ACTA ARITHMETICA
LXXIX.3 (1997)

The asymptotic distribution of Kloosterman sums

by

S. J. Patterson (Göttingen)

To Ian Cassels on his 75th birthday

1. Introduction. In this paper we shall discuss the asymptotic distribution of a wide class of generalized Kloosterman sums. To define these we let k be a global field and S a finite set of places of k containing the infinite ones, if there are any. Let $k_S = \prod_{v\in S} k_v$ where k_v denotes, as usual, the completion of k at v. Let R be the ring of S-integers of k; it is a discrete, cocompact subring of k_S. Let $n \in \mathbb{N}$ and $\mu_n(k) = \{\zeta \in k : \zeta^n = 1\}$. We shall assume that $\mu_n(k)$ has n elements and that all the divisors of n lie in S. This means that n is invertible in R. Let $e : k_S \to \mathbb{C}^\times$ be a non-trivial additive character, trivial on R. We denote the fractional ideal $\{x \in k : e|xR = 1\}$ by $\mathbf{d}(e)^{-1}$; then $\mathbf{d}(e)$ is an ideal of R.

Let $\big(\,-\,\big)_n$ denote the nth order Legendre symbol in R. For $a, b \in R$ and coprime we can write the reciprocity law as follows:

$$\left(\frac{a}{b}\right)_n = \left(\frac{b}{a}\right)_n (a,b)_S$$

where $(\,\cdot\,,\,\cdot\,)_S : k_S^\times \times k_S^\times \to \mu_n(k)$ is the Hilbert symbol. Let $\varepsilon : \mu_n(k) \to \mathbb{C}^\times$ be an injective character. Then the Kloosterman sums in question are defined by

$$S_\varepsilon(a,b;c) = \sum_{\substack{x,\overline{x} \,(\mathrm{mod}\, c)\\ x\overline{x}\equiv 1 \,(\mathrm{mod}\, c)}} \varepsilon\left(\left(\frac{x}{c}\right)_n\right) e\left(\frac{ax + b\overline{x}}{c}\right)$$

for $a, b, c \in R,\ c \neq 0$. If a or b is divisible by c then this sum will reduce to a Gauss sum. When $k = \mathbb{Q}$ and $S = \{\infty\}$ sums of this sort where first introduced by Kloosterman [13] in the context of the theory of modular forms of integral weight. They are important in estimating the number of representations of a number by a quadratic form. In view of this application Kloosterman was led to conjecture an estimate for these sums, namely

$|S_1(a,b;p)| \leq 2\sqrt{p}$ ($R = \mathbb{Z}, ab \not\equiv 0 \pmod{p}$ and $n = 1$) for a prime modulus p. After initial work by Estermann [5], Davenport [3] and Salié [19] this estimate was proved in 1948 by A. Weil [23]. In fact, this estimate is a fairly direct consequence of the Riemann hypothesis for curves over a finite field applied to Artin–Schreier extensions. For composite moduli one can also give very good estimates for $S_1(a,b;c)$ in this case (see [6]).

The case $n = 2$ was studied intensively by A. V. Malyshev [17] in the context of modular forms of half-integral weight, and so also in that of the representation of an integer by a quadratic form in an odd number of variables. Here again one can prove analogous, indeed simpler, estimates. In fact, Kloosterman sums have many properties analogous to Bessel functions and, just as the Bessel functions of half-integral order are elementary, so also are the Kloosterman sums with $n = 2$. We shall return to this theme later. It is worth noting that the coefficients of the Hardy–Ramanujan–Rademacher formula for the partition function are essentially Kloosterman sums of this type and are therefore the earliest appearance of these functions (see [16], p. 351f.).

The more general Kloosterman sums of order n are associated with metaplectic forms on GL_2 of order n in the sense of [12]. We shall not go into this aspect of the theory here.

As we have indicated in the definition of the $S_\varepsilon(a,b;c)$ we shall be interested in Kloosterman sums in a global context. We shall show that when $\gcd(c,ab) = (1)$ we can represent $S_\varepsilon(a,b;c)$ as a sum of $2^{\omega(c)}$ complex numbers of modulus $N(c)^{1/2}$ satisfying one additional condition. Here $\omega(c)$ denotes the number of prime divisors of c in R. The problem which we shall discuss is whether this set of numbers is uniformly distributed when normalized to lie on the unit circle. We shall not be able to answer this question but we shall attempt to formulate it more precisely.

One can understand Kloosterman sums in a number of contexts; they arise in the representation theory of GL_2 over finite and non-archimedean local fields (see, for example [8], 2.2.9, 2.3.2). There is also a wide-ranging development of Weil's proof, due principally to Deligne and Katz (see [11]). Although this appears to be a global method it yields information about the distribution of Kloosterman sums defined over finite fields. Indeed, in Weil's proof of the fundamental estimate for Kloosterman sums yields these as a sum over all the places of $\mathbb{F}_q(t)$ with given degree. It follows that this sort of equidistribution is quite different from that with which we shall be concerned here.

As will be pointed out later, the case $n = 3$ has an especial interest. This is the subject of a joint research programme with R. Livné (Jerusalem) whom I would like to thank for stimulating my interest in this problem. I

would also like to thank the Israel Science Foundation which is supporting this project.

Finally, I would like to acknowledge here my personal debt to Ian Cassels for his help and encouragement in many respects, not least for having suggested that I might find Kubota's theory of metaplectic forms interesting.

2. Basic properties of Kloosterman sums. Although Kloosterman sums are not multiplicative they can be represented as a product over the primes dividing c. Our first objective will be to reduce the Kloosterman sum $S_\varepsilon(a,b;c)$ to the corresponding local case and to make some deductions from this.

Let w be a place of k, $w \notin S$; let r_w be the ring of integers in k_w, the completion of k at w. Let π_w be a uniformizer of k_w. We shall let q_w denote the order of the residue class field at w; we have $q_w \equiv 1 \pmod n$. We define a character $\varepsilon_w : r_w^\times \to \mu_n(\mathbb{C})$ by demanding that $\varepsilon_w(x) = \varepsilon(\zeta)$ where ζ is determined by $x^{(q_w-1)/n} \equiv \zeta \pmod{\pi_w}$ and $\zeta \in \mu_n(k)$. We shall let ord_w be the order (or valuation) function at w. Let the order of $\mathbf{d}(e)$ at w be d_w. We recall that we can define an additive character e_w on k_w, trivial on $\pi_w^{-d_w} \cdot r_w$ but not on $\pi_w^{-d_w-1} \cdot r_w$, so that for any $x \in k$,

$$e_S(x) = \prod_{w \notin S} e_w(x).$$

Note that the product on the right is finite for any x in the sense that all but a finite number of terms are 1.

We shall now investigate the Kloosterman sum as defined above. Note that the definition makes sense for $a, b \in \mathbf{d}(e)^{-1}$ and we shall make use of this extension. Let $t_w = \mathrm{ord}_w(c)$. Then a standard application of the Chinese Remainder Theorem yields

$$S_\varepsilon(a,b;c) = \prod_{\substack{w \notin S \\ w|c}} S_w(a,b;t_w;c)$$

with

$$S_w(a,b;t;c) = \sum_{\substack{x,\overline{x} \pmod c \\ x\overline{x} \equiv 1 \pmod c}} \varepsilon_w(x)^t e_w\left(\frac{ax+b\overline{x}}{c}\right)$$

where now the congruences are taken in r_w. It is now a standard matter to analyze this sum.

PROPOSITION 2.1. *Assume that the residual characteristic of k_w is not* 2.

(i) *If* $\min(\mathrm{ord}_w(a), \mathrm{ord}_w(b)) \geq \mathrm{ord}_w(c) - d_w$ *then*

$$S_w(a,b;t;c) = \begin{cases} 0 & \text{if } t \not\equiv 0 \pmod{n}, \\ q^{\mathrm{ord}_w(c)}(1 - q_w^{-1}) & \text{if } t \equiv 0 \pmod{n}. \end{cases}$$

(ii) *If* $\mathrm{ord}_w(a) = \mathrm{ord}_w(b) = \mathrm{ord}_w(c) - d_w - 1$ *then*

$$S_w(a,b;t;c) = q_w^{\mathrm{ord}_w(c)-1} \cdot (\alpha + \alpha')$$

where $\alpha, \alpha' \in \mathbb{C}, |\alpha'| = q_w^{1/2}$ *and* $\alpha \cdot \alpha' = q_w \cdot \varepsilon_w(-ab^{-1})^t$.

(iii) *If* $\mathrm{ord}_w(a) = \mathrm{ord}_w(b) = \mathrm{ord}_w(c) - d_w - k$ *with* $k > 1$ *and* k *even then*

$$S_w(a,b;t;c) = q_w^{\mathrm{ord}_w(c)-k/2} \sum_{u:u^2 \equiv ba^{-1} \,(\mathrm{mod}\, \pi_w^k)} \varepsilon_w(u)^{t_w} e_w\left(\frac{2ua}{c}\right).$$

(iv) *If* $\mathrm{ord}_w(a) = \mathrm{ord}_w(b) = \mathrm{ord}_w(c) - d_w - k$ *with* $k > 1$ *and* k *odd then*

$$S_w(a,b;t;c) = q_w^{\mathrm{ord}_w(c)-(k+1)/2} \sum_{u:u^2 \equiv ba^{-1} \,(\mathrm{mod}\, \pi_w^k)} \varepsilon_w(u)^{t_w} e_w\left(\frac{2ua}{c}\right) \times \sum_{\xi(\pi_w)} e_w\left(\frac{au}{c}\pi_w^{k-1}\xi^2\right).$$

(v) *If* $\mathrm{ord}_w(a) \neq \mathrm{ord}_w(b)$ *and*

$$\min(\mathrm{ord}_w(a), \mathrm{ord}_w(b)) < \mathrm{ord}_w(c) - d_w - 1$$

then $S_w(a,b;t;c) = 0$.

Although the proof of this proposition might safely be left to the reader, because of its significance for our discussion we shall prove it here.

P r o o f. (i) is trivial. The sum in (ii) reduces to

$$q_w^{\mathrm{ord}_w(c)-1} \cdot \sum_{\substack{x,\overline{x} \,(\mathrm{mod}\, \pi_w) \\ x\overline{x} \equiv 1 \,(\mathrm{mod}\, \pi_w)}} \varepsilon_w(x)^t \cdot e_w\left(\frac{1}{\pi_w^{d_w}} \cdot \frac{\alpha x + \beta \overline{x}}{\pi_w}\right)$$

where $\alpha = a \cdot \pi^{d_w+1}/c$, $\beta = b \cdot \pi^{d_w+1}/c$ are now units. Now we can apply Weil's estimate [23], top of page 207, but to obtain what we need we let χ (in Weil's notation) be the character of order 2 times $\varepsilon_w^{t_w}$. This shows that we have a representation of the type asserted for S_w. Finally, we have

$$\overline{S_w(a,b;t;c)} = \varepsilon_w(-\beta\alpha^{-1})^t S_w(a,b;t;c)$$

after making the substitution $x \mapsto -\overline{x} \cdot \beta\alpha^{-1}$. This means that $(\varepsilon_w(-ba^{-1})^t)^{1/2}(\alpha + \alpha')$ is real; also $\alpha\overline{\alpha} = q_w, \alpha' \cdot \overline{\alpha'} = q_w$ so that

$$((\varepsilon_w(-ba^{-1})^t)^{\varepsilon'})(\alpha + \alpha') = ((\varepsilon_w(-ba^{-1})^t)^{\varepsilon'})^{-1}(\overline{\alpha} + \overline{\alpha'})$$

or

$$\alpha + \alpha' = \varepsilon_w(-ba^{-1})^{-t}(\alpha^{-1} + \alpha^{-1})q_w.$$

If $\alpha+\alpha'\neq 0$ then we have

$$\alpha\alpha' = q_w\cdot\varepsilon_w(-ba^{-1})^{-t}$$

as required. If $\alpha+\alpha'=0$ then we take α and α' to be the two square roots of $q_w\varepsilon_w(-ba^{-1})^{-t_w}$.

(iii) We fix a set U of representatives of $r_v/\pi_v^{k/2}r_v$. We see that $S_w(a,b;t;c)$ is

$$q_w^{\mathrm{ord}_w(c)-k}\sum_{\substack{u,\overline{u},v,\overline{v}\,(\mathrm{mod}\,\pi^{k/2}),\,u\in U\\ (u+\pi^{k/2}v)(\overline{u}+\pi^{k/2}\overline{v})\equiv 1\,(\pi_w^k)}}\varepsilon_w(u+\pi^{k/2}v)^t \times e_w\left(\frac{1}{\pi_w^{d_w}}\cdot\frac{\alpha(u+\pi^{k/2}v)+\beta(\overline{u}+\pi^{k/2}\overline{v})}{\pi_w^k}\right).$$

We take $\overline{u}$ so that $u\overline{u}\equiv 1\pmod{\pi_w^k}$. This means that $\overline{v}$ is determined by $\overline{u}v+\overline{v}u\equiv 0\pmod{\pi_w^{k/2}}$.

We see that the exponential term is

$$e_w\left(\frac{1}{\pi_w^{d_w}}\left(\frac{\alpha u+\beta\overline{u}}{\pi_w^k}+\frac{\alpha-\beta\overline{u}^2}{\pi_w^{k/2}}v\right)\right)$$

and the ε_w-term does not depend on v. The sum over v forces u to satisfy $u^2\equiv\beta\alpha^{-1}\pmod{\pi_v^{k/2}}$ and substituting back we obtain the formula asserted.

(iv) This case is similar to the previous one, but is a little more involved. This time we consider $u,\overline{u}\pmod{\pi_w^{(k+1)/2}}$ and $v,v'\pmod{\pi_w^{(k-1)/2}}$ and so we see that u is restricted by $u^2\alpha\equiv\beta\pmod{\pi^{(k+1)/2}}$. We now fix u_0 so that $u_0^2\equiv\alpha^{-1}\beta\pmod{\pi^k}$ and consider $u=u_0(1+\pi_w^{(k-1)/2}\cdot\xi)$ where ξ runs modulo π_w. We have $\overline{u}\equiv\overline{u}_0(1-\pi_w^{(k-1)/2}\xi+\pi_w^{k-1}\xi^2)$. It now follows that

$$S_w(a,b;t;c)=q_w^{\mathrm{ord}_w(c)-(k+1)/2}\sum_{u_0:u_0^2\equiv\alpha^{-1}\beta\,(\mathrm{mod}\,\pi_w^k)}\varepsilon_w(u_0)^t \times\sum_{\xi\,(\mathrm{mod}\,\pi_w)}e_w\left(\frac{1}{\pi_w^{d_w}}\left(\frac{2\alpha\cdot u_0}{\pi_w^k}+\frac{\beta\overline{u}_0}{\pi_w}\xi^2\right)\right)$$

which is a variant form of what is asserted.

(v) In this case the sums vanish just as in the case of Gauss sums taken to a modulus exceeding the conductor of the character.

It is worth noting that when the residual characteristic of k_w is 2 the estimate in (ii) does not hold. Also in (iv) the inner sum, over ξ, will no longer be a quadratic Gauss sum but an ordinary character sum and so will be either q_w or 1. Moreover, the outer sums, over u, will have a different structure.

In order to bring our results into a more convenient form we let T be the set of places of k not in S so that

(a) if $w \mid 2$ then $w \in T$,
(b) if $w \mid \mathbf{d}(e)$ then $w \in T$,
(c) if $\mathrm{ord}_w(a) \neq \mathrm{ord}_w(b)$ then $w \in T$.

In view of the last condition T depends on the pair a, b. We can extend e_S to $k_{S\cup T}$ by

$$e_{S\cup T}|k_S = e_S, \quad e_{S\cup T}|k_w = \overline{e}_w \quad (w \in T).$$

We let R^* be the ring of $(S \cup T)$-integers in k. We let $\big(\,-\,\big)_n^*$ be the corresponding Legendre symbol and $S_\varepsilon^*(a, b; c)$ the corresponding Kloosterman sum. We let

$$K_\varepsilon(a, b; c) = \prod_{w\in T} S_w(a, b; t_w; c);$$

then K_ε is a continuous function on its domain of definition in the topology on $\prod_{w\in T} k_w$. Then we have shown that

$$S_\varepsilon(a, b; c) = K_\varepsilon(a, b; c) \cdot S_\varepsilon^*(a, b; c)$$

where $S_\varepsilon^*(a, b; c)$ is of the form

$$(N^*(a) \cdot N^*(c))^{1/2} \sum_{\alpha\in A} \alpha$$

where A is a multiset of $2^{\omega^*(c)}$ elements with $|\alpha| = 1$ which are paired by an operation $' : A \to A$ so that $\alpha\alpha' = \varepsilon\big(\big(\frac{-ab^{-1}}{c}\big)_n^*\big)$. Here N^* denotes the norm in R^* and $\omega^*(c)$ is the number of prime factors of c in R^*.

3. Two special formulae. In this section we shall prove two special formulae, one for the case $n = 2$, analogous to the elementary expression for $K_{1/2}$, and for $n = 3$, analogous to the Wirtinger–Nicolson formula for Airy's integral (see [22], 188–190). These two formulae are known in various forms. The case $n = 2$ goes back to Davenport and Salié. It seems very possible that the form given here can be found in the works of Malyshev but I have been unable to locate it. A more local version of the case $n = 3$ has been given by Duke and Iwaniec [4]. Other proofs are due to Elkies, Katz and Livné. R. Livné informs me that he has proved a version of the $n = 3$ case for general c by a rather different method. The point of the following theorem is that it brings out the global aspects of the formulae and these are important for understanding questions of uniform distribution. The formulae in question are given in the following theorem.

THEOREM 3.1. *Suppose that* $\gcd(ab,c)=1$ *and that* $\gcd(\mathbf{d}(e),c)=(1)$. *Then if* $n=2$ *we have*

$$S_\varepsilon(a,b;c)=\sum_{x^2\equiv a\cdot b\,(\mathrm{mod}\,c)} e\Big(\frac{2x}{c}\Big)\cdot\varepsilon\bigg(\Big(\frac{a}{c}\Big)_2\bigg)\cdot G(c)$$

where

$$G(c)=\sum_{y\,(\mathrm{mod}\,c)} e\Big(\frac{y^2}{c}\Big)$$

if $\gcd(2,c)=(1)$. *If* $n=3$ *and* $\gcd(3,c)=(1)$ *then*

$$S_\varepsilon(a,b;c)=\sum_{x\,(\mathrm{mod}\,c)} e\Big(\frac{Ax^3+Bx}{c}\Big)\cdot\varepsilon\bigg(\Big(\frac{a}{c}\Big)_3\bigg)\varepsilon\bigg(\Big(\frac{b}{c}\Big)_3\bigg)^{-1}$$

when A,B *are also coprime to* c *and* $27abA+B^3\equiv 0\ (\mathrm{mod}\ c)$.

P r o o f. The proofs of these two formulae are similar. First of all, in view of the Chinese Remainder Theorem, we may enlarge S so that the prime ideals in the decomposition of c become principal. We shall then verify the formulae when c is a prime power. It is then easy, as in Section 2, to deduce the general case using the primary decomposition. As always in this connection the cases $c=\pi$ (π prime) and $c=\pi^k$ ($k\geq 2$) behave rather differently and we treat them separately.

We begin with $c=\pi$ and $n=2$. We shall consider the group X of characters on $(R/\pi R)^*$. Let, for $\chi\in X$,

$$\tau(\chi)=\sum_{x\,(\mathrm{mod}\,\pi)}\chi(x)e(x/\pi)$$

be the local Gauss sum. Then

$$e(b\overline{x}/\pi)=\sum_\chi \overline{\chi}(b\overline{x})\tau(\chi)/(N(\pi)-1).$$

This gives

$$S_\varepsilon(a,b;c)=(N(\pi)-1)^{-1}\sum_{\chi\in X}\sum_x \overline{\chi}(b\overline{x})\eta(x)e\Big(\frac{ax}{\pi}\Big)\tau(\chi)$$

where η is the character of order 2 (i.e. $\eta(x)=\varepsilon\big(\big(\frac{x}{\pi}\big)_2\big)$. The inner sum is $\overline{\chi}\eta(a)\tau(\chi\eta)$. By the Davenport–Hasse theorem ([3], [15], Theorem 10.1) we have

$$\tau(\chi\eta)\cdot\tau(\chi)=\chi(2)^{-2}\tau(\chi^2)\tau(\eta).$$

In this case $\tau(\eta) = G(\pi)$ and so we obtain

$$\begin{aligned} S_\varepsilon(a,b;c) &= (N(\pi)-1)^{-1}\sum_{\chi}\overline{\chi}(b)\chi\eta(a)\tau(\chi^2) \\ &= (N(\pi)-1)^{-1}\eta(a)G(\pi) \\ &\quad\times\sum_{\chi\in X}\chi(a\overline{b})\cdot\chi(4)^{-1}\sum_{\substack{\xi(\pi)\\ \xi\not\equiv 0\,(\mathrm{mod}\,\pi)}}\chi(\xi^2)e\left(\frac{\xi}{\pi}\right). \end{aligned}$$

The sum over χ can now be carried out and we obtain

$$\eta(a)\sum_{\xi^2\equiv 4\overline{a}b\,(\mathrm{mod}\,\pi)}e\left(\frac{\xi}{\pi}\right)G(\pi)$$

which is clearly equivalent to the formula asserted. When $n=3$ the formula has been proven with $k=1$, $a=1$, $b=-\overline{27}A$, $B=1$ by Duke and Iwaniec [4] (see also [24]). The general case follows by a simple change of variables. If we now consider the case $c=\pi^k$ with $k\geq 2$ then for the case $n=2$ we have only to apply Proposition 2.1. In the case $n=3$ we have to analyze

$$\sum_{x\,(\mathrm{mod}\,c)}e\left(\frac{Ax^3+Bx}{c}\right)$$

asymptotically. Again there is a distinction between the cases where c is an even and an odd power. Let us suppose first that $c=\pi^{2k}$ $(k\geq 1)$. Then we fix a set of representations of R/π^k and with x_1 in this set we see that the sum above is

$$\begin{aligned} &\sum_{x_1}\sum_{x_2\,(\mathrm{mod}\,\pi^k)}e\left(\frac{a(x_1+\pi^k x_2)^3+B(x_1+\pi^k x_2)}{\pi^{2k}}\right) \\ &\qquad=\sum_{x_1}\sum_{x_2\,(\mathrm{mod}\,\pi^k)}e\left(\frac{Ax_1^3+Bx_1}{\pi^{2k}}+\frac{3Ax_1^2x_2+Bx_2}{\pi^k}\right). \end{aligned}$$

The summation over x_2 can be carried out. We obtain

$$N(\pi)^k\sum_{x_1:\,3Ax_1^2\not\equiv B\equiv 0\,(\mathrm{mod}\,\pi^k)}e\left(\frac{Ax_1^3+Bx_1}{\pi^{2k}}\right).$$

We can assume, by the appropriate choice of the set of representations, that $3Ax_1^2+B\equiv 0\ (\mathrm{mod}\ \pi^{2k})$. We obtain now

$$N(\pi)^k\sum_{x_1:\,3Ax_1^2+B\equiv 0\,(\mathrm{mod}\,\pi^k)}e\left(\frac{-2Ax_1}{\pi^{2k}}\right).$$

In this case also $G(\pi^{2k}) = N(\pi)^k$ so that we have now proved the assertion in this case.

Now consider the case $c = \pi^{2k+1}$ where $k \geq 1$. Here we let x_1 run through a set of representatives modulo π^{k+1}. By the same argument as before we obtain for our sum

$$N(\pi)^k \sum_{x_1 :\, 3Ax_1^2 + B \equiv 0 \,(\mathrm{mod}\, \pi^k)} e\left(\frac{Ax_1^3 + Bx_1}{\pi^{2k+1}}\right).$$

We let $x_1 = x_0 + \pi^k \xi$ where x_0 satisfies $3Ax_0^2 + B \equiv 0 \pmod{\pi^{2k+1}}$ (if the congruence is soluble). In this case we obtain

$$N(\pi)^k \sum_{x_0} \sum_{\xi} e\left(\frac{Ax_0^3 + Bx_0}{\pi^{2k+1}} + \frac{3\xi^2 A}{\pi}\right)$$
$$= N(\pi)^k G(\pi) \varepsilon\left(\left(\frac{3A}{\pi}\right)_2\right) \sum_{x_0} e\left(\frac{Ax_0^3 + Bx_0}{\pi^{2k+1}}\right).$$

We have $G(\pi^{2k+1}) = N(\pi)^k G(\pi)$ and so the result follows as before.

4. Uniform distribution. In this section we shall formulate a problem, the question as to the uniform distribution of the Kloosterman sums. The characteristic feature of the sums $S_\varepsilon(a, b; c)$ is that on altering c by a unit the value of $S_\varepsilon(a, b; c)$ is changed in an unpredictable fashion. In order to take account of this we shall consider a family of compact sets B_j in k_S so that $B_1 \subset B_2 \subset \ldots$ These will be assumed to grow in a uniform fashion. The type of example we have in mind is, for $w \in S$,

$$B_j = \{x \in k_w : |x|_w \leq j\} \times B^0$$

where B^0 is fixed and compact in $\prod_{v \in S - \{w\}} k_v$. Another class of examples could be

$$B_j = \prod_{v \in S} \{x \in k_v : |x|_v \leq \Phi_v(j)\}$$

where the Φ_v are increasing functions of j. We shall suppose that

$$\mathrm{Card}(B_j \cap R) = \mathrm{Vol}(B_j)/\,\mathrm{Vol}(k_S/R) + O(\mathrm{Vol}(B_j)^\Theta)$$

for some $\Theta < 1$. The volume is taken with respect to some additive Haar measure on k_S; the formulation above does not depend on the choice. Let us now suppose that S is taken so that $\mathbf{d}(e) = (1)$ and that a and b are units. Then in view of Proposition 2.1 we see that $S_\varepsilon(a, b; c)$ is a sum of $2^{\omega(c)}$ numbers which can be paired by an involution $'$ so that if α is one of these numbers then

$$\alpha\alpha' = N(c) \cdot \varepsilon\left(\left(\frac{-a^{-1}b}{c}\right)_n\right).$$

Under this condition the numbers are unique. Here $\omega(c)$ is, as before, the number of prime factors of c in R. We can now define

$$S_\varepsilon^{(k)}(a,b;c) = \sum_\alpha \alpha^k$$

where the α represent the numbers above.

PROBLEM 4.1. *For which families $B_1 \subset B_2 \subset \dots$ as above do we have, for $k \neq 0$,*

$$\sum_{c \in B_j \cap R} S_\varepsilon^{(k)}(a,b;c) \cdot N(c)^{-k/2} = o(\mathrm{Vol}(B_j))?$$

Since $2^{\omega(c)} = O(N(c)^\varepsilon)$ for any $\varepsilon > 0$ this would mean that we would have equidistribution of the α's as long as

(a) $\mathrm{Vol}(B_{j+1}) \sim \mathrm{Vol}(B_j)$ and
(b) $\sum_{c \in (B_{j+1} - B_j) \cap R} N(c)^\varepsilon = o(\mathrm{Vol}(B_j))$

for some $\varepsilon > 0$. This is a non-lacunarity condition on the B_j. Note that (b) implies the analogous estimate with $\varepsilon = 0$. It would then follow from our regularity condition

$$\mathrm{Card}(B_j \cap R) = \mathrm{Vol}(B_j)/\mathrm{Vol}(k_S/R) + O(\mathrm{Vol}(B_j)^\Theta)$$

that

$$\begin{aligned}\mathrm{Vol}(B_{j+1} - B_j)/\mathrm{Vol}(k_S/R) &= O(\mathrm{Vol}(B_j)^\Theta) + o(\mathrm{Vol}(B_j)) \\ &= o(\mathrm{Vol}(B_j)).\end{aligned}$$

From this condition (a) follows. The reverse implication would follow if the B_j are so constructed that $c \in (B_{j+1} - B_j) \cap R$ implies that $N(c) = O(\mathrm{Vol}(B_j)^{\Theta'})$ for a $\Theta' > 0$ and $\mathrm{Vol}(B_{j+1} - B_j) = O(\mathrm{Vol}(B_j)^{\Theta''})$ for $\Theta'' < 1$ (which is rather stronger than (a)). In this case we would have, for $\varepsilon > 0$,

$$\sum_{c \in (B_{j+1} - B_j) \cap R} N(c)^\varepsilon = O(\mathrm{Vol}(B_j)^{\max(\Theta, \Theta'') + \varepsilon\Theta'}).$$

Thus we would have the implication for

$$\varepsilon < \frac{1}{\Theta}(1 - \max(\Theta, \Theta'')).$$

The second, and naturally accompanying problem is:

PROBLEM 4.2. *For which families $B_1 \subset B_2 \subset \dots$ do we have, for $k \neq 0$,*

$$\sum_{\substack{c \in B_j \cap R \\ c \text{ prime}}} S_\varepsilon^{(k)}(a,b;c) \log N(c) \cdot N(c)^{-k/2} = o(\mathrm{Vol}(B_j))?$$

This is perhaps the problem which is of greater general interest and applicability. One would hope that it would follow by a sieve argument from a solution to Problem 4.1.

In order to analyze these further we have to examine what happens when we enlarge S. Let then v be a valuation of k not in S. Let $\widehat{S} = S \cup \{v\}$ and let $\widehat{S}_\varepsilon(a,b;c)$ be the corresponding Kloosterman sum. We then have, by the multiplicativity of S_ε,

$$S_\varepsilon(a,b;c) = S_v(a,b,\mathrm{ord}_v(c);c)\cdot \widehat{S}_\varepsilon(a,b;c).$$

In fact, this also yields, with an obvious notation,

$$S_\varepsilon^{(k)}(a,b;c) = S_v^{(k)}(a,b;\mathrm{ord}_v(c);c)\cdot \widehat{S}_\varepsilon^{(k)}(a,b;c).$$

Suppose now that we know that Problem 4.1 has a positive solution for $\widehat{S}$ and a family of domains of the form

$$B_j \times \{x \in k_v : \mathrm{ord}_v(x) = u,\ x \equiv \xi \pmod{\pi_v^v}\}.$$

The function $c \mapsto S_v^{(k)}(a,b;\mathrm{ord}_v(c);c)$, considered as a v-adic function, becomes very singular as $c \to 0$ in k_v. If $\mathrm{ord}_v(c) > 1$ then we have

$$\begin{aligned} &S_v^{(k)}(a,b;\mathrm{ord}_v(c);c)\\ &\qquad = N(c)^{k/2}\left\{\varepsilon(c,A)_w^k e_v\left(\frac{2Ak}{c}\right) + \varepsilon(c_1 - A)_w^k e_v\left(\frac{-2Ak}{c}\right)\right\}\\ &\qquad\quad \times \varepsilon((c,a)_w^{-1})^k \end{aligned}$$

where A is a v-adic solution to $A^2 = ab$ if there are any, and $S_v^{(k)}(a,b;\mathrm{ord}_v(c);c) = 0$ otherwise.

It follows from this that if Problem 4.1 has a positive solution with a fair degree of regularity in the family B_j for $S \cup \{v\}$ then it will also hold for S. This should also be the case in the context of Problem 4.2. These observations are important as they show that we can avoid the problems caused by the factors of a and b in c (and also of any common factors of c and $\mathbf{d}(e)$).

5. Discussion. At the present time it is not realistic to expect a solution to the problems formulated in Section 4. The most significant results at present known are those concerning the case $k = 1$. Here the method of Kuznetsov [14] can be applied; it is much more convenient to use the simplified version due to Goldfeld–Sarnak [9]. In its original form this applies only to the case where $k = \mathbb{Q}$, $S = \{\infty\}$ and $n = 1$. It is not difficult to extend this method to the case of imaginary-quadratic fields k (see [20] and [1] for an even wider generalization). All of these results apply to the case where S consists of one place ∞ and $B_j = \{x : |x|_\infty \le j\}$. One obtains non-trivial

esitmates for

$$\sum_{N(c)\le j} S(a,b;c)N(c)^{-1/2}.$$

If we use the theory of metaplectic groups, that is, in this context, the Kubota homomorphism (see [12]), we can extend such results to the case of general n. The question of relaxing the conditions on the set S, by far the most restrictive condition, has been approached by Cogdell and Piatetski-Shapiro in [2], Part II, where they treat the function-field case. In their main result they show how, essentially, one can treat the case $S = \{\infty, v_1, v_2, \dots, v_t\}$ and

$$B_j = \{x \in k_\infty : |x|_\infty \le j\} \times C_1 \times \ldots \times C_t$$

with $C_1, \dots, C_t$ fixed. They prove in fact a rather different result in which the characteristic functions of $C_1, \dots, C_t$ are replaced by the matrix coefficients of fixed cuspidal representations. In Theorem 4.3 they investigate the case where $S = \{v_1, \dots, v_t\}$ and $B_j = \prod_{i=1}^t \{x \in k_{v_i} : |x|_{v_i} \le j_i\}$ where now j is a multi-index $(j_1, \dots, j_t)$.

These methods can be extended directly to the case of metaplectic groups, so that general n can be investigated. One curious feature of this case is that a leading term appears from the generalized theta functions indicating a certain bias in the distribution of the Kloosterman sums. To obtain precise results involves not only very delicate calculations but also the application of the generalized Shimura correspondence ([7]) and an analysis of the corresponding local representations. These results will appear elsewhere.

For the case $k \ne 1$ the situation is more delicate. First of all we should note that from $\alpha\alpha' = \varepsilon\big(\big(\frac{-ab^{-1}}{c}\big)_n\big)$ we have

$$S_\varepsilon^{(k)}(a,b;c) = \varepsilon\left(\left(\frac{-ab^{-1}}{c}\right)_n\right)^k S_\varepsilon^{(-k)}(a,b;c).$$

Since under our assumptions a and b are units we see that

$$\varepsilon\left(\left(\frac{-ab^{-1}}{c}\right)_n\right)\cdot\varepsilon((-ab^{-1},c)_S)^k$$

where $(\cdot,\cdot)_S$ denotes the Hilbert symbol on $k_S^\times \times k_S^\times$. This is a locally constant function and it follows that we can, by a suitable modification of the B_j, deduce the case of negative k from that for positive k. For this reason we shall restrict our attention to the case of $k > 0$.

Unfortunately, with one exception, no analytic representation is known for the $S_\varepsilon^{(k)}(a,b;c)$ with $k > 1$ (cf. [18], [21]). It is not impossible that one will be found as there is still little clarity on what general functions of this type can be obtained from the theory of automorphic forms on groups of

higher rank. The one case where such results do exist is when $n = 2$; by Theorem 3.1 we have

$$S_\varepsilon^{(k)}(a,b;c) = S_\varepsilon^{(1)}(ka,kb;c)$$

for $\gcd(k,c) = (1)$. Hence in this case Kuznetsov's method will suffice to deal with the general case. This would mean, for example, that if we take a non-square $a \in \mathbb{N}$ and we consider, for $X > 1$,

$$T(X) = \{x/c : 0 < x < c,\, x^2 \equiv a \pmod c,\ c \le X\}$$

then $T(X)$ will be uniformly distributed in [0,1] for $X \to \infty$. In fact, this has been proved by C. Hooley [10] using a weaker estimate for the averages of the $S_\varepsilon^{(k)}(a,b;c)$ than we can obtain. Apparently this argument has been carried out in detail by I. Vardi in his MIT Ph.D. Thesis (see the remarks in [9] and [20]) but I have not had access to it. I am grateful to J. Brüdern for pointing out Hooley's work to me. (Added in proof: see also W. Duke, J. B. Friedlander and H. Iwaniec, *Equidistribution of roots of a quadratic congruence to prime moduli*, Ann. of Math. 141 (1995), 423–441.)

Finally, we shall examine the case $k = 2$. Suppose that c is a prime power and $\gcd(ab,c) = 1$. Then we have

$$S_\varepsilon^{(2)}(a,b;c) = S_\varepsilon(a,b;c)^2 - 2N(c)\cdot\varepsilon\left(\left(\frac{-\overline{a}b}{c}\right)_n\right).$$

Now

$$S_\varepsilon(a,b;c)^2 = \sum_{x,y}\varepsilon\left(\left(\frac{xy}{c}\right)_n\right)e\left(\frac{a(x+y)+b(\overline{x}+\overline{y})}{c}\right).$$

We substitute $u = x+y,\ v = xy$. In order that $T^2 - uT + v \equiv 0 \pmod c$ be soluble we require that $u^2 - 4v$ should be a quadratic residue $(\mathrm{mod}\, c)$. We deduce that (writing QNR to indicate a quadratic non-residue)

$$\begin{aligned} S_\varepsilon(a,b;c)^2 = 2 &\sum_{\substack{u,v\\ v\not\equiv 0 \pmod c}} \varepsilon\left(\left(\frac{v}{c}\right)_n\right)e\left(\frac{au+bu\overline{v}}{c}\right)\\ &- 2\sum_{u^2-4v \text{ QNR}} \varepsilon\left(\left(\frac{v}{c}\right)_n\right)e\left(\frac{au+bu\overline{v}}{c}\right)\\ &- \sum_{\substack{u^2-4v\equiv 0 \pmod c\\ v\not\equiv 0 \pmod c}} \varepsilon\left(\left(\frac{v}{c}\right)_n\right)e\left(\frac{au+buv}{c}\right) \end{aligned}$$

where the factor 2 accounts for the fact that if $u^2 - 4v \not\equiv 0 \pmod c$ we have two distinct summands in the original sum and the last term corrects for the exceptional case. The first term is easily computed; it is $2N(c)\varepsilon\big(\big(\frac{-\overline{a}b}{c}\big)_n\big)$.

The last term is $-S_{\varepsilon^2}(2a, 2b; c)$. Thus we have

$$S_\varepsilon^{(2)}(a, b; c) = -2 \sum_{u^2-4v \text{ QNR}} \varepsilon\left(\left(\frac{v}{c}\right)_n\right) e\left(\frac{au + bu\overline{v}}{c}\right) - S_{\varepsilon^2}(2a, 2b; c).$$

If we fix a non-trivial quadratic non-residue δ then we can write this last sum as

$$S_\varepsilon^{(2)}(a, b; c) = - \sum_{\substack{u,v,w \,(\mathrm{mod}\, c) \\ u^2-4v \equiv \delta w^2 \,(\mathrm{mod}\, c)}} \varepsilon\left(\left(\frac{v}{c}\right)_n\right) \varepsilon\left(\frac{au + bu\overline{v}}{c}\right).$$

An analogous result holds for arbitrary c when δ is taken to be a quadratic non-residue modulo all divisors of c and the factor (-1) is replaced by $(-1)^{\omega(c)}$.

Analogous formulae hold for higher $S^{(k)}(a, b; c)$ but they are more complicated. From the point of view of collecting numerical evidence about the problems of Section 4 they might be of some use—unfortunately, the evaluation of these sums for composite c is very involved.

References

[1] R. W. Bruggeman and R. J. Miatello, *Estimates of Kloosterman sums for groups of real rank one*, Duke Math. J. 80 (1995), 105–137.

[2] J. W. Cogdell and I. I. Piatetski-Shapiro, *The Arithmetic and Spectral Analysis of Poincaré Series*, Perspect. Math. 13, Academic Press, 1990.

[3] H. Davenport, *On certain exponential sums*, J. Reine Angew. Math. 169 (1933), 158–176.

[4] W. Duke and H. Iwaniec, *A relation between cubic exponential and Kloosterman sums*, in: Contemp. Math. 143, Amer. Math. Soc., 1993, 255–258.

[5] T. Estermann, *Vereinfachter Beweis eines Satzes von Kloosterman*, Abh. Math. Sem. Hamburg 7 (1929), 82–98.

[6] —, *On Kloosterman's sum*, Mathematika 2 (1961), 83–86.

[7] Y. Z. Flicker, *Automorphic forms on covering groups of $GL(2)$*, Invent. Math. 57 (1980), 119–182.

[8] I. M. Gel'fand, M. I. Graev and I. I. Piatetski-Shapiro, *Representation Theory and Automorphic Functions*, W. B. Saunders, 1969.

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

[10] C. Hooley, *On the distribution of the roots of polynomial congruences*, Mathematika 11 (1964), 39–49.

[11] N. Katz, *Gauss Sums, Kloosterman Sums and Monodromy Groups*, Ann. of Math. Study 116, Princeton Univ. Press, 1988.

[12] D. A. Kazhdan and S. J. Patterson, *Metaplectic forms*, Publ. Math. I.H.E.S. 59 (1984), 35–142.

[13] H. D. Kloosterman, *Asymptotische Formeln für die Fourierkoeffizienten ganzer Modulformen*, Abh. Math. Sem. Hamburg 5 (1927), 337–352.

[14] N. V. Kuznetsov, *Peterson's hypothesis for parabolic forms of weight zero and Linnik's hypothesis. Sums of Kloosterman sums*, Mat. Sb. 111 (1980), 334–383 (in Russian).
[15] S. Lang, *Cyclotomic Fields*, Grad. Texts in Math. 59, Springer, 1978.
[16] J. Lehner, *Discontinuous Groups and Automorphic Functions*, Amer. Math. Soc., 1964.
[17] A. V. Malyshev, *Generalized Kloosterman sums and their applications*, Vestnik Leningrad. Univ. 13 (1960), 59–75 (in Russian).
[18] I. Piatetski-Shapiro, *Invariant theory and Kloosterman sums*, in: Algebraic Groups, Proceedings of a Symposium in Utrecht, Lecture Notes in Math. 1271, Springer, 1987, 229–236.
[19] H. Salié, *Zur Abschätzung der Fourierkoeffizienten ganzer Modulformen*, Math. Z. 36 (1932), 263–278.
[20] P. Sarnak, *Additive number theory and Maass forms*, in: Lecture Notes in Math. 1052, Springer, 1984, 286–309.
[21] G. Stevens, *Poincaré series on $GL(r)$ and Kloosterman sums*, Math. Ann. 277 (1987), 25–51.
[22] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge Univ. Press, 1944.
[23] A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U.S.A. 34 (1948), 204–207 = Collected Papers, Vol. I, 1948, 386–389.
[24] D. J. Wright, *Cubic character sums of cubic polynomials*, Proc. Amer. Math. Soc. 100 (1987), 409–413.

Mathematisches Institut
Bunsenstr. 3-5
D-37073 Göttingen, Germany
E-mail: sjp@cfgauss.uni-math.gwdg.de

Received on 25.6.1996 (3007)