

References

[1] A. Adolphson and S. Sperber, *Exponential sums and Newton polyhedra*, Bull. Amer. Math. Soc. 16 (1987), 282–286.

[2] — *Exponential sums and Newton polyhedra: Cohomology and estimates*, to appear.

[3] A. O. L. Atkin and J. Lehner, *Hecke operators on $\Gamma_0(m)$*, Math. Ann. 185 (1970), 134–160.

[4] E. Bombieri, *On exponential sums in finite fields*, Invent. Math. 47 (1978), 29–39.

[5] P. Buser, *Riemannsche Flächen mit Eigenwerten in (0, 1/4)*, Comment. Math. Helv. 52 (1977), 25–34.

[6] P. Deligne, *La conjecture de Weil. I*, Publ. Math. I.H.E.S 43 (1974), 273–307.

[7] J.-M. Deshouillers and H. Iwaniec, *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. Math. 70 (1982), 219–288.

[8] — *Accumulation of the eigenvalues of the hyperbolic Laplacian at 1/4*, preprint, 1983.

[9] B. Dwork, *On the rationality of the zeta function of an algebraic variety*, Amer. J. Math. 82 (1962), 631–648.

[10] S. S. Gelbart and H. Jacquet, *A relation between automorphic representations of* GL(2) *and* GL(3), Ann. Sci. École Norm. Sup. 11 (1978), 471–542.

[11] D. Goldfeld and P. Sarnak, *Sums of Kloosterman sums*, Invent. Math. 71 (1983), 243–250.

[12] D. Hejhal, *The Selberg Trace Formula*, Vol. 2, Lecture Notes in Math. 1001, Springer-Verlag, 1983.

[13] C. Hooley, *On exponential sums and certain of their applications*, in *Journées Arithmétiques 1980*, Cambridge Univ. Press, 1982, 92–122.

[14] M. Huxley, *Introduction to Kloostermania*, in Banach Center Publ. 17, PWN, 1985, 217–306.
[15] H. Iwaniec, *Non-holomorphic modular forms and their applications*, in *Modular Forms*, Proc. 1983, Halsted Press, New York 1984, 157–196.
[16] — *Character sums and small eigenvalues for $\Gamma_0(p)$*, Glasgow Math. J. 27 (1985), 99–116.
[17] — *Fourier coefficients of modular forms of half-integral weight*, Invent. Math. 87 (1987), 385–401.
[18] — *Selberg's lower bound of the first eigenvalue for congruence groups*, in *Number Theory, Trace Formulas and Discrete Groups*, Proc. 1987, Academic Press, San Diego 1989, 371–375.
[19] H. Iwaniec and J. Szmidt, *Density theorems for exceptional eigenvalues of Laplacian for congruence groups*, in Banach Center Publ. 17, PWN, 1985, 317–331.
[20] R. Langlands, *Problems in the theory of automorphic forms*, in Lecture Notes in Math. 170, Springer-Verlag, 1970, 18–61.
[21] W. Li, *Newforms and functional equations*, Math. Ann. 212 (1975), 285–315.
[22] C. Moreno and F. Shahidi, *The L-functions $L(s, \mathrm{Sym}^m(r), \pi)$*, Canad. Math. Bull. 28 (4) (1985), 405–410.
[23] R. Murty, *On the estimation of eigenvalues of Hecke operators*, Rocky Mountain J. Math. 15 (2) (1985), 521–533.
[24] W. Roelcke, *Über die Wellengleichung bei Grenzkreisgruppen erster Art*, S.-B. Heidelberger Akad. Wiss. Math. Nat. Kl., 1956, 4. Abh.
[25] H. Salié, *Über die Kloostermanschen Summen $S(u, v; q)$*, Math. Z. 34 (1934), 91–109.
[26] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. 20 (1956), 47–87.
[27] — *On the estimation of Fourier coefficients of modular forms*, in Proc. Sympos. Pure Math. 8, A.M.S., Providence, R.I., 1965, 1–15.
[28] J.-P. Serre, *Abelian l-Adic Representations and Elliptic Curves*, Benjamin, New York–Amsterdam 1968.
[29] — Letter to J.-M. Deshouillers of August 1981.
[30] A. Smith, *The L^2-norm of Maass wave functions*, Proc. Amer. Math. Soc. 82 (1981), 179–182.
[31] L. A. Takhtadzhyan and A. I. Vinogradov, *The Gauss–Hasse hypothesis on real quadratic fields with class number one*, J. Reine Angew. Math. 335 (1982), 40–86.
[32] — — *An estimate of the residuum of Rankin's L-series* (in Russian), Dokl. Akad. Nauk SSSR 267 (1) (1982), 30–34.
[33] K. S. Williams, *Note on the Kloosterman sum*, Proc. Amer. Math. Soc. 30 (1) (1971), 61–62.

DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
Hill Center for the Mathematical Sciences
Bush Campus
New Brunswick, New Jersey 08903, USA

Received on 17.3.1989 (1915)

ACTA ARITHMETICA
LVI (1990)

A matrix paraphrase of Kloosterman sums

by

D. H. Lehmer (Berkeley, Calif.)

1. Introduction. In 1967 Lehmer and Lehmer [3] showed that there was a strong connection between the cyclotomic periods and the ordinary Kloosterman sums

(1) $$s_h = \sum_{x=1}^{p-1} e^{2\pi i(x+h\bar{x})/p} \quad (h = 0, 1, \ldots, p-1)$$

where $\bar{x} \equiv 1/x \pmod p$ and the Gaussian periods

$$\sum_{y=0}^{f-1} e^{2\pi i g^{ey+h}/p}$$

where $p = ef+1$ and g is a primitive root of the odd prime p. In this paper we exploit this connection to give a matrix paraphrase of the Kloosterman sum and its periods.

2. Notation. Throughout the paper capital letters are reserved for matrices. The matrices will be of special kind known as circulants. A *circulant* is an n by n matrix of the form

$$M = \begin{bmatrix} a_0 & a_1 & a_2 \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 \cdots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 \cdots & a_{n-3} \\ \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & a_3 \cdots & a_0 \end{bmatrix}.$$

The matrix M depends only on its first row. To save space we will write M as follows:

(2) $$M = \mathrm{cir}(a_0, a_1, \ldots, a_{n-1}).$$

We number the rows and columns of M from 0 to $n-1$ to allow the use of residue classes modulo n. If we denote the element in the ith row and jth column by α_{ij} we have

(3) $$\alpha_{ij} = a_{j-i}$$