

The number of ideals in a quadratic field

M N HUXLEY and N WATT

School of Mathematics, University of Wales College of Cardiff, Senghenydd Road, Cardiff
Wales CF2 4YH, UK

Dedicated to the memory of Professor K G Ramanathan

Abstract. Let K be a quadratic field, and let $R(N)$ be the number of integer ideals in K with norm at most N . Let χ with conductor k be the quadratic character associated with K . Then

$$|R(N) - NL(1, \chi)| \leq Bk^{50/73} N^{23/73} (\log N)^{461/146}$$

for $N \geq Ak$, where A and B are constants. For $N \geq Ak^C$, C sufficiently large, the factor $k^{50/73}$ may be replaced by $(d(k))^{4/73} k^{46/73}$.

Keywords. Exponential sums; character sums; ideals; quadratic field.

1. Introduction

Let K be a quadratic number field, and let $r(n)$ be the number of integer ideals in K whose norm is n . Then

$$r(n) = \sum_{d|n} \chi(d),$$

where $\chi(d)$ is a real primitive character whose conductor k is the absolute value of the discriminant of K . Let $R(N)$ be the number of integer ideals with norm at most N . Dirichlet (see [1, 6]) showed that

$$R(N) = \sum_{n \leq N} r(n) = NL(1, \chi) + O(kN^{1/2}); \quad (1.1)$$

the factor k in the error term can be reduced to $k^{1/2} \log k$ using the Polya–Vinogradov theorem. For the Gaussian field $Q(i)$, the sum $R(N)$ is the number of lattice points in a quarter-circle. The remainder term in (1.1) has been studied in this case. Recently Iwaniec and Mozzochi [5] used a powerful new method to reduce the exponent of N in (1.1) to $7/22 + \varepsilon$. This work was generalized and taken further by Huxley [2, 3], who obtained the error term

$$O(N^{23/73} (\log N)^{315/146}). \quad (1.2)$$

For a general quadratic field K we can apply the lattice-point method separately to each ideal class, with a remainder that depends on N as in (1.2) in the complex case, with the power of $\log N$ increased by one in the real case, and also on the ideal class

by way of the maximum radius of curvature of the ellipse or hyperbolic segment that contains the lattice points, and in the real case, also on the fundamental unit. In this paper we obtain results that depend on K only as a power of k , even for real fields; in fact χ can be any primitive character.

Theorem 1. *There are absolute constants A and B such that for $N \geq Ak$ we have*

$$|R(N) - NL(1, \chi)| \leq Bk^{50/73} N^{23/73} (\log N)^{461/146}.$$

Theorem 2. *There are absolute constants A , B and C such that for $N \geq Ak^C$ we have*

$$|R(N) - NL(1, \chi)| \leq B(d(k))^{4/73} k^{46/73} N^{23/73} (\log N)^{461/146}.$$

The constants A , B and C could be calculated effectively. Theorem 2 comes from an upper bound with several terms involving different powers of $d(k)$, k , N and $\log N$. The other terms involve smaller powers of N , but powers of k which may be closer to one. To find the infimum of those C for which Theorem 2 can be proved with some A and B would involve a large number of cases and alternative arguments.

2. Preparation

We obtain Theorem 1 from the following lemmas.

Lemma 1. (Accelerated convergence for $L(1, \chi)$). For any non-trivial character mod k we have

$$L(1, \chi) = \sum_1^N \frac{\chi(n)}{n} \left(1 - \frac{n^2}{N^2}\right) + O\left(\frac{k^{3/2}}{N^2}\right).$$

If $\chi(-1) = 1$ and $k|N$, then

$$L(1, \chi) = \sum_1^N \frac{\chi(n)}{n} + O\left(\frac{k^{3/2}}{N^2}\right).$$

Proof. Let

$$\rho(t) = [t] - t + \frac{1}{2}, \quad \sigma(t) = \int_0^t \rho(x) dx.$$

Since $\sigma(t)$ has a Lipschitz condition, the weighted Polya–Vinogradov bound gives

$$\sum_{a \bmod k} \chi(a) \sigma\left(\frac{x-a}{k}\right) = O(\sqrt{k})$$

uniformly in x . Now

$$\sum_{N+1}^M \frac{\chi(n)}{n} = \sum_{a \bmod k} \chi(a) \int_{N+1/2}^{M+1/2} \frac{1}{x} d\left[\frac{x-a}{k}\right]$$

$$\begin{aligned}
 &= \sum_a \chi(a) \int_{N+1/2}^{M+1/2} \left(\frac{1}{kx} dx + \frac{1}{x} d\rho\left(\frac{x-a}{k}\right) \right) \\
 &= \sum_a \chi(a) \left[\frac{1}{x} \rho\left(\frac{x-a}{k}\right) \right]_{N+1/2}^{M+1/2} + \sum_a \chi(a) \int_{N+1/2}^{M+1/2} \frac{1}{x^2} \rho\left(\frac{x-a}{k}\right) dx \\
 &= -\frac{1}{N+1/2} \sum_a \chi(a) \rho\left(\frac{N+1/2-a}{k}\right) + O\left(\frac{\sqrt{k \log k}}{M}\right) \\
 &\quad + \left[\frac{k}{x^2} \sum_a \chi(a) \sigma\left(\frac{x-a}{k}\right) \right]_{N+1/2}^{M+1/2} + \int_{N+1/2}^{M+1/2} \frac{2k}{x^3} \sum_a \chi(a) \sigma\left(\frac{x-a}{k}\right) dx.
 \end{aligned} \tag{2.1}$$

For large M all terms after the first term on the right of (2.1) are $O(k^{3/2}/N^2)$. Similarly

$$\begin{aligned}
 \sum_1^N \frac{n\chi(n)}{N^2} &= \sum_{a \bmod k} \chi(a) \int_0^{N+1/2} \frac{x}{N^2} d\left[\frac{x-a}{k}\right] \\
 &= \sum_a \chi(a) \frac{1}{N^2} \int_0^{N+1/2} \left(\frac{x dx}{k} + x d\rho\left(\frac{x-a}{k}\right) \right) \\
 &= \frac{1}{N^2} \sum_a \chi(a) \left[x\rho\left(\frac{x-a}{k}\right) - k\sigma\left(\frac{x-a}{k}\right) \right]_0^{N+1/2} \\
 &= \frac{N+1/2}{N^2} \sum_a \chi(a) \rho\left(\frac{N+1/2-a}{k}\right) + O\left(\frac{k^{3/2}}{N^2}\right),
 \end{aligned}$$

which cancels with (2.1) up to $O(k^{3/2}/N^2)$. If $\chi(-1) = 1$ and $k|N$, then

$$\sum_1^N n\chi(n) = \sum_1^N (N-n)\chi(N-n) = \sum_1^N (N-n)\chi(n);$$

but the right hand and left hand sides sum to zero. □

Lemma 2. (Dissection of the remainder term). Let $\chi(n)$ be a nontrivial character mod k , and let

$$r(n) = \sum_{d|n} \chi(d).$$

The sum function $R(N)$ of $r(n)$ satisfies

$$R(N) = \sum_1^N r(n) = NL(1, \chi) + R_1 + R_2 + O(\sqrt{k \log k}),$$

with

$$\begin{aligned}
 R_1 &= \sum_{d \leq \sqrt{kN}} \chi(d) \rho\left(\frac{N}{d}\right), \\
 R_2 &= \sum_{a \bmod k} \chi(a) \sum_{e \leq \sqrt{N/k}} \rho\left(\frac{N/e-a}{k}\right).
 \end{aligned}$$

Proof. $R(N)$ is the sum of $\chi(d)$ over pairs of integers d, e with $de \leq N$. If $\chi(d) \neq 0$, then d is not a multiple of k , and either $d < ke$ or $d > ke$. Hence

$$\begin{aligned} R(N) &= \sum_{d \leq \sqrt{kN}} \chi(d) \sum_{d/k < e \leq N/d} 1 + \sum_{e \leq \sqrt{N/k}} \sum_{ke < d \leq N/e} \chi(d) \\ &= \sum_{d \leq \sqrt{kN}} \chi(d) \left(\frac{N}{d} - \frac{d}{k} + \rho\left(\frac{N}{d}\right) - \rho\left(\frac{d}{k}\right) \right) \\ &\quad + \sum_{e \leq \sqrt{N/k}} \sum_a \chi(a) \left(\frac{N}{ek} - \frac{a}{k} - \left(e - \frac{a}{k} \right) + \rho\left(\frac{N/e - a}{k}\right) - \rho\left(e - \frac{a}{k}\right) \right). \end{aligned}$$

Since $\rho(e - a/k) = -\rho(a/k)$ when $\chi(a)$ is nonzero, we have

$$\begin{aligned} R(N) &= N \sum_{d \leq \sqrt{kN}} \frac{\chi(d)}{d} \left(1 - \frac{d^2}{kN} \right) + R_1 + R_2 \\ &\quad + \sum_{a \pmod k} \chi(a) \rho\left(\frac{a}{k}\right) \left(\sum_{e \leq \sqrt{N/k}} 1 - \sum_{\substack{d \leq \sqrt{kN} \\ d \equiv a \pmod k}} 1 \right). \end{aligned}$$

The first term is $N\mathcal{L}(1, \chi) + O(\sqrt{k})$ by Lemma 1, and the last term is $O(\sqrt{k} \log k)$ by the Polya–Vinogradov theorem. □

There are three ways of proceeding.

1. Split R_1 into sums with a condition $n \equiv a \pmod k$ for $a = 1, \dots, k - 1$, and consider each value of a separately in R_1 and R_2 . Theorem 1 follows at once from Theorem 4 of [3], which is Theorem 5.2.4 of [4].
2. Split R_1 and R_2 into $k - 1$ sums corresponding to nonzero residue classes mod k as above. They form a congruence family in the sense of Lemma 4.3.6 of [4]. A slightly better bound holds on average for the sums of a congruence family.
3. Modify the method of [3] to take the character in R_1 and R_2 through all the Poisson summation steps.

Method 3 should be the most powerful. However the calculations produce characters of shifted arguments. In order to separate the variables in readiness for the large sieve, we must either subdivide or endure extra Gauss sums as factors in the upper bound. Also, the rank of the bilinear form in the large sieve is multiplied by a power of k .

For methods 2 and 3 we expand $\rho(t)$ as a finite Fourier series [4, Lemma 2.1.9]:

$$\rho(t) = \sum_{h \neq 0} \frac{c(h) e(ht)}{c(0) 2\pi i h} + O\left(\frac{1}{H} + \min\left(1, \frac{1}{H^3 \|t\|^3}\right)\right), \tag{2.2}$$

where $c(h)$ are the coefficients of the second Fejer kernel, expressed in terms of binomial coefficients by:

$$c(\pm h) = \begin{cases} 2\binom{H-h+1}{h} C_3 - 4\binom{H-h+1}{h-1} C_3 & \text{for } 0 \leq h \leq H-2, \\ 2\binom{H-h+1}{h} C_3 & \text{for } H-1 \leq h \leq 2H-2, \\ 0 & \text{for } h \geq 2H-1. \end{cases}$$

Thus R_1 can be written as

$$R_1 = \sum_{m \leq \sqrt{(kN)}} \chi(m) \sum_{h \neq 0} \frac{c(h)}{c(0)} \frac{e(hN/m)}{2\pi i h} + O\left(\frac{\sqrt{(kN)}}{H}\right) + O\left(\sum_{m \leq \sqrt{(kN)}} \min\left(1, \frac{1}{H^3 \|N/m\|^3}\right)\right). \tag{2.3}$$

Similarly

$$R_2 = \sum_{b \bmod k} \chi(b) \sum_{n \leq \sqrt{(N/k)}} \sum_{h \neq 0} \frac{c(h)}{c(0)} \frac{1}{2\pi i h} e\left(\frac{hN}{kn} - \frac{bh}{k}\right) + O\left(\frac{\sqrt{(kN)}}{H}\right) + O\left(\sum_b \sum_{n \leq \sqrt{(N/k)}} \min\left(1, \frac{1}{H^3 \|N/kn - b/k\|^3}\right)\right). \tag{2.4}$$

In (2.3) the number of values of m in a range $M \leq m < 2M$ for which $\|N/m\| \leq \delta$ is

$$O(\delta M + N^{23/73}(\log N)^{315/146})$$

by Theorem 2 of [3], modified in the same way that Theorem 3 of [3] was modified to produce Theorem 4. When we sum M through powers of two, then the third error term in (2.3) is

$$O\left(\frac{\sqrt{(kN)}}{H} + N^{23/73}(\log N)^{461/146}\right). \tag{2.5}$$

We obtain the same estimate for the error term in (2.4).

This account is over-simplified. It is better to divide R_1 and R_2 into blocks in which m or n have fixed order of magnitude, $M \leq n < 2M$ for some M , and then to choose $H = H(M)$ in (2.2) to be constant within the block, but different for different blocks. In each block, the error term in (2.2) can also be estimated by double exponential sums (without characters) as in [3].

3. Congruence families of sums

A sum $\sum \rho(g(m))$ over some range $M \leq m < M_2$ corresponds to the remainder term in counting lattice points (m, n) in a region partly bounded by a curve $y = g(x)$. Counting points with $m \equiv \ell \pmod{k}$ corresponds to a sum

$$\sum \rho(g(km + \ell)) = \sum \rho\left(f\left(m + \frac{\ell}{k}\right)\right) \tag{3.1}$$

between suitable limits, with $f(x) = g(kx)$. Counting points with $n \equiv \ell \pmod{k}$ corresponds to a sum

$$\sum \rho\left(\frac{g(m) - \ell}{k}\right) = \sum \rho\left(f(m) - \frac{\ell}{k}\right), \tag{3.2}$$

where $f(x) = g(x)/k$. The family of sums of the form (3.1) or (3.2) as ℓ varies is called a congruence family. Theorem 8 of [3], which deals with a family of sums given by

different values of a parameter, does not apply, as the parameter must occur non-trivially, not as a linear shift. The saving occurs because changing the parameter changes the first derivative of the argument of $\rho(t)$. In a congruence family the change in the derivative is negligible, but the function itself changes in a predictable way. This idea was developed by Watt [7] for simple exponential sums. We obtain results that correspond to Theorems 7 and 8 of [3].

Lemma 3. (Congruence families of double exponential sums). Let $F(x)$ be a real function with four continuous derivatives for $1 \leq x \leq 2$, and let $g(x), G(x)$ be bounded functions of bounded variation on $1 \leq x \leq 2$. Let C_0, \dots, C_5 be real numbers ≥ 1 . Let H and M (integers) and T (real) be large parameters. Suppose that

$$|F^{(r)}(x)| \leq C_r,$$

for $r = 1, \dots, 4$, that

$$|F^{(r)}(x)| \geq 1/C_r, \tag{3.3}$$

for $r = 1, 2$, and that either case 1 or case 2 holds:

Case 1. $M \leq C_0 T^{1/2}$ and (3.3) holds for $r = 3$ also.

Case 2. $M \geq C_0^{-1} T^{1/2}$ and

$$|F'F^{(3)} - 3F''^2| \geq 1/C_5.$$

Let k be a fixed positive integer, and for $\ell = 0, \dots, k-1$ let S_ℓ denote either the sum

$$S_\ell = \sum_{h=H}^{2H-1} g\left(\frac{h}{H}\right) \sum_{m=M}^{2M-1} G\left(\frac{m}{M}\right) e\left(\frac{hT}{M} F\left(\frac{m}{M} + \frac{\ell}{kM}\right)\right) \tag{3.4}$$

or the sum

$$S_\ell = \sum_{h=H}^{2H-1} g\left(\frac{h}{H}\right) \sum_{m=M}^{2M-1} G\left(\frac{m}{M}\right) e\left(\frac{hT}{M} F\left(\frac{m}{M}\right) - \frac{h\ell}{k}\right). \tag{3.5}$$

Then there are constants C_6, C_7 and C_8 constructed from C_0, \dots, C_5 such that if

$$C_6 T^{1/3} \leq M \leq C_6^{-1} T^{2/3}$$

and

$$H \leq C_7 \min(M^{3/2}/T^{1/2}, M^{1/2}, MT^{-7/27}),$$

then we have bounds of the form

$$\sum_{\ell=0}^{k-1} |S_\ell|^2 = O(EkH^2 T(\log T)^{9/2}), \tag{3.6}$$

where the constant in the upper bound is constructed from C_0, \dots, C_5 , from the bounds for the functions $g(x)$ and $G(x)$.

Case (a). In cases 1 and 2, for

$$\begin{aligned} & \left(\frac{d^3(k)HT}{k^3M}\right)^{1/5} + \left(\frac{MT^{1/8}}{H}\right)^{1/4} + \frac{d(k)}{k} \left(\frac{HT^{1/3}}{M}\right)^4 + \left(\frac{HT^{1/3}}{M}\right)^{5/2} \\ & \geq C_8^{-1} \min\left(\frac{T}{M^2} + \frac{M^2}{T}, \frac{H^5 T^2}{M^5}\right), \end{aligned} \tag{3.7}$$

we have (3.6) with

$$\begin{aligned} E &= \left(\frac{d(k)}{k}\right)^{4/35} \left(\frac{H}{M}\right)^{3/35} \frac{1}{T^{12/35}} + \frac{1}{T^{3/8}} \\ &+ \left(\frac{d(k)}{k}\right)^{1/2} \left(\frac{H}{M}\right)^{15/4} T^{3/4} + \left(\frac{H}{M}\right)^3 T^{1/2}. \end{aligned} \tag{3.8}$$

Case (b). In case 1 we have (3.6) with

$$\begin{aligned} E &= \left(\frac{d(k)}{k}\right)^{2/7} \frac{H^{1/7} M^{3/7}}{T^{4/7}} + \frac{M^{9/11}}{H^{1/11} T^{8/11}} + \frac{1}{H^{1/9} M^{1/3} T^{2/9}} \\ &+ \left(\frac{d(k)}{k}\right)^{1/2} \frac{H}{T^{1/2}} + \frac{H^{1/4} M^{3/4}}{T^{3/4}} \end{aligned} \tag{3.9}$$

Case (c). In case 2 we have (3.6) with

$$\begin{aligned} E &= \left(\frac{d(k)}{k}\right)^{2/7} \frac{H^{1/7}}{M^{5/7}} + \frac{1}{H^{1/11} M^{7/11}} + \frac{M^{5/9}}{H^{1/9} T^{2/3}} \\ &+ \left(\frac{d(k)}{k}\right)^{1/2} \frac{HT^{1/2}}{M^2} + \frac{H^{1/4} T^{1/4}}{M^{5/4}}. \end{aligned} \tag{3.10}$$

Proof. The proof is a variation on that of Theorem 7 in [3]. The sum over m is divided into short intervals labelled by rational numbers (Farey arcs), with an approximate equivalence relation that we call resonance. Approximate or fuzzy equivalence means that transitivity weakens the approximation. In [3] the extra structure given by the parameter is used only to compare corresponding Farey arcs in different sums of the family. Here we use the congruence structure in the same way. The congruence structure is simpler, so there is less constraint on the length of the short intervals. The comparison occurs differently, and the possible saving is less. The second term and the last term in the bounds for E dominate in cases when the maximum saving occurs. The other terms correspond to terms and cases in Theorem 7 of [3]. □

We make some changes of notation in order to apply Lemma 3 to the sums R_1 and R_2 . In R_1 we must classify the values of m into residue classes $\ell \pmod{k}$, so that $m = kn + \ell$ for some n . In R_2 we merely write ℓ for b . The variable n in R_1 and R_2 becomes m for Lemma 3. We write H_1 for $H(M)$ in (2.2), so that we can use H as a

parameter for blocks $H \leq h < 2H$ of the finite sum. We only consider positive h ; negative h gives the complex conjugate sum.

The block sums of R_1 take the form (3.4), and those of R_2 take the form (3.5), with

$$F(x) = 1/x, \quad T = N/k,$$

and with $g(x)$ a continuous function with

$$g\left(\frac{h}{H}\right) = \frac{H}{2\pi h} \frac{c(h)}{c(0)}$$

(the factor H is inserted for homogeneity), and

$$G\left(\frac{x}{M}\right) = \begin{cases} 1 & \text{for } x \leq \sqrt{(N/k)}, \\ 0 & \text{for } x > \sqrt{(N/k)}. \end{cases}$$

To prove Theorem 2 we need

$$H_1 = C_9 \left(\frac{k}{d(k)}\right)^{4/73} MT^{-23/73} (\log T)^{-315/146} \tag{3.11}$$

for some C_9 (which affects the constant B), to overcome the term $1/H_1$ in (2.2), and

$$\sum_{\ell} |S_{\ell}|^2 = O\left(\left(\frac{d(k)}{k}\right)^{8/73} kH^2 T^{46/73} (\log T)^{315/73}\right) \tag{3.12}$$

for each block sum. There are $O(\log^2 T)$ different block sums, for different size ranges of H and M . The various cases of Lemma 3 give ranges $H_2(M) \leq H \leq H_3(M)$ in which (3.12) holds, actually with an extra factor of the form

$$(H_2(M)/H)^{\delta_1} + (H/H_3(M))^{\delta_2}$$

for some positive δ_1 and δ_2 . The sum over blocks of h gives a constant factor, not a logarithmic one. We always have $M = O(\sqrt{T})$. The terms in case (a) of Lemma 3 have H to a power greater than two. If the first term in (3.8) gives the order of magnitude of E , then (3.12) holds for $H \leq H_1$. The second term is smaller for

$$k/d(k) \leq C_{10} T^{3/64} (\log T)^{-9/16}, \tag{3.13}$$

and the third and fourth terms in (3.8) do not matter for $H \leq H_1$. For

$$C_{11} \left(\frac{k}{d(k)}\right)^{1/6} T^{7/16} \leq M \leq C_{11}^{-1} \left(\frac{d(k)}{k}\right)^{1/6} T^{9/16} \tag{3.14}$$

the condition (3.7) is satisfied for all H . For smaller M we use case (b) for small H . The order of magnitude of E changes smoothly as we pass from case (a) to case (b). The terms in (3.9) with H in the denominator may make (3.12) fail for small H . We must also consider H and M below the ranges permitted in Lemma 3.

As in [3], for small H or M we use the simple exponential sum bound from the exponent-pair (2/7, 4/7) to get

$$S_{\ell} = O(HT)^{2/7},$$

which implies (3.12) for

$$H \leq H_0 = \left(\frac{d(k)}{k}\right)^{14/73} T^{15/146} (\log T)^{2205/292}.$$

This range contains $H \leq H_1$ for

$$M \leq C_{12} \left(\frac{d(k)}{k}\right)^{18/73} T^{61/146} (\log T)^{2835/292}. \quad (3.15)$$

We find that blocks with $H > H_0$ satisfy (3.12) by case (b) of Lemma 3 for

$$\begin{aligned} C_{13} \left(\frac{k}{d(k)}\right)^{86/219} T^{179/438} (\log T)^{-573/292} &\leq M \\ &\leq C_{13} \left(\frac{d(k)}{k}\right)^{34/219} T^{589/1314} (\log T)^{179/292}. \end{aligned} \quad (3.16)$$

For

$$\frac{k}{d(k)} \leq C_{14} T^{1/70} (\log T)^{639/35}, \quad (3.17)$$

the range of M in (3.16) overlaps the ranges (3.14) and (3.15), and we have covered all cases. Since $T = N/k$, we have proved Theorem 2 for any $C > 71$, with A chosen so that (3.13) holds. The lower bound for C can be reduced by using deeper bounds for simple exponential sums, which would increase H_0 , and relax (3.15), (3.16) and (3.17). However we must have $C > 67/3$ to satisfy (3.13).

References

- [1] Davenport H, *Multiplicative number theory* (1967) (Chicago: Markham)
- [2] Huxley M N, Exponential sums and lattice points, *Proc. London Math. Soc.* **60** (1990) 471–502
Huxley M N, Corrigendum, *Proc. London Math. Soc.* **66** (1993) 70
- [3] Huxley M N, Exponential sums and lattice points II, *Proc. London Math. Soc.* **66** (1993) 279–301
- [4] Huxley M N, Area, lattice points and exponential sums, (to appear)
- [5] Iwaniec H and Mozzochi C J, On the circle and divisor problems, *J. Number Theory* **29** (1988) 60–93
- [6] Landau E, *Vorlesungen über Zahlentheorie* (1927) (Leipzig: Hirsel)
- [7] Watt N, A hybrid bound for Dirichlet L -functions on the critical line, *Proc. Amalfi Conference in analytic number theory* (Salerno U P) (1992) 387–392