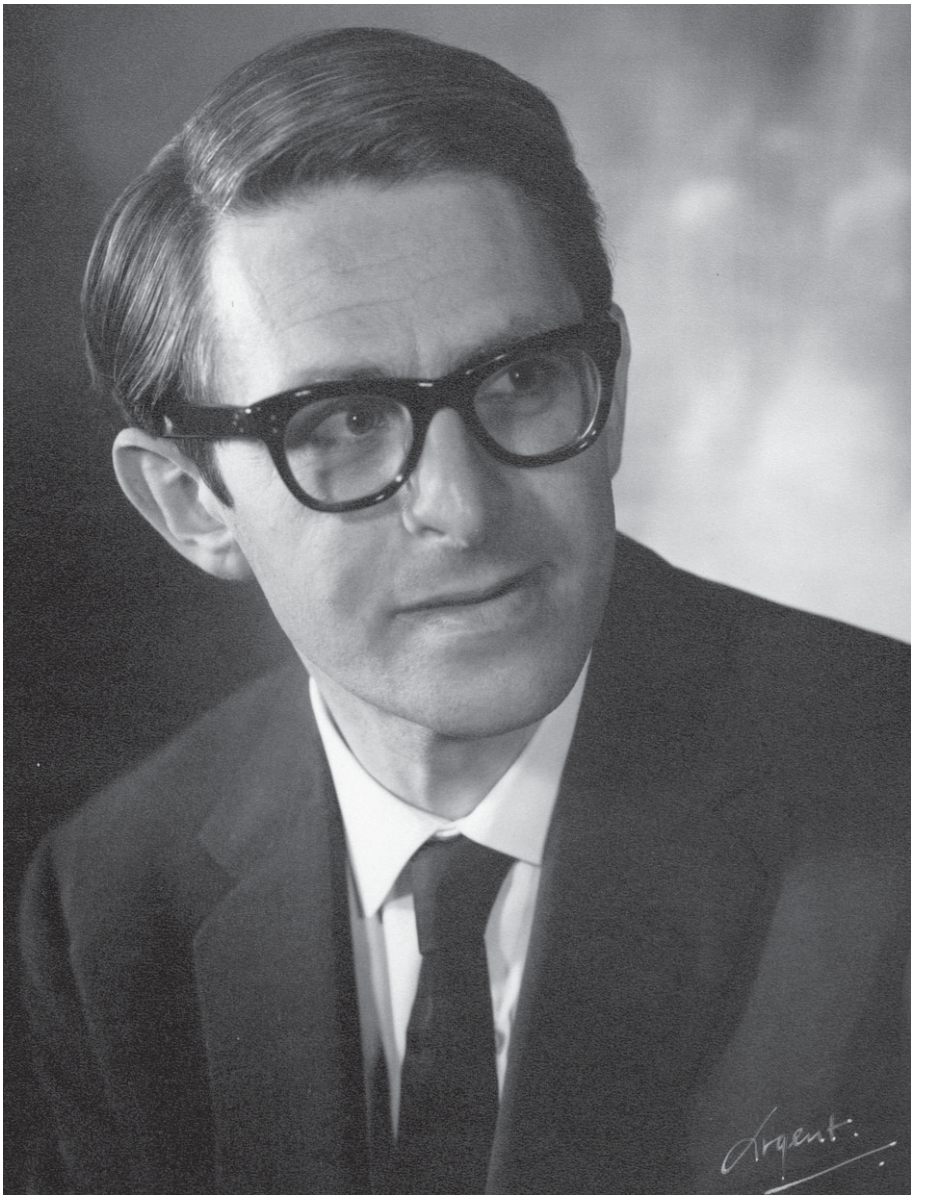
KLAUS FRIEDRICH ROTH

29 October 1925 — 10 November 2015



K. F. Roth.

KLAUS FRIEDRICH ROTH

29 October 1925 — 10 November 2015

Elected FRS 1960

By William Chen[1]* and Robert Vaughan FRS[2]**

[1]*Macquarie University, Sydney, New South Wales, Australia*
[2]*Pennsylvania State University, University Park, Pennsylvania, USA*

Klaus Friedrich Roth, who died in Inverness on 10 November 2015 aged 90, made fundamental contributions to different areas of number theory, including diophantine approximation, the large sieve, irregularities of distribution and what is nowadays known as arithmetic combinatorics. He was the first British winner of the Fields Medal, awarded in 1958 for his solution in 1955 of the famous Siegel conjecture concerning approximation of algebraic numbers by rationals. He was elected a Fellow of the Royal Society in 1960, and received its Sylvester Medal in 1991. He was also awarded the De Morgan Medal of the London Mathematical Society in 1983, and elected Fellow of University College London in 1979, Honorary Fellow of Peterhouse in 1989, Honorary Fellow of the Royal Society of Edinburgh in 1993 and Fellow of Imperial College London in 1999.

Life and career

Klaus Roth, son of Franz and Matilde (née Liebrecht), was born on 29 October 1925 in the German city of Breslau, in Lower Silesia, Prussia, now Wrocław in Poland. To escape from Nazism, he and his parents moved to England in 1933 and settled in London. He would recall that the flight from Berlin to London took eight hours and they landed in Croydon. Franz, a solicitor by training, had suffered from gas poisoning during the First World War and he died a few years after their arrival in England.

Roth studied at St Paul's School between 1937 and 1943, during which time the school was relocated to Easthampstead Park, near Crowthorne in Berkshire, as part of the wartime

* william.chen@mq.edu.au ** rcv4@psu.edu

http://dx.doi.org/10.1098/rsbm.2017.0014

evacuation of London. There, he excelled in mathematics and chess, and one master, Mr Dowswell, observed interestingly that he possessed complete intellectual honesty. As extracurricular activity, Roth was deeply interested in the Air Training Corps, but his efforts to be a member were thwarted for a long time because of his German nationality, until special permission was finally given towards the end of his time at St Paul's. His badly coordinated muscular movements ensured that his wish to become a pilot was never going to be realized.

Roth proceeded to read mathematics at the University of Cambridge, and became a student at Peterhouse. He also played first board for the university chess team. However, he had many unhappy and painful memories of his two years in Cambridge as an undergraduate. Uncontrollable nerves would seriously hamper his examination results, and he graduated with third class honours.

After this not too distinguished start to his academic career, Roth then did his war time service as an alien and became a junior master at Gordonstoun, where he divided his spare time between roaming the Scottish countryside on a powerful motorcycle and playing chess with Robert Combe. On the first day of the first British Chess Championships after the war, Roth famously went up to Hugh Alexander, the reigning champion, to tell him that he would not retain his title. This was of course right—the previously largely unknown Robert Combe became the new British champion.

Peterhouse did not support Roth's return to Cambridge after his war service, and his tutor, John Charles Burkill (FRS 1953), had suggested instead that he pursued 'some commercial job with a statistical bias'. Fortunately, his real ability and potential, particularly his problem solving skills, had not escaped the eyes of Harold Davenport (FRS 1940), who subsequently arranged for him to pursue mathematical research at University College London, funded by the, then, highest leaving exhibition ever awarded by his old school. Although Theodor Estermann was officially his thesis advisor, Roth was heavily influenced by Davenport during this period, and indeed into the mid 1960s. He completed his PhD work, which Estermann considered good enough for a DSc, and also joined the staff of the Department of Mathematics.

Davenport's influence clearly cultivated Roth's interest in diophantine approximation. Significant work had already been done by Dirichlet (FMemRS 1855), Liouville (FMemRS 1850), Thue, Siegel, Dyson (FRS 1952) and Gelfond. Indeed, a crucial exponent was believed to depend on the degree of the algebraic number under consideration, but Siegel had conjectured that it should be 2. In 1955, Roth showed precisely that. In a letter to Davenport, Siegel commented that this result 'will be remembered as long as mankind is interested in mathematics'. For this, Roth was awarded the Fields Medal in 1958. In speaking of Roth's work at the opening ceremony of the International Congress of Mathematicians in 1958, Davenport said, 'The achievement is one that speaks for itself: it closes a chapter, and a new chapter is opened. Roth's theorem settles a question which is both of a fundamental nature and of extreme difficulty. It will stand as a landmark in mathematics for as long as mathematics is cultivated.' He ended with the following words. 'The Duchess, in *Alice in Wonderland*, said that there is a moral in everything if only you can find it. It is not difficult to find the moral of Dr Roth's work. It is that the great unsolved problems may still yield to direct attack however difficult and forbidding they appear to be, and however much effort has already been spent on them.'

While most mathematicians consider Roth's result on diophantine approximation as his most famous, it is in fact another problem that gave him the greatest satisfaction. At about the same time, he became interested in the question of the impossibility of a just distribution

for any sequence in the unit interval, conjectured by van der Corput in 1935. Van Aardenne-Ehrenfest obtained the first quantitative estimate in 1949. By reformulating the problem in a geometric setting, Roth obtained in 1954 the best possible lower bound for the mean squares of the discrepancy function. This geometric setting paved the way for what is now known as geometric discrepancy theory, a subject at the crossroads of harmonic analysis, combinatorics, approximation theory, probability theory and even group theory. Once asked why he considered this his best work, Roth replied, 'But I started a subject!' He was particularly pleased that Burkill, with whom he had remained on good terms, offered the same opinion.

Further recognition came. Roth was elected Fellow of the Royal Society in 1960 and also promoted to a professorship at the University of London in 1961. He was very proud that the Fields Medal, the Fellowship of the Royal Society and the professorship came *in reverse order*.

The close relationship between Roth and Davenport in those days can be illustrated by a charming incident some time in the 1950s, and which Heini Halberstam recalled with great delight. Early one Sunday morning, Davenport went to his bathroom and switched on the light. The phone rang, and it was Roth. Could he possibly come over and explain the proof of a new result? Davenport suggested that Roth should come after breakfast, but as soon as he put the phone down, the door bell rang. Roth had been so eager that he had spent much of the early morning waiting in the telephone kiosk across the street.

It was also during this time at University College London that Roth met his wife, Melek Khaïry. Melek and her sister Hoda, daughters of Senator Khaïry Pacha in Egypt, had defied the wishes of their old-fashioned family, apart from their father, and come to study in London. It was the first ever university lecture given by Roth and the first ever university lecture attended by Melek. After the lecture, Roth had asked Halberstam whether he had noticed the young lady on the front row. 'I will marry her,' he claimed. By the end of that year, Roth had felt unsuitable to mark Melek's examination script, claiming that he felt 'unable to be impartial', much to the amusement of his colleagues. Another problem at the time was that during their courtship, Hoda often tagged along, much to Roth's annoyance. To counter that, Roth brought along his best friend, Laimons Ozolins, the Latvian born architect and a fellow pupil at St Paul's School, as a distraction for Hoda. Ozolins took to his assignment with great gusto, and indeed married Hoda.

In the mid 1960s, Roth had planned to emigrate to the United States to take up the offer of a position at the Massachusetts Institute of Technology, where he had spent a year a decade earlier. Imperial College and Walter Hayman (FRS 1956) intervened, and agreement was reached in the middle of a reception at the Soviet Embassy in London. Roth recalled that Sir Patrick Linstead (FRS 1940), then Rector of Imperial College, told him that he needed to make an application, but reassured him that there would be no other applicant. So Roth joined Imperial College in 1966 after a sabbatical at the MIT, and remained there until his retirement.

Following his retirement in 1987, Roth moved with Melek to Inverness. Melek's death in 2002 was a great setback, and Roth never recovered from this loss. In later years, he became increasingly disappointed at the services and facilities available to old people in Inverness, and subsequently left the bulk of his estate towards improving these.

Roth was an excellent lecturer. He explained his points so clearly that a good student could often just sit there and listen, and only had to record the details afterwards in the evening. However, he occasionally would have a bad day, and he warned his students at the beginning of the year that they would notice these very easily. One of us recalls that, on one occasion,

Roth wrote down a very complicated expression on the blackboard, then retired to the back of the room. A lot of thought was followed by an equal sign, and he retired to the back of the room again. After a long time he came once more up to the board and wrote down the same complicated expression on the right hand side. The audience held their collective breath at this profound assertion. But the best was yet to come. He then proceeded to write down $+O(1)$, at which point all burst into laughter. Roth looked at his masterpiece again, turned to the class and protested, 'But it is correct, isn't it?'

Outside mathematics, Roth enjoyed Latin American dancing, and would elegantly jive away the evening with Melek. They took this very seriously, to the point that they had a room in their house in Inverness specially fitted for dancing practice. For many years while they were in London, they had dancing lessons with Alan Fletcher, who, with wife Hazel, was five-time world Latin American dancing champion. Indeed, Roth dedicated one of his research papers to Fletcher. He explained that he had been bothered by a problem which he could not solve and was therefore not dancing very well. Fletcher had annoyed him so much by asking him week after week without fail whether he had solved his problem. So, to get Fletcher off his back, he just had to crack the problem, and when he did, he needed to acknowledge Fletcher for having provided the annoyance.

Roth maintained great modesty throughout his life. He felt very privileged to have been given the opportunity to pursue what he loved, and very lucky that he had some 'moderate success'. He was always very generous to his colleagues, and had inspired many to achieve good results.

Roth received the De Morgan Medal of the London Mathematical Society in 1983 and the Sylvester Medal of the Royal Society in 1991. He was also elected Fellow of University College London in 1979, Honorary Fellow of Peterhouse in 1989, Honorary Fellow of the Royal Society of Edinburgh in 1993 and Fellow of Imperial College London in 1999. He and Melek had no children.

A VERY BRIEF SUMMARY OF MATHEMATICAL WORK

Roth's work in the very early part of his career concerns the application of the Hardy–Littlewood method to study certain additive questions in number theory. His subsequent work can be described as a career-long fascination with, and repeated efforts at, understanding the limitations to the degree of regularity possible in various discrete systems, often making very clever use of artificial orthogonality or quasi-orthogonality, and punctuated by a small number of spectacular digressions, including his seminal contribution to diophantine approximation (14)* and to the large sieve (18).

In the mid 1950s, with encouragement and advice from Paul Erdős (FMemRS 1989), Roth and his close colleague, Heini Halberstam, began to write the influential volume *Sequences* (19). The effort took nearly 10 years, and Roth particularly enjoyed writing about Rényi's version of the large sieve; he worked on it even after the completion of the book, culminating in his own remarkable contribution to the subject. Halberstam recalled fondly that Roth carried the completed manuscript by hand all the way to the offices of the Clarendon Press in Oxford.

* Numbers in this form refer to the bibliography at the end of the text.

Questions on regularity occupy the bulk of Roth's writings, and these can be divided roughly into three areas: irregularities of integer sequences in arithmetic progressions, irregularities of point distribution and Heilbronn's triangle problem.

Notation. Throughout this article, we adopt the O, o as well as Vinogradov notation $\ll$ and $\gg$. Thus for any function f and any positive function g, we write $f = O(g)$ or $f \ll g$ to denote that there exists a positive constant c such that $|f| \leqslant cg$, and write $f = o(g)$ to denote that $|f/g| \to 0$. Furthermore, if f is also a positive function, then we write $f \gg g$ to denote $g \ll f$, and write $f \asymp g$ to indicate that $f \ll g$ and $f \gg g$ both hold. The symbols O, o, $\ll$ and $\gg$ may have subscripts if the constant c in question depends on the variables represented by those subscripts. The cardinality of a finite set $\mathscr{A}$ is denoted by $|\mathscr{A}|$.

Early work

Squarefree and k-free numbers

A squarefree number is a positive integer with no repeated prime factors. They have some properties analogous to prime numbers, but are generally less demanding to understand and thus often provide a good testing ground for techniques.

Estermann (1931) had obtained an asymptotic formula for the number of representations of a large natural number as the sum of a square and a squarefree number. In his first research paper, Roth (1) extended this result to the situation in which the square was restricted to being the square of a squarefree number. While not a technically demanding problem, it nevertheless provided a good introduction to the methodology applied to questions in analytic number theory.

His second paper on squarefree numbers (4) is of considerable importance, and deals with gaps between squarefree numbers.

It is an elementary exercise to show that the number $Q(x)$ of squarefree numbers not exceeding x satisfies

$$Q(x) = \frac{6}{\pi^2}x + O(x^{1/2}).$$

Thus if q_n is the n-th squarefree number in order of magnitude, then it follows that

$$q_{n+1} - q_n = O(n^{1/2}).$$

Generally it is believed that

$$q_{n+1} - q_n = O(n^{\varepsilon})$$

for every fixed positive ε, and this would follow from the ABC conjecture; see Granville (1998).

Fogels (1941) had reduced the exponent $1/2$ to $2/5 + \varepsilon$. Estermann and Roth had found different methods to further reduce the exponent to $1/3$. Estermann's simple elementary argument was outlined by Roth in his paper (4), which also outlined a small further improvement shown to him by Davenport. However, Roth also made substantial further improvements. He first gave a simple proof that

$$q_{n+1} - q_n = O(n^{1/4}),$$

and then went on to combine this with a method of van der Corput (1928) to show that

$$q_{n+1} - q_n = O(n^{3/13}(\log n)^{4/13}).$$

There followed papers by Richert (1954), Rankin (1955), Schmidt (1964), Graham and Kolesnik (1988), Trifonov (1988, 1989) and Filaseta (1990), many of them quite technical.

The current best bound is due to Filaseta and Trifonov (1990), and their method leans heavily on the elementary method of Roth. They showed that

$$q_{n+1} - q_n = O(n^{1/5} \log n),$$

and said that their core lemma, namely Lemma 1, was 'essentially contained in Roth's paper (4)', and the final section of the paper was entitled 'The use of Roth's method'.

On the other hand, in collaboration with Halberstam, Roth developed his original ideas to treat k-free numbers, namely, those integers with no more than $k-1$ repeated prime factors, in their paper (5).

There is also a substantial literature on gaps between k-free numbers, and also on k-free values of polynomials, much of it stimulated by the ideas in (5). For an article with a good overview of the subject, see Filaseta (1993).

The Hardy–Littlewood method

The Hardy–Littlewood method is a technique in additive number theory, developed in the 1920s from a famous paper of Hardy (FRS 1910) and Ramanujan (FRS 1918) (Hardy & Ramanujan 1918) and by Hardy and Littlewood (FRS 1916) in a series of papers (Hardy & Littlewood 1920a, 1920b, 1921, 1923, 1922, 1924, 1925, 1928).

There had been many important developments by Davenport and I.M. Vinogradov (FMemRS 1942), and Estermann was also considered a leading expert on the method. Thus it is not surprising that, with both Davenport and Estermann as mentors at University College London, several chapters in Roth's PhD thesis should involve applications of the method.

Roth had been given by Davenport the problem of showing that almost every natural number n could be expressed as the sum of a square, a cube, a fourth power and a fifth power, in the sense that the number of exceptional $n \leqslant N$ is $o(N)$. He met Estermann one day and announced that actually he did not need the fifth power. This result in (2) was quite sensational. At the time there had been relatively little done on ternary additive problems. Of course, there was the classical theorem of Gauss–Legendre that every natural number not of the form $4^j(8k+7)$ could be expressed as the sum of three squares. There were also two papers by Davenport and Heilbronn (FRS 1951) showing that almost every natural number could be expressed as the sum of two squares and a k-th power with k odd (Davenport & Heilbronn 1938b), and that almost every natural number could be expressed as the sum of a square and two cubes (Davenport & Heilbronn 1938a). However, these were the extent of the known results. Thus Roth's theorem pushed the envelope of what was known.

Estermann, however, insisted that the result with the four terms should be written up as well, on the grounds that Davenport must have had a reason for wanting the fifth power! That the reason was to make the question amenable to a beginning postgraduate student seemed to have escaped Estermann. Anyway, a variation of the problem was eventually found and, in (7),

Roth showed that every sufficiently large natural number could be expressed in the form

$$\sum_{j=1}^{s} x_j^{j+1}$$

with $s = 50$. This question was worked on subsequently by several researchers. Thanigasalam, Vaughan (FRS 1990) and Brüdern successively reduced the size of s, and the best current result is due to Ford (1996) with $s = 14$.

The earlier paper (2) also stimulated quite a lot of later work. Suppose that $2 \leqslant k_1 \leqslant k_2 \leqslant k_3$. Then one can ask about the solubility for large n of the diophantine equation

$$n = x_1^{k_1} + x_2^{k_2} + x_3^{k_3}, \qquad [1]$$

with the integral variables x_1, x_2, x_3 all positive. For this to hold for almost all n, one needs

$$\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3} > 1.$$

In this case, it is readily seen that there are only four possible configurations of exponents, namely

$$k_1 = k_2 = 2, \ k_3 \text{ arbitrary}, \qquad [2]$$

$$k_1 = 2, \ k_2 = 3, \ k_3 = 3, \qquad [3]$$

$$k_1 = 2, \ k_2 = 3, \ k_3 = 4, \qquad [4]$$

$$k_1 = 2, \ k_2 = 3, \ k_3 = 5. \qquad [5]$$

A further necessary condition for solubility of these equations is that they be soluble modulo q for every modulus q, and this is essentially the same as requiring that there be a non-singular solution in each p-adic field. In particular, since there are a positive proportion of n which cannot be represented as the sum of three squares, we may suppose in the case [2] that k_3 is odd.

We now know that in many of these configurations there are infinitely many natural numbers n for which local solubility is not sufficient. The first such examples are due to Jagy and Kaplansky (1995); see also Vaughan (1997: Chapter 8, Exercise 5). More recently, others have been added to the list by Dietmann and Elsholtz (2008, 2016) and by Gundlach (2013). Apparently, in each case, the counter-example can be interpreted as a Brauer–Manin obstruction. However, these failures of the local to global principle only occur for a thin set of natural numbers n, and this enhances the interest of results which show solubility for almost all natural numbers n, and in particular in estimates for the size of any exceptional set.

It has been possible to show in each case [2]–[5] that almost all natural numbers n can be represented in the form [1]. As mentioned above, the first results of this kind are due to Davenport and Heilbronn, and Roth added to this with his work on a square, a cube and a fourth power. The picture was completed by Vaughan (1980), who established that almost all natural numbers could be expressed as the sum of a square, a cube and a fifth power.

Of enhanced interest, following the work of Jagy and Kaplansky and of others, is the size of the exceptional set. Davenport and Heilbronn (1938b) gave the bound

$$\ll X(\log X)^{-3/5+\varepsilon}$$

for the number of exceptional natural numbers not exceeding X and not representable as the sum of two squares and a cube. They (Davenport & Heilbronn 1938a) also gave the bound

$$\ll X(\log X)^{-1/11}$$

for the number of natural numbers not exceeding X and not representable as the sum of a square and two cubes. In the case of a square, a cube and a fourth power, Roth (2) obtained the bound

$$\ll X(\log X)^{-1/20}$$

for the number of exceptions.

The first improvement on results of this kind was given by Vaughan (1980) who showed the existence of a positive number δ such that the number of exceptional natural numbers n not exceeding X and not representable in the form [1] for the case [5] could not exceed

$$\ll X^{1-\delta}.$$

The crucial new ideas stemmed from the large sieve (see section *The Large Sieve*) and could also be applied to the other cases considered here.

There is a considerable body of work on adapting these methods to situations in which one or more of the variables are restricted in some way. See Vaughan (2015) for a review of this material.

Davenport had used the Hardy–Littlewood method to show (Davenport 1939) that every sufficiently large natural number could be expressed as the sum of eight cubes and, more significantly, that almost every natural number could be expressed as the sum of four cubes. Roth then showed in (6) that in each case all but one of the variables could be taken to be prime. This is somewhat routine, although there are some technical difficulties to be overcome. However this paper clearly led to an interest in Vinogradov's methods for estimating exponential sums and resulted in the translation into English by Roth and Anne Davenport of Vinogradov's monograph (Vinogradov 1954) on exponential sums. Roth added extensive notes to each of the chapters, and Vinogradov told him at the International Congress of Mathematicians in Edinburgh in 1958 that serious consideration should be given to translating the book back into Russian!

The work on cubes also attracted a large body of modern work, leading, for example, to Kawada's result (1997) that the non-prime variable could be replaced by a number having at most three prime factors. It would be of great interest if the non-prime could be replaced by a prime. There is a brief survey of this area in Vaughan (2013).

Roth also has a paper (15) with Davenport which uses the Hardy–Littlewood method. This concerns the following question. Suppose that $\lambda_1, \dots, \lambda_s$ are real numbers, not all of the same sign and not all in rational ratio. Given k, η and ε, how small can $s = s(k)$ be taken as a function of k so that the inequality

$$\left|\eta + \lambda_1 x_1^k + \cdots + \lambda_s x_s^k\right| < \varepsilon$$

has infinitely many solutions in positive integers $x_1, \dots, x_s$?

When $\eta = 0$ and $k = 2$, Davenport and Heilbronn (1946) had shown that $s = 5$ was permissible and it was clear that their method would establish the desired conclusion for general k and η with $s = 2^k + 1$. In their paper, Davenport and Roth obtained a result with $s(k)$ such that $\limsup s(k)/(k \log k) = 6$. For joint work by two of the most powerful analytic

number theorists of the era, this is a surprisingly ordinary result. In dealing with the somewhat more onerous situation in which the variables were assumed to be prime, Vaughan (1974) was able to obtain $s(k)$ with $\limsup s(k)/(k\log k) = 4$. Later, applying the techniques developed in Waring's problem by Vaughan (1989) and Wooley (FRS 2007) (Wooley 1992), Li (1995) was able to obtain the desired conclusion with an $s(k)$ satisfying $\limsup s(k)/(k\log k) = 1$.

The method used here depends on rational approximations a/q to one of the irrational ratios, for example λ_r/λ_s, as given by the continued fraction expansion. In particular, the range for the variables $x_1, \dots, x_s$ depends on the size of q. However, suitably good rational approximations a/q can be very rare, so that the denominator q_n of the n-th convergent can grow extremely rapidly as a function of n. This can happen if, for example, the ratio is a Liouville number. Thus the method does not allow one to localize the solutions, in the sense that given a large parameter X, one cannot guarantee that there is a solution with, say, $\sqrt{X} < \max_j x_j \leqslant X$. There has been a considerable blossoming of work in this area since the problem of localization of solutions was overcome in a groundbreaking paper of Bentkus and Götze (1999). This was followed by papers by Freeman (2002) and Wooley (2003). A comprehensive review of this area is given in the paper of Brüdern, Kawada and Wooley (2009).

There is one other paper which can also be considered in the classical Hardy–Littlewood method *milieu*. The paper (13) with Szekeres on generalized partition functions has been largely overlooked. It restricts to the case when the summands in a partition are distinct but the method applies more generally. In particular, the method is easily adapted to give an asymptotic formula for the number of partitions of a large number into primes, a result which, until quite recently, with the appearance of Vaughan (2008), experts in the area had thought could not be obtained in the current state of knowledge.

It can be said that Roth's early work on the Hardy–Littlewood method is not his most important work. Yet it is unlikely that, without this introductory phase, Roth would have considered using a variant of the Hardy–Littlewood method to treat sets having no three terms in arithmetic progression, with all that which followed. See section *Distribution of Integer Sequences in Arithmetic Progression*.

Diophantine approximation

Questions of diophantine approximation have attracted the attention of the world's leading mathematicians for at least four centuries, and continue to do so. Roth's theorem on diophantine approximation settled a central question which had already been heavily worked over by leading researchers. To appreciate his achievement, it is necessary to give a short review of the earlier work.

Given a real irrational number α, how small can one make

$$\left|\alpha - \frac{a}{q}\right|,$$

where a is an integer, q is a positive integer and $(a, q) = 1$, as a function of q? The continued fraction algorithm shows that for infinitely many q, one has

$$\left|\alpha - \frac{a}{q}\right| \leqslant \frac{1}{q^2}.$$

When α is a quadratic irrational, it is readily seen that this cannot be much improved. For example, from the irrationality of $\sqrt{2}$, it follows that for any a and q, one has $|2q^2 - a^2| \geqslant 1$, and so there is a positive constant c such that

$$\left|\sqrt{2} - \frac{a}{q}\right| > \frac{1}{cq^2}.$$

A similar observation holds for any quadratic surd α. On the other hand, Liouville (1844) constructed irrational numbers, known nowadays as Liouville numbers, such as

$$\lambda = \sum_{n=1}^{\infty} 2^{-n!},$$

which have approximations of the form

$$\left|\lambda - \frac{a}{q}\right| < \frac{1}{q^{\kappa}}$$

with κ arbitrarily large. A number α is algebraic of degree d if it is the root of a polynomial of degree d with integer coefficients and d is the smallest possible degree of such a polynomial. When one supposes that the coefficients do not have a common factor and the leading coefficient is positive, then the polynomial is unique and is known as the minimal polynomial. If the leading coefficient is 1, then α is called an algebraic integer.

Liouville had also shown by a generalization of the argument above that for any α algebraic of degree $d \geqslant 2$, one could find a positive constant $C(\alpha)$ such that

$$\left|\alpha - \frac{a}{q}\right| > \frac{1}{C(\alpha)q^d}.$$

Thus *inter alia*, Liouville numbers are transcendental.

When $d > 2$, it was generally believed that equations such as $2a^d - q^d = 1$ and, more generally, equations of the form $q^d f(a/q) = c$, with $f(x) = c_d x^d + c_{d-1}x^{d-1} + \cdots + c_0$ irreducible over $\mathbb{Q}$ and integer coefficients c and $c_0, \ldots, c_d$, had only a finite number of solutions in integers a and q with $(a, q) = 1$. This was established by Thue (1909) who showed that for any α algebraic of degree d and any $\kappa > 1 + d/2$, one could find a positive constant $C(\alpha)$ such that

$$\left|\alpha - \frac{a}{q}\right| < \frac{1}{C(\alpha)q^{\kappa}} \qquad [6]$$

had only a finite number of solutions.

Suppose that a_1/q_1 satisfies [6] with q_1 sufficiently large and a_2/q_2 also satisfies [6] with $q_2 > q_1$. Then one can construct a family of polynomials $P_n(x_1, x_2)$ of two variables with integer coefficients which vanish to high order at the point (α, α) but only to low order at $(a_1/q_1, a_2/q_2)$. This is used to show that q_2 is excluded from an interval of the kind

$$I_n = \{x : c^n q_1^{\theta(n)} \leqslant x \leqslant c^{-n} q_1^{\phi(n)}\}$$

for suitable functions $\theta(n)$ and $\phi(n)$. Thue's proof is then completed by showing that, for large n, these intervals overlap and cover the line segment $[C, \infty)$. There is a nice account of this in Davenport (1968).

The permissible size of κ was refined by Siegel (1921) to

$$\kappa > \min\left\{\frac{d}{m} + m - 1 : m = 1, 2, \ldots, d\right\},$$

so that $\kappa > 2\sqrt{d}$, and independently by Dyson (1947) and Gelfond to $\kappa > \sqrt{2d}$. In overall structure the proofs follow Thue's. Siegel had also conjectured that for any $\kappa > 2$, the inequality [6] could have only a finite number of solutions.

It was generally understood that the way forward was to make use of polynomials $P(x_1, \ldots, x_k)$ in many variables. However, it was not until 1955 that Roth caused a sensation by proving Siegel's conjecture. The crucial part of the proof is the construction of such polynomials which, while vanishing to high order at $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)$, only vanish to low order at $\boldsymbol{\rho} = (a_1/q_1, \ldots, a_k/q_k)$. This is a considerable escalation in the difficulty of the proof. Consider a general polynomial

$$P(\mathbf{x}) = \sum_{r_1=0}^{e_1} \ldots \sum_{r_k=0}^{e_k} b(\mathbf{r}) x_1^{r_1} \ldots x_k^{r_k}, \qquad [7]$$

with integer coefficients $b(\mathbf{r})$, the generalized derivative

$$D^{\mathbf{i}} = \frac{1}{i_1! \ldots i_k!} \frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_k}}{\partial x_k^{i_k}},$$

and the associated index

$$I(\boldsymbol{\alpha}, \mathbf{e}) = \min\left\{\sum_{j=1}^{k} \frac{i_j}{e_j} : P_{\mathbf{i}}(\boldsymbol{\alpha}) = 0\right\},$$

where $\mathbf{r} = (r_1, \ldots, r_k)$, $\mathbf{i} = (i_1, \ldots, i_k)$ and $\mathbf{e} = (e_1, \ldots, e_k)$. The positive integers $e_1, \ldots, e_k$ are at our disposal but are typically chosen to make $q_1^{s_1}, \ldots, q_k^{s_k}$ roughly equal. Also, let the height of P be defined by

$$H(P) = \max\{|b(\mathbf{r})| : \mathbf{r}\}.$$

It suffices to suppose that α is an algebraic integer of degree $d \geqslant 2$. The first step is the index theorem. This states that if α is an algebraic integer, $k \geqslant d/2\varepsilon^2$ and $e_1, \ldots, e_k$ are positive integers, then there is a polynomial of the form [7], not identically zero, such that P has degree at most e_j in the variable x_j,

$$I(\boldsymbol{\alpha}, \mathbf{e}) \geqslant k(1 - \varepsilon)/2 \quad \text{and} \quad H(P) \leqslant C(\alpha)^{e_1 + \ldots + e_k}.$$

The core of the argument is Roth's lemma. This shows that if P is as in [7] and satisfies $H(P) \leqslant q_1^{e_1\theta}$, where $\theta < 1$ depends at most on k and ε, then $I(\boldsymbol{\rho}, \mathbf{e}) \leqslant \varepsilon$. The proof is based on the properties of generalized Wronskians.

Given a non-negative integer i, choose $\mathbf{i}$ so that $i_1 + \ldots i_{k-1} \leqslant i$ and put

$$\triangle_{ij} = \triangle_{ij}(\mathbf{x}) = D^{\mathbf{i}} \frac{\partial^{j-1}}{(j-1)! \partial x_k^{j-1}} P(\mathbf{x}).$$

When i is large, there are many possible choices of $\mathbf{i}$. Hence, when ℓ is large, there are many possible generalized Wronskians

$$W(\mathbf{x}) = (\det \triangle_{ij})_{1 \leqslant i,j \leqslant \ell}.$$

Thus it is not completely surprising that it can be shown that at least one can be factored into two polynomials, each with fewer independent variables. Moreover, Wronskians preserve the dichotomy of having a high order zero at $\boldsymbol{\alpha}$ and a low order zero at $\boldsymbol{\rho}$. This enables an induction to be performed on the number of variables. This realization must have been the Eureka moment in Roth's discovery.

The proof is completed by using Roth's lemma to obtain a contradiction against the index theorem.

The method was extended by Baker (FRS 1973) (Baker 1964) to show, for example, that the transcendental (Champernowne) number $0.1234567890111213\ldots$ is not a U-number in the sense of Mahler.

Although Roth's theorem on diophantine approximation is essentially best possible, this is not the end of the story, only perhaps the end of the beginning. The underlying ideas have been highly influential and have had far reaching consequences. Most notably Schmidt generalized the result to establish his theorem on simultaneous diophantine approximation (Schmidt 1970a) and his subspace theorem (Schmidt 1989). The latter states that if $L_1, \ldots, L_n$ are linear forms in n variables with algebraic coefficients which are linearly independent and if ε is any positive real number, then the non-zero integer points $\mathbf{x}$ with

$$|L_1(\mathbf{x}) \ldots L_n(\mathbf{x})| < |\mathbf{x}|^{-\varepsilon}$$

lie in a finite number of proper subspaces of $\mathbb{Q}^n$.

While Roth's theorem gives no bound on the size of possible q occurring in the inequality, Davenport and Roth (16) were able to obtain a bound for the number of pairs a, q which satisfy an inequality of the type

$$\left|\alpha - \frac{a}{q}\right| < \frac{1}{2q^{\kappa}}$$

when $\kappa > 2$.

The large sieve

The large sieve was introduced by Linnik (1941) with the aim of considering questions in which, at least on average, one could sieve out a large number of residue classes modulo each prime p.

Let $\mathscr{A}$ be a subset of Z integers in the interval $[1, N]$. For every $n = 1, \ldots, N$, let $c_n = 1$ if $n \in \mathscr{A}$ and $c_n = 0$ if $n \notin \mathscr{A}$. Then

$$Z = \sum_{n=1}^{N} c_n.$$

Given a positive integer q, let

$$Z(q, a) = \sum_{\substack{n=1 \\ n \equiv a \bmod q}}^{N} c_n$$

denote the number of members of $\mathscr{A}$ in the residue class a modulo q. An easy calculation shows that the variance

$$V(q) = \sum_{a=1}^{q} \left(Z(q,a) - \frac{Z}{q} \right)^2$$

satisfies

$$qV(q) = \sum_{a=1}^{q-1} \left| S\left(\frac{a}{q}\right) \right|^2,$$

where

$$S(\alpha) = \sum_{n=1}^{N} c_n e(\alpha n).$$

Here $e(z) = \mathrm{e}^{2\pi \mathrm{i} z}$ for every $z \in \mathbb{R}$.

Linnik related the quantity $|S(a/q)|^2$ to the integral

$$(2\delta)^{-1} \int_{-\delta}^{\delta} \left| S\left(\frac{a}{q} + \beta\right) \right|^2 \mathrm{d}\beta$$

for a suitably small positive δ. Then, as long as the various intervals $(a/q - \delta, a/q + \delta)$ do not overlap, Parseval's identity

$$\int_0^1 |S(\alpha)|^2 \, \mathrm{d}\alpha = \sum_{n=1}^{N} |c_n|^2,$$

which in this instance is equal to Z, can be used to obtain an upper bound. In this way, Linnik obtained a non-trivial upper bound for the sum

$$\sum_{p \leqslant Q} pV(p)$$

in terms of N, Q and Z. If $Z(p,a) = 0$ for many values of a, in other words, if one is sieving out a large number of residue classes modulo p for each prime p, then this enables one to show that Z is strikingly small. Thus Linnik (1942) was able to show that, for any fixed $\delta > 0$, writing $n(p)$ for the least quadratic non-residue modulo p, the number P of primes not exceeding X for which $n(p) > p^{\delta}$ would satisfy $P \ll \log\log X$.

This work was then developed by Rényi in a long series of papers (Rényi 1947, 1948, 1949a, 1949b, 1949c, 1949d, 1950, 1958a, 1958b, 1959); see also Rényi (1962). Perhaps the most important application made by Rényi was as an aid to *small* sieves which led to theorems of the kind that every sufficiently large even number could be expressed as the sum of a prime and a number having a bounded number of prime factors.

Rényi seemed to have been seduced by the probabilistic nature of the sum $V(q)$, but this did not really lead to any further profound advances. The state of the art before the seminal papers of Roth (18) and Bombieri (1965) was summarized by Barban (1961, 1966). At this point, the methods were quite effective when $Q \leqslant N^{1/3}$, but less effective for larger values of Q. However, by the time the survey article by Barban (1966) appeared, it was obsolescent.

Roth (18) was the first to realize that harmonic analysis was the key and that one could fruitfully use Fourier (FMemRS 1823) analysis on an interval to obtain bounds close to best

possible. He was able to obtain a bound which in particular gave

$$\sum_{p\leqslant Q} pV(p) \ll (N + Q^2 \log Q) \sum_{n=1}^{N} |c_n|^2.$$

That this should remain effective even with Q close to $N^{1/2}$ was a huge breakthrough. This, together with the almost contemporaneous paper of Bombieri (1965), led to a huge amount of activity through the 1970s, and continues to have a profound impact on analytic number theory today.

In order to describe the modern developments, one should view c_n as an arbitrary sequence of complex numbers with support on $[1, N]$. Also, to enable a comparison with Roth's result, one should observe that

$$\sum_{p\leqslant Q} pV(p) = \sum_{p\leqslant Q} \sum_{a=1}^{p-1} \left| S\left(\frac{a}{p}\right)\right|^2 \leqslant \sum_{q\leqslant Q} \sum_{\substack{a=1\\(a,q)=1}}^{q} \left| S\left(\frac{a}{q}\right)\right|^2.$$

Thus a non-trivial bound of the type

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \leqslant \Delta(N, \delta) \sum_{n=1}^{N} |c_n|^2,$$

where the points $\alpha_1, \ldots, \alpha_R$ satisfy

$$\min_{r\neq s} \min_{z\in\mathbb{Z}} |\alpha_r - \alpha_s - z| \geqslant \delta,$$

has become known as the large sieve inequality. Bombieri and Davenport (1968) showed that in general $\Delta(N, \delta)$ could not be taken smaller than

$$\Delta(N, \delta) = N - 1 + \delta^{-1}, \qquad [8]$$

and Selberg (1991) showed that $\Delta(N, \delta)$ could indeed be given by [8]. Selberg's method represents the ultimate development of the Fourier analysis attack introduced by Roth. These methods were adapted by Vaaler (1985) and others for a variety of applications in analytic number theory and related areas. See, for example, Montgomery (1994).

Perhaps the most significant development stemming from this work is the following result of Bombieri (1965) and A.I. Vinogradov (1965), stated here in a slightly sharper form due to Bombieri. Suppose that A is any given positive real number. Then there are two positive real numbers $B = B(A)$ and $X_0 = X_0(A)$ such that for every $X > X_0$, we have

$$\sum_{q\leqslant X^{1/2}(\log X)^{-B}} \max_{\substack{1\leqslant a\leqslant q\\(a,q)=1}} \sup_{Y\leqslant X} \left|\pi(Y; q, a) - \frac{1}{\phi(q)} \int_2^Y \frac{\mathrm{d}t}{\log t}\right| \ll \frac{X}{(\log X)^A}.$$

Here $\pi(Y; q, a)$ denotes the number of primes not exceeding Y in the residue class a modulo q, and ϕ is Euler's function. This result can often be used in place of the as yet unestablished generalized Riemann hypothesis, perhaps the most important unsolved problem in mathematics, and the exponent $1/2$ here is crucial in many applications. Thus the recent work of Goldston, Pintz and Yıldırım (Goldston *et al.* 2009), Zhang (2014), Maynard (2015)

and Tao (FRS 2007), showing that there are bounded gaps in the primes, would not have been possible without many of the developments stemming from the large sieve.

MISCELLANY

Lattice coverings

Roth essayed only one paper in convexity and the geometry of numbers. This is especially surprising since this was a major interest of Davenport, and Roth was ideally equipped to work in the area. Rogers (FRS 1959) was already making great progress and undoubtedly Roth was only too happy to let him get on with it. However, a visit by Bambah stimulated his interest.

A lattice Λ in n-dimensional space is called a covering lattice for a symmetrical convex body K if every point of the space belongs to a body of the type $K + P$, where P is a point of the lattice. Let $V(K)$ denote the volume of K, and let $d(\Lambda)$ denote the discriminant of Λ. The lattice covering density $\vartheta(K)$ of K is defined by

$$\vartheta(K) = \inf_{\Lambda} \frac{V(K)}{d(\Lambda)},$$

where the infimum is taken over all covering lattices Λ for K.

Hlawka (1949) had shown that $\vartheta(K) \leqslant n^n$, and Rogers (1950) had lowered this bound to $\vartheta(K) \leqslant 3^n$. By use of the Brunn–Minkowski theorem, Bambah and Roth (9) established that

$$\vartheta(K) \leqslant \frac{\pi n^n}{n!\sqrt{27}}.$$

This is at most $\pi e^n/\sqrt{27}$, so is clearly stronger and asymptotically $e^n(\pi/(54n))^{1/2}$. This question then attracted a series of improvements by Rogers (1958) and Schmidt (1959), culminating in the bound

$$\vartheta(K) \leqslant \exp\left(\frac{\log n}{\log 2}\log\log n + c\log n\right)$$

by Rogers (1959), apparently still the best that is known.

Cosine polynomials

There are two intriguing questions that involve trigonometrical polynomials with integer frequencies.

Let $\mathcal{N}$ denote a set of N distinct natural numbers, and define

$$E(\alpha) = \sum_{n\in\mathcal{N}} e(\alpha n),$$

where $e(\beta) = e^{2\pi i\beta}$. Furthermore, put

$$C(\alpha) = \sum_{n\in\mathcal{N}} \cos(2\pi n\alpha),$$

so that $C(\alpha) = \frac{1}{2}E(\alpha) + \frac{1}{2}E(-\alpha)$, and write

$$A(\mathcal{N}) = \int_0^1 |E(\alpha)|\,\mathrm{d}\alpha \quad \text{and} \quad B(\mathcal{N}) = -\min_{\alpha\in[0,1]} C(\alpha).$$

Littlewood, in a list of problems circulated privately, had conjectured the existence of a positive constant c such that

$$A(\mathcal{N}) \geqslant c \log N. \qquad [9]$$

If true, this would be best possible, as could be seen by taking the set $\mathcal{N}$ to consist of an arithmetic progression.

It is clear that the identity

$$\int_0^1 C(\alpha)\,\mathrm{d}\alpha = 0,$$

together with the value $C(0) = N$, implies that $B(\mathcal{N}) > 0$. Ankeny and Chowla (Chowla 1952) had conjectured that

$$B(\mathcal{N}) > f(N), \qquad [10]$$

where $f(N) \to \infty$ as $N \to \infty$.

There is a connection between the two problems, since

$$\int_0^1 |C(\alpha)|\,\mathrm{d}\alpha = \int_0^1 (|C(\alpha)| - C(\alpha))\,\mathrm{d}\alpha \leqslant 2B(\mathcal{N})$$

and

$$|C(\alpha)| = \frac{1}{2}|(E(\alpha) + E(-\alpha))e(\alpha M)|,$$

and so if M is large enough, then

$$(E(\alpha) + E(-\alpha))e(\alpha M) = \sum_{n\in\mathcal{M}} e(\alpha n),$$

where $\mathcal{M}$ is a set of $2N$ distinct natural numbers. Thus

$$A(\mathcal{M}) \leqslant 4B(\mathcal{N}).$$

Roth (26), by a unique method, directly established the existence of a positive constant c such that

$$B(\mathcal{N}) > c\left(\frac{\log N}{\log\log N}\right)^{1/2}.$$

At the time the best that was known for $A(\mathcal{N})$, and hence for $B(\mathcal{N})$, was the bound

$$A(\mathcal{N}) > c\left(\frac{\log N}{\log\log N}\right)^{1/4},$$

obtained by Davenport (1960) following earlier work of Cohen (1960).

Later, following progress by Pichorides and Fournier, Konyagin (1981) and McGehee, Pigno and Smith (McGehee *et al.* 1981) independently and by different methods proved Littlewood's conjecture [9]. For a comprehensive survey of the area, see Odlyzko (1982).

With regard to the Ankeny–Chowla problem, it is still an open question as to how large one can take $f(N)$ in [10]. Chowla (1965) conjectured that there might even be a positive constant

c such that $f(N) = cN^{1/2}$ would hold. If true, this would be best possible; see Pichorides (1977).

Square packing

The story behind this is that, some time in the late spring of 1977, Hugh Montgomery had mentioned to one of us [Vaughan] the question of packing unit squares into a large square of side length ℓ not equal to an integer, say $\ell = n + \theta$ where $0 < \theta < 1$.

Perhaps not surprisingly, just lining up as many unit squares as possible with the sides of the large square, which leaves an area uncovered of $2n\theta + \theta^2$, is not very efficient, at least when θ is not very small, and rather more efficient packings are known. For example, Chung and Graham (2009) could exhibit a packing with the uncovered area at most $C\ell^{\phi} \log \ell$, where $\phi = \left(3 + \sqrt{2}\right)/7 = 0.6306\ldots$.

The more interesting question is whether one can show that a certain amount of space can never be covered however ingenious the packing.

The Imperial College Mathematics Department held its annual examiners meeting in June, which went on all day. That year it was, as usual, quite tedious. One had to stay alert in case something came up about one's own tutees, and occasionally there would be some discussion about borderline or exceptional cases, but otherwise it could be a bit of a bore. Anyway, at lunch that day, in order to lighten the mood, the problem was described to Roth. He visibly brightened, and a few days later came up with a very clever argument which eventually led to a joint paper (27) showing that the amount of waste space, whatever the packing, was always at least $c(n \min\{\theta, 1-\theta\})^{1/2}$. Afterwards he claimed that this had been a much more interesting examiners meeting than usual!

The principal idea is to think of a ray entering the large square from the left and to suppose that on reaching the side of a unit square it is 'refracted' to a direction orthogonal to that face. If the little squares through which it passes are not skewed very much, then it will have to pass through waste space to a distance of roughly θ. If there is an appreciable amount of skewing, then there will be triangular pieces of waste between successive squares the area of which can be approximated in terms of the angles of skew. Thus the amount of waste space can be bounded below by a sum of cosines. There are complications if rays cross each other, but in principle one can obtain a lower bound for the amount of waste area.

The question then arises whether $1/2$ is the right exponent. Probably it is not. The proof was 'one dimensional'. If one could take advantage of both dimensions, then perhaps the exponent could be made larger. Nevertheless the theorem is still the best that is known.

Distribution of integer sequences in arithmetic progression

Roth wrote eight papers (8, 10, 11, 17, 20, 21, 22, 23) on the distribution of sequences of natural numbers in arithmetic progressions and related subjects. They were motivated by the following fundamental result established in 1927.

Theorem (van der Waerden 1927)**.** *If the natural numbers are partitioned into any finite number of classes, then at least one of the classes must contain arbitrarily long arithmetic progressions.*

In particular, he was fascinated by the following conjecture made in 1936.

Conjecture (Erdős & Turán 1936)**.** *If a strictly increasing sequence of natural numbers has positive upper density, then it contains arbitrarily long arithmetic progressions.*

The solution of this conjecture in 1975 is one of the cornerstones of combinatorics.

Theorem (Szemerédi 1975)**.** *Let $r_k(N)$ denote the greatest number of natural numbers that can be selected from $1, \ldots, N$ to form a set that does not contain any arithmetic progression of length k. Then $r_k(N) = o_k(N)$ as $N \to \infty$.*

Roth's interest in this area was primarily analytic in nature. He solved the special case $k = 3$ first (8) in 1952, and then showed (10) in 1953 that

$$N^{-1} r_3(N) \ll (\log \log N)^{-1} \qquad [11]$$

as $N \to \infty$. The method involved is the famous Hardy–Littlewood method, but applied in a very unusual and novel way. Let S be a largest set of natural numbers selected from $1, \ldots, N$ with no arithmetic progression of length 3. If S is 'dense', then it has considerable regularity of distribution, both in position and among residue classes to any modulus. These regularity features prove enough to make the Hardy–Littlewood method work. It was the first instance that additive properties of an unknown sequence were studied using the Hardy–Littlewood method, which was first developed with specific sequences such as the primes or the k-th powers in mind.

In 1954, Roth (11) considered a generalization to systems of linear equations. Let $A = (a_{\mu\nu})$ be an $\ell \times n$ matrix with integer entries. A set $\mathcal{U}$ of natural numbers is called an A-set if there are no distinct integers $x_1, \ldots, x_n \in \mathcal{U}$ such that

$$\sum_{\nu=1}^{n} a_{\mu\nu} x_\nu = 0, \quad \mu = 1, \ldots, \ell. \qquad [12]$$

Denote by $A(N)$ the greatest number of natural numbers that can be selected from $1, \ldots, N$ to form an A-set. Under the following conditions on the matrix A, Roth could prove that $A(N) = o(N)$ as $N \to \infty$.

(i) The columns of A add up to form the zero column, i.e.

$$\sum_{\nu=1}^{n} a_{\mu\nu} = 0, \quad \mu = 1, \ldots, \ell;$$

(ii) A has ℓ linearly independent columns. Furthermore, if one of these ℓ linearly independent columns of A is omitted from A, then the remaining $n - 1$ columns of A can be divided into two sets so that there are ℓ linearly independent columns in each of these two sets.

It is quite easy to understand the motivation behind this study. For the special case

$$A = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix},$$

clearly $A(N) = r_3(N)$. However, the requirement that $n > 2\ell$ unfortunately excludes the case

$$A = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix}, \qquad [13]$$

which would have led to a solution for $A(N) = r_4(N)$. Roth would return to this case later.

In 1964, Roth (17) continued his study and considered irregularities of distribution of integer sequences. He showed that a sequence neither very thin nor very dense could not be well distributed simultaneously *among* and *within* all congruence classes. More precisely, let $\mathscr{S}$ be a subset of the set $\{1, \dots, N\}$ of natural numbers, and let $\eta = N^{-1}|\mathscr{S}|$ denote the proportion of these natural numbers that are in $\mathscr{S}$. Since $\mathscr{S}$ is neither very thin nor very dense, η is close to neither 0 nor 1, thus $\eta(1-\eta)$ is not too small. For any arithmetic progression

$$\mathscr{A} = \{n, n+q, \dots, n+(\ell-1)q\}$$

of length ℓ, where $1 \leqslant n < n+(\ell-1)q \leqslant N$, a suitable discrepancy function is defined by

$$D[\mathscr{S}; \mathscr{A}] = |\mathscr{S} \cap \mathscr{A}| - \ell\eta,$$

where $\ell\eta$ is the expectation for the cardinality of the set $\mathscr{S} \cap \mathscr{A}$. Here Roth showed that for any such subset $\mathscr{S}$, one could find an arithmetic progression $\mathscr{A}$ such that

$$|D[\mathscr{S}; \mathscr{A}]| \gg (\eta(1-\eta))^{1/2} N^{1/4}.$$

This result is often affectionately known as Roth's 1/4-theorem. The technique used to establish it is an exponential sums method which can be interpreted as a Fourier transform argument. This observation inspired Beck to develop his Fourier transform approach to problems in irregularities of point distribution, leading to many spectacular results.

It is worthwhile to note that Roth's 1/4-theorem is sharp, as shown by Matoušek and Spencer (1996).

On the other hand, one-sided discrepancy problems, seeking to establish the existence of an arithmetic progression $\mathscr{A}$ for which $D[\mathscr{S}; \mathscr{A}]$ is large and non-negative, is of particular interest. Although this was not discussed in his paper in 1964, Roth (30) commented in a survey written in 2000 that the following was equivalent to the conjecture of Erdős and Turán, nowadays known as Szemerédi's theorem.

Proposition. *Let $c \in (0, 1/2)$ be given. Then there exists $\epsilon > 0$ such that for all sufficiently large ℓ and N, given any subset $\mathscr{S}$ of the set $\{1, \dots, N\}$ satisfying $c < N^{-1}|\mathscr{S}| < 1 - c$, there exists an arithmetic progression $\mathscr{A}$ of length ℓ such that $D[\mathscr{S}; \mathscr{A}] > \epsilon\ell$.*

The conjecture of Erdős and Turán motivated Roth (20, 21) to study irregularities of distribution of integer sequences in arithmetic progressions. Let ℓ be a fixed large natural number. Suppose that $s_1, \dots, s_N$ is a sequence of real numbers, with sum L and average value

LN^{-1}. For natural numbers n and q satisfying the conditions

$$1 \leqslant n < n + (\ell - 1)q \leqslant N, \qquad [14]$$

consider the arithmetic progression $n, n + q, \ldots, n + (\ell - 1)q$ of indices. For any such index $n + \nu q$, the quantity $s_{n+\nu q} - LN^{-1}$ represents the difference of the term $s_{n+\nu q}$ with the average value. Consider now the cumulative difference

$$\sum_{\nu=0}^{\ell-1} (s_{n+\nu q} - LN^{-1}),$$

taken over the ℓ-term arithmetic progression of indices. Here a simplification can be made through the observation that we can take $L = 0$ without loss of generality, for we can simply replace each s_n by $s_n - LN^{-1}$. Thus we consider the sum

$$\sum_{\nu=0}^{\ell-1} s_{n+\nu q}. \qquad [15]$$

Using the exponential sums method of his 1964 paper, Roth (20) showed that for sufficiently large N in terms of the length ℓ, one could find natural numbers n and q satisfying [14] such that the sum [15] would be large in absolute value. However, to obtain any result relevant to the conjecture of Erdős and Turán, it would be necessary to establish one-sided estimates such as a lower bound for the sum [15], and this proved considerably more difficult and would also require the assumption that $L = 0$, as well as the existence of a fixed number Δ such that $1 \leqslant |s_j| \leqslant \Delta$ for every $j = 1, \ldots, N$. Roth (20, 21) made a detailed study of this question, and obtained bounds in terms of the length ℓ. Much later, Sárközy (1978) showed that much better one-sided estimates could be obtained in terms of the upper bound Q of the moduli q.

Motivated by Szemerédi's solution (Szemerédi 1969) of the conjecture of Erdős and Turán in the case $k = 4$, Roth proceeded to develop a new method, embodying a number of Szemerédi's ideas but nevertheless analytic in nature, for proving the same result and certain generalizations of it. In two papers (22, 23) in 1970 and 1972, he considered again $\ell \times n$ integer matrices $A = (a_{\mu\nu})$ and the solvability of the system [12] of linear equations, first studied in 1954. He was able to relax the earlier condition $n > 2\ell$ to include the case $n = 2\ell$, thus including the matrix [13], leading to a solution for $A(N) = r_4(N)$. He remarked that the method could be adopted to give quantitative results regarding the rate at which $N^{-1}r_4(N)$ would tend to 0 as $N \to \infty$, but the proofs would then become complicated and the resulting estimates would be poor. Indeed, Szemerédi's eventual proof in 1975 of the conjecture of Erdős and Turán for arbitrary k left open the problem of quantitative results that were not 'hopelessly weak' regarding the rate at which $N^{-1}r_k(N)$ would tend to 0 as $N \to \infty$.

On the other hand, for $k = 3$, Roth's original exponential sums method in 1953 gave a reasonable bound [11], and subsequent improvements to this by Heath-Brown (FRS 1993) (Heath-Brown 1987) and by Szemerédi (1990), with the bound

$$N^{-1}r_3(N) \ll (\log N)^{-c}, \qquad [16]$$

for some positive absolute constant c, also used exponential sums. The strongest result in this direction currently in the literature, due to Bloom (2006), replaces the right hand side of [16] by $(\log N)^{-1}(\log\log N)^4$.

There remained the question of whether Roth's technique could be extended to achieve reasonable bounds for $N^{-1}r_k(N)$ for $k \geqslant 4$.

A major breakthrough was achieved by Gowers (FRS 1999) (Gowers 1998, 2001) who showed the existence of positive constants $c(k)$, depending only on k, such that

$$N^{-1}r_k(N) \ll_k (\log\log N)^{-c(k)},$$

first in 1998 for $k = 4$ and then in 2001 for arbitrary k. More recently, stronger bounds were obtained for the special case $k = 4$ by Green (FRS 2010) and Tao (Green & Tao 2009, forthcoming), first in 2009 with the estimate

$$N^{-1}r_4(N) \ll \exp(-c(\log\log N)^{1/2}),$$

for some positive absolute constant c, again using Roth's ideas, and later with a bound of the same strength as [16] in a recent paper.

We close this section with a quote from Tao on the impact of Roth's pioneering result.

First of all, we state Roth's original theorem in an alternative form.

Theorem. *Let $\mathscr{A}$ be a set of natural numbers of positive upper density, so that*

$$\limsup_{N\to\infty} \frac{|\mathscr{A} \cap \{1, \ldots, N\}|}{N} > 0.$$

Then $\mathscr{A}$ contains infinitely many arithmetic progressions of three distinct terms.

> At the heart of Roth's elegant argument was the following (surprising at the time) dichotomy: if $\mathscr{A}$ had some moderately large density within some arithmetic progression $\mathscr{P}$, either one could use Fourier-analytic methods to detect the presence of an arithmetic progression of length three inside $\mathscr{A} \cap \mathscr{P}$, or else one could locate a long subprogression $\mathscr{P}'$ of $\mathscr{P}$ on which $\mathscr{A}$ had increased density. Iterating this dichotomy by an argument now known as the *density increment argument*, one eventually obtains Roth's theorem, no matter which side of the dichotomy actually holds. This argument (and the many descendants of it), based on various 'dichotomies between structure and randomness', became essential in many other results of this type, most famously perhaps in Szemerédi's proof of his celebrated theorem on arithmetic progressions that generalized Roth's theorem to progressions of arbitrary length. More recently, my recent work on the Chowla and Elliott conjectures that was a crucial component of the solution of the Erdős discrepancy problem, relies on an *entropy decrement argument* which was directly inspired by the density increment argument of Roth.

Irregularities of point distribution

Roth's work on irregularities of point distribution was motivated by the work of van Aardenne-Ehrenfest (1945, 1949) on a conjecture of van der Corput (1935a, 1935b), that for any infinite sequence $x_1, x_2, x_3, \ldots$ of numbers in the unit interval $[0, 1]$ and for any given positive number κ, one could find two subintervals, of equal length, and a natural number n such that among the terms $x_1, \ldots, x_n$, the number of terms that fell into one subinterval would differ from the number of terms that fell into the other subinterval by more than κ.

Van Aardenne-Ehrenfest (1949) gave a quantitative result on this question. Consider a sequence $x_1, \ldots, x_N$ of numbers in $[0, 1]$. For every natural number n satisfying $1 \leqslant n \leqslant N$

and any number $a \in [0, 1]$, the discrepancy of the subset $\{x_1, \ldots, x_n\}$ with respect to the subinterval $[0, a)$ is denoted by

$$D(n, a) = |\{x_1, \ldots, x_n\} \cap [0, a)| - na.$$

Then

$$\sup_{\substack{1 \leqslant n \leqslant N \\ a \in [0,1]}} |D(n, a)| \gg \frac{\log \log N}{\log \log \log N}. \qquad [17]$$

In what he considered part of his best work, Roth (12) reformulated the problem in 1954 by replacing the discrete parameter n by a continuous one. Consider a finite set $\mathscr{P}$ of N points in the unit square $[0, 1]^2$. For any point $\mathbf{a} = (a_1, a_2) \in [0, 1]^2$, the discrepancy of the set $\mathscr{P}$ with respect to the aligned rectangular box $B(\mathbf{a}) = [0, a_1) \times [0, a_2)$ is denoted by

$$D[\mathscr{P}; B(\mathbf{a})] = |\mathscr{P} \cap B(\mathbf{a})| - N a_1 a_2.$$

Then the two quantities

$$\sup_{\mathbf{a} \in [0,1]^2} |D[\mathscr{P}; B(\mathbf{a})]| \quad \text{and} \quad \sup_{\substack{1 \leqslant n \leqslant N \\ a \in [0,1]}} |D(n, a)|$$

share lower bounds, as functions of N, of the same order of magnitude. Furthermore,

$$\sup_{\mathbf{a} \in [0,1]^2} |D[\mathscr{P}; B(\mathbf{a})]| \gg (\log N)^{1/2}, \qquad [18]$$

vastly superior to the earlier estimate [17].

Indeed, Roth considered the reformulation of the problem in the more general setting of the unit cube $[0, 1]^k$ in the k-dimensional euclidean space, with k-dimensional rectangular boxes $B(\mathbf{a}) = [0, a_1) \times \ldots \times [0, a_k)$, where $\mathbf{a} = (a_1, \ldots, a_k) \in [0, 1]^k$. Consider the L^2-norm

$$\|D_k[\mathscr{P}]\|_2 = \left(\int_{[0,1]^k} |D[\mathscr{P}; B(\mathbf{a})]|^2 \, \mathrm{d}\mathbf{a} \right)^{1/2} \qquad [19]$$

and the L^∞-norm

$$\|D_k[\mathscr{P}]\|_\infty = \sup_{\mathbf{a} \in [0,1]^k} |D[\mathscr{P}; B(\mathbf{a})]|. \qquad [20]$$

Then for every finite set $\mathscr{P}$ of N points in the unit cube $[0, 1]^k$,

$$\|D_k[\mathscr{P}]\|_2 \gg_k (\log N)^{(k-1)/2}, \qquad [21]$$

leading also to the estimate

$$\|D_k[\mathscr{P}]\|_\infty \gg_k (\log N)^{(k-1)/2}, \qquad [22]$$

the multi-dimensional analogue of the estimate [18]. On the other hand, Halton (1960) showed that for every integer $N \geqslant 2$, one could find sets $\mathscr{P}$ of N points in the unit cube

$[0, 1]^k$ such that

$$\|D_k[\mathscr{P}]\|_\infty \ll_k (\log N)^{k-1}. \qquad [23]$$

Schmidt (1972a) showed that the lower bound [18] could be improved to

$$\|D_2[\mathscr{P}]\|_\infty \gg \log N.$$

In view of Halton's result [23], this is therefore essentially best possible. However, efforts to bridge the gap between the lower bound [22] and upper bound [23] when $k > 2$ has met with less success, and this is known as the *Great open problem* in the subject. The current best improvement to the lower bound [22] is due to Bilyk, Lacey and Vagharshakyan (Bilyk *et al.* 2008), with the estimate

$$\|D_k[\mathscr{P}]\|_\infty \gg_k (\log N)^{(k-1)/2+\delta(k)} \qquad [24]$$

for some $\delta(k) \in (0, 1/2)$.

To obtain the lower bound [21], Roth first tried to find the necessary estimates from those parts of the unit cube near the discontinuities of the discrepancy function $D[\mathscr{P}; B(\mathbf{a})]$ at the points of $\mathscr{P}$. However, the arbitrary nature of the point set $\mathscr{P}$ gave him no precise information to work on. Instead, he found many subsets of the unit cube devoid of points of $\mathscr{P}$ and with *trivial discrepancies*. If one partitions the unit cube into a disjoint union of $2N$ subsets of roughly equal volume, then since the set $\mathscr{P}$ contains only N points, at least half of these $2N$ subsets must therefore contain no points of $\mathscr{P}$. From among these subsets, one can then find many subsets of volume αN^{-1}, say, where α is a very small but fixed positive number. These subsets contain no points of $\mathscr{P}$, and so trivially must have discrepancy $-\alpha$, a small quantity but, crucially, bounded away from zero. Roth then achieved his aim by partitioning the unit cube dyadically in roughly $(\log N)^{k-1}$ different ways and then using a system of orthogonal Rademacher functions, modified suitably to eschew the contributions from those parts of the unit cube near the points of $\mathscr{P}$, to pick up these trivial discrepancies, with each of these $(\log N)^{k-1}$ different ways of partition contributing a small positive fixed quantity to the integral in [21].

Much of Roth's work on upper bounds in the second half of the 1970s is centred on showing that the estimate [21] is essentially best possible, and involves the introduction of deep probabilistic ideas into the subject. This entails establishing the existence of point sets that satisfy the required bounds, and there are two main approaches.

One of these approaches is based on the use of badly approximable numbers, and the idea goes back to the 1920s in the work of Hardy and Littlewood on counting the number of lattice points in right angled triangles. Using this, Davenport (1956) showed that the estimate [21] in the special case $k = 2$ was essentially best possible, using lattices and a novel reflection principle. However, the natural extension of this approach to the special case $k = 3$ would require the falsity of the famous conjecture of Littlewood on simultaneous diophantine approximation. In 1979, Roth (28) replaced the reflection principle and devised instead a probabilistic approach through the introduction of a translation variable and obtained an alternative proof of Davenport's result using the same basic construction. Furthermore, by using sheets of lattices on top to each other and arranged in an ingenious way, he was able to extend the argument to the special case $k = 3$, showing that the estimate [21] was also essentially best possible. However, there appears to be no clear way of extending this approach to the cases $k > 3$.

A second approach to upper bounds is based on the generalization of dyadic point sets constructed from the famous van der Corput sequence, first used by Halton to establish the upper bound [23]. Although Halton's proof is remarkably simple, no better upper bound has ever been established. On the other hand, if one tries to use Halton's construction without any modification, then one obtains the estimate

$$\int_{[0,1]^k} |D[\mathscr{P}; B(\mathbf{a})]|^2 \, \mathrm{d}\mathbf{a} \asymp_k (\log N)^{2k-2},$$

the square of what one desires, as observed by Halton and Zaremba (1969). Indeed, this rather large estimate is caused solely by the requirement that all the rectangular boxes $B(\mathbf{a})$ are anchored at the origin. Thus one may suspect that some *average version* of the Halton construction may lead to a better upper bound. In 1980, Roth (29) devised a variant of his earlier probabilistic approach, again using a translation variable. As the Halton construction involves the use of the van der Corput sequence in coprime bases and the Chinese Remainder theorem, this translation variable has to be applied in an ingenious way. The translation variable plays the role of creating some artificial orthogonality. If one takes simply the Halton construction, then the discrepancy function $D[\mathscr{P}; B(\mathbf{a})]$ is a sum of roughly $(\log N)^{k-1}$ functions, each bounded by 1 but collectively without orthogonality or quasi-orthogonality. The corresponding functions modified by the translation variable t each remain bounded by 1 but collectively are now quasi-orthogonal as functions of the parameter t. Thus Roth showed that the lower bound [21] was essentially best possible for every $k \geqslant 2$.

A variant of the second approach was introduced by Faure (1982) with a different generalization of point sets constructed from the van der Corput sequence. These do not possess the periodicity property of the Halton construction, and Roth's translation technique fails. However, Chen (1983) noted that the effects of the Roth translation technique could be achieved also by a system of digit shifts, and that this could be applied to the constructions of Halton as well as the constructions of Faure. Thus he gave an alternative proof that the lower bound [21] was essentially best possible for every $k \geqslant 2$. This digit shift technique led to significant recent progress which we now describe.

Schmidt (1977) extended Roth's lower bound [21] to the L^W-norm analogous to the L^2-norm [19] for every $W > 1$, and showed that for every finite set $\mathscr{P}$ of N points in the unit cube $[0, 1]^k$,

$$\|D_k[\mathscr{P}]\|_W \gg_{k,W} (\log N)^{(k-1)/2}. \qquad [25]$$

This was complemented by a result of Chen (1980), who showed that for every $W > 0$ and every integer $N \geqslant 2$, one could find sets $\mathscr{P}$ of N points in the unit cube $[0, 1]^k$ such that

$$\|D_k[\mathscr{P}]\|_W \ll_{k,W} (\log N)^{(k-1)/2}. \qquad [26]$$

For $W = 1$, Halász (1981) showed that for every finite set $\mathscr{P}$ of N points in the unit cube $[0, 1]^k$,

$$\|D_k[\mathscr{P}]\|_1 \gg_k (\log N)^{1/2}.$$

Clearly there is a rather large gap between this and the upper bound [26] when $k > 2$. Furthermore, no non-trivial lower bound is known for the L^W-average for any W satisfying $0 < W < 1$. Meanwhile, the lower bound [24] of Bilyk, Lacey and Vagharshakyan suggests

the possibility that perhaps for every finite set $\mathscr{P}$ of N points in the unit cube $[0,1]^k$,

$$\|D_k[\mathscr{P}]\|_\infty \gg_k (\log N)^{k/2}. \qquad [27]$$

In recent quite spectacular work, Skriganov (2016) combined the digit shift technique of Chen (1983) with Khinchin's inequality and showed that for every finite set $\mathscr{P}$ of N points in the unit cube $[0,1]^k$, one could find a finite collection $\mathcal{T}_W$ of dyadic digit shifts, depending only on N, k and W, and containing a shift $\mathbf{t} \in \mathcal{T}_W$ such that for the shifted set $\mathscr{P} \oplus \mathbf{t}$,

$$\|D_k[\mathscr{P} \oplus \mathbf{t}]\|_W \gg_{k,W} (\log N)^{(k-1)/2},$$

as well as a finite collection $\mathcal{T}_\infty$ of dyadic digit shifts, depending only on N and k, and containing a shift $\mathbf{t} \in \mathcal{T}_\infty$ such that for the shifted set $\mathscr{P} \oplus \mathbf{t}$,

$$\|D_k[\mathscr{P} \oplus \mathbf{t}]\|_\infty \gg_k (\log N)^{k/2}.$$

On the other hand, the estimates [25] and [26] for $\|D_k[\mathscr{P}]\|_W$, together with the estimate [24] and the conjectured estimate [27] for $\|D_k[\mathscr{P}]\|_\infty$, clearly show that the quantities $\|D_k[\mathscr{P}]\|_W$ and $\|D_k[\mathscr{P}]\|_\infty$ have different orders of magnitude, leading to questions concerning norms between these two. In very recent work, Dick, Hinrichs, Markhasin and Pillichshammer (Dick *et al.* forthcoming) gave sharp lower and upper bounds for the BMO-norm of the discrepancy function. They also studied the analogous questions in Besov, Sobolev and Triebel–Lizorkin spaces, and established sharp lower and upper bounds for the discrepancy function.

Roth's reformulation of van Aardenne-Ehrenfest's problem to a more geometric setting opened the subject of irregularities of distribution to many very interesting questions, by replacing the collection of aligned rectangular boxes in the classical problem by other collections of geometric objects. This led to a number of contributions from Schmidt (1969a, 1969b, 1969c, 1970b, 1975). The ultimate challenge here is then to understand how discrepancy is related to the geometry of these collections, leading to some of Beck's seminal contributions.

The difficulty is that the discrepancy function is a somewhat complicated function which contains information about the geometry, through the characteristic function of the geometric objects under investigation, as well as the measure, since discrepancy is the difference between the discrete counting measure of the points of a finite set $\mathscr{P}$ and a continuous measure arising from the volume. To understand this point, consider a set A of finite volume in k-dimensional euclidean space $\mathbb{R}^k$. Let $\mathscr{P}$ be a set of N points in the unit cube $[0,1]^k$. Then an appropriate discrepancy function for the set A is given by

$$D[\mathscr{P}; A] = |\mathscr{P} \cap A| - N\mu_0(A),$$

where μ_0 denotes the usual volume in $\mathbb{R}^k$ restricted to $[0,1]^k$. This can be written in the form

$$D[\mathscr{P}; A] = \int_{\mathbb{R}^k} \chi_A(\mathbf{y})(\mathrm{d}Z_0(\mathbf{y}) - N\mathrm{d}\mu_0(\mathbf{y})),$$

where Z_0 denotes the counting measure of the set $\mathscr{P}$. Let us consider the translate $A + \mathbf{x}$ of A, where $\mathbf{x} \in \mathbb{R}^k$. Then

$$D[\mathscr{P}; A + \mathbf{x}] = \int_{\mathbb{R}^k} \chi_{A+\mathbf{x}}(\mathbf{y})(\mathrm{d}Z_0(\mathbf{y}) - N\mathrm{d}\mu_0(\mathbf{y}))$$
$$= \int_{\mathbb{R}^k} \chi_A(\mathbf{x} - \mathbf{y})(\mathrm{d}Z_0(\mathbf{y}) - N\mathrm{d}\mu_0(\mathbf{y})),$$

if, for simplicity, we make the further assumption that A is symmetric across the origin. In other words, discrepancy is a convolution of the characteristic function χ_A and the discrepancy measure $\mathrm{d}Z_0 - N\mathrm{d}\mu_0$. The characteristic function χ_A is purely geometric in nature, depending only on the set A and not on the finite set $\mathscr{P}$ at all. On the other hand, the discrepancy measure $\mathrm{d}Z_0 - N\mathrm{d}\mu_0$ depends only on the finite set $\mathscr{P}$ and not on the set A at all. If, for simplicity, we write $D(\mathbf{x}) = D[\mathscr{P}; A + \mathbf{x}]$, then

$$D = \chi_A * (\mathrm{d}Z_0 - N\mathrm{d}\mu_0). \qquad [28]$$

Recall that Roth's 1/4-theorem on integer sequences described earlier was established by an exponential sums technique which could be viewed as a Fourier transform approach. This observation was the catalyst that propelled Beck to arguably the most fascinating results in irregularities of distribution. Passing over to the Fourier transform, the convolution [28] becomes

$$\widehat{D} = \widehat{\chi_A} \cdot \widehat{(\mathrm{d}Z_0 - N\mathrm{d}\mu_0)},$$

an ordinary product of the Fourier transforms of the geometric part and of the measure part, permitting them to be studied separately. For lower bounds, since the finite sets $\mathscr{P}$ are arbitrary, we have little useful information on the measure term, so we concentrate on the term $\widehat{\chi_A}$ or, more precisely, certain averages of $\widehat{\chi_A}$ over sets A belonging to some collection $\mathscr{A}$ with respect to some integral geometric measure. For upper bounds, we have good information on the finite sets $\mathscr{P}$, so we have better control over the measure term $\widehat{(\mathrm{d}Z_0 - N\mathrm{d}\mu_0)}$.

Using the Fourier transform technique, Beck was able to establish, among others, the following two quite remarkable results.

Consider the k-dimensional unit cube $[0, 1]^k$, for convenience treated as a torus. Let A be a compact and convex set in $[0, 1]^k$ satisfying a further mild technical condition, and consider all similar copies $A(\lambda, \tau, \mathbf{x})$ obtained from A by contraction $\lambda \in [0, 1]$, proper orthogonal transformation $\tau \in \mathscr{T}$ and translation $\mathbf{x} \in [0, 1]^k$, where $\mathscr{T}$ denotes the group of all proper orthogonal transformations in $\mathbb{R}^k$, with normalized measure $\mathrm{d}\tau$ so that the total measure is equal to 1. Let

$$D[\mathscr{P}; A(\lambda, \tau, \mathbf{x})] = |\mathscr{P} \cap A(\lambda, \tau, \mathbf{x})| - N\mu_0(A(\lambda, \tau, \mathbf{x}))$$

denote the discrepancy of $\mathscr{P}$ in $A(\lambda, \tau, \mathbf{x})$. Beck (1987) established the lower bound

$$\int_{[0,1]^k} \int_{\mathscr{T}} \int_0^1 |D[\mathscr{P}; A(\lambda, \tau, \mathbf{x})]|^2 \, \mathrm{d}\lambda \, \mathrm{d}\tau \, \mathrm{d}\mathbf{x} \gg_A N^{1-1/k}. \qquad [29]$$

Furthermore, this lower bound is sharp, as shown by Beck and Chen (1990). Thus the order of the magnitude of the mean squares discrepancy is independent of the geometry of the objects under consideration.

Next, consider the unit square $[0,1]^2$, for convenience again treated as a torus. Let A be a compact and convex set in $[0,1]^2$ again satisfying a further mild technical condition, and consider all homothetic copies $A(\lambda, \mathbf{x})$ obtained from A by contraction $\lambda \in [0,1]$ and translation $\mathbf{x} \in [0,1]^2$. Let

$$D[\mathscr{P}; A(\lambda, \mathbf{x})] = |\mathscr{P} \cap A(\lambda, \mathbf{x})| - N\mu_0(A(\lambda, \mathbf{x}))$$

denote the discrepancy of $\mathscr{P}$ in $A(\lambda, \mathbf{x})$. Beck (1988) established the lower bound

$$\int_{[0,1]^2} \int_0^1 |D[\mathscr{P}; A(\lambda, \mathbf{x})]|^2 \, d\lambda \, d\mathbf{x} \gg_A \max\{\log N, \xi_N(A)\}, \qquad [30]$$

where $\xi_N(A)$ depends on the boundary curve ∂A of A. Roughly speaking, the function $\xi_N(A)$ varies from being a constant, in the case when A is a convex polygon, to being a power of N, in the case when A is a circular disc. In fact, it is some sort of measure of how well A can be approximated by an inscribed polygon with not too many sides. Also, the term $\log N$ on the right hand side of [30] should be compared to the estimate [21] in the classical problem with $k = 2$.

Roth also contributed to the subject of irregularities of point distribution by asking a particular question which led to a number of new ideas as well as very strong and even surprising results.

Suppose that $\mathscr{P}$ is a set of N points in U_0, the closed disc of unit area and centred at the origin. For every real number $r \in \mathbb{R}$ and every angle θ satisfying $0 \leqslant \theta \leqslant 2\pi$, let $S(r, \theta)$ denote the closed half-plane

$$S(r, \theta) = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \cdot \mathrm{e}(\theta) \geqslant r\}.$$

Here $\mathrm{e}(\theta) = (\cos\theta, \sin\theta)$ and $\mathbf{x} \cdot \mathbf{y}$ denotes the scalar product of $\mathbf{x}$ and $\mathbf{y}$. Let

$$D[\mathscr{P}; S(r, \theta)] = |\mathscr{P} \cap S(r, \theta)| - N\mu(S(r, \theta) \cap U_0),$$

where μ denotes the usual measure in $\mathbb{R}^2$, denote the discrepancy of $\mathscr{P}$ in $S(r, \theta)$. Write

$$F(N) = \inf_{|\mathscr{P}|=N} \sup_{\substack{0 \leqslant r \leqslant \pi^{-1/2} \\ 0 \leqslant \theta \leqslant 2\pi}} |D[\mathscr{P}; S(r, \theta)]|.$$

Roth asked whether $F(N) \to \infty$ as $N \to \infty$. Here the supremum is taken over all disc-segments in U_0, and the infimum is taken over all sets $\mathscr{P}$ of N points in U_0.

This question was resolved by Beck (1983), who used a very clever adaptation of the Fourier transform technique, first introduced by Roth in 1964, to show that $F(N) \gg N^{1/4}(\log N)^{-7/2}$. Alexander (1990) then devised an ingenious integral geometric approach and improved this to $F(N) \gg N^{1/4}$. On the other hand, it can be shown, using large deviation techniques in probability theory, that $F(N) \ll N^{1/4}(\log N)^{1/2}$. However, closing the gap between the lower and upper bounds proved a real challenge, and the eventual solution by Matoušek (1995), who showed that $F(N) \ll N^{1/4}$, remains one of the most remarkable feats in this subject.

Beck and Alexander basically studied the L^2-norm of the discrepancy function $D[\mathscr{P}; S(r, \theta)]$. Then it can be proved that for every set $\mathscr{P}$ of N points in U_0,

$$\int_0^{2\pi} \int_0^{\pi^{-1/2}} |D[\mathscr{P}; S(r, \theta)]|^2 \, \mathrm{d}r \, \mathrm{d}\theta \gg N^{1/2}. \qquad [31]$$

It can be shown, using probabilistic techniques, that this estimate is sharp. Subsequently, Beck and Chen (1993) showed that the discrepancy function $D[\mathscr{P}; S(r, \theta)]$ possessed very surprising behaviour, in that it could have very large absolute value occasionally but very rarely. More precisely, they showed that for every integer $N \geqslant 2$, one could find sets $\mathscr{P}$ of N points in U_0 such that

$$\int_0^{2\pi} \int_0^{\pi^{-1/2}} |D[\mathscr{P}; S(r, \theta)]| \, \mathrm{d}r \, \mathrm{d}\theta \ll (\log N)^2. \qquad [32]$$

Indeed, they applied a variant of Roth's probabilistic technique which essentially randomized the origin, with the bonus that the point sets $\mathscr{P}$ constructed could be explicitly given. Note that such sets satisfy [31] and [32] simultaneously.

Heilbronn's triangle problem

One day in the 1940s, Heilbronn looked out of his window and saw a group of soldiers. They were very dispirited and seemed not to be marching in formation. He thus set out to investigate how badly they could possibly do.

Let

$$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N \qquad [33]$$

be a distribution of N points in a closed disc of unit area, such that the minimum of the areas of the triangles $\mathbf{p}_i\mathbf{p}_j\mathbf{p}_k$, taken over $1 \leqslant i < j < k \leqslant N$, assumes its maximum value $\Delta = \Delta(N)$. Heilbronn conjectured that $\Delta(N) \ll N^{-2}$, and Erdős showed that, if true, this would be best possible.

It is almost trivial that $\Delta(N) \ll N^{-1}$. The first improvement to this was due to Roth (3) who proved in 1951 that $\Delta(N) \ll N^{-1}(\log\log N)^{-1/2}$. Schmidt (1972b), using a different method involving weights, improved this to $\Delta(N) \ll N^{-1}(\log N)^{-1/2}$. Roth (24, 25), using weights in a different way, proved in 1972, first that $\Delta(N) \ll N^{-\mu+\epsilon}$, where $\mu = 2 - \sqrt{4/5} = 1.105\ldots$, and then that $\Delta(N) \ll N^{-\mu'+\epsilon}$, where $\mu' = \left(17 - \sqrt{65}\right)/8 = 1.117\ldots$.

Let

$$D = \{\mathbf{x} = (x, y) \in \mathbb{R}^2 : x^2 + y^2 \leqslant \pi^{-1}\}$$

denote the closed disc of unit area containing the points [33]. For any pair $\tau = (\mathbf{p}_i, \mathbf{p}_j)$ of distinct points in [33], we write $d(\tau) = |\mathbf{p}_i - \mathbf{p}_j|$ to denote their euclidean distance. If

$$y\cos\theta - x\sin\theta = a,$$

where $0 \leqslant \theta < \pi$, is the line joining these two points, let $H_{\tau,w}$ denote the open strip

$$a - \frac{w}{2} < y\cos\theta - x\sin\theta < a + \frac{w}{2}$$

of width $w > 0$ about this line.

If $\mathscr{A}$ is any subset of the plane, let $\mathscr{A}(\mathbf{x})$ denote the characteristic function of $\mathscr{A}$, and let $|\mathscr{A}|$ denote the number of points of [33] in $\mathscr{A}$.

The fact that no three points of the set [33] form a triangle of area less than Δ implies that each pair τ satisfies $|H_{\tau,4\Delta/d(\tau)}| = 2$, but with the corresponding statistical expectation

$$N \int_{\mathbb{R}^2} D(\mathbf{x}) H_{\tau,4\Delta/d(\tau)}(\mathbf{x})\, \mathrm{d}\mathbf{x},$$

roughly having the order of magnitude $N\Delta/d(\tau)$.

Here, and in the sequel, we ignore any complications that can arise in connection with pairs τ close to the boundary of D, so that some statements may not stand up to closer scrutiny.

Suppose that $N\Delta$ is *not too small*, contrary to what we wish to prove. Then $N\Delta/d(\tau)$ is *large* if $d(\tau)$ is *very small*. Thus for suitable u and w'', Roth established a result of the following kind.

Assertion A. *All the strips $H_{\tau,w''}$ for which $d(\tau) \leqslant u$ are very deficient of points of* [33].

By a reasonably simple argument, Roth also established, for suitable w', results of the following nature.

Assertion B. *A significant proportion of all those pairs τ for which $d(\tau)$ is appropriately restricted are such that the strips $H_{\tau,w'}$ are not unduly deficient of points of* [33].

He went on to show that if the restriction on $d(\tau)$ in Assertion B was taken to be identical to that in Assertion A, namely $d(\tau) \leqslant u$, then the respective premises would require that the order of magnitude of w' was large compared to that of w''. He subsequently devised a technique for deducing from Assertion A, on the assumption that $N\Delta$ was not small, that almost all the wider strips $H_{\tau,w'}$, with $d(\tau) \leqslant u$, were deficient of points of [33], thus contradicting Assertion B.

Indeed, he used weights to construct a system of quasi-orthogonal functions, and made use of a generalization of Bessel's inequality, due to Selberg in the course of investigating the large sieve and applicable to quasi-orthogonal systems, that for elements $f, \psi^{(1)}, \ldots, \psi^{(R)}$ of an inner product space over the complex numbers,

$$\sum_{r=1}^{R} |\langle f, \psi^{(r)}\rangle|^2 \left(\sum_{s=1}^{R} |\langle \psi^{(r)}, \psi^{(s)}\rangle|\right)^{-1} \leqslant \|f\|^2, \qquad [34]$$

with an inner product of the type

$$\langle f, g\rangle = \int_{\mathbb{R}^2} f(\mathbf{x}) g(\mathbf{x})\, \mathrm{d}\mathbf{x}.$$

Observe that for any two strips H_{τ_1,w_1} and H_{τ_2,w_2}, the integral

$$\int_{\mathbb{R}^2} H_{\tau_1,w_1}(\mathbf{x}) H_{\tau_2,w_2}(\mathbf{x})\, \mathrm{d}\mathbf{x},$$

if finite, is equal to the area of the parallelogram $H_{\tau_1,w_1} \cap H_{\tau_2,w_2}$, given by cw_1w_2, where the constant $c = c(\tau_1, \tau_2)$ depends only on τ_1 and τ_2.

For any $w' > w'' > 0$, write

$$\phi_{\tau,w',w''}(\mathbf{x}) = \frac{H_{\tau,w'}(\mathbf{x})}{w'} - \frac{H_{\tau,w''}(\mathbf{x})}{w''}. \qquad [35]$$

Then for any two pairs τ^* and τ^{**} of points of [33], in view of the observation above, the integral

$$\int_{\mathbb{R}^2} \phi_{\tau^*,w',w''}(\mathbf{x})\phi_{\tau^{**},w',w''}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0$$

whenever it is finite. Note here that this orthogonality is achieved by using suitable weights on the two functions $H_{\tau,w'}(\mathbf{x})$ and $H_{\tau,w''}(\mathbf{x})$ in [35], weighted in inverse proportion to the widths of the strips.

To use Selberg's inequality [34], we must work with functions of finite norms. We may consider functions derived from $\phi_{\tau,w',w''}(\mathbf{x})$ by replacing its value by zero outside the disc D; in other words, functions of the form

$$\Phi_{\tau,w',w''}(\mathbf{x}) = D(\mathbf{x})\phi_{\tau,w',w''}(\mathbf{x}).$$

Then for any two pairs τ^* and τ^{**} of points of [33], the integral

$$\int_{\mathbb{R}^2} \Phi_{\tau^*,w',w''}(\mathbf{x})\Phi_{\tau^{**},w',w''}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0$$

unless the common parallelogram of the two strips $H_{\tau^*,w'}$ and $H_{\tau^{**},w'}$ intersects the boundary circle of the disc D. Thus we achieve quasi-orthogonality in some sense.

Now let

$$f(\mathbf{x}) = \sum_{i=1}^{N} D_{\mathbf{p}_i}(\mathbf{x}),$$

where for every $i = 1, \ldots, N$, $D_{\mathbf{p}_i}$ is the closed disc with centre $\mathbf{p}_i$, radius $w''/2$ and area

$$A = \pi\left(\frac{w''}{2}\right)^2 = \int_{\mathbb{R}^2} D_{\mathbf{p}_i}(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

We can think of $f(\mathbf{x})$ as an approximation to a mass distribution of mass A at each point $\mathbf{p}_i$ of [33], obtained by spreading out this mass A over each small disc $D_{\mathbf{p}_i}$ instead, in order to reduce the norm. The choice of small radius $w''/2$ ensures that the total mass falling into a typical strip of width w'' is roughly proportional to the number of points of [33] in the strip.

Ignoring the error arising from the overlap of the discs $D_{\mathbf{p}_i}$, we envisage that for each pair τ, the integral

$$\int_{\mathbb{R}^2} \Phi_{\tau,w',w''}(\mathbf{x})\, f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

is approximately equal to

$$\pi\left(\frac{w''}{2}\right)^2\left(\frac{|H_{\tau,w'}|}{w'} - \frac{|H_{\tau,w''}|}{w''}\right).$$

Selberg's inequality [34] is now applied in the form

$$\sum_{\tau^*:\, d(\tau^*)\leqslant u} |\langle f, \Phi_{\tau^*}\rangle|^2 \left(\sum_{\tau^{**}:\, d(\tau^{**})\leqslant u} |\langle \Phi_{\tau^*}, \Phi_{\tau^{**}}\rangle| \right)^{-1} \leqslant \| f \|^2,$$

where Φ_τ denotes the function $\Phi_{\tau,w',w''}(\mathbf{x})$.

It is quite clear from Assertion A that the manner in which the strips $H_{\tau,w''}$ overlap is relevant to the problem. Some quantitative estimates corresponding to this observation can be formulated, on the assumption that $N\Delta$ is *not too small*. These can now be used to estimate, for each fixed τ^*, the sum

$$\sum_{\tau^{**}:\, d(\tau^{**})\leqslant u} |\langle \Phi_{\tau^*}, \Phi_{\tau^{**}}\rangle|.$$

Clearly, only those τ^{**} for which orthogonality breaks down contribute to this sum. We thus obtain a good bound for this sum of $N\Delta$ is *not too small*. On the other hand, estimating $\| f \|^2$ presents no difficulties. We thus obtain a good upper bound for the sum

$$\sum_{\tau:\, d(\tau)\leqslant u} \left(\frac{|H_{\tau,w'}|}{w'} - \frac{|H_{\tau,w''}|}{w''} \right)^2, \qquad [36]$$

a bound small compared to the expectation of roughly

$$N^2 \sum_{\tau:\, d(\tau)\leqslant u} 1$$

for the sum

$$\sum_{\tau:\, d(\tau)\leqslant u} \left(\frac{|H_{\tau,w'}|}{w'} \right)^2.$$

Such an estimate for the sum [36] enables us to deduce, from the fact that $|H_{\tau,w''}|/w''$ is always small compared to its expected value, that $|H_{\tau,w'}|/w'$ is nearly always small compared to its expected value, contradicting Assertion B.

Heilbronn's conjecture was disproved by Komlós, Pintz and Szemerédi (Komlós *et al.* 1982), who used combinatorial methods to show that $\Delta(N) \gg N^{-2}\log N$. Earlier, they (Komlós *et al.* 1981) had used a small refinement of Roth's method to give a better upper bound $\Delta(N) \ll N^{-\mu''+\epsilon}$, where $\mu'' = 8/7 = 1.142\ldots$.

Acknowledgements

The authors thank St Paul's School, London, and its librarian and archivist Alexandra Aslett for their kind permission and generous help in accessing the school records of Klaus Roth, as well as Myra Daridan for her help in providing some family background of Klaus and Melek Roth.

The authors have borrowed liberally from the masterly description of Roth's work on distribution of integer sequences in arithmetic progressions written by their colleagues András Sárközy and Cameron Stewart (Sárközy & Stewart 2009) and published in the eightieth birthday volume of Klaus Roth. They also thank Terence Tao for permission to quote from his blog.

Last, but definitely not least, the authors thank Trevor Stuart (FRS 1974) for his encouragement to undertake this project and his support throughout, and Benjamin Baumslag, József Beck, Frank Berkshire, David Cox (FRS

1973), Harold Diamond, Freeman Dyson, Ben Green, Roger Heath-Brown, David Larman, David Masser (FRS 2005), Wolfgang Schmidt, Eira Scourfield, Peter Shiu, Endre Szemerédi, Terry Tao, Lynda White and Trevor Wooley for their comments to an earlier draft of this article. The frontispiece was taken by Godfrey Argent and is © Godfrey Argent Studio.

Author profiles

William Chen and Robert Vaughan

William Chen is Emeritus Professor of Mathematics at Macquarie University Sydney. Robert Vaughan is Professor of Mathematics at the Pennsylvania State University. They both held positions at Imperial College and were colleagues of Roth, before moving to Australia and the USA in 1991 and 1997, respectively.

William Chen was the last PhD student of Roth. His main work is in irregularities of distribution, or geometric discrepancy theory, a subject started by Roth in 1954 with the publication of his pioneering paper. Robert Vaughan, like Roth, was a PhD student of Theodor Estermann. He works in analytic number theory and diophantine approximation, and is an expert on the Hardy–Littlewood method, a technique in additive number theory central to Roth's seminal work on the distribution of integer sequences in arithmetic progression.

William Chen and Robert Vaughan were joint organizers of a symposium in analytic number theory at Imperial College in July 1985, in honour of Roth's sixtieth birthday. They were also two of the editors, with Timothy Gowers, Heini Halberstam and Wolfgang Schmidt, of the volume *Analytic Number Theory: Essays in Honour of Klaus Roth*, a collection of essays celebrating the eightieth birthday of Roth published in 2009 by the Cambridge University Press.

Robert Vaughan was elected Fellow of the Royal Society in 1990.

References to other authors

Alexander, J. R. 1990 Geometric methods in the study of irregularities of distribution. *Combinatorica* **10**, 115–136.

Baker, A. 1964 On Mahler's classification of transcendental numbers. *Acta Math.* **111**, 97–120.

Barban, M. B. 1961 New applications of the 'great sieve' of Ju.V. Linnik (Russian). *Akad. Nauk Uzbek. SSR Trudy Inst. Mat.* **22**, 1–20.

Barban, M. B. 1966 The 'large sieve' method and its application to number theory (Russian). *Uspehi Mat. Nauk* **21**, 51–102.

Beck, J. 1983 On a problem of K. F. Roth concerning irregularities of point distribution. *Invent. Math.* **74**, 477–487.

Beck, J. 1987 Irregularities of point distribution I. *Acta Math.* **159**, 1–49.

Beck, J. 1988 Irregularities of point distribution II. *Proc. Lond. Math. Soc.* **56**, 1–50.

Beck, J. & Chen, W. W. L. 1990 Note on irregularities of distribution II. *Proc. Lond. Math. Soc.* **61**, 251–272.

Beck, J. & Chen, W. W. L. 1993 Irregularities of point distribution relative to half-planes I. *Mathematika* **40**, 102–126.

Bentkus, V. & Götze, F. 1999 Lattice point problems and distribution of values of quadratic forms. *Ann. Math.* **150**, 977–1027.

Bilyk, D., Lacey, M. T. & Vagharshakyan, A. 2008 On the small ball inequality in all dimensions. *J. Funct. Anal.* **254**, 2470–2502.

Bloom, T. F. 2016 A quantitative improvement for Roth's theorem on arithmetic progressions. *J. Lond. Math. Soc.* **93**, 643–663.

Bombieri, E. 1965 On the large sieve. *Mathematika* **12**, 201–225.

Bombieri, E. & Davenport, H. 1968 On the large sieve method. In *Abhandlungen aus Zahlentheorie und Analysis. Zur Erinnerung an Edmund Landau (1877–1938)*, pp. 11–22. Berlin: VEB Deutscher Verlag der Wissenschaften.

Brüdern, J., Kawada, K. & Wooley, T. D. 2009 Additive representation in thin sequences VIII: diophantine inequalities in review. In *Number Theory, Dreaming in Dreams: Proceedings of the 5th China–Japan Seminar* (ed. T. Aoki, S. Kanemitsu & J.Y. Liu), pp. 20–79. World Scientific.

Chen, W. W. L. 1980 On irregularities of distribution. *Mathematika* **27**, 153–170.

Chen, W. W. L. 1983 On irregularities of distribution II. *Q. J. Math. Oxf.* **34**, 257–279.

Chowla, S. 1952 The Riemann zeta and allied functions. *Bull. Am. Math. Soc.* **58**, 287–305.

Chowla, S. 1965 Some applications of a method of A. Selberg. *J. Reine Angew. Math.* **217**, 128–132.

Chung, F. & Graham, R. 2009 Packing equal squares into a large square. *J. Combin. Theory Ser. A* **116**, 1167–1175.

Cohen, P. J. 1960 On a conjecture of Littlewood and idempotent measures. *Am. J. Math.* **82**, 191–212.

Davenport, H. 1939 On Waring's problem for cubes. *Acta Math.* **71**, 123–143.

Davenport, H. 1956 Note on irregularities of distribution. *Mathematika* **3**, 131–135.

Davenport, H. 1960 On a theorem of P. J. Cohen. *Mathematika* **7**, 93–97.

Davenport, H. 1968 A note on Thue's theorem. *Mathematika* **15**, 76–87.

Davenport, H. & Heilbronn, H. 1938a On Waring's problem: two cubes and one square. *Proc. Lond. Math. Soc.* **43**, 73–104.

Davenport, H. & Heilbronn, H. 1938b Note on a result in the additive theory of numbers. *Proc. Lond. Math. Soc.* **43**, 142–151.

Davenport, H. & Heilbronn, H. 1946 On indefinite quadratic forms in five variables. *J. Lond. Math. Soc.* **21**, 185–193.

Dick, J., Hinrichs, A., Markhasin, L. & Pillichshammer, F. In press. Discrepancy of second order digital sequences in function spaces with dominating mixed smoothness. *Mathematika*.

Dietmann, R. & Elsholtz, C. 2008 Sums of two squares and one biquadrate. *Funct. Approx. Comment. Math.* **38**, 233–234.

Dietmann, R. & Elsholtz, C. 2016 Sums of two squares and a power. In *From arithmetic to zeta-functions* (ed. J. Sander, J. Steuding & R. Steuding), pp. 103–108. Springer.

Dyson, F. J. 1947 The approximation of algebraic numbers by rationals. *Acta Math. Acad. Sci. Hungar.* **79**, 225–240.

Erdős, P. & Turán, P. 1936 On some sequences of integers. *J. Lond. Math. Soc.* **11**, 261–264.

Estermann, T. 1931 Einige Sätze über quadratfreie Zahlen. *Math. Ann.* **105**, 653–662.

Faure, H. 1982 Discrépance de suites associées à un système de numération (en dimension *s*). *Acta Arith.* **41**, 337–351.

Filaseta, M. 1990 Short interval results for squarefree numbers. *J. Number Theory* **35**, 128–149.

Filaseta, M. 1993 Short interval results for *k*-free values of irreducible polynomials. *Acta Arith.* **64**, 249–270.

Filaseta, M. & Trifonov, O. 1990 On gaps between squarefree numbers. In *Analytic number theory: proceedings of a conference in honor of Paul T. Bateman* (ed. B. C. Berndt, H. G. Diamond, H. Halberstam & A. J. Hildebrand), pp. 235–253. Birkhäuser, *Prog. Math.* **85**.

Fogels, E. 1941 On the average values of arithmetic functions. *Math. Proc. Camb. Phil. Soc.* **37**, 358–372.

Ford, K. 1996 The representation on numbers as sums of unlike powers II. *J. Am. Math. Soc.* **9**, 919–940; addendum and corrigendum, **12**, 1213.

Freeman, D. E. 2002 Asymptotic lower bounds and formulas for diophantine inequalities. In *Number theory for the millennium* (ed. M. A. Bennett, B. C. Berndt, N. Boston, H. G. Diamond, A. J. Hildebrand & W. Philipp), vol. 2, pp. 57–74. A. K. Peters.

Goldston, D. A., Pintz, J. & Yıldırım, C. Y. 2009 Primes in tuples I. *Ann. Math.* **170**, 819–862.

Gowers, W. T. 1998 A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.* **8**, 529–551.

Gowers, W. T. 2001 A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11**, 465–588.

Graham, S. W. & Kolesnik, G. 1988 On the difference between consecutive squarefree integers. *Acta Arith.* **49**, 435–447.

Granville, A. 1998 ABC allows us to count squarefrees. *Int. Math. Res. Notices* **1998**(19), 991–1009.

Green, B. J. & Tao, T. 2009 New bounds for Szemerédi's theorem II: a new bound for $r_4(N)$. In *Analytic number theory: essays in honour of Klaus Roth* (ed. W. W. L. Chen, W. T. Gowers, H. Halberstam, W. M. Schmidt & R. C. Vaughan), pp. 180–204. Cambridge University Press.

Green, B. J. & Tao, T. Submitted. New bounds for Szemerédi's theorem III: a polylogarithmic bound for $r_4(N)$. *Mathematika*.

Gundlach, F. 2013 Integral Brauer–Manin obstructions for sums of two squares and a power. *J. Lond. Math. Soc.* **88**, 599–618.

Halász, G. 1981 On Roth's method in the theory of irregularities of point distributions. In *Recent progress in analytic number theory* (ed. H. Halberstam & C. Hooley), vol. 2, pp. 79–94. Academic Press.

Halton, J. H. 1960 On the efficiency of certain quasirandom sequences of points in evaluating multidimensional integrals. *Num. Math.* **2**, 84–90.

Halton, J. H. & Zaremba, S. K. 1969 The extreme and L^2 discrepancies of some plane sets. *Monatsh. Math.* **73**, 316–328.

Hardy, G. H. & Littlewood, J. E. 1920a A new solution of Waring's problem. *Q. J. Math. Oxf.* **48**, 272–293.

Hardy, G. H. & Littlewood, J. E. 1920b Some problems of 'Partitio Numerorum' I: a new solution of Waring's problem. *Göttinger Nachrichten* **1920**, 33–54.

Hardy, G. H. & Littlewood, J. E. 1921 Some problems of 'Partitio Numerorum' II: proof that every large number is the sum of at most 21 biquadrates. *Math. Z.* **9**, 14–27.

Hardy, G. H. & Littlewood, J. E. 1922 Some problems of 'Partitio Numerorum' IV: the singular series in Waring's problem and the value of the number $G(k)$. *Math. Z.* **12**, 161–188.

Hardy, G. H. & Littlewood, J. E. 1923 Some problems of 'Partitio Numerorum' III: on the expression of a number as a sum of primes. *Acta Math.* **44**, 1–70.

Hardy, G. H. & Littlewood, J. E. 1924 Some problems of 'Partitio Numerorum' V: a further contribution to the study of Goldbach's problem. *Proc. Lond. Math. Soc.* **22**, 46–56.

Hardy, G. H. & Littlewood, J. E. 1925 Some problems of 'Partitio Numerorum' VI: further researches in Waring's problem. *Math. Z.* **23**, 1–37.

Hardy, G. H. & Littlewood, J. E. 1928 Some problems of 'Partitio Numerorum' VIII: the number $\Gamma(k)$ in Waring's problem. *Proc. Lond. Math. Soc.* **28**, 518–542.

Hardy, G. H. & Ramanujan, S. 1918 Asymptotic formulae in combinatory analysis. *Proc. Lond. Math. Soc.* **17**, 75–115.

Heath-Brown, D. R. 1987 Integer sets containing no arithmetic progressions. *J. Lond. Math. Soc.* **35**, 385–394.

Hlawka, E. 1949 Ausfüllung und Überdeckung konvexer Körper durch konvexe Körper. *Monatsh. Math.* **53**, 81–131.

Jagy, W. C. & Kaplansky, I. 1995 Sums of squares, cubes and higher powers. *Exp. Math.* **4**, 169–173.

Kawada, K. 1997 Note on the sum of cubes of primes and an almost prime. *Arch. Math. (Basel)* **69**, 13–19.

Komlós, J., Pintz, J. & Szemerédi, E. 1981 On Heilbronn's triangle problem. *J. Lond. Math. Soc.* **24**, 385–396.

Komlós, J., Pintz, J. & Szemerédi, E. 1982 A lower bound for Heilbronn's problem. *J. Lond. Math. Soc.* **25**, 13–24.

Konyagin, S. V. 1981 On a problem of Littlewood (Russian). *Izv. Akad. Nauk SSSR Ser. Mat.* **35**, 243–265.

Li, H. 1995 The solubility of certain diophantine inequalities. *Acta Math. Sin. (Engl. Ser.)* **11**, 137–145.

Linnik, Ju. V. 1941 The large sieve. *C. R. (Dokl.) Acad. Sci. URSS (N.S.)* **30**, 292–294.

Linnik, Ju. V. 1942 A remark on the least quadratic non-residue. *C. R. (Dokl.) Acad. Sci. URSS (N.S.)* **36**, 119–120.

Liouville, J. 1844 Sur des classes très-étendues de quantités dont la irrationelles algébrigues. *C. R. Acad. Sci. Paris* **18**, 883–885; 910–911.

Matoušek, J. 1995 Tight upper bounds for the discrepancy of halfspaces. *Discrete Comput. Geom.* **13**, 593–601.

Matoušek, J. & Spencer, J. 1996 Discrepancy in arithmetic progressions. *J. Am. Math. Soc.* **9**, 195–204.

Maynard, J. 2015 Small gaps between primes. *Ann. Math.* **181**, 383–413.

McGehee, O. C., Pigno, L. & Smith, B. 1981 Hardy's inequality and the L^1 norm of exponential sums. *Ann. Math.* **113**, 613–618.

Montgomery, H. L. 1971 *Topics in multiplicative number theory*. Springer, *Lecture Notes in Mathematics* **227**.

Montgomery, H. L. 1978 The analytic principle of the large sieve. *Bull. Am. Math. Soc.* **84**, 547–567.

Montgomery, H. L. 1994 *Ten lectures on the interface between analytic number theory and harmonic analysis*. American Mathematical Society, *CBMS Regional Conference Series in Mathematics* **84**.

Montgomery, H. L. & Vaughan, R. C. 1973 The large sieve. *Mathematika* **20**, 119–134.

Montgomery, H. L. & Vaughan, R. C. 1974 Hilbert's inequality. *J. Lond. Math. Soc.* **8**, 73–82.

Odlyzko, A. M. 1982 Minima of cosine sums and maxima of polynomials on the unit circle. *J. Lond. Math. Soc.* **26**, 412–420.

Pichorides, S. K. 1977 Norms of exponential sums. *Publ. Math. d'Orsay* **77–73**, 69p.

Rankin, R. A. 1955 Van der Corput's method and the theory of exponent pairs. *Q. J. Math. Oxf.* **6**, 147–153.

Rényi, A. 1947 On the representation of an even number as the sum of a single prime and a single almost-prime number (Russian). *Dokl. Acad. Nauk SSSR* **56**, 455–458.

Rényi, A. 1948 On the representation of an even number as the sum of a single prime and a single almost-prime number (Russian). *Izv. Acad. Nauk SSSR Ser. Mat.* **12**, 57–78.

Rényi, A. 1949a Probability methods in number theory. *Publ. Math. Collectae Budapest* **1**(21), 9 p.

Rényi, A. 1949b Un nouveau théorème concernant les fonctions indépendantes et ses applications à la théorie des nombres. *J. Math. Pures Appl.* **28**, 137–149.

Rényi, A. 1949c On a theorem of the theory of probability and its application in number theory (Russian. Czech summary). *Časopis Pěst. Mat. Fys.* **74**, 167–175.

Rényi, A. 1949d Sur un théorème général de probabilité. *Ann. Inst. Fourier (Grenoble)* **1**, 43–52.

Rényi, A. 1950 On the large sieve of Ju. V. Linnik. *Compositio Math.* **8**, 68–75.

Rényi, A. 1958a Probability methods in number theory (Chinese). *Adv. Math.* **4**, 465–510.

Rényi, A. 1958b On the probabilistic generalization of the large sieve of Linnik (Hungarian and Russian summaries). *Magyar Tud. Akad. Mat. Kutató Int. Közl* **3**, 199–206.

Rényi, A. 1959 New version of the probabilistic generalization of the large sieve (Russian summary). *Acta Math. Acad. Sci. Hung.* **10**, 217–226.

Rényi, A. 1962 On the representation of an even number as the sum of a prime and an almost prime. *Am. Math. Soc. Transl.* **19**, 299–321.

Richert, H. -E. 1954 On the difference between consecutive squarefree numbers. *J. Lond. Math. Soc.* **29**, 16–20.

Rogers, C. A. 1950 A note on coverings and packings. *J. Lond. Math. Soc.* **25**, 327–331.

Rogers, C. A. 1958 Lattice coverings of space: the Minkowski–Hlawka theorem. *Proc. Lond. Math. Soc.* **8**, 447–465.

Rogers, C. A. 1959 Lattice coverings of space. *Mathematika* **6**, 33–39.

Sárközy, A. 1978 Some remarks concerning irregularities of distribution of sequences of integers in arithmetic progressions III. *Period. Math. Hung.* **9**, 127–144.

Sárközy, A. & Stewart, C. L. 2009 Irregularities of sequences relative to long arithmetic progressions. In *Analytic number theory: essays in honour of Klaus Roth* (ed. W. W. L. Chen, W. T. Gowers, H. Halberstam, W. M. Schmidt & R. C. Vaughan), pp. 389–401. Cambridge University Press.

Schmidt, P. G. 1964 *Abschätzungen bei unsymmetrischen Gitterpunktproblemen*. Dissertation, Georg-August-Universität Göttingen.

Schmidt, W. M. 1959 Masstheorie in der Geometrie der Zahlen. *Acta Math.* **102**, 159–224.

Schmidt, W. M. 1969a Irregularities of distribution II. *Trans. Am. Math. Soc.* **136**, 347–360.

Schmidt, W. M. 1969b Irregularities of distribution III. *Pacific J. Math.* **29**, 225–234.

Schmidt, W. M. 1969c Irregularities of distribution IV. *Invent. Math.* **7**, 55–82.

Schmidt, W. M. 1970a Simultaneous approximation to algebraic numbers by rationals. *Acta Math.* **125**, 189–201.

Schmidt, W. M. 1970b Irregularities of distribution V. *Proc. Am. Math. Soc.* **25**, 608–614.

Schmidt, W. M. 1972a Irregularities of distribution VII. *Acta Arith.* **21**, 45–50.

Schmidt, W. M. 1972b On a problem of Heilbronn. *J. Lond. Math. Soc.* **4**, 545–550.

Schmidt, W. M. 1975 Irregularities of distribution IX. *Acta Arith.* **27**, 385–396.

Schmidt, W. M. 1977 Irregularities of distribution X. In *Number theory and algebra* (ed. H. Zassenhaus), pp. 311–329. Academic Press.

Schmidt, W. M. 1989 The subspace theorem in diophantine approximations. *Compositio Math.* **69**, 121–173.

Selberg, A. 1991 Lectures on sieves. In *Collected papers of Atle Selberg*, vol. 2, pp. 45–247. Springer.

Siegel, C. L. 1921 Approximationen algebraischer Zahlen. *Math. Z.* **10**, 173–213.
Skriganov, M. M. 2016 Dyadic shift randomization in classical discrepancy theory. *Mathematika* **62**, 183–209.
Szemerédi, E. 1969 On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.* **20**, 89–104.
Szemerédi, E. 1975 On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27**, 199–245.
Szemerédi, E. 1990 Integer sets containing no arithmetic progressions. *Acta Math. Hung.* **56**, 155–158.
Thue, A. 1909 Über Annäherungswerte algebraischer Zahlen. *J. Reine Angew. Math.* **135**, 284–305.
Trifonov, O. 1988 On the squarefree problem. *C. R. Acad. Bulgare Sci.* **41**, 37–40.
Trifonov, O. 1989 On the squarefree problem II. *Math. Balkanica* **3**, 284–295.
Vaaler, J. D. 1985 Some extremal functions in Fourier analysis. *Bull. Am. Math. Soc.* **12**, 183–216.
van Aardenne-Ehrenfest, T. 1945 Proof of the impossibility of a just distribution of an infinite sequence of points over an interval. *Proc. Kon. Ned. Akad. v. Wetensch.* **48**, 266–271.
van Aardenne-Ehrenfest, T. 1949 On the impossibility of a just distribution. *Proc. Kon. Ned. Akad. v. Wetensch.* **52**, 734–739.
van der Corput, J. G. 1928 Zahlentheoretische Abschätzungen mit Anwendungen auf Gitterpunktprobleme. *Math. Z.* **28**, 301–310.
van der Corput, J. G. 1935a Verteilungsfunktionen I. *Proc. Kon. Ned. Akad. v. Wetensch.* **38**, 813–821.
van der Corput, J. G. 1935b Verteilungsfunktionen II. *Proc. Kon. Ned. Akad. v. Wetensch.* **38**, 1058–1066.
van der Waerden, B. L. 1927 Beweis einer Baudetschen Vermutung. *Nieuw Arch. Wiskd.* **15**, 212–216
Vaughan, R. C. 1974 Diophantine approximation by prime numbers II. *Proc. Lond. Math. Soc.* **28**, 385–401.
Vaughan, R. C. 1980 A ternary additive problem. *Proc. Lond. Math. Soc.* **41**, 516–532.
Vaughan, R. C. 1989 A new iterative method in Waring's problem. *Acta Math.* **162**, 1–71.
Vaughan, R. C. 1997 *The Hardy–Littlewood method*. Cambridge University Press, *Tracts Math.* **125**.
Vaughan, R. C. 2008 On the number of partitions into primes. *Ramanujan J.* **15**, 109–121.
Vaughan, R. C. 2013 Some problems of 'Partitio Numerorum': hybrid expressions. In *The legacy of Srinivasa Ramanujan* (ed. B. C. Berndt & D. Prasad), pp. 363–385. Lecture Notes Series 20. Ramanujan Mathematical Society.
Vaughan, R. C. 2015 Squares: additive questions and partitions. *Int. J. Number Theory* **11**, 1367–1409.
Vinogradov, A. I. 1965 The density hypothesis for Dirichlet L-series. *Izv. Akad. Nauk SSSR Ser. Mat.* **29**, 903–934; corrigendum, **30**, 719–720.
Vinogradov, I. M. 1954 *The method of trigonometrical sums in the theory of numbers* (translated, revised and annotated by K. F. Roth & A. Davenport). London: Interscience.
Wooley, T. D. 1992 Large improvements in Waring's problem. *Ann. Math.* **135**, 131–164.
Wooley, T. D. 2003 On diophantine inequalities: Freeman's asymptotic formulae. In *Proceedings of the session in analytic number theory and diophantine equations* (ed. D. R. Heath-Brown & B. Z. Moroz), 32 pp. Bonner, *Mathematische Schriften* **360**.
Zhang, Y. 2014 Bounded gaps between primes. *Ann. Math.* **179**, 1121–1174.

Bibliography

The following publications are those referred to directly in the text. A full bibliography is available as electronic supplementary material at http://dx.doi.org/10.1098/rsbm.2017.0014 or via https://doi.org/10.6084/m9.figshare.c.3792406.

(1) 1947 A theorem involving squarefree numbers. *J. Lond. Math. Soc.* **22**, 231–237. (doi:10.1112/jlms/s1-22.3.231)
(2) 1949 Proof that almost all positive integers are sums of a square, a positive cube and a fourth power. *J. Lond. Math. Soc.* **24**, 4–13. (doi:10.1112/jlms/s1-24.1.4)
(3) 1951 On a problem of Heilbronn. *J. Lond. Math. Soc.* **26**, 198–204. (doi:10.1112/jlms/s1-26.3.198)

(4) On the gaps between squarefree numbers. *J. Lond. Math. Soc.* **26**, 263–268. (doi:10.1112/jlms/s1-26.4.263)
(5) (With H. Halberstam) On the gaps between consecutive k-free integers. *J. Lond. Math. Soc.* **26**, 268–273. (doi:10.1112/jlms/s1-26.4.268)
(6) On Waring's problem for cubes. *Proc. Lond. Math. Soc.* **53**, 268–279. (doi:10.1112/plms/s2-53.4.268)
(7) A problem in additive number theory. *Proc. Lond. Math. Soc.* **53**, 381–395. (doi:10.1112/plms/s2-53.5.381)
(8) 1952 Sur quelques ensembles d'entiers. *C. R. Acad. Sci. Paris* **234**, 388–390.
(9) (With R. P. Bambah) A note on lattice coverings. *J. Ind. Math. Soc. (N.S.)* **16**, 7–12.
(10) 1953 On certain sets of integers. *J. Lond. Math. Soc.* **28**, 104–109. (doi:10.1112/jlms/s1-28.1.104)
(11) 1954 On certain sets of integers II. *J. Lond. Math. Soc.* **29**, 20–26. (doi:10.1112/jlms/s1-29.1.20)
(12) On irregularities of distribution. *Mathematika* **1**, 73–79. (doi:10.1112/S0025579300000541)
(13) (With G. Szekeres) Some asymptotic formulae in the theory of partitions. *Q. J. Math. Oxf.* **5**, 241–259. (doi:10.1093/qmath/5.1.241)
(14) 1955 Rational approximations to algebraic numbers. *Mathematika* **2**, 1–20; corrigendum, 168. (doi:10.1112/S0025579300000644)
(15) (With H. Davenport) The solubility of certain diophantine inequalities. *Mathematika* **2**, 81–96. (doi:10.1112/S0025579300000723)
(16) (With H. Davenport) Rational approximation to algebraic numbers. *Mathematika* **2**, 160–167. (doi:10.1112/S0025579300000814)
(17) 1964 Remark concerning integer sequences. *Acta Arith.* **9**, 257–260.
(18) 1965 On the large sieves of Linnik and Rényi. *Mathematika* **12**, 1–9. (doi:10.1112/S0025579300005088)
(19) 1966 (With H. Halberstam) *Sequences*. Oxford: Clarendon Press.
(20) 1967 Irregularities of sequences relative to arithmetic progressions. *Math. Ann.* **169**, 1–25. (doi:10.1007/BF01399529)
(21) Irregularities of sequences relative to arithmetic progressions II. *Math. Ann.* **174**, 41–52. (doi:10.1007/BF01363122)
(22) 1970 Irregularities of sequences relative to arithmetic progressions III. *J. Number Theory* **2**, 125–142. (doi:10.1016/0022-314X(70)90013-2)
(23) 1972 Irregularities of sequences relative to arithmetic progressions IV. *Period. Math. Hung.* **2**, 301–326. (doi:10.1007/BF02018670)
(24) On a problem of Heilbronn II. *Proc. Lond. Math. Soc.* **25**, 193–212. (doi:10.1112/plms/s3-25.2.193)
(25) On a problem of Heilbronn III. *Proc. Lond. Math. Soc.* **25**, 543–549. (doi:10.1112/plms/s3-25.3.543)
(26) 1973 On cosine polynomials corresponding to sets of integers. *Acta Arith.* **24**, 87–98.
(27) 1978 (With R. C. Vaughan) Inefficiency in packing squares with unit squares. *J. Combin. Theory Ser. A* **24**, 170–186. (doi:10.1016/0097-3165(78)90005-5)
(28) 1979 On irregularities of distribution III. *Acta Arith.* **35**, 373–384.
(29) 1980 On irregularities of distribution IV. *Acta Arith.* **37**, 67–75.
(30) 2000 Limitations to regularity. In *Mathematics: frontiers and perspectives* (ed. V. I. Arnold, M. F. Atiyah, P. D. Lax & B. Mazur), pp. 235–250. Providence: American Mathematical Society.