Invent. math. 71, 243–250 (1983)

Inventiones mathematicae

Sums of Kloosterman Sums

D. Goldfeld[1] and P. Sarnak[2]

[1] Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Courant Institute of Mathematical Sciences, New York, NY 10012, USA

1. Introduction

The classical Kloosterman sum is defined by

$$S(m,n,c)=\sum_{\substack{a=1\\(a,c)=1\\ad\equiv 1\,(\mathrm{mod}\,c)}}^{c} e\left(\frac{am+dn}{c}\right)$$

where $m,n\in Z^+$ and $e(z)=e^{2\pi iz}$. It is known to satisfy [see 14]

$$|S(m,n,c)|\leqq d(c)\,c^{1/2}(m,n,c)^{1/2} \tag{1.1}$$

where (m,n,c) denotes the greatest common divisor, and $d(c)$ the number of divisors of c.

Many problems in number theory, especially additive problems may be reduced to estimating sums of the type

$$\sum_{c\leqq x}\frac{S(m,n,c)}{c}.$$

One expects considerable cancellation in this sum, and Linnik [8], and Selberg [12] have conjectured that

$$\sum_{c\leqq x}\frac{S(m,n,c)}{c}=O(x^{\varepsilon})\qquad\forall\,\varepsilon>0$$

and $x>(m,n)^{\frac{1}{2}+\varepsilon}$.

Recently, Kuznetsov [7] was able to show that

$$\sum_{c\leqq x}\frac{S(m,n,c)}{c}\underset{\varepsilon,m,n}{\ll} x^{1/6+\varepsilon}\qquad(\varepsilon>0).$$

0020-9910/83/0071/0243/$01.60

This result was generalized by Deshouillers and Iwaniec [4] to Kloosterman sums associated to congruence subgroups, and to ones associated to a general discrete subgroup with finite volume by Proskurin [9]. In the last case the 1/6 is replaced by 1/3 while in both generalizations exceptional eigenvalues (see (2.1)) may create additional terms. All of the above results are based on a general trace formula developed in [7] and [1].

In the following Γ will denote a subgroup of $SL_2(Z)$ of finite index, though there is no need to make this restriction. Let q be the smallest positive integer such that

$$\begin{pmatrix} 1 & q \\ 0 & 1 \end{pmatrix} \in \Gamma$$

and put

$$\Gamma_\infty = \left\{ \begin{pmatrix} 1 & nq \\ 0 & 1 \end{pmatrix} : n \in Z \right\}$$

As in [12] let $k \in R^+$, and χ be a multiplier for this k. Then for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\gamma' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in \Gamma$ we must have the consistency condition

$$\chi(-1) = e^{ik\pi}$$
$$j_{\gamma\gamma'}(z)^k \, \chi(\gamma\gamma') = j_\gamma(\gamma'(z))^k \, j_{\gamma'}(z)^k \, \chi(\gamma) \, \chi(\gamma')$$

where $j_\gamma(z) = cz + d$. We also fix $0 \leqq \alpha < 1$ so that

$$\chi\left(\begin{pmatrix} 1 & q \\ 0 & 1 \end{pmatrix}\right) = e^{-2\pi i\alpha}$$

The notation is as in [12].

We define the generalized Kloosterman sum, $c > 0$

$$S(m, n, c, \chi) = \sum_{\substack{0 \leqq a < qc \\ 0 \leqq d < qc \\ \gamma \in \Gamma}} \overline{\chi(\gamma)} \, e\left(\frac{(m-\alpha)\,a + (n-\alpha)\,d}{qc}\right)$$

and Selberg's Kloosterman zeta function

$$Z_{m,n}(s, \chi) = \sum_{c>0} \frac{S(m, n, c, \chi)}{c^{2s}}$$

Our aim in this paper is to prove the following theorems.

Theorem 1. *The function $Z_{m,n}(s, \chi)$ is meromorphic in* $\mathrm{Re}(s) > \frac{1}{2}$ *with at most a finite number of simple poles all in $(\frac{1}{2}, 1)$, and satisfies the growth condition*

$$|Z_{m,n}(s, \chi)| = O\left(\frac{|s|^{1/2}}{\sigma - \frac{1}{2}}\right)$$

for $s = \sigma + it$, $\sigma > \frac{1}{2}$ as $t \to \infty$. The implied constant depends on Γ, χ, k, m, n.

Theorem 2. *Let*

$$\beta = \overline{\lim_{c\to\infty}} \frac{\log|S(m,n,c,\chi)|}{\log c}$$

then

$$\sum_{c \leqq x} \frac{S(m,n,c,\chi)}{c} = \sum_{j=1}^{l} \tau_j x^{\alpha_j} + O(x^{\beta/3+\varepsilon}) \qquad (\varepsilon > 0)$$

for certain constants τ_j *and real* $0<\alpha_j<1$. *The implied constant depends only on* Γ, χ, k, m, n.

Remarks and Further Results. 1. The meromorphic continuation of $Z_{m,n}(s,\chi)$ to the entire complex plane was first obtained by Selberg [12]. The estimate on $Z_{m,n}(s,\chi)$ for large t, which is usually difficult to develop for these types of functions, is new. By a careful choice of constants in our proof we may replace the $O\left(\frac{|s|^{1/2}}{\sigma-\frac{1}{2}}\right)$ by $O\left(\frac{|mn|}{q^2}\mathrm{Vol}(\Gamma\backslash\mathscr{H})\frac{|s|^{1/2}}{|\sigma-\frac{1}{2}|}\right)$ with the implied constant depending at most on k. It is likely that this is the true order of growth.

2. The number τ_j appears only if there are exceptional eigenvalues in the interval $(0,\frac{1}{4})$ for the Laplacian on $\Gamma\backslash\mathscr{H}$ [see (2.1)].

The τ_j's are computed explicitly in (3.3). It follows from Weil's estimate (1.1) that $\beta \leqq \frac{1}{2}$ for congruence subgroups. In general all we know is that $\beta \leqq 1$. Thus Theorem 2 agrees with the results in [7, 3 and 9]. As in Remark 1, the O-constant can be changed similarly.

3. Interesting examples are the groups $\Gamma = \Gamma_0(4N)$, $N \in Z^+$ and $k=\frac{1}{2}$.

It follows by [13] that $\beta = \frac{1}{2}$ for these sums.

Indeed in these cases we have been able to show that aside from the exceptional eigenvalue at $\lambda = \frac{3}{16}$, due to θ-functions all exceptional eigenvalues are greater than or equal to 15/64. It follows from Theorem 2 that

$$\sum_{c \leqq x} \frac{S(m,n,c,\chi)}{c} = A x^{1/2} + O(x^{1/4})$$

$(A = A(m,n,\chi))$.

Using further work of Vardi [13], one finds that the last asymptotic relation is closely related to the distribution of solutions of certain quadratic congruences.

4. In the case of weight $k=1$, there are no exceptional eigenvalues [11]. Let $N \in \mathbb{N}$, and let χ be an odd Dirichlet character to modulus N, i.e. $\chi(-1) = -1$. Define the Kloosterman sums (see [2])

$$S(m,n,Nc,\chi) = \sum_{\substack{a \bmod Nc \\ d \bmod Nc \\ ad \equiv 1 \pmod{Nc}}} \chi(d)\, e\left(\frac{ma+nd}{c}\right), \qquad m,n,c>0$$

It follows from Weil's estimates that $\beta = \frac{1}{2}$, and since these sums come up in odd weight, we obtain using Theorem 2,

Theorem 4.

$$\sum_{c \leq x} \frac{S(m,n,Nc,\chi)}{c} \ll x^{\frac{1}{6}+\varepsilon} \qquad \forall \varepsilon > 0.$$

2. Proof of Theorem 1

Let $L^2(\Gamma \backslash \mathscr{H}, \chi, k)$ be the Hilbert space of functions $g: \mathscr{H} \to C$ satisfying

(i) $$g(\gamma z) = \chi(\gamma) \left(\frac{cz+d}{|cz+d|}\right)^k g(z), \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$$

(ii) $$\iint_{D_\Gamma} |g(z)|^2 \frac{dx\,dy}{y^2} < \infty$$

where D_Γ is a fundamental domain for $\Gamma \backslash \mathscr{H}$.

The operator

$$\Delta_k = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - iky \frac{\partial}{\partial x} \tag{2.1}$$

has a self acjoint extension to $L^2(\Gamma \backslash \mathscr{H}, x, k)$, with real spectrum. As pointed out in [12], the spectrum below 1/4 is finite dimensional. Let $u_1, u_2, \ldots, u_l$ and $\lambda_1, \ldots, \lambda_l < \frac{1}{4}$, with

$$\Delta_k u_j + \lambda_j u_j = 0, \qquad 1 \leq j \leq l$$

be the corresponding normalized eigenfunctions and eigenvalues. By separation of variables, such an eigenfunction has a Fourier expansion of the type

$$u_j(z) = \sum_{n=-\infty}^{\infty} \hat{u}_j(x,y)\, e\left(\frac{n-\alpha}{q} x\right) \tag{2.2}$$

where

$$\hat{u}_j(n,y) = \rho_j(n)\, W_{\frac{k}{2}\operatorname{sgm}\left(\frac{n-\alpha}{q}\right),\, s_j - \frac{1}{2}}\left(4\pi \left|\frac{n-\alpha}{q}\right| y\right),$$

if $n - \alpha \neq 0$ and

$$\hat{u}_j(0,y) = \rho_j(0)\, y^{s_j} + \rho_j'(0)\, y^{1-s_j}, \qquad \text{if } n = \alpha = 0,$$

where $\lambda_j = s_j(1-s_j)$, $\rho_j(n)$ are constants, and $W_{\beta,\mu}(y)$ is the Whittaker function which decays exponentially in y as $y \to \infty$ and satisfies the ordinary differential equation

$$\frac{d^2 W}{dy^2} + \left(-\frac{1}{4} + \frac{\beta}{y} + \frac{\frac{1}{4} - \mu^2}{y^2}\right) W = 0$$

We shall have frequent recourse to the following Mellin transform [5, p. 860]

$$\int_0^\infty e^{-2\pi N y} y^s W_{\beta,\mu}(4\pi N y) \frac{dy}{y} = (4\pi N)^{-s} \frac{\Gamma(s+\frac{1}{2}+\mu)\,\Gamma(s+\frac{1}{2}-\mu)}{\Gamma(s-\beta+1)} \tag{2.3}$$

which holds for $N > 0$ and $\operatorname{Re}(s + \frac{1}{2} \pm \mu) > 0$.

Following Selberg [12], we introduce the nonholomorphic Poincaré series for $m>0$, $m\in Z$,

$$P_m(z,s,\chi)=P_m(z,s,\chi,k)$$
$$=\sum_{\gamma\in\Gamma_\infty\setminus\Gamma}\overline{\chi(\gamma)}\left(\frac{cz+d}{|cz+d|}\right)^{-k}e\left(\frac{m-\alpha}{q}\gamma(z)\right)\frac{y^s}{|cz+d|^{2s}}$$

where $\gamma=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$.

Just as in the case of ordinary Eisenstein series the above converges absolutely and uniformly in $\mathrm{Re}(s)>1$. Since $m>0$ it is also clear that $P_m(z,s,x)\in L^2(\Gamma\backslash\mathscr{H},\chi,k)$. Furthermore, $P_m(z,s,\chi)$ satisfies the following recursion relation,

$$P_m(z,s,\chi)=-4\pi\left(\frac{m-\alpha}{q}\right)\left(s-\frac{k}{2}\right)R_{s(1-s)}(P_m(z,s+1,\chi)) \tag{2.4}$$

for $\mathrm{Re}(s)>1$, where

$$\mathscr{R}_\lambda=(\Delta_k+\lambda)^{-1}$$

is the resolvent of Δ_k. By the remark following (2.1) it is clear that $R_{s(1-s)}$ is holomorphic in $\mathrm{Re}(s)>\frac{1}{2}$ except possibly at the points s_j, $j=1,2,\dots,l$. It then follows from (2.4) that $P_m(z,s,\chi)$ may be meromorphically continued to $\mathrm{Re}(s)>\frac{1}{2}$ with at most a finite number of simple poles at the points $s=s_j$ for $\frac{1}{2}<s_j<1$.

The residue is given by the following

$$\operatorname*{Res}_{s=s_j}\langle P_m(\cdot,s,\chi),u_j\rangle\, u_j(z)$$

Here

$$\langle P_m(\cdot,s,\chi),u_j\rangle=\iint_D P_m(z,s,\chi)\,\overline{u_j(z)}\,\frac{dx\,dy}{y^2}$$
$$=q\overline{\rho_j(m)}\int_0^\infty W_{\frac{k}{2},s_j-\frac{1}{2}}\left(\pi\left(\frac{m-\alpha}{q}\right)y\right)e^{-2\pi\left(\frac{m-\alpha}{q}\right)y}\,y^{s-1}\frac{dy}{y}$$
$$=q\overline{\rho_j(m)}\left(\frac{4\pi(m-\alpha)}{q}\right)^{1-s}\frac{\Gamma(s+s_j-1)\,\Gamma(s-s_j)}{\Gamma(s-k/2)}$$

after using (2.3). Hence

$$\operatorname*{Res}_{s=s_j}P_m(z,s,\chi)=q\overline{\rho_j(m)}\,4\pi\left(\frac{m-\alpha}{q}\right)^{1-s_j}\frac{\Gamma(2s_j-1)}{\Gamma(s_j-k/2)}u_j(z) \tag{2.5}$$

Lemma 1. *Let $s=\sigma+it$, for $\frac{1}{2}<\sigma\leqq 2$ and $|t|>1$*

$$\iint_{D_\Gamma}|P_m(z,s,\chi)|^2\frac{dx\,dy}{y^2}=O\left(\frac{m^2}{(\sigma-\frac{1}{2})^2}\right)$$

where the implied constant depends only on Γ, χ and k.

Proof. First for $\frac{3}{2} \leqq \sigma \leqq 3$

$$|P_m(z,s)| = O(1)$$

uniformly in z and t since the series converges absolutely.

It then follows from (2.4) and the bound [10, pp. 342]

$$|\mathscr{R}_\lambda| \leqq \frac{1}{\text{distance } (\lambda, \text{ spectrum } (\Delta_k))}$$

that

$$\left(\iint_D |P_m(z,s,\chi)|^2 \frac{dx\,dy}{y^2}\right)^{1/2} = O\left(\frac{m-\alpha}{q}\right) \frac{\left|s-\frac{k}{2}\right|}{|t|\,|2\sigma-1|}$$

since $\operatorname{dist}(s(1-s), \text{spectrum } \Delta_k) \geqq \operatorname{Im}(s(1-s)) = |t(2\sigma-1)|$.

Lemma 2. *For $m,n>0$, $\sigma>\frac{1}{2}$, we have*

$$\iint_{D_\Gamma} P_m(z,s,\chi)\,\overline{P_n(z,\bar{s}+2,\chi)}\,\frac{dx\,dy}{y^2}$$
$$= (-i)^k\, 4^{-s-1}\, \pi^{-1} \left(\frac{n-\alpha}{q}\right)^{-2} \frac{\Gamma(2s+1)}{\Gamma(s+k/2)\,\Gamma(s-k/2+2)}\, Z_{m,n}(s,\chi) + R(s),$$

where $R(s)$ is holomorphic in $\sigma>\frac{1}{2}$ and $|R(s)| = O(1/\sigma-\frac{1}{2}|)$ in this region.

Proof. We compute

$$\iint_{D_\Gamma} P_m(z,s,\chi)\,\overline{P_n(z,\omega,\chi)}\,\frac{dx\,dy}{y^2} = \delta_{m,n}\left(\frac{4\pi(n-\alpha)}{q}\right)^{1-s-\bar{\omega}} \Gamma(s+\bar{\omega}-1)$$
$$+\sum_{c\neq 0} \frac{S(m,n,c,\chi)}{|c|^{2s}} \int_0^\infty \int_{-\infty}^\infty \frac{y^{\bar{\omega}-s}}{(x^2+1)^s} \left(\frac{x+i}{(x^2+1)^{1/2}}\right)^{-k}$$
$$\cdot e\left[-\left(\frac{m-\alpha}{q}\right)\left(\frac{1}{yc^2(x+i)}\right) - \left(\frac{n-\alpha}{q}\right)(xy-iy)\right]\cdot\frac{dx\,dy}{y}$$

Using the formula [5, p. 321],

$$\int_{-\infty}^\infty \frac{(x+i)^{-k}}{(x^2+1)^{s-k/2}}\, e^{-2\pi i\left(\frac{m-\alpha}{q}\right)xy}\, dx = \frac{-\pi(-i)^k \left(y\pi\left(\frac{m-\alpha}{q}\right)\right)^{s-1}}{\Gamma(s+k/2)}\, W_{\frac{k}{2},s-\frac{1}{2}}\left(\frac{4\pi(m-\alpha)}{q}y\right)$$

in conjunction with (2.3) it follows on setting $\omega = \bar{s}+2$ that

$$\langle P_m(\cdot,s,\chi), P_n(\cdot,\bar{s}+2,\chi)\rangle = \delta_{m,n}\left(\frac{4\pi(n-\alpha)}{q}\right)^{-2s-1}\Gamma(2s+1)$$
$$+(-i)^k\, 4^{-s-1}\,\pi^{-1}\left(\frac{n-\alpha}{q}\right)^{-2}\frac{\Gamma(2s+1)}{\Gamma(s+k/2)\,\Gamma(s-k/2+2)}\, Z_{m,n}(s,\chi)$$
$$+\sum_{c\neq 0}\frac{S(m,n,c)}{|c|^{2s}}\, R_{m,n}(s,c)$$

where

$$R_{m,n}(s,c)=\int_0^\infty \int_{-\infty}^\infty \frac{y^2}{(x^2+1)^s}\left(\frac{x+i}{(x^2+1)^{1/2}}\right)^{-k} \cdot\left(\exp\left\{-2\pi i\left(\frac{m-\alpha}{q}\right)\frac{x-i}{yc^2(x^2+1)}\right\}-1\right) \cdot\exp\left\{-2\pi i\left(\frac{n-\alpha}{q}\right)(xy-iy)\right\}\frac{dx\,dy}{y}$$

Now

$$\int_0^\infty y\left|\exp\left\{-2\pi i\left(\frac{m-\alpha}{q}\right)\frac{x-i}{yc^2(x^2+1)}\right\}-1\right|\exp\left\{-2\pi\left(\frac{n-\alpha}{q}\right)y\right\}dy \ll \int_0^{c^{-2}} y\,dy+\int_{c^{-2}}^\infty y\frac{\exp\left\{-2\pi\left(\frac{n-\alpha}{q}\right)y\right\}}{c^2 y}dy \ll c^{-2}$$

Therefore $|R_{m,n}(s,c)|\ll c^{-2}/|\sigma-\frac{1}{2}|$ which implies that

$$\sum_{c\neq 0}\frac{S(m,n,c,\chi)}{|c|^{2s}}R_{m,n}(s,c)$$

is holomorphic in $\sigma>\frac{1}{2}$ and $O(1/\sigma-\frac{1}{2})$ in this region. Q.E.D.

Theorem 1 follows easily from Lemmas 1 and 2 and Stirling's formula

$$\left|\frac{\Gamma(2s+1)}{\Gamma(s+k/2)\,\Gamma(s-k/2+2)}\right|\sim\frac{|t|^{-1/2}}{\sqrt{2\pi}},\qquad |t|\to\infty.$$

3. Proof of Theorem 2

Choose $\varepsilon>0$. By Theorem 1 and the Phragmén Lindelöf principle [6, Th. 14] it follows that

$$\left|Z_{m,n}\left(\frac{1+s}{2},\chi\right)\right|\ll|t|^{\frac{1}{2}-\frac{\sigma}{2\beta}+\varepsilon} \tag{3.1}$$

for $0<\varepsilon\leqq\sigma\leqq\beta+\varepsilon$. Now proceeding as in the proof of the prime number theorem [3, p. 104]

$$\sum_{c\leqq x}\frac{S(m,n,c,\chi)}{c}=\frac{1}{2\pi i}\int_{+\beta-iT}^{+\beta-iT} Z_{m,n}\left(\frac{s+1}{2},\chi\right)\frac{x^s}{s}ds+O\left(\frac{x^{\beta+\varepsilon}}{T}\right)$$

As was shown before, the function $Z_{m,n}\left(\frac{s+1}{2},\chi\right)$ has poles at $s=2s_j-1$ and a computation shows

$$\operatorname*{Res}_{s=2s_j-1} Z_{m,n}\left(\frac{1+s}{2},\chi\right)=\tau_j(m,n)$$

$$=\frac{q^2\overline{\rho_j(m)}\,\rho_j(n)\left(\left(\frac{m-\alpha}{q}\right)\left(\frac{n-\alpha}{q}\right)\right)^{-s_j+1}\Gamma(s_j+k/2)\,\Gamma(2s_j-1)}{(-i)^k\,4^{2s_j}\,\pi^{3s_j+1}\,\Gamma(s_j-k/2)} \tag{3.2}$$

Now shifting the line of integration in (3.2) to $\operatorname{Re}(s)=\varepsilon$, using (3.1) and (3.3) we obtain

$$\sum_{|c|\leq x}\frac{S(m,n,c)}{|c|}=\sum_{\frac{1}{2}<s_j<1}\tau_j(m,n)\frac{x^{2s_j-1}}{2s_j-1}+O\left(x^{\varepsilon}\,T^{1/2+\varepsilon}+\frac{x^{\beta+\varepsilon}}{T}\right)$$

Choosing $T=x^{2\beta/3}$ proves the theorem.

Acknowledgements. The authors would like to thank Professor D. Héjhal for his remarks and suggestions.

References

1. Bruggeman, R.W.: Fourier coefficients of cusp forms. Invent. Math. **45**, 1-18 (1978)
2. Davenport, H.: On certain exponential sums. J. für Mathematik **169**, 158 (1933)
3. Davenport, H.: Multiplicative number theory. Berlin-Heidelberg-New York: Springer 1980
4. Deshouillers, J.M., Iwaniec, H.: Kloosterman sums and Fourier coefficients of cusp forms. Preprint 1981
5. Gradshteyn, I.S., Ryzhik, I.M.: Table of integrals, series, and products. New York: Academic Press 1980
6. Hardy, G., Riesz, M.: Dirichlet series. Cambridge: Cambridge University Press 1915
7. Kuznetsov, N.V.: Peterson hypothesis, for parabolic forms of weight zero and Linnik hypothesis. Sums of Kloosterman sums. Math. Sbornik **111**, (153, no. 3) 334-383 (1980)
8. Linnik, Y.V.: Additive problems and eigenvalues of the modular operators. Proc. Internat. Congr. Math. Stockholm 1962, pp. 270-284
9. Proskurin, N.V.: Summation formulas for generalized Kloosterman sums. Zap. Navcn. Sem. Leningrad Otdel. Mat. Inst. Steklov **82**, 103-135 (1979)
10. Riesz, F., Nagy, B.: Functional analysis. New York: Frederick Ungar Publishing Company 1978
11. Roelcke, W.: Das Eigenwertproblem der automorphen Formen.... Math. Ann. **167**, 292-337 (1966)
12. Selberg, A.: On the estimation of Fourier coefficients of modular forms. Proc. Symposia in Pure Math. VIII A.M.S., Providence 1965, pp. 1-15
13. Vardi, I.: Mass. Inst. Tech. Ph.D. Thesis, 1982
14. Weil, A.: On some exponential sums. Proc. Natl. Acad. Sci. USA **34**, 204-207 (1948)

Oblatum 5-IV-1982 & 10-V-1982