

Some Old Problems and New Results about Quadratic Forms

William Duke

Introduction

It may be a challenging problem to describe the integer solutions to a polynomial equation in several variables. Which integers, for example, are represented by a quadratic polynomial? This question has a rich and complex history, and the theory it has motivated is still flourishing today. My aim in this article is to describe some recent developments in this theory, especially about positive quadratic forms in three variables, and to introduce the nonexpert to some of its basic themes. Let us begin with a classical and easily described example.

An Assertion of Fermat

As early as 1638 Fermat made the statement that every number is a sum of at most three triangular numbers, four squares, five pentagonal numbers, and so on.¹ In a letter to Pascal in 1654 he describes it as being so far his most important result. Much doubt has been cast upon his claim for a proof, especially for the case of

three triangular numbers.² Two hundred years ago Gauss began his mathematical diary, and among the early entries is his famous Archimedean allusion dated July 10, 1796,

$$**EYPHKA \quad num = \triangle + \triangle + \triangle.$$

Gauss is recording that he has found a proof that every positive integer m can be written as the sum of three triangular numbers. The triangular numbers are $0, 1, 3, 6, 10, 15, \dots, (1/2)n(n+1), \dots$, so Gauss's statement means that every number m may be represented by the particular quadratic polynomial in three variables

$$m = \frac{n_1^2 + n_1}{2} + \frac{n_2^2 + n_2}{2} + \frac{n_3^2 + n_3}{2}$$

¹Triangular numbers are the numbers $1, 3, 6, 10, \dots, (1/2)(n^2 + n), \dots$; squares are $1, 4, 9, 16, \dots, n^2, \dots$; pentagonal numbers are $1, 5, 12, 22, \dots, (1/2)(3n^2 - n), \dots$; and hexagonal numbers are $1, 6, 15, 28, \dots, 2n^2 - n, \dots$. In general, the n th polygonal number of order k is given by the quadratic polynomial $(1/2)[(k-2)n^2 - (k-4)n]$. To simplify statements, we shall include 0 as a polygonal number.

²For an authoritative discussion of the early history and references, see [24].

Research supported in part by NSF grant DMS-9500797. William Duke is associate professor of mathematics at Rutgers University. His e-mail address is duke@math.rutgers.edu.

with nonnegative integers n_1, n_2 , and n_3 . This is equivalent to the equation

$$8m + 3 = (2n_1 + 1)^2 + (2n_2 + 1)^2 + (2n_3 + 1)^2,$$

so Gauss's statement is equivalent to the statement that every number of the form $8m + 3$ is a sum of three odd squares. The more general theorem that a number is a sum of three squares precisely when it is not of the form $4^b(8m + 7)$ for $b \geq 0$ was first published by Legendre in 1798, and this fundamental result was given a definitive proof by Gauss in his *Disquisitiones* in 1801. That every number is a sum of four squares was proved earlier by Lagrange in 1772 building on work of Euler. In 1813 Cauchy gave the first proof of Fermat's assertion in total by deriving it in an elementary (but involved) way from the three triangular number theorem. Cauchy's theorem is sharp in the sense that there are numbers which cannot be represented by fewer than k polygonal numbers of order k , for example, $2k - 1$.

Some time later Dirichlet gave a beautiful formula for the number of ways in which m can be expressed as the sum of three triangular numbers. In the special case that $8m + 3$ is a prime it says that this number is the excess of the number of quadratic residues (squares modulo $8m + 3$) over nonresidues modulo $8m + 3$ in the interval from 1 to $4m + 1$.³ A triangular number is an example of a polygonal number of order three. The n th polygonal number of order k may be defined as the sum of the first n terms of an arithmetic progression with first term 1 and common difference $k - 2$, and so is given by the quadratic polynomial $(1/2)[(k - 2)n^2 - (k - 4)n]$. As their name implies, polygonal numbers have a geometric origin. This may be seen in the accompanying figure (Figure 1). Quite a lot of early number theory was concerned with various properties of these and other geometrically related sequences [4].

It has recently been pointed out by Guy [11] that some interesting and difficult problems about representing integers as sums of polygonal numbers are still open. Without going into details, I will describe an example where some progress has recently been made on one of these old problems. Legendre, in the third edition of his *Théorie des Nombres* of 1830, proved by elementary means that every number larger than 1,791 is a sum of four hexagonal numbers. The question arose whether or not three hexagonal

numbers eventually suffice. Theorem 1 of [10] has the following application to this question.

Theorem. Every sufficiently large number is a sum of three hexagonal numbers.

Since the n th hexagonal number is $n(2n - 1)$, which is also the $(2n - 1)$ st triangular number, this result is a kind of strengthening of the three triangular number theorem of Gauss. However, it has the drawback of being noneffective in the sense that an explicit bound for the largest number which is not a sum of three hexagons cannot be given at present, unless one is willing to assume conjectures like the Riemann hypothesis. We shall see the origin of this defect later.

Before going into the general problem of representation and some of the methods used to study it, I will first describe two more recent results about quadratic forms placed in their historical context.

A Paper of Ramanujan

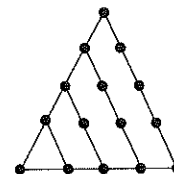
In 1917 Ramanujan [16] published a paper which was to have a big impact on subsequent research on representations by quadratic forms. He considered the problem of finding all integers $0 \leq a \leq b \leq c \leq d$ for which every positive integer is represented in the form

$$ax_1^2 + bx_2^2 + cx_3^2 + dx_4^2.$$

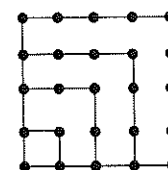
An elementary (and amusing) case-by-case analysis shows that in order for 1, 2, 3, 5 to be represented, the first three terms (a, b, c) must be $(1, 1, 1)$, $(1, 1, 2)$, $(1, 1, 3)$, $(1, 2, 2)$, $(1, 2, 3)$, $(1, 2, 4)$, or $(1, 2, 5)$. None of the associated ternary forms $ax_1^2 + bx_2^2 + cx_3^2$ represents all numbers, the smallest exceptions being, respectively, 7, 14, 6, 7, 10, 14, and 10. This leaves 55 possible quaternary forms, and, based on simple rules for the integers represented by the above ternaries which Ramanujan discovered empirically, he concluded that these 55 forms actually do rep-

Figure 1. Polygonal numbers figured prominently in problems of early number theory.

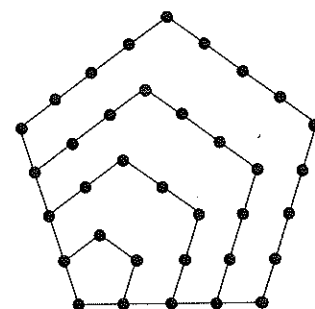
Triangular numbers arise by counting the number of points in a triangular array.



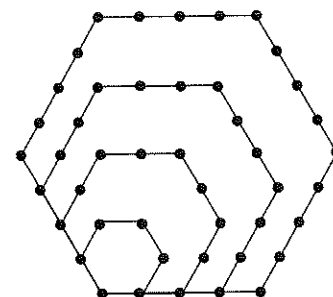
The sequence of triangular numbers is thus 1, 3, 6, 10, 15, ...



Similarly, the sequence of squares is 1, 4, 9, 16, 25, ...



Pentagonal numbers: 1, 5, 12, 22, 35, ...



Hexagonal numbers: 1, 6, 15, 28, 45, ...

The n th polygonal number of order k is given by the formula $1/2[(k - 2)n^2 - (k - 4)n]$.

³For instance, when $m = 2$ so $8m + 3 = 19$, the quadratic residues modulo 19 up to $9 = 4m + 1$ are 1, 4, 5, 6, 7, 9 and the nonresidues are 2, 3, 8, so the excess is 3, which is the number of ways of expressing 2 as a sum of three triangular numbers: $2 = 1 + 1 + 0 = 1 + 0 + 1 = 0 + 1 + 1$.

represent all numbers.⁴ It is natural to generalize Ramanujan's problem to other quadratic forms. An integer-valued m -ary positive quadratic form (or just a form) is a homogenous quadratic polynomial $Q(x) = Q(x_1, x_2, \dots, x_m)$ with integral coefficients which satisfies $Q(x) > 0$ for real $x \neq 0$. Such a form may be represented in matrix notation by $Q(x) = x^t A x$ where $A = \frac{1}{2} \frac{\partial^2 Q(x)}{\partial x_i \partial x_j}$ is a positive symmetric matrix. If A has integer entries, then $Q(x)$ may be called an integer-matrix form, for example, the diagonal form $ax_1^2 + bx_2^2 + cx_3^2 + dx_4^2$. Perhaps the final word on the subject of forms which represent all positive numbers has been given very recently by Conway and Schneeberger, who have provided the following elegant characterizations.

Theorem [2]. If a positive integer-matrix quadratic form represents each of

$$1, 2, 3, 5, 6, 7, 10, 14, 15,$$

then it represents all positive integers.

Conjecture. If a positive integer-valued quadratic form represents each of

$$\begin{aligned} &1, 2, 3, 5, 6, 7, 10, 13, 14, 15, 17, 19, 21, 22, \\ &23, 26, 29, 30, 31, 34, \\ &35, 37, 42, 58, 93, 110, 145, 203, 290, \end{aligned}$$

then it represents all positive integers.

These statements are sharp in the sense that the form $x_1^2 + 2x_2^2 + 5x_3^2 + 5x_4^2$ represents all numbers but 15, while $4x_1^2 + x_1x_2 + x_2^2 + x_2x_3 + 2x_3^2 + 29x_4^2 + 29x_4x_5 + 29x_5^2$ represents all but 290. They expect that their conjecture will soon be a theorem.

In his paper [16] Ramanujan also introduces the problem of finding the $ax_1^2 + bx_2^2 + cx_3^2 + dx_4^2$ that represents all sufficiently large integers, a problem he refers to as being *much more difficult and interesting*. This problem was essentially solved by Kloosterman [14] in 1926. Kloosterman's paper represents a major breakthrough, for in it he refined the circle method in a way that allowed him to obtain a qualitative result about representations by quaternary forms from a certain estimate for what is now known as a Kloosterman sum. In its simplest incarnation this is the finite sum

$$K(n, p) = \sum_{d=1}^{p-1} e^{2\pi i n(\bar{d}+d)/p}$$

⁴Ten years later Dickson [5] observed that one of them, the form corresponding to (1,2,5,5), does not represent 15, but proved that otherwise Ramanujan was correct, in particular about the rules for the special ternaries.

for prime p and integral n not divisible by p , where $\bar{d}$ is the multiplicative inverse of d modulo p . Kloosterman's estimate, which was enough for the application to quaternary quadratic forms, was later superseded by Weil's best possible estimate

$$|K(n, p)| \leq 2\sqrt{p}$$

obtained as a consequence of the Riemann Hypothesis for curves.

Ramanujan's paper has also stimulated much work on the still-more-difficult theory for ternary forms. In a footnote he wrote that *the even numbers which are not of the form $x^2 + y^2 + 10z^2$ are the numbers $4^k(16\mu + 6)$, while the odd numbers that are not of that form, viz.,*

$$\begin{aligned} &3, 7, 21, 31, 33, 43, 67, 79, \\ &87, 133, 217, 219, 223, 253, 307, 391 \dots \end{aligned}$$

do not seem to obey any simple law. Dickson confirmed the observation about even numbers by a simple argument, but the problem of whether there are infinitely many odd numbers that are not represented remained open until recently. It follows from [10] that

Theorem. The set of odd numbers not represented by Ramanujan's form

$$x_1^2 + x_2^2 + 10x_3^2$$

is finite.

Once again, the proof of this result does not yield an explicit bound for the number of exceptions. Actually, Ramanujan's list of exceptions is not complete and two more exist: 679 and 2,719. In an impressive recent paper, Ono and Soundararajan [15] have shown that if one assumes certain Riemann Hypotheses, then these are actually all.

The Problem of Representation

The above examples belong to the general problem of understanding which positive integers n are represented by a given integer-valued positive form $Q(x_1, \dots, x_m)$ for integral vectors x , how many such vectors there are, and how these vectors are distributed.

A necessary condition for the integral solvability of $n = Q(x)$ is that the congruence $Q(x) \equiv n \pmod{q}$ have a solution for all positive integers q or, in other words, that n be represented over the p -adic integers for all p . The local representability of n by $Q(x)$, that is, the solvability of these congruences, does not in general guarantee the existence of an *integral* solution to $Q(x) = n$ unless we replace the individual form by a certain equivalence class of forms, the genus. The notion of a genus of forms was first introduced by Gauss in the binary case. To de-

fine it, first say that two forms Q_1 and Q_2 are equivalent over $\mathbb{Z}$ if there is a $U \in GL(m, \mathbb{Z})$ so that $U^t A_1 U = A_2$ for the associated matrices. Two (positive) forms Q_1 and Q_2 are said to belong to the same genus if, for each q , Q_1 is equivalent over $\mathbb{Z}$ to a form which is congruent modulo q to Q_2 . Another way of saying this is that Q_1 and Q_2 are equivalent over the p -adic integers for all p . Two forms equivalent over $\mathbb{Z}$ represent the same integers, so for questions of representation one may consider as basic objects $\mathbb{Z}$ -equivalence classes of forms. It is an important fact that a genus of forms consists of only finitely many $\mathbb{Z}$ -equivalence classes. The following result shows the importance of this concept for us.

Theorem [1]. If, for every q , the congruence $Q(x) \equiv n \pmod{q}$ has a solution, then some form in the same genus as $Q(x)$ represents n integrally.

Although local representability seems to involve infinitely many congruences, in fact it suffices to take a single value of q depending on the determinant $D = \det 2A$ and possibly n . If the genus happens to consist of a single class, then this theorem provides a most satisfactory answer to the question of which integers are represented. This is quite rare, however, and it cannot happen if m or D is too big. Ramanujan's form $x_1^2 + x_2^2 + 10x_3^2$ is in a genus consisting of two classes, the other being represented by $2x_1^2 + 2x_2^2 + 3x_3^2 - 2x_1x_3$. Thus, although it may be clear which integers are represented by at least one form in the genus, the difficulty is in deciding which. Despite this, one might hope that different forms in the same genus tend to represent the same integers. This is not true for binary forms, whose deeper representation properties comprise a part of class field theory and will not be considered here. It is essentially true for forms in three or more variables, and we will attempt to make that more precise and point out the special obstacles that arise for quaternary and ternary forms.

The Analytic Approach

A natural approach is to turn to the problem of counting the number of representations. The power of analysis enters this problem through the *theta series*

$$\vartheta(z) = \sum_{\alpha \in \mathbb{Z}^m} e(zQ(\alpha)) = \sum_{n \geq 0} r(n)e(nz)$$

where $e(z) = e^{2\pi iz}$, which is a generating (Fourier) series for the number of representations $r(n) = r_Q(n) = \#\{\alpha \in \mathbb{Z}^m; Q(\alpha) = n\}$. This is seen by collecting together the terms in the first sum with the same value of $Q(\alpha)$.

Jacobi was the first to exploit theta functions for representation problems. For $Q_4(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2$, for example, he gave the nice identity

$$\vartheta(z) = 1 + 8 \sum_{n \geq 1} \frac{ne(nz)}{1 + (-1)^n e(nz)}$$

obtained by using the theory of elliptic functions. By expanding each term in the sum as a geometric series and then collecting together the coefficients of $e(nz)$, it can be seen that the number of representations of n as a sum of four squares is eight times the sum of those divisors of n which are not multiples of four. In particular, it is never zero!

A theta series defines a holomorphic function on the upper-half plane $\mathcal{H}$ having the real line as a natural boundary. While its Fourier coefficients encode the integral representation properties of $Q(x)$, its behavior near the rational points encodes the local representation properties of $Q(x)$. The basic idea of the circle method is to approximate $\vartheta(z)$ by a function with similar behavior at the rational points a/c and then compare Fourier expansions. This idea leads one to attempt to approximate $r(n)$ for $n > 0$ by an infinite series

$$\rho(n) = c_Q n^{m/2-1} \sum_{c>0} A_c(n) \quad (1)$$

where $A_c(n)$ is a certain finite exponential sum and c_Q is a positive constant depending on the form Q .

This series is called the *singular series*, since it arises from the rational singularities of $\vartheta(z)$. It has a remarkable relation with the local solvability of $Q(x) = n$. Let $r(n, q)$ be the number of solutions of the congruence $Q(x) \equiv n \pmod{q}$. It is shown that there is a product representation $\sum A_c(n) = \prod_p \alpha_p(n)$ where $\alpha_p(n)$ is the p -adic density of representation which is given by the stabilizing limit $\alpha_p(n) = \lim_{q=p^k \rightarrow \infty} r(n, q)/q^{m-1}$. The series and product converge (absolutely if $m > 3$) to a nonzero limit exactly when we have local solvability.

In general, the analytic approach to the problem of integral representation is to write

$$r(n) = \rho(n) + a(n)$$

and try to estimate $a(n)$ from above and $\rho(n)$ from below when it is not zero.

It turns out that the only obstruction to $\rho(n)$ being essentially as large as $n^{m/2-1}$ when it is not zero is the possible existence of primes p for which $Q(x)$ does not represent zero p -adically, which means that for some a , $Q(x) \equiv 0 \pmod{p^a}$ implies that $x \equiv 0 \pmod{p}$. No such p exists when $m > 4$, so one has $\rho(n) \geq cn^{m/2-1}$, for some positive constant c ,

[3] when he proved the Ramanujan conjecture (formulated by Ramanujan in just our context in terms of Q_{24} !) giving $a(n) \ll n^{m/4-1/2+\epsilon}$.

It is perhaps less well known that for odd m —that is, for the half-integral weight $m/2$ —the same bound is conjectured to hold for square-free n . This half-integral weight Ramanujan conjecture is open, but it does follow from the Riemann Hypothesis for certain global L -functions. Until 1987 the best-known estimate for odd m was that corresponding to Weil's bound for Kloosterman sums:

$$a(n) \ll n^{m/4-1/4+\epsilon}.$$

Any improvement of this is precisely what is needed for the representation problem for ternary forms, since the exponent $m/4 - 1/4 = 1/2$ just fails to beat the lower bound of Siegel for the singular series. In 1987 a breakthrough was made by Iwaniec [12] in the half-integral weight case. In this paper he succeeded in reducing the exponent $m/4 - 1/4$ for odd $m > 3$ and square-free n . Shortly thereafter, his methods were extended to cover the case $m = 3$, and applications to ternary quadratic forms, both definite and indefinite, were made [6, 7]. An excellent exposition of Iwaniec's method is given by Sarnak in [18].

In fact, there exist cusp forms of weight $3/2$ that attain the bound $|a(n)| \geq cn^{1/2}$, but they are supported on finitely many square classes. This new difficulty was discovered in the context of the quadratic forms by Jones and Pall [13]. The concept of spinor genus was used to identify these pseudo-cusp forms locally in [19], and a spinor genus version of the representation theorem for ternary forms is given in [10]. The following is an easily stated corollary.

Theorem. Every sufficiently large square-free integer that is represented by a ternary form (*mod* D^2) is integrally represented.

To reiterate, there is no known way to estimate effectively the largest possible exception without assuming certain Riemann hypotheses. This result was conjectured by Ross and Pall [17] and also by Watson [23]. Watson established several interesting results about the integers that are locally represented but not integrally represented by a ternary form.

As alluded to earlier, a rather different approach to the ternary representation problem is possible by relating the Fourier coefficients to special values of L -functions. After the fundamental work of Shimura and Waldspurger, reducing the Weil exponent for the Fourier coefficients amounts to *breaking convexity* for a twisted automorphic L -function. In this context, breaking convexity means reducing the estimate

provided by the functional equation and the Phragmén-Lindelöf theorem for the size of the L -function at the central critical point in terms of the twist. The papers [8] do break convexity for these L -functions as well as others and thus yield new proofs of the ternary representation theorems. Yet another approach has been given in [9].

The third aspect of the problem of representation entails understanding the way in which the representing vectors x are distributed on the ellipsoid $Q(x) = n$ and in arithmetic progressions to a fixed modulus. This is what is needed, for example, to prove results about representing numbers as sums of polygonal numbers. The methods described in this article also apply to this distribution problem, for the appropriate harmonics needed to detect such distribution may be built into the theta series. Roughly speaking, the general result is that, even when restricted to certain progressions, the representing vectors become uniformly distributed on the ellipsoid $Q(x) = n$ as n gets large, provided they are numerous enough. For references on the distribution aspect of the representation problem and details on the ternary case, the reader may consult [10].

Acknowledgments

I would like to thank J. H. Conway, H. Rossi, P. Sarnak, and R. Schulze-Pillot for valuable comments on an earlier version of this paper.

References

- [1] J. W. S. CASSELS, *Rational quadratic forms*, Academic Press, 1978.
- [2] J. H. CONWAY and W. SCHNEEBERGER, to appear.
- [3] P. DELIGNE, *La conjecture de Weil*. I, *Inst. Hautes Études Sci. Publ. Math.* 43 (1973), 273–308.
- [4] L. E. DICKSON, *History of the theory of numbers*, II, Chelsea Publishing, 1971.
- [5] ———, *Integers represented by positive ternary quadratic forms*, *Bull. Amer. Math. Soc.* 33 (1927), 63–70.
- [6] W. DUKE, *Hyperbolic distribution problems and half-integral weight Maass forms*, *Invent. Math.* 92 (1988), 73–90.
- [7] ———, *Lattice points on ellipsoids*, *Sém. Théor. Nombres Bordeaux* (2), 1987–88, No. 37, pp. 1–6.
- [8] W. DUKE, J. FRIEDLANDER, and H. IWANIEC, *Bounds for automorphic L -functions* I, II, *Invent. Math.* 112, 115 (1993, 1994), 1–8, 219–239.
- [9] ———, *Bilinear forms with Kloosterman fractions*, to appear in *Invent. Math.*
- [10] W. DUKE and R. SCHULZE-PILLOT, *Representations of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids*, *Invent. Math.* 99 (1990), 49–57.
- [11] R. K. GUY, *Every number is expressible as a sum of how many polygonal numbers?*, *Amer. Math. Monthly* 101 (1994), 169–172.

when it does not vanish. (It is conventional in analytic number theory to write this as $\rho(n) \gg n^{m/2-1}$, the positive constant c then being referred to as the *implied constant*.) For $m = 3$ and $m = 4$ such p may exist, but they are readily determined divisors of D . If we restrict the powers to which such primes divide n , then for $\rho(n) \neq 0$ we have the bound $\rho(n) \gg n/\log(\log n)$ when $m = 4$, and for any $\epsilon > 0$ the bound $\rho(n) \gg n^{1/2-\epsilon}$ when $m = 3$. The latter is a consequence of a theorem of Siegel [21] and is noneffective in the sense that no way of specifying the implied constant for a given $\epsilon > 0$ is known. This is the origin of the noneffectivity in our applications to ternary forms, and to remove it is a major unsolved problem in number theory. If one assumes the Riemann hypothesis for Dirichlet L -functions, then an effective lower bound can be given for the case $m = 3$.

As for upper bounds for $a(n)$, the circle method quickly yields the bound $|a(n)| \ll n^{m/4}$, and this is enough to give the following result of Tartakowsky.

Theorem [22]. For forms in five or more variables, every sufficiently large number n that is represented (*mod* D) is integrally represented.

For $m = 4$ the above bound for $|a(n)|$ no longer suffices (just), and it is here that Kloosterman made his breakthrough by improving the estimate for $a(n)$ and thus extending the above theorem to quaternary forms, provided that the powers to which certain primes divide n are restricted or that only primitive representations are considered. The (primitive) representation theorem for forms in four or more variables has also been proved using algebraic methods (see [1]). For ternary forms a deeper barrier exists, and to overcome it so far only the analytic approach has been fully successful. To this end it is natural to exploit the relation between quadratic forms and modular forms.

Fourier Coefficients of Modular Forms

The central role of modular forms in the study of quadratic forms, after being hinted at by Hardy and Ramanujan, was made very clear by Hecke and Siegel. This is based on the fact that the theta function $\vartheta(z)$ satisfies the transformation rule (for $\gamma \in \Gamma$, a certain congruence subgroup of the modular group)

$$\vartheta\left(\frac{az+b}{cz+d}\right) = \chi(\gamma)(cz+d)^{m/2}\vartheta(z),$$

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

for a certain multiplier $\chi(\gamma)$, which may be given explicitly. The same is true for the function

$$E(z) = \sum_{n \geq 0} \rho(n)e(nz), \quad \rho(0) = 1,$$

which has by construction the same behavior as the theta function at the rationals. This means that the difference $f(z) = \vartheta(z) - E(z) = \sum_{n \geq 1} a(n)e^{2\pi inz}$ is a holomorphic function on $\mathcal{H}$ that satisfies the same transformation rules but which has the property that the function $y^{m/4}|f(z)|$, which is well defined on the quotient $\Gamma \backslash \mathcal{H}$, is bounded there. In other words, $f(z)$ is a *cusp form of weight* $m/2$. The problem of bounding $a(n)$ from above is the problem of bounding a Fourier coefficient of a cusp form. The mere boundedness of $y^{m/4}|f(z)|$ is enough to give the estimate $a(n) \ll n^{m/4}$, which is also the trivial bound from the circle method. It was found independently by Rankin, Selberg, and Petersson that there is an elegant reformulation of the circle method based on Poincaré series and simple L^2 theory which encompasses the refinement made by Kloosterman. It gives the following estimate, which invites comparison with the singular series (1):

$$|a(n)|^2 \ll n^{m/2-1} \sum_{c>0} c^{-1} K(n, c) J(n/c). \quad (2)$$

Here $K(n, c)$ is the Kloosterman sum

$$K(n, c) = \sum \bar{\chi}(\gamma) e(n(a+d)/c),$$

the sum being over $\gamma \in \Gamma$ with $0 \leq a, d < c$, and J is a Bessel function, of which the Kloosterman sum is a finite version. Now the Kloosterman sum is over roughly c terms, and any bound of the form $K(n, c) \ll c^{1-\delta}$ with $\delta > 0$ yields, upon splitting the sum in (2) at $c = n$ and using $J(x) \ll \min(x^{m/2-1}, x^{-\frac{1}{2}})$, the estimate $a(n) \ll n^{m/4-\delta/2}$. For even m the multiplier $\chi = 1$ and each Kloosterman sum which occurs in (2) can be written

$$K(n, c) = \sum_{d(\text{mod } c), (c, d)=1} e(n(\bar{d} + d)/c)$$

where $\bar{d}$ is the multiplicative inverse of d modulo c . For its estimation Kloosterman showed that one may take any $\delta < 1/4$, and later Weil showed that one may take any $\delta < \frac{1}{2}$ as a consequence of the Riemann Hypothesis for curves, this result being best possible. For m odd the multiplier $\chi \neq 1$ and this circumstance actually allows any $\delta < \frac{1}{2}$ to be obtained rather easily, as Salié observed.

Another way to get a nontrivial bound for $a(n)$ is by the Rankin-Selberg method [20]. For even m (i.e., integral weights) the problem of estimating $a(n)$ was completely solved by Deligne

[3] when he proved the Ramanujan conjecture (formulated by Ramanujan in just our context in terms of Q_{241}) giving $a(n) \ll n^{m/4-1/2+\epsilon}$.

It is perhaps less well known that for odd m —that is, for the half-integral weight $m/2$ —the same bound is conjectured to hold for square-free n . This half-integral weight Ramanujan conjecture is open, but it does follow from the Riemann Hypothesis for certain global L -functions. Until 1987 the best-known estimate for odd m was that corresponding to Weil's bound for Kloosterman sums:

$$a(n) \ll n^{m/4-1/4+\epsilon}.$$

Any improvement of this is precisely what is needed for the representation problem for ternary forms, since the exponent $m/4 - 1/4 = 1/2$ just fails to beat the lower bound of Siegel for the singular series. In 1987 a breakthrough was made by Iwaniec [12] in the half-integral weight case. In this paper he succeeded in reducing the exponent $m/4 - 1/4$ for odd $m > 3$ and square-free n . Shortly thereafter, his methods were extended to cover the case $m = 3$, and applications to ternary quadratic forms, both definite and indefinite, were made [6, 7]. An excellent exposition of Iwaniec's method is given by Sarnak in [18].

In fact, there exist cusp forms of weight $3/2$ that attain the bound $|a(n)| \geq cn^{1/2}$, but they are supported on finitely many square classes. This new difficulty was discovered in the context of the quadratic forms by Jones and Pall [13]. The concept of spinor genus was used to identify these pseudo-cusp forms locally in [19], and a spinor genus version of the representation theorem for ternary forms is given in [10]. The following is an easily stated corollary.

Theorem. Every sufficiently large square-free integer that is represented by a ternary form (*mod* D^2) is integrally represented.

To reiterate, there is no known way to estimate effectively the largest possible exception without assuming certain Riemann hypotheses. This result was conjectured by Ross and Pall [17] and also by Watson [23]. Watson established several interesting results about the integers that are locally represented but not integrally represented by a ternary form.

As alluded to earlier, a rather different approach to the ternary representation problem is possible by relating the Fourier coefficients to special values of L -functions. After the fundamental work of Shimura and Waldspurger, reducing the Weil exponent for the Fourier coefficients amounts to *breaking convexity* for a twisted automorphic L -function. In this context, breaking convexity means reducing the estimate

provided by the functional equation and the Phragmén-Lindelöf theorem for the size of the L -function at the central critical point in terms of the twist. The papers [8] do break convexity for these L -functions as well as others and thus yield new proofs of the ternary representation theorems. Yet another approach has been given in [9].

The third aspect of the problem of representation entails understanding the way in which the representing vectors x are distributed on the ellipsoid $Q(x) = n$ and in arithmetic progressions to a fixed modulus. This is what is needed, for example, to prove results about representing numbers as sums of polygonal numbers. The methods described in this article also apply to this distribution problem, for the appropriate harmonics needed to detect such distribution may be built into the theta series. Roughly speaking, the general result is that, even when restricted to certain progressions, the representing vectors become uniformly distributed on the ellipsoid $Q(x) = n$ as n gets large, provided they are numerous enough. For references on the distribution aspect of the representation problem and details on the ternary case, the reader may consult [10].

Acknowledgments

I would like to thank J. H. Conway, H. Rossi, P. Sarnak, and R. Schulze-Pillot for valuable comments on an earlier version of this paper.

References

- [1] J. W. S. CASSELS, *Rational quadratic forms*, Academic Press, 1978.
- [2] J. H. CONWAY and W. SCHNEEBERGER, to appear.
- [3] P. DELIGNE, *La conjecture de Weil*, I, Inst. Hautes Études Sci. Publ. Math. 43 (1973), 273–308.
- [4] L. E. DICKSON, *History of the theory of numbers*, II, Chelsea Publishing, 1971.
- [5] ———, *Integers represented by positive ternary quadratic forms*, Bull. Amer. Math. Soc. 33 (1927), 63–70.
- [6] W. DUKE, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. 92 (1988), 73–90.
- [7] ———, *Lattice points on ellipsoids*, Sémin. Théor. Nombres Bordeaux (2), 1987–88, No. 37, pp. 1–6.
- [8] W. DUKE, J. FRIEDLANDER, and H. IWANIEC, *Bounds for automorphic L -functions*, I, II, Invent. Math. 112, 115 (1993, 1994), 1–8, 219–239.
- [9] ———, *Bilinear forms with Kloosterman fractions*, to appear in Invent. Math.
- [10] W. DUKE and R. SCHULZE-PILLOT, *Representations of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids*, Invent. Math. 99 (1990), 49–57.
- [11] R. K. GUY, *Every number is expressible as a sum of how many polygonal numbers?*, Amer. Math. Monthly 101 (1994), 169–172.