Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

C. HOOLEY

Some Recent Advances in Analytical Number Theory

The realm of the analytical theory of numbers is nowadays too vast for one to attempt a complete survey within an article of this length. We therefore mainly restrict ourselves to those aspects of the additive theory that are associated with the author's recent work.

The circle method of Hardy and Littlewood plays a dominant rôle in the analytic part of the additive theory of numbers. Familiar though this method is to experts in the field, it is appropriate in an expository article that we should give a brief description of the underlying procedure in order that we should be aware of its limitations and of the relevance to it of recent mathematical developments.

Avoiding complete generality for the sake of brevity and clarity, we can indicate the nature of the method by considering its formal application to the problem of determining whether an indeterminate equation

$$f(l_1, l_2, \ldots, l_r) = 0 \tag{1}$$

is soluble, where $f(x_1, \ldots, x_r)$ is a polynomial with rational integral coefficients. It being inherent in the technique that normally it should only be applied when the answer to the proposed question is thought to be in the affirmative, the method usually not only settles the problem of existence but supplies an estimate for the number $\nu(x)$ of solutions of (1) that lie in some large appropriate region R_x, where x is a parameter tending to infinity. In many, but by no means all, of the more important problems the natural form of this region is inherent in the other data and is therefore not the subject of a special definition; this, for example, is the situation in Waring's problem when we consider the representation of large numbers N as the sum of s k-th non-negative powers.

The genesis of the method, as modified by Vinogradov, is the observation that

$$r(x) = \int_0^1 \sum_{(l_1,\dots,l_r)\in R_x} e^{2\pi i f(l_1,\dots,l_r)\theta}\, d\theta.$$

To treat the integral the range of integration is split up into intervals (or arcs as they are usually called, since the procedure is easily interpreted in terms of the circumference of the unit circle) that are in some sense centred by rational numbers (Farey fractions) of the form h/k, where

$$(h,k) = 1, \quad 0 \leqslant h < k, \quad k \leqslant X,$$

and where X is a suitable function of x. When θ is at the "centre" h/k of an arc and k is small, the integrand can be estimated with great accuracy because

$$e^{2\pi i f(l_1,\dots,l_r)h/k}$$

is a periodic function in $l_1, \dots, l_r$ with small periods; consequently, by partial summation or some equivalent process, the integrand can also be satisfactorily calculated when θ is close to h/k. Thus part of the integral can be adequately treated, while the form of the calculations suggests that the residual part is negligible in circumstances where we may reasonably expect there to be an asymptotic formula for $r(x)$.

To validate the asymptotic formula thus suggested it is requisite to overcome the difficulties encountered when k is large or when θ is far from the centre of the arc it lies in (these two possibilities are partly interchangeable because there is usually some latitude in the choice of X). There are two main lines of development here. The first is for us to refine the calculations already made so that they are applicable to the entire range, endeavouring in some places to gain improvements by shewing there is some cancellation between contributions from arcs corresponding to a common denominator k. However, matters of such nicety intervene that it is seldom that the programme succeeds. This approach was first used by Kloosterman in his investigation on quaternary quadratic forms, whence flows the present custom of designating such a procedure by the term Kloosterman refinement.

The second and more common technique is applicable to problems that are additive or that can be made additive by a suitable transform-

ation. The Waring's problem about the number $\nu(N)$ of solutions of

$$l_1^k + \ldots + l_s^k = N$$

typifying the situation to be covered by this routine, the integrand is now

$$f^s(\theta)e^{-2\pi iN\theta},$$

where

$$f(\theta) = \sum_{l \leqslant N^{1/k}} e^{2\pi i l^k \theta}. \tag{2}$$

Over the set $\mathscr{M}$ of minor arcs, which the residual range of integration is termed, the integrand is bounded by

$$\overline{(bd}_{\theta \in \mathscr{M}} |f(\theta)|)^g \int_0^1 |f(\theta)|^{s-g} d\theta,$$

where $0 < g < s$. In favourable circumstances, which alas do not too often occur, the integral above can be estimated because it has a natural arithmetical meaning. The upper bound, on the other hand, has often been satisfactorily estimated through the work of Weyl, Weil, and Vinogradov. Impressive developments in Waring's and other problems have been achieved by these means. There is, however, the substantial shortcoming that the method is inapplicable whenever the order of magnitude of $\nu(N)$ is not larger than N. Consequently, it cannot deal with important unsolved problems such as the Goldbach problem or Waring's problem for four cubes. Similar remarks relate to the more general earlier context when $\nu(x)$ is small in terms of x, where as before R_x is chosen in a natural way. Finally, many variations in this line of attack have been introduced by various writers, and we refer the interested reader to the several treatises on the subject for further details.

Enough has been said already to see the potential relevance of exponential sums of the type

$$\sum_{l_1, \ldots, l_r \leqslant u} e^{2\pi i f(l_1, \ldots, l_r)h/k} \tag{3}$$

to the circle method. The study of such sums can be easily reduced to that of complete sums

$$\sum_{0 < l_1, l_2, \ldots, l_r \leqslant k} e^{2\pi i f(l_1, \ldots, l_r)h/k}, \tag{4}$$

which themselves are a specialization of sums of the form

$$\sum_{\substack{0<l_1,\dots,l_r\leqslant k\\ g(l_1,\dots,l_r)=0}} e^{2\pi i f(l_1,\dots,l_r)h/k}, \tag{5}$$

where the significance of the notation f may change as we go from (3) to (5). Although the importance of (4) has been long understood, it has been perhaps less appreciated that there are a multitude of ways in which (5) might conceivably be of assistance.

Best possible estimates for (3) in the case $r = 1$ were made available by Weil's work and allowed considerable progress to be made in additive number theory (both by the circle method and by other means). It was therefore to be expected that Deligne's fundamental and far reaching proof of the generalized Weil conjectures should lead to further advances as soon as it could be shewn how his work was applicable to the sums (5). Estimates for special cases having been obtained by various writers, best possible estimates in the general case were in fact first obtained by the speaker in 1979 by a very simple method ([7], [10], and in a paper shortly to appear), while shortly afterwards Katz obtained similar estimates by a more recondite method, which also shed much light on the structure of the L-functions over algebraic varieties [14].

Recently, a striking advance has been made by Heath-Brown using a Kloosterman refinement and the estimates for multiple exponential sums. As the culmination of a series of important papers, Davenport shewed that an n-ary cubic form f with integral coefficients had a non-trivial integral zero provided that $n \geqslant 16$ ([1], [2], [3]). The result is false for $n = 9$ even when f is non-singular but it had been conjectured that it is always true for $n = 10$. Heath-Brown [4] has proved its truth for $n = 10$ when the form is non-singular, an achievement that settles the situation for the most important category of cubic forms.

Notwithstanding the potential relevance of the Deligne estimates to the circle method, no other significant advance has yet been made through this order of ideas. This is due in part to certain deficiencies in the circle method to which we shall later allude and also to the fact that in many of the more important outstanding problems the expected value of $r(x)$ is too small for the method to be applied in any but the most abstruse manner.

Yet there is a further possible avenue of advance through the circle method that seems not yet to have been exploited. This is to go beyond the Kloosterman refinement and to consider possible cancellations be-

tween contributions from integrals corresponding to different values of k Serious arithmetical and analytic difficulties, not yet normally capable of resolution, lie athwart this path. But the author has been successful in directing this idea to the theory of indefinite and definite ternary quadratic forms, in which the cardinality of the representations of numbers is too small for a Kloosterman refinement to be adequate. Interesting though this development may be from a methodological angle, it enables no real progress to be made since the theory of ternary forms has already been successfully treated by other more appropriate methods.

We turn now to some recent progress in additive number theory that has been made by alternative methods. First, we mention the mixed problem of representing numbers as the sum of squares and non-negative cubes, the history of which goes back to Hardy and Littlewood. Although it is conjectured that all large numbers are both the sum of one square and two cubes and of two squares and one cube, the best that was known through the circle method about two squares in this context was that they and four cubes sufficed to represent all large numbers. Not so long ago, however, Linnik [15] proved by his ergodic method that, if $r(n)$ is the number of representations of n as the sum of two squares and three non-negative cubes, then

$$r(n) > n^{2/3-\varepsilon}$$

for $n > n_0$, thus shewing that all large numbers are expressible in the proposed manner. But this work neither supplied an asymptotic formula for $r(n)$ nor even shewed that $r(n)$ was of the expected order of magnitude. We therefore propose to sketch briefly how we proved the asymptotic formula

$$r(n) \sim \frac{1}{27}\pi\Gamma^3\left(\frac{1}{3}\right)\mathfrak{S}(n)n,$$

where $\mathfrak{S}(n)$ is the singular series, thus demonstrating that the theory for two squares and three cubes conforms to the traditional pattern of results in Waring's problem [8]. Note that this is only the third genuine example of an asymptotic formula in Waring's problem where the cardinality of representations of n does not essentially exceed n in order of magnitude (the other two are the explicit formulae for three squares and for four squares), a fact that is related to our avoidance of the circle method.

The source of the method is that, if $r(\mu)$ denotes the number of ways of expressing μ as the sum of two squares, then

$$r(\mu) = 4 \sum_{l|\mu} \chi(l) \qquad (\mu \neq 0) \tag{6}$$

and

$$v(n) = \sum_{X^3+Y^3+Z^3 \leqslant n} r(n-X^3-Y^3-Z^3). \tag{7}$$

Substituting (6) in (7), we find that $v(n)$ is expressible as a combination of sums such as

$$\sum_{\substack{X^3+Y^3+Z^3 \leqslant n \\ X^3+Y^3+Z^3 \equiv n, \bmod k}} 1,$$

where $k \leqslant n^{1/2}$. The latter sums in turn can be evaluated by complicated transformations in terms of the exponential sums

$$\sum_{\substack{X^3+Y^3+Z^3 \equiv n, \bmod k \\ 0<X,Y,Z \leqslant k}} e^{2\pi i(aX+bY+cZ)/k},$$

to which our estimates through Deligne's theory are applicable. The formulae thus obtained almost, but not quite, suffice, a very complicated argument involving a deep theory of elliptic curves over finite fields being needed to complete the proof.

A somewhat surprising lacuna in the theory of these mixed problems has been the absence of known asymptotic formulae for the representations of numbers as the sum of three squares and a non-negative k-th power when k is greater than 2. Notwithstanding the existence of an exact formula for the number of ways of expressing a number as the sum of three squares, this question turns out to be unexpectedly difficult for the larger values of k, and it is only now that the asymptotic formulae have been derived by the author by exploiting relatively recent developments in the theory of the Dirichlet's L-functions [11].

We next consider the classical Diophantine equation

$$X^h+Y^h = Z^h+W^h \qquad (h>2), \tag{8}$$

which was studied, in particular, by Fermat and Euler. Although these scholars obtained rational parametric solutions when h is 3 or 4, it has been conjectured that the equation has no non-trivial solutions whenever

$h \geqslant 5$. This speculation being obviously extraordinarily difficult to treat in view of its connection with Fermat's Last Theorem, it is of interest to ponder some associated questions involving the expression of a number as the sum of two h-th powers whose resolution would provide some guidance about the matter. Let $r_h(n)$ be the number of ways of expressing n as the sum of two h-th powers (positive or negative, order being relevant), let $N_h(x)$ be the number of positive integers n not exceeding x for which $r_h(n) > 0$, and let $\nu_h(x)$ be the number of those integers for which $r_h(n) > 2$, noting that $\nu_h(x) = 0$ for $h \geqslant 5$ if the conjecture is true. Then we have been able to shew that

$$N_h(x) \sim A(h)x^{2/h} \tag{9}$$

and

$$\nu_h(x) = O(x^{5/(3h-1)+\varepsilon}),$$

thus demonstrating that it is certainly exceptional for a number expressible in the given form to be thus expressible in more than essentially one way. This goes some way in the required direction and is actually true for all $h \geqslant 3$, although so far we have only supplied the full details for the case where h is odd ([6], [7], in which are supplied references to relevant earlier writings by Erdös, Mahler, Greaves, and the author). This work also furnishes an analytic theory of the representation of numbers as the sum of two h-th powers for $h > 2$, a theory that is seen to contrast markedly with the classical theory for the case $h = 2$.

Considerations relating to the density of representations, to which we have previously alluded, preclude the application of the circle method to the additive equation (8), which indeed is even beyond the theoretical powers of that method whenever $h > 3$. Yet we cannot tarry long enough to describe our method in detail on account of its complication. It must therefore suffice to indicate briefly the ideas involved by referring to the case $h = 3$, in which (8) takes the form

$$r(r^2+3s^2) = \varrho(\varrho^2+3\sigma^2) \tag{10}$$

after a simple transformation. Since only solutions with $X+Y \neq Z+W$ serve to give a bound for $\nu_3(x)$, we are led to study (10) subject to $r < \varrho$ and other appropriate conditions. Now (10) is contained in the equation

$$r(r^2+3s^2) = \varrho l, \tag{11}$$

which, being of the form

$$r(r^2+3s^2) \equiv 0 \bmod \varrho,$$

can be studied with great accuracy by the theory of exponential sums in a manner akin to that used in the two squares and three cubes problem. The cardinality of solutions of (11) is too large, however, and it is necessary to take into account the special nature of the number l by means of a sieve method that exploits the idea that, for any prime p, a square σ^2 is *not* a quadratic non-residue, mod$p$. The calculations involved in this refinement are somewhat complicated and involve our estimates for multiple exponential sums of type (5) for $r = 3$.

There is an application of these ideas to the study of the number $\varrho(n)$ of representations of n as the sum of four non-negative cubes. It being at present impossible to find an asymptotic formula or even a positive lower bound for $\varrho(n)$, it is not without interest to elicit as keen an upper bound as possible for $\varrho(n)$. Here our method gives [5]

$$\varrho(n) = O(n^{11/18+\varepsilon}),$$

which represents an improvement on the trivial bound $O(n^{2/3+\varepsilon})$.

It had been guessed by Davenport and others that there is a positive density of numbers expressible as the sum of two cubes of rational numbers, and this was proved by Stephens [17] on the assumption that the Birch–Swinnerton-Dyer conjectures for certain elliptic curves are true. At the level of unconditional results, if $M(x)$ is the number of positive integers up to x that are the sum of two rational cubes, then our result (9) gives

$$M(x) \geqslant N_3(x) > A_1 x^{2/3}$$

with an explicit value for A_1. But our method can be adapted to take meaningful account of the changed circumstances with the consequence that we can shew that [12]

$$M(x) > A_2 x^{2/3} \log x.$$

The calculations involved also shed other light on the conjecture and suggest that it can only be true if the elliptic equation

$$X^3+Y^3 = nZ^3 \tag{12}$$

frequently has a smallest solution in which Z is almost exponentially large in terms of n.

Another interesting question in this field is whether a polynomial $f(x)$ equalling a sum of two integral h-th powers for every integer x is identically of the form

$$\{f_1(x)\}^h + \{f_2(x)\}^h. \tag{13}$$

Our method cannot so far resolve this matter but can at least shew that such polynomials $f(x)$ have certain properties that are consistent with their having the proposed form (13). We should also observe that Schinzel [16] has actually shewn that the answer is in the affirmative provided that certain far-reaching generalizations of the prime-twins conjecture are true.

Our thesis has tended to shew up certain shortcomings in the powerful circle method in spite of the suggestions we have made concerning its improvement. Apart from theoretical limitations, these deficiencies fall into a number of categories. For example, the method is in some respects not very flexible in adapting itself to the peculiar circumstances of individual problems, a penalty no doubt of the wide ranging scope of the machinery. Moreover, for the deeper problems the analysis becomes very complicated, a situation that is brought about in part by the need to consider exponential sums at arguments other than the arithmetically natural values h/k. In view of these facts and our present inability to make further substantial progress with Waring's problem through the circle method, it seems worthwhile to devise an alternative method of some generality that might incorporate some of the features of the special methods already mentioned. We therefore go on to describe a procedure developed by the author [9] that is applicable in principle to Waring's problem for any exponent and that has already been successful in isolating some new results. In some respects it has a potential for further refinement that is denied the circle method, although we have not succeeded in using it to resolve any of the deeper unsettled questions. Furthermore, the method has nowhere the same universality as that of the circle method.

We hint at the method by considering its relevance to problems involving the equation

$$l^2 + \varphi(l_1, \ldots, l_r) = n, \tag{14}$$

in which there is always a square present and in which $\varphi(l_1, \ldots, l_r)$ is a sum of powers. The underlying idea, implemented in practice with rather more refinement than our remarks here might suggest, is to split up $\varphi(l_1, \ldots, l_r)$ into two sums of powers $\varphi_1(l_1, \ldots, l_s)$, $\varphi_2(l_{s+1}, \ldots, l_r)$ in a suitable

way and to prove that the expected asymptotic expression $L(m)$ for the number $N(m)$ of solutions of

$$l^2+\varphi_1(l_1, \ldots, l_s) = m$$

is in fact always valid save possibly for a small exceptional set of m. If this can be achieved, then one can estimate the number of solutions of (14) by considering

$$\sum N\{n-\varphi_2(l_{s+1}, \ldots, l_r)\}$$

provided that $r-s$ is not too small.

The connection between $N(m)$ and $L(m)$ is treated by attempting to shew that the variance

$$\sum_{m\leqslant x} \{N(m)-L(m)\}^2$$

is small, to which end we require a good asymptotic formula for

$$\sum_{m\leqslant x} N^2(m).$$

Now the latter sum is obviously equal to the number of solutions in certain integers of the equation

$$l^2-\lambda^2 = \varphi_1(\lambda_1, \ldots, \lambda_s)-\varphi_1(l_1, \ldots, l_s) = \psi(\lambda_1, \ldots, \lambda_s, l_1, \ldots, l_s),$$

say, and hence of

$$\varrho\sigma = \psi(\lambda_1, \ldots, \lambda_s, l_1, \ldots, l_s),$$

where in particular ϱ, σ are of the same parity. For given ϱ, this gives rise to the condition

$$\psi(\lambda_1, \ldots, \lambda_s, l_1, \ldots, l_s) \equiv 0 \bmod \varrho,$$

which can be treated by means of the exponential sums

$$\sum_{\lambda_1,\ldots,\lambda_s,l_1,\ldots,l_s} e^{2\pi i\psi(\lambda_1,\ldots,\lambda_s,l_1,\ldots,l_s)h/k} \qquad (k|\varrho)$$

by a variation of earlier methods described. The analysis is then completed by using, inter alia, many of the properties of these sums that were previously developed in connection with the circle method, it being notable that we now need only work with trigonometrical sums corresponding to rational arguments.

We notice that our procedure consists partly of reducing our problem to another one in which one of the unknowns occurs linearly. If the lowest exponent occurring in the given problem is greater than two, then several transformations are needed to secure a linear problem and the details can become very formidable. However, a simple proof of the asymptotic formula in the nine cubes problem can be derived in this manner.

When examined systematically, our method is seen to have many links with the circle method in spite of the different genesis, the occurrence of similar exponential sums being a case in point. But the exponential sums in our method are shorn of arithmetically irrelevant analytic complications, thus lightening the potential task of effecting Kloosterman type refinements when these might be relevant or possible.

Our mention of the sizes of the solutions of the Diophantine equation (12) gives us an opening to introduce our final topic. This is the Pellian equation

$$T^2 - DU^2 = 1,$$

whose fundamental solution $\eta_D = T + \sqrt{D}U$ is known to satisfy

$$2\sqrt{D} < \eta_D < e^{A\sqrt{D}\log D}$$

for positive (non-square) determinants D. Since these inequalities have more or less represented the full extent of our knowledge, the author [13] has evolved a lattice point method that determines the distribution of the determinants D for which η_D is limited by small functions of D. Although the results obtained can only be rigorously substantiated for the smaller limits, the author in fact believes they are true for much larger limits. If we were right in this opinion, and in our reasons for holding it, then some interesting facts concerning the class number $h(D)$ of properly primitive indefinite binary quadratic forms

$$ax^2 + 2bxy + cy^2$$

of determinant $D = b^2 - ac$ would emerge. For example, we could obtain the asymptotic formula

$$\sum_{D \leqslant x} h(D) \sim (25/12\pi^2) x \log^2 x,$$

which would settle a matter that has been open since it was first raised by Gauss in the Disquisitiones Arithmeticae (V, Art. 304). As it is, we

can obtain unconditional lower bounds for the above sum that advance our knowledge. Moreover, we are led to conjecture that

$$\sum_{\substack{p \leqslant x \\ p \equiv 1, \mathrm{mod} 4}} h(p) \sim \frac{x}{6}$$

and that, if $\tau(\beta, x)$ is the number of determinants $p \equiv 1 \bmod 4$ for which $h(p) > \beta$, then

$$\lim_{x \to \infty} \frac{\tau(\beta, x)}{x/\log x} \sim \frac{1}{3\beta} \tag{15}$$

as $\beta \to \infty$. Impressive corroboration of these ideas comes from the present work of Henri Cohen, who simultaneously has been led by entirely different considerations of a more algebraic nature to enunciate conjectures about the behaviour of $h(p)$. His work, which is strongly supported by numerical evidence, agrees with ours in all areas where the subjects of investigation coincide (in particular, equation (15)), although it should be stressed that he and the author by no means study the same questions overall. Conditional work on similar matters has also been described in a recent paper by Takhtajan and Vinogradov [18]. As with the earlier matters discussed, this topic shews there is much life left in many of the important questions in number theory that were first raised centuries ago.

References

[1] Davenport H., Cubic Forms in Thirty Two Variables, *Proc. Trans. Roy. Soc London Ser. A* **251** (1959), pp. 193–232.
[2] Davenport H., Cubic Forms in 29 Variables:, *ibid.* **266** (1962), pp. 287–298.
[3] Davenport H., Cubic Forms in Sixteen Variables, *ibid.* **272** (1963), pp. 258–303.
[4] Heath-Brown D. R., Cubic Forms in 10 Variables, to appear.
[5] Hooley C., On the Representations of a Number as the Sum of Four Cubes, *Proc. London Math. Soc.* **36** (3) (1978), pp. 117–140.
[6] Hooley C., On the Numbers that Are Expressible as the Sum of Two Cubes, *J. Reine Angew. Math.* **314** (1980), pp. 146–173.
[7] Hooley C., On Another Sieve Method and the Numbers that Are a Sum of Two h-th Powers, *Proc. London Math. Soc.* **43** (3) (1981), pp. 73–109.
[8] Hooley C., On Waring's Problem for two Squares and Three Cubes, *J. Reine Angew. Math.* **328** (1981), pp. 161–207.
[9] Hooley C., On a New Approach to Various Problems of Waring's Type, *Recent Progess in Analytic Number Theory*, Vol. 1, Academic Press, 1981.
[10] Hooley C., On Exponential Sums and Certain of Their Applications, Journées Arithmétiques 1980, *London Math. Soc. Lecture Notes Series* **56**, Cambridge, 1982.
[11] Hooley C., *On Waring's Problem for Three Squares and an h-th Power*, to appear.

[12] Hooley C., *On the Numbers that Are a Sum of Two Rational Cubes*, to appear.
[13] Hooley C., *On the Pellian Equation and the Class Number of Indefinite Binary Quadratic Forms*, to appear.
[14] Katz N. M., Sommes exponentielles, *Astérisque* **79**, Société Mathématique de France (1980).
[15] Linnik Ju. V., Additive Problems Involving Squares, Cubes, and Almost Primes, *Acta Arith.* **21** (1972), pp. 413–422.
[16] Schinzel A., On the Relation between Two Conjectures on Polynomials, *Acta Arith.* **38** (1982), pp. 285–322.
[17] Stephens N. M., A Corollary to a Conjecture of Birch and Swinnerton-Dyer, *J. London Math. Soc.* **43** (1968), pp. 146–148.
[18] Takhtajan I. A., and Vinogradov A. I., The Gauss–Hasse Hypothesis on Real Quadratic Fields with Class Number One, *J. Reine Angew. Math.* **335** (1982), pp. 40-86.

DEPARTMENT OF PURE MATHEMATICS
UNIVERSITY COLLEGE
CARDIFF, GREAT BRITAIN