10. SOME OF ROTH'S IDEAS IN DISCREPANCY THEORY

WILLIAM CHEN AND GIANCARLO TRAVAGLINI

Dedicated to Klaus Roth with our very best wishes

Abstract. We give a brief survey on some of the main ideas that Klaus Roth introduced into the study of irregularities of point distribution and, through a small selection of results, indicate how some of these ideas have been developed by him and others to obtain better understanding of this intriguing subject.

1. The Classical Problem

Although parts of it border on harmonic analysis, combinatorics and probability theory, irregularities of point distribution began as a branch of the theory of uniform distribution, and may sometimes be described as a quantitative form of the theory. It originated from a conjecture of van der Corput [20, 21] in 1935 that expresses the fact that no infinite sequence in $[0,1]$ can, in a certain sense, be too evenly distributed; see [6, page 3] for a precise statement of the conjecture. This was confirmed in 1945 by van Aardenne-Ehrenfest [1], who also gave a quantitative version [2] of this in 1949, with a relatively weak bound.

In 1954, Roth [30] showed that van Aardenne-Ehrenfest's quantitative version of the problem is equivalent to a geometric discrepancy problem concerning the distribution of a finite set of points in the unit square $[0,1]^2$. We shall now describe the multi-dimensional version of this geometric discrepancy problem. In the sequel, the letter k will denote a positive integer greater than 1.

Let $\mathcal{P}$ be a distribution of N points, not necessarily distinct, in the unit cube $[0,1]^k$. For any point $\mathbf{x}=(x_1,\ldots,x_k)$, let[1] $B(\mathbf{x})=[0,x_1)\times\ldots\times[0,x_k)$ denote the rectangular box anchored at the origin and with opposite vertex $\mathbf{x}$. Let $Z[\mathcal{P};B(\mathbf{x})]$ denote the number of points of $\mathcal{P}$ that lie in $B(\mathbf{x})$, and consider the discrepancy

$$D[\mathcal{P};B(\mathbf{x})]=Z[\mathcal{P};B(\mathbf{x})]-Nx_1\ldots x_k.$$

Roth showed that[2]

$$\int_{[0,1]^k}|D[\mathcal{P};B(\mathbf{x})]|^2\,\mathrm{d}\mathbf{x}\gg_k(\log N)^{k-1}, \tag{1}$$

[1] The assumption that the boxes $B(\mathbf{x})$ are half-open is introduced purely for convenience.

[2] Throughout, we adopt Vinogradov notation $\ll$ and $\gg$. For any two functions f and g, we write $f\ll g$ to denote $|f|\leqslant Cg$ for some positive absolute constant C. If f and g are non-negative, then we write $f\gg g$ to denote $f\geqslant Cg$ for some positive absolute constant C. Furthermore, we use the notation $\ll$ and $\gg$ with subscripts; in this case, the implicit constant C may then depend on those subscripts. Any deviation from this convention will be indicated beforehand.

from which it follows easily that

$$\sup_{\mathbf{x}\in[0,1]^k} |D[\mathcal{P}; B(\mathbf{x})]| \gg_k (\log N)^{(k-1)/2}. \tag{2}$$

Roth's deduction of the inequality (1) contains two crucial ideas. For the benefit of the reader, we illustrate them in the special case $k = 2$.

Trivial Discrepancy. Since the point set $\mathcal{P}$ is arbitrary, we have essentially no information about it, making it hard if not impossible to extract the discrepancy from those parts of the square $[0,1]^2$ that are near the points of $\mathcal{P}$. On the other hand, those parts of the square that are short of points of $\mathcal{P}$ give rise to "trivial discrepancies" which we then exploit. One can make parts of the square deficient of points of $\mathcal{P}$ in a very simple way. If we partition the square into more than $2N$ subsets, then given that there are only N points, at least half of these subsets are devoid of points of $\mathcal{P}$. More precisely, we can partition the square into similar rectangles of area 2^{-n}, where the integer n is chosen to satisfy $2N \leqslant 2^n < 4N$. Then there are 2^n such rectangles, at least half of which contain no points of $\mathcal{P}$. Roth then proceeded to extract the discrepancy from such "empty" rectangles.

A typical rectangle of area 2^{-n} under consideration is of the form

$$B = [m_1 2^{-r_1}, (m_1+1)2^{-r_1}) \times [m_2 2^{-r_2}, (m_2+1)2^{-r_2}) \subset [0,1]^2, \tag{3}$$

where m_1, m_2, r_1, r_2 are non-negative integers satisfying $r_1 + r_2 = n$. Consider the smaller rectangle of area 2^{-n-2} given by

$$B' = [m_1 2^{-r_1}, (m_1+\tfrac{1}{2})2^{-r_1}) \times [m_2 2^{-r_2}, (m_2+\tfrac{1}{2})2^{-r_2}) \subset [0,1]^2,$$

made up of the bottom left quarter of B. For any $\mathbf{x} = (x_1, x_2) \in B'$, the rectangle

$$B'(\mathbf{x}) = [x_1, x_1 + 2^{-r_1-1}) \times [x_2, x_2 + 2^{-r_2-1})$$

is similar to B' and contained in B, and so does not contain any point of $\mathcal{P}$ if neither does B. In this case, the rectangle $B'(\mathbf{x})$ has trivial discrepancy $N2^{-n-2}$.

A device to pick up this trivial discrepancy is given by the Rademacher function defined locally on such an empty rectangle B by

$$R_{r_1,r_2}(\mathbf{x}) = \begin{cases} +1 & \text{if } \mathbf{x} \in B', \\ -1 & \text{if } \mathbf{x} \in B' + (2^{-r_1-1}, 0), \\ -1 & \text{if } \mathbf{x} \in B' + (0, 2^{-r_2-1}), \\ +1 & \text{if } \mathbf{x} \in B' + (2^{-r_1-1}, 2^{-r_2-1}), \end{cases}$$

depending on which quadrant of B the point $\mathbf{x}$ falls into. Now

$$\int_B x_1 x_2 R_{r_1,r_2}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{B'} \sum_{\epsilon_1=0}^{1} \sum_{\epsilon_2=0}^{1} (-1)^{\epsilon_1+\epsilon_2} (x_1 + \epsilon_1 2^{-r_1-1})(x_2 + \epsilon_2 2^{-r_2-1})\,\mathrm{d}\mathbf{x}.$$

But then the integrand on the right hand side is equal to 2^{-n-2}, the area of the rectangle $B'(\mathbf{x})$. It follows that writing $D(\mathbf{x}) = D[\mathcal{P}; B(\mathbf{x})]$, we have

$$\int_B D(\mathbf{x}) R_{r_1,r_2}(\mathbf{x})\,\mathrm{d}\mathbf{x} = -N2^{-2n-4}.$$

Note that the function R_{r_1,r_2} has picked out the trivial discrepancy created by the rectangle $B'(\mathbf{x})$ not containing any point of $\mathcal{P}$.

Let us keep r_1 and r_2 fixed. For every rectangle B of the form (3), we write

$$f_{r_1,r_2}(\mathbf{x}) = \begin{cases} -R_{r_1,r_2}(\mathbf{x}) & \text{if } B \cap \mathcal{P} = \emptyset, \\ 0 & \text{if } B \cap \mathcal{P} \neq \emptyset. \end{cases}$$

Then

$$\int_B D(\mathbf{x}) f_{r_1,r_2}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \begin{cases} N2^{-2n-4} & \text{if } B \cap \mathcal{P} = \emptyset, \\ 0 & \text{if } B \cap \mathcal{P} \neq \emptyset. \end{cases}$$

Summing over all similar rectangles B, we obtain[3]

$$\int_{[0,1]^2} D(\mathbf{x}) f_{r_1,r_2}(\mathbf{x})\,\mathrm{d}\mathbf{x} = N2^{-2n-4} \#\{B : B \cap \mathcal{P} = \emptyset\} \gg 1, \tag{4}$$

since $2N \leqslant 2^n < 4N$ and at least 2^{n-1} rectangles B satisfy $B \cap \mathcal{P} = \emptyset$.

Orthogonality. But then there are $n+1$ choices of non-negative integers r_1 and r_2 satisfying $r_1 + r_2 = n$. Roth then proceeded to construct the auxiliary function

$$F(\mathbf{x}) = \sum_{\substack{r_1,r_2 \geqslant 0 \\ r_1+r_2=n}} f_{r_1,r_2}(\mathbf{x}). \tag{5}$$

Note that $F(\mathbf{x})$ depends on the distribution $\mathcal{P}$. The Cauchy—Schwarz inequality then gives

$$\left| \int_{[0,1]^2} D(\mathbf{x}) F(\mathbf{x})\,\mathrm{d}\mathbf{x} \right| \leqslant \left(\int_{[0,1]^2} |D(\mathbf{x})|^2\,\mathrm{d}\mathbf{x} \right)^{1/2} \left(\int_{[0,1]^2} |F(\mathbf{x})|^2\,\mathrm{d}\mathbf{x} \right)^{1/2}. \tag{6}$$

In view of (4) and (5), we clear have a lower bound

$$\int_{[0,1]^2} D(\mathbf{x}) F(\mathbf{x})\,\mathrm{d}\mathbf{x} \gg n+1,$$

so an upper bound of the form

$$\int_{[0,1]^2} |F(\mathbf{x})|^2\,\mathrm{d}\mathbf{x} \ll n+1 \tag{7}$$

will complete the proof of (1) in the case $k = 2$. Here we observe that the functions in the summand of (5) are orthogonal, in the sense that

$$\int_{[0,1]^2} f_{r_1',r_2'}(\mathbf{x}) f_{r_1'',r_2''}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0$$

for distinct pairs (r_1', r_2') and (r_1'', r_2'') satisfying $r_1' + r_2' = r_1'' + r_2'' = n$. The inequality (7) follows easily.

We remark that the lower bound (2) in the special case $k = 2$ was improved by Schmidt [35] in 1972 from $(\log N)^{1/2}$ to $\log N$; see also [6, Sections 4.1–4.2]. This is best possible, in view of an old result of Lerch [27]. However, Schmidt studied the problem in van Aardenne-Ehrenfest's formulation which is more combinatorial in nature. A

[3]We use $\#\mathcal{S}$ to denote the cardinality of a finite set $\mathcal{S}$.

different proof of Schmidt's result was obtained by Halász [23], using Roth's method with an auxiliary function of the form

$$H(\mathbf{x}) = \prod_{\substack{r_1,r_2\geqslant 0\\ r_1+r_2=n}} (1+\alpha f_{r_1,r_2}(\mathbf{x})) - 1,$$

where $\alpha > 0$ is a suitably chosen constant and the functions $f_{r_1,r_2}(\mathbf{x})$ are precisely those used by Roth. In other words, Halász further exploited the orthogonality of these functions.

Also, Schmidt [36] extended Roth's lower bound (1) to

$$\int_{[0,1]^k} |D[\mathcal{P}; B(\mathbf{x})]|^p \,\mathrm{d}\mathbf{x} \gg_{k,p} (\log N)^{(k-1)p/2}, \tag{8}$$

for every fixed exponent $p \in (1,\infty)$, using an auxiliary function of the form (5) but with the functions $f_{r_1,r_2}(\mathbf{x})$ defined slightly differently from those of Roth, but nevertheless still using orthogonality to ensure that higher moment analogues of the inequality (7) hold; see also [6, Lemma 2.4]. This enabled him to employ the Hölder inequality generalization of the Cauchy–Schwarz inequality (6).

Roth's orthogonal function technique also features in the various improvements to the lower bound (2), first in dimension $k = 3$ by Beck [5], and more recently in dimension $k = 3$ again by Bilyk and Lacey [8], and then in all dimensions $k \geqslant 3$ by Bilyk, Lacey and Vagharshakyan [9], the last with the bound

$$\sup_{\mathbf{x}\in[0,1]^k} |D[\mathcal{P}; B(\mathbf{x})]| \gg_k (\log N)^{(k-1)/2+\delta(k)}$$

for some constant $\delta(k) > 0$ depending only on the dimension k.

It has been conjectured [6, page 6] for a long time that the exponent above may be replaced by $k-1$. However, more detailed analysis recently suggests that perhaps the correct bound may instead be

$$\sup_{\mathbf{x}\in[0,1]^k} |D[\mathcal{P}; B(\mathbf{x})]| \gg_k (\log N)^{k/2}.$$

2. Measure and Geometry

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. The challenge here is then to understand how discrepancy is related to the geometry of these collections.

The difficulty here 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 $\mathcal{P}$ 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 $\mathcal{P}$ be a distribution of N points in $[0,1]^k$. Then an appropriate discrepancy function for the set A is given by

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

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

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

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

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

if, for simplicity, we make the further assumption[4] 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 distribution $\mathcal{P}$ at all, whereas the discrepancy measure $\mathrm{d}Z_0 - N\mathrm{d}\mu_0$ depends only on the distribution $\mathcal{P}$ and not on the set A at all. If we write $D(\mathbf{x}) = D[\mathcal{P}; A+\mathbf{x}]$, then

$$D = \chi_A * (\mathrm{d}Z_0 - N\mathrm{d}\mu_0). \tag{9}$$

Fourier Transform. In his famous study of integer sequences relative to long arithmetic progressions[5], Roth [31] established a result sometimes affectionately known as his 1/4-theorem, through the use of Fourier transform. This observation is the catalyst that propelled József Beck to arguably the most fascinating results in irregularities of distribution. Passing over to Fourier transform, the convolution (9) 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 distributions $\mathcal{P}$ 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 $\mathcal{A}$ with respect to some integral geometric measure. Like in the classical problem, one searches for trivial discrepancy, and seeks ways to blow them up.

For upper bounds, we have good information on the distributions $\mathcal{P}$, 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 amongst others the following 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 technical condition[6], and consider all similar copies $A(\lambda, \tau, \mathbf{x})$ obtained from A by contraction $\lambda \in [0,1]$, proper orthogonal transformation $\tau \in \mathcal{T}$ and translation $\mathbf{x} \in [0,1]^k$, where $\mathcal{T}$ denotes the group

[4]If we do not make this assumption, then we need to study instead the characteristic function of the image of A across the origin.

[5]For a nice account of this work, the reader is referred to the paper of Sárközy and Stewart [34] in this volume.

[6]The technical condition simply requires that A is not too thin; we omit the details here.

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

$$D[\mathcal{P}; A(\lambda,\tau,\mathbf{x})] = \#(\mathcal{P} \cap A(\lambda,\tau,\mathbf{x})) - N\mu(A(\lambda,\tau,\mathbf{x}))$$

denote the discrepancy of $\mathcal{P}$ in $A(\lambda,\tau,\mathbf{x})$. Beck [3] proved in 1987 the lower bound

$$\int_{[0,1]^k} \int_{\mathcal{T}} \int_0^1 |D[\mathcal{P}; A(\lambda,\tau,\mathbf{x})]|^2 \,\mathrm{d}\lambda \mathrm{d}\tau \mathrm{d}\mathbf{x} \gg_A N^{1-1/k}. \tag{10}$$

As there is only a very mild restriction on the set A and hence its similar copies, this result suggests that the large lower bound is due essentially entirely to the presence of the proper orthogonal transformations $\tau \in \mathcal{T}$.

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 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

$$D[\mathcal{P}; A(\lambda,\mathbf{x})] = \#(\mathcal{P} \cap A(\lambda,\mathbf{x})) - N\mu(A(\lambda,\mathbf{x}))$$

denote the discrepancy of $\mathcal{P}$ in $A(\lambda,\mathbf{x})$. Beck [4] proved in 1988 the lower bound

$$\int_{[0,1]^2} \int_0^1 |D[\mathcal{P}; A(\lambda,\mathbf{x})]|^2 \,\mathrm{d}\lambda \mathrm{d}\mathbf{x} \gg_A \max\{\log N, \xi_N^2(A)\}, \tag{11}$$

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 (11) should be compared to the estimate (1) in the classical problem with $k = 2$.

3. Some Upper Bounds

Not long after Roth's initial breakthrough in 1954, Davenport [22] showed in 1956 that Roth's lower bound (1) is best possible in the case $k = 2$. Davenport made use of the sequence of fractional parts $\{n\theta\}$, where θ is a badly approximable number like $\sqrt{2}$, and showed that the sawtooth function $\psi(x) = x - [x] - 1/2$, where $[x]$ denotes the largest integer not exceeding x, played a vital rôle in describing the discrepancy function. In his proof, Davenport encountered a technical difficulty which he eventually overcame by introducing an ingenious reflection argument. However, Davenport appeared to have missed, or at the very least ignored, a more natural way to overcome the technical difficulty that he encountered. But then[7] it took more than twenty years for anyone to realize that.

Periodicity and a Probabilistic Technique. Observing that the function $\psi(x)$ is periodic, Roth [32] introduced a probabilistic variable into Davenport's argument, and obtained an average of the square of the discrepancy function over an interval corresponding to the period of $\psi(x)$. This probabilistic variable, $t \in [0,1]$ say, is translation

[7]Perhaps one of the authors is getting a little mischievous here!

in nature, and gives rise to a translated distribution $\mathcal{P}(t)$ from the original distribution $\mathcal{P}$. Thus Roth was able to show that in dimension $k=2$, there is a distribution $\mathcal{P}$ of N points in $[0,1]^2$ such that

$$\int_0^1 \int_{[0,1]^2} |D[\mathcal{P}(t);B(\mathbf{x})]|^2 \,\mathrm{d}\mathbf{x}\mathrm{d}t \ll \log N,$$

from which one concludes that there exists $t^* \in [0,1]$ such that

$$\int_{[0,1]^2} |D[\mathcal{P}(t^*);B(\mathbf{x})]|^2 \,\mathrm{d}\mathbf{x} \ll \log N.$$

Of course, this is purely an existence proof, since we have no information on the value of t^*. Nevertheless, it provides an alternative proof of Davenport's result that the lower bound (1) is best possible in the case $k=2$. Indeed, using an elaboration of his idea, Roth was able to show in the same paper that the lower bound (1) is also best possible in the case $k=3$.

Unfortunately, neither Davenport's construction nor Roth's variant of it could be made to work in higher dimensions. Indeed, it would require the falsity of a conjecture of Littlewood concerning the existence, or otherwise, of a pair of irrational numbers ϑ and φ with the property[8] that $\nu\|\nu\vartheta\| \cdot \|\nu\varphi\|$ is bounded away from zero for all positive integers ν.

To show that the lower bound (1) is best possible for every $k \geqslant 2$, Roth made use of the famous van der Corput sequence. Consider the set of 2^n points in $[0,1]^2$, given in dyadic expansion by

$$(0.a_1 \ldots a_n, 0.a_n \ldots a_1), \quad a_1, \ldots, a_n \in \{0,1\}. \tag{12}$$

Such sets are very useful in showing that Schmidt's lower bound

$$\sup_{\mathbf{x}\in[0,1]^2} |D[\mathcal{P};B(\mathbf{x})]| \gg \log N$$

is best possible. On the other hand, multidimensional generalizations of these sets, using the Chinese remainder theorem, by Halton [24] are very useful in showing that the inequality[9]

$$\sup_{\mathbf{x}\in[0,1]^k} |D[\mathcal{P};B(\mathbf{x})]| \gg_k (\log N)^{k-1},$$

if true, would be best possible.

A natural first step would be to determine whether distributions $\mathcal{P}$ of $N=2^n$ points such as (12) in $[0,1]^2$ would satisfy an upper bound of the type

$$\int_{[0,1]^2} |D[\mathcal{P};B(\mathbf{x})]|^2 \,\mathrm{d}\mathbf{x} \ll n. \tag{13}$$

[8] See [12] for an account of this problem. Also, $\|x\|$ denotes the distance of x to the nearest integer.

[9] Present wisdom is that this inequality is too strong to be true. Perhaps the exponent should instead be $k/2$. See the discussion at the end of Section 1 on attempts from below.

However, with the set $\mathcal{P}$ of $N = 2^n$ points given by (12) in $[0,1]^2$, Halton and Zaremba [25] showed instead that[10]

$$\int_{[0,1]^2} |D[\mathcal{P}; B(\mathbf{x})]|^2 \, \mathrm{d}\mathbf{x} = 2^{-6} n^2 + O(n). \tag{14}$$

Probability and Quasi-Orthogonality. In 1980, Roth [33] finally succeeded in showing that the lower bound (1) is best possible for every $k \geqslant 2$. We shall briefly explain his ideas by concentrating on the case $k = 2$ and using the set $\mathcal{P}$ of $N = 2^n$ points given by (12). Let us consider a rectangle $B(\mathbf{x}) = B(x_1, x_2)$, where x_1 is an integer multiple of 2^{-n}, and let us keep x_1 fixed. Then we can write $D[\mathcal{P}; B(x_1, x_2)]$ in the form

$$D[\mathcal{P}; B(x_1, x_2)] = \sum_{i \in \mathcal{I}(x_1)} \left(a_i - \psi\left(\frac{x_2 + b_i}{2^{i-n}} \right) \right), \tag{15}$$

where $\mathcal{I}(x_1) \subset \{0, 1, \ldots, n\}$ is a collection of indices depending on x_1 only and the real numbers a_i, b_i are fixed. Since we need to study $|D[\mathcal{P}; B(x_1, x_2)]|^2$, the constants a_i give rise to terms of the form $a_{i'} a_{i''}$, where $i', i'' \in \mathcal{I}(x_1)$, and these ultimately lead to the term $2^{-6} n^2$ in (14). To make progress, Roth proceeded in a way which led ultimately to the removal of these problematic terms a_i. It is not difficult to check that the functions

$$\psi\left(\frac{x_2 + b_i}{2^{i-n}} \right), \quad i \in \mathcal{I}(x_1),$$

treated as functions of the single variable x_2, are *quasi-orthogonal*.

Roth introduced, as in his earlier work in [32], a translation variable $t \in [0,1]$, this time in the x_2-direction. This has the effect of translating the points in $\mathcal{P}$ in the x_2-direction modulo 1 to create sets $\mathcal{P}(t)$. Then

$$D[\mathcal{P}(t); B(x_1, x_2)] = \sum_{i \in \mathcal{I}(x_1)} \left(\psi\left(\frac{z_i + t}{2^{i-n}} \right) - \psi\left(\frac{w_i + t}{2^{i-n}} \right) \right), \tag{16}$$

where the real numbers z_i, w_i are fixed and depend on x_2. But now the functions

$$\psi\left(\frac{z_i + t}{2^{i-n}} \right) - \psi\left(\frac{w_i + t}{2^{i-n}} \right), \quad i \in \mathcal{I}(x_1),$$

are quasi-orthogonal in the new variable t. Squaring $D[\mathcal{P}(t); B(x_1, x_2)]$ and then integrating over $t \in [0,1]$, we note that the off-diagonal terms contribute no more than the diagonal terms, giving rise to an inequality of the form

$$\int_0^1 |D[\mathcal{P}(t); B(x_1, x_2)]|^2 \, \mathrm{d}t \ll \#(\mathcal{I}(x_1)) \ll n.$$

This leads eventually to an upper bound of the desired form in the case $k = 2$.

There are other ways that one can overcome the difficulty created by those problematic terms a_i in (15); for a brief description of these techniques, see the survey article [16, Section 5]. However, they do not all take advantage of the periodicity of the sawtooth function. For instance, in Chen [14], digit shifts were introduced, taking advantage instead of some group-like structure in the sets of the form (12) and many of their generalizations. We shall return to this in Section 5.

[10] Halton and Zaremba actually calculated the precise value of the integral!

Using different constructions and probabilistic techniques involving translation, Skriganov [37, 38] was also able to show that the lower bound (1) is best possible for every $k \geqslant 2$.

Using an elaboration of Roth's ideas in [33] that involves substantial generalization of the point sets under consideration, Chen [13] was able to show that Schmidt's lower bound (8) is best possible for every $k \geqslant 2$ and every fixed exponent $p \in (1, \infty)$; an alternative proof was also given by Chen [14], using digit shifts.

To have a better understanding of some of these probabilistic techniques, we shall return to these problems in Section 5, where we shall consider explicit constructions of good point sets and also briefly discuss their relationship with probabilistic techniques that involve digit shifts.

4. Other Probabilistic and Fourier Analytic Techniques

In the previous section, we discussed probabilistic techniques for obtaining upper bounds in cases of small discrepancy, of order equal to powers of $\log N$, where N is the number of points. Now we turn our attention to upper bounds in cases of large discrepancy, of order a power of N.

We begin with the lower bound (10) concerning similar copies of a given compact and convex set in the torus $[0,1]^k$. In particular, it follows from (10) that

$$\sup_{\substack{\lambda\in[0,1]\\ \tau\in\mathcal{T}\\ \mathbf{x}\in[0,1]^k}} |D[\mathcal{P}; A(\lambda,\tau,\mathbf{x})]| \gg_A N^{1/2-1/2k}. \tag{17}$$

For simplicity, let us assume that $N = M^k$ for some integer M. We can partition the torus $[0,1]^k$ into M^k small cubes in a natural way, and place a random point in each of these small cubes. We ensure that each random point is uniformly distributed within its own small cube, and independent of the distribution of the random points in other small cubes. Using a probabilistic model of this type, Beck was able to use large deviation type results in probability theory to show that the lower bound (17) is sharp to within a factor no more than $(\log N)^{1/2}$. Moreover, the technique gives a very good explanation for the exponent $1/2 - 1/2k$ in the bounds. Discrepancy of good distributions tend to occur near the boundary of sets. The quantity $M^{k-1} = N^{1-1/k}$ is the order of magnitude of the number of little cubes that intersect the boundary surface of a set $A(\lambda,\tau,\mathbf{x})$, and so gives a trivial upper bound for the supremum of the discrepancy.

More detailed and careful analysis of the construction turns out to be sufficient to show that the lower bound (10) is best possible for all $k \geqslant 2$, as shown by Beck and Chen [7]. For a simpler proof and a stronger result, see [15].

Besides the probabilistic argument, one can also show that (10) is best possible for all $k \geqslant 2$ by using a deterministic point set $\mathcal{P} = N^{-1/k}\mathbb{Z}^k \cap [0,1)^k$ when $N = M^k$ where M is an integer. Then periodicity and a change of variables lead us to a lattice point problem in $\mathbb{Z}^k$, namely an L^2 upper bound, originally studied by Kendall [26], for the function

$$\mathbf{x} \mapsto D^*(\mathbf{x},\tau) = \#(\mathbb{Z}^k \cap N^{1/k} A(1,\tau,\mathbf{x})),$$

which has Fourier series

$$N \sum_{0\neq\mathbf{l}\in\mathbb{Z}^k} \widehat{\chi}_A(N^{1/k}\tau(\mathbf{l}))\mathrm{e}^{2\pi\mathrm{i}\mathbf{l}\cdot\mathbf{x}}.$$

We have

$$\int_{[0,1]^k}\int_{\mathcal{T}} |D^*(\mathbf{x},\tau)|^2 \,\mathrm{d}\tau\mathrm{d}\mathbf{x} = N^2 \sum_{0\neq\mathbf{l}\in\mathbb{Z}^k} \int_{\mathcal{T}} |\widehat{\chi}_A(N^{1/k}\tau(\mathbf{l}))|^2 \,\mathrm{d}\tau$$
$$\ll_A N^2 \sum_{0\neq\mathbf{l}\in\mathbb{Z}^k} N^{1-1/k}|\mathbf{l}|^{-k-1} = cN^{1-1/k},$$

by the Parseval identity and the L^2 average estimate [11]

$$\int_{\mathcal{T}} |\widehat{\chi}_A(\tau(\boldsymbol{\xi}))|^2 \,\mathrm{d}\tau \ll_A (1+|\boldsymbol{\xi}|)^{-k-1}, \quad \boldsymbol{\xi}\in\mathbb{R}^k.$$

The average decay of the Fourier transform of χ_A, with respect to rotation and possibly to contraction, is a basic ingredient in the argument for the probabilistic upper bounds as well as for lower bounds such as (10) and others that were established by Beck [3] and by Montgomery [28]. We may say that the study of the L^2 discrepancy for contracted, rotated and translated copies of a given convex set is essentially equivalent to the study of the L^2 average decay of the Fourier transform of the characteristic function of the set; see [10, 41].

The two choices of distributions $\mathcal{P}$ in $[0,1]^k$, of random and of deterministic points in small cubes of volume N^{-1} when $N = M^k$ where M is an integer, which we have used to establish upper bounds complementing the lower bound (10), can be seen as particular cases of the following construction, which we now describe in the special case when A is a ball in $[0,1]^k$. Let $N = M^k$, where M is an integer. Write

$$\mathcal{P}_N = \{\mathbf{p}_1,\ldots,\mathbf{p}_N\} = N^{-1/k}\mathbb{Z}^k \cap [0,1)^k,$$

and let $\mathrm{d}\mu$ be a probability measure on $[0,1]^k$, treated as a torus. For every $j = 1,\ldots,N$, let $\mathrm{d}\mu_j$ denote the translation of $\mathrm{d}\mu$ by one of the points $\mathbf{p}_j$ in $\mathcal{P}_N$, so that for any integrable function f on $[0,1]^k$, we have

$$\int_{[0,1]^k} f(\mathbf{t})\,\mathrm{d}\mu_j = \int_{[0,1]^k} f(\mathbf{t}-\mathbf{p}_j)\,\mathrm{d}\mu.$$

We now average the discrepancy function $D[\mathcal{P}; A+\mathbf{x}]$ in $L^2([0,1]^k, \mathrm{d}\mu_j)$ for every $j = 1,\ldots,N$, defining

$$D^2_{\mathrm{d}\mu}[\mathcal{P}; A] = \int_{[0,1]^k}\cdots\int_{[0,1]^k}\int_{[0,1]^k} |D[\mathcal{P}; A+\mathbf{x}]|^2 \,\mathrm{d}\mathbf{x}\mathrm{d}\mu_1\ldots\mathrm{d}\mu_N.$$

Note that we have not considered contractions here and that, since A is a ball, we need not consider orthogonal transformations. If we take $\mathrm{d}\mu = \delta_0$, the Dirac measure concentrated at $\mathbf{0}$, then we have the discrepancy for the deterministic point set $\mathcal{P} = \mathcal{P}_N$. On the other hand, if we take $\mathrm{d}\mu = \mathrm{d}\lambda$, where $\mathrm{d}\lambda = N\chi_{[0,N^{-1/k}]^k}\,\mathrm{d}\mu_0$, the normalized uniform measure in one of the small cubes we have described, then we are considering the random points in the small cubes. Of course both $D^2_{\delta_0}[\mathcal{P}; A]$ and $D^2_{\mathrm{d}\lambda}[\mathcal{P}; A]$ have the same order of growth $N^{1-1/k} = M^{k-1}$ with respect to $N = M^k$, but it is not true that one is always greater than the other. Indeed a Fourier analytic argument shows that for small $k \not\equiv 1 \bmod 4$, we have

$$D^2_{\delta_0}[\mathcal{P}; A] < D^2_{\mathrm{d}\lambda}[\mathcal{P}; A] \quad \text{for large values of } M,$$

while for large $k \not\equiv 1 \bmod 4$, we have

$$D^2_{\mathrm{d}\lambda}[\mathcal{P};A] < D^2_{\delta_0}[\mathcal{P};A] \quad \text{for large values of } M.$$

When $k \equiv 1 \bmod 4$, the situation becomes much more complicated, caused by the unusual distribution of lattice points in balls in euclidean space $\mathbb{R}^k$ for these dimensions k; see [19] for the discrepancy result and [29] for a discussion on the distribution of lattice points in euclidean space.

5. Fourier–Walsh Analysis

To show that the inequality (1) is essentially best possible, we need to find distributions $\mathcal{P}$ of N points in $[0,1]^k$ such that the upper bound

$$\int_{[0,1]^k} |D[\mathcal{P};B(\mathbf{x})]|^2 \,\mathrm{d}\mathbf{x} \ll_k (\log N)^{k-1} \tag{18}$$

holds. Roth's proof of this upper bound, as well as subsequent attempts by Chen and Skriganov, amongst others, are all probabilistic in nature and do not give rise to an explicit point set $\mathcal{P}$ satisfying (18).

Recall that near the end of Section 3, we mentioned group-like structure in the sets of the form (12) and many of their generalizations. Indeed, the set (12) is isomorphic to the group $\mathbb{Z}_2^k$, so it is natural to study its distribution via the characters of such groups, and these are the classical Walsh functions, defined on $[0,1]$ and which can be generalized to $[0,1]^k$ in a natural way.

On the other hand, if we replace $\mathbb{Z}_2$ by a finite field $\mathbb{Z}_p$ for some prime p, then Skriganov [39] has shown that point distributions that possess the structure of vector spaces over $\mathbb{Z}_p$ are distributed very uniformly in the cube $[0,1]^k$ with respect to the supremum norm of the discrepancy function, provided that the corresponding vector spaces have large weights relative to a special metric. Chen and Skriganov [17] then extended this argument and showed that such point sets in fact satisfy the upper bound (18), provided that the corresponding vector spaces have large weights simultaneously relative to two special metrics, the well known Hamming metric and a new non-Hamming metric arising recently in coding theory. Furthermore, they showed that the large weights can be guaranteed if the prime p is taken large enough to satisfy $p \geqslant 2k^2$. The problem was studied later in greater detail by Chen and Skriganov [18].

To study the discrepancy of such distributions, one naturally appeals to the corresponding characters. These are the Walsh functions base p, also known as Chrestenson or Chrestenson–Levy functions if $p > 2$. They are defined on $[0,1]$ and take as their values p–th roots of unity. Furthermore, their definitions can be generalized to $[0,1]^k$ in a natural way. For a fixed p, the collection of all Walsh functions in $[0,1]^k$ form an orthonormal basis for $L^2([0,1]^k)$, enabling us to do Fourier–Walsh analysis on $[0,1]^k$.

Consider a distribution $\mathcal{P}$ of $N = p^n$ points in $[0,1]^k$ which possesses the vector space structure that we mentioned earlier. Chen and Skriganov showed that a good approximation of the discrepancy function $D[\mathcal{P};B(\mathbf{x})]$ is given by some function

$$E[\mathcal{P};B(\mathbf{x})] = N \sum_{\mathbf{l}\in\mathcal{L}} \phi_{\mathbf{l}}(\mathbf{x}),$$

where $\mathcal{L}$ is a finite set depending on $\mathcal{P}$ and each term $\phi_{\mathbf{l}}(\mathbf{x})$ is a product of certain coefficients of the Fourier–Walsh series of the characteristic function of intervals of the type $[0, x_i)$, where $\mathbf{x} = (x_1, \ldots, x_k)$. In [17], it was shown that if p is sufficiently large, then the corresponding set $\mathcal{L}$ gives rise to a collection of functions $\phi_{\mathbf{l}}(\mathbf{x})$ that are quasi-orthogonal. In [18], it was shown that under the same condition for p, this collection of functions is in fact orthogonal, so that

$$\int_{[0,1]^k} |E[\mathcal{P}; B(\mathbf{x})]|^2 \, \mathrm{d}\mathbf{x} = N^2 \sum_{\mathbf{l} \in \mathcal{L}} \int_{[0,1]^k} |\phi_{\mathbf{l}}(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x}. \tag{19}$$

On the other hand, if p is not large enough to guarantee the orthogonality of the collection of functions $\phi_{\mathbf{l}}(\mathbf{x})$, then consider a group $\mathcal{T}$ of digit shifts $\mathbf{t}$, as for instance in Chen [14]. Then it was shown in [18] that for each modified point set $\mathcal{P} \oplus \mathbf{t}$ under a digit shift $\mathbf{t}$, we have

$$E[\mathcal{P} \oplus \mathbf{t}; B(\mathbf{x})] = N \sum_{\mathbf{l} \in \mathcal{L}} \overline{W_{\mathbf{l}}(\mathbf{t})} \phi_{\mathbf{l}}(\mathbf{x}),$$

where $W_{\mathbf{l}}(\mathbf{t})$ is a k-dimensional Walsh function. It follows easily that

$$\sum_{\mathbf{t} \in \mathcal{T}} |E[\mathcal{P} \oplus \mathbf{t}; B(\mathbf{x})]|^2 = N^2 \sum_{\mathbf{l}', \mathbf{l}'' \in \mathcal{L}} \left(\sum_{\mathbf{t} \in \mathcal{T}} \overline{W_{\mathbf{l}'}(\mathbf{t})} W_{\mathbf{l}''}(\mathbf{t}) \right) \phi_{\mathbf{l}'}(\mathbf{x}) \overline{\phi_{\mathbf{l}''}(\mathbf{x})}.$$

By the orthogonality property

$$\sum_{\mathbf{t} \in \mathcal{T}} \overline{W_{\mathbf{l}'}(\mathbf{t})} W_{\mathbf{l}''}(\mathbf{t}) = \begin{cases} \#(\mathcal{T}) & \text{if } \mathbf{l}' = \mathbf{l}'', \\ 0 & \text{otherwise}, \end{cases} \tag{20}$$

of the Walsh functions, we conclude that

$$\frac{1}{\#(\mathcal{T})} \sum_{\mathbf{t} \in \mathcal{T}} |E[\mathcal{P} \oplus \mathbf{t}; B(\mathbf{x})]|^2 = N^2 \sum_{\mathbf{l} \in \mathcal{L}} |\phi_{\mathbf{l}}(\mathbf{x})|^2. \tag{21}$$

In other words, if the orthogonality property (19) is not satisfied, then one can still achieve orthogonality by probabilistic techniques and create orthogonality (21) by importing it from the orthogonality property (20) that is actually present in the construction.

We conclude by mentioning that Skriganov [40] has succeeded in taking these ideas further to obtain an explicit construction $\mathcal{P}$ that shows that Schmidt's lower bound (8) is best possible for every $k \geqslant 2$ and every fixed exponent $p \in (1, \infty)$, superseding the earlier probabilistic efforts of Chen [13, 14].

References

[1] T. van Aardenne-Ehrenfest. Proof of the impossibility of a just distribution of an infinite sequence of points over an interval. *Proc. Kon. Ned. Akad. v. Wetensch.*, 48 (1945), 266–271.

[2] T. van Aardenne-Ehrenfest. On the impossibility of a just distribution. *Proc. Kon. Ned. Akad. v. Wetensch.*, 52 (1949), 734–739.

[3] J. Beck. Irregularities of distribution I. *Acta Math.*, 159 (1987), 1–49.

[4] J. Beck. Irregularities of distribution II. *Proc. London Math. Soc.*, 56 (1988), 1–50.

[5] J. Beck. A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution. *Compos. Math.*, 72 (1989), 269–339.

[6] J. Beck, W.W.L. Chen. *Irregularities of distribution* (Cambridge Tracts in Mathematics 89, Cambridge University Press, 1987).

[7] J. Beck, W.W.L. Chen. Note on irregularities of distribution II. *Proc. London Math. Soc.*, 61 (1990), 251–272.

[8] D. Bilyk, M.T. Lacey. On the small ball inequality in three dimensions. *Duke Math. J.*, 143 (2008), 81–115.

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

[10] L. Brandolini, L. Colzani, G. Travaglini. Average decay of Fourier transforms and integer points in polyhedra. *Ark. Mat.*, 35 (1997), 253–275.

[11] L. Brandolini, S. Hofmann, A. Iosevich. Sharp rate of average decay of the Fourier transform of a bounded set. *Geom. Funct. Anal.*, 13 (2003), 671–680.

[12] J.W.S. Cassels, H.P.F. Swinnerton-Dyer. On the product of three homogeneous linear forms and indefinite ternary quadratic forms. *Philos. Trans. R. Soc. London Ser. A*, 248 (1955), 73–96.

[13] W.W.L. Chen. On irregularities of distribution. *Mathematika*, 27 (1980), 153–170.

[14] W.W.L. Chen. On irregularities of distribution II. *Q. J. Math.*, 34 (1983), 257–279.

[15] W.W.L. Chen. On irregularities of distribution IV. *J. Number Theory*, 80 (2000), 44–59.

[16] W.W.L. Chen. Fourier techniques in the theory of irregularities of point distribution. *Fourier Analysis and Convexity* (L. Brandolini, L. Colzani, A. Iosevich, G. Travaglini, eds.), pp. 59–82 (Applied and Numerical Harmonic Analysis, Birkhäuser-Verlag, 2004).

[17] W.W.L. Chen, M.M. Skriganov. Explicit constructions in the classical mean squares problem in irregularities of point distribution. *J. reine angew. Math.*, 545 (2002), 67–95.

[18] W.W.L. Chen, M.M. Skriganov. Orthogonality and digit shifts in the classical mean squares problem in irregularities of point distribution. *Diophantine Approximation: Festschrift for Wolfgang Schmidt* (H.-P. Schlickewei, K. Schmidt, R.F. Tichy, eds.), pp. 141–159 (Developments in Mathematics 16, Springer-Verlag, 2008).

[19] W.W.L. Chen, G. Travaglini. Deterministic and probabilistic discrepancies. *Ark. Mat.* (to appear).

[20] J.G. van der Corput. Verteilungsfunktionen I. *Proc. Kon. Ned. Akad. v. Wetensch.*, 38 (1935), 813–821.

[21] J.G. van der Corput. Verteilungsfunktionen II. *Proc. Kon. Ned. Akad. v. Wetensch.*, 38 (1935), 1058–1066.

[22] H. Davenport. Note on irregularities of distribution. *Mathematika*, 3 (1956), 131–135.

[23] G. Halász. On Roth's method in the theory of irregularities of point distributions. *Recent Progress in Analytic Number Theory*, volume 2 (H. Halberstam, C. Hooley, eds.), pp. 79–94 (Academic Press, 1981).

[24] J.H. Halton. On the efficiency of certain quasirandom sequences of points in evaluating multidimensional integrals. *Numer. Math.*, 2 (1960), 84–90.

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

[26] D.G. Kendall. On the number of lattice points in a random oval. *Q. J. Math.*, 19 (1948), 1–26.

[27] M. Lerch. Question 1547. *L'Intermediaire Math.*, 11 (1904), 144–145.

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

[29] L. Parnovski, A.V. Sobolev. On the Bethe–Sommerfeld conjecture for the polyharmonic operator. *Duke Math. J.*, 107 (2001), 209–223.

[30] K.F. Roth. On irregularities of distribution. *Mathematika*, 1 (1954), 73–79.

[31] K.F. Roth. Remark concerning integer sequences, *Acta Arith.*, 9 (1964), 257–260.

[32] K.F. Roth. On irregularities of distribution III. *Acta Arith.*, 35 (1979), 373–384.

[33] K.F. Roth. On irregularities of distribution IV. *Acta Arith.*, 37 (1980), 67–75.

[34] A. Sárközy, C.L. Stewart. Irregularities of sequences relative to long arithmetic progressions. *Analytic Number Theory: Essays in Honour of Klaus Roth* (W.W.L. Chen, W.T.Gowers, H. Halberstam, W.M. Schmidt, R.C. Vaughan, eds.), pp. 389–401 (Cambridge University Press, 2008).

[35] W.M. Schmidt. Irregularities of distribution VII. *Acta Arith.*, 21 (1972), 45–50.

[36] W.M. Schmidt. Irregularities of distribution X. *Number Theory and Algebra* (H. Zassenhaus, ed.), pp. 311–329 (Academic Press, 1977).

[37] M.M. Skriganov. Lattices in algebraic number fields and uniform distribution modulo 1. *Algebra i Analiz*, 1 (1989), 207–228; English translation in *Leningrad Math. J.*, 1 (1990), 535–558.

[38] M.M. Skriganov. Constructions of uniform distributions in terms of geometry of numbers. *Algebra i Analiz*, 6 (1994), 200–230; reprinted in *St. Petersburg Math. J.*, 6 (1995), 635–664.

[39] M.M. Skriganov. Coding theory and uniform distributions. *Algebra i Analiz*, 13 (2001), 220–273; English translation in *St. Petersburg Math. J.*, 13 (2002), 301–337.

[40] M.M. Skriganov. Harmonic analysis on totally disconnected groups and irregularities of point distributions. *J. reine angew. Math.*, 600 (2006), 25–49.

[41] G. Travaglini. Average decay of the Fourier transform. *Fourier Analysis and Convexity* (L. Brandolini, L. Colzani, A. Iosevich, G. Travaglini, eds.), pp. 245–268 (Applied and Numerical Harmonic Analysis , Birkhäuser-Verlag, 2004).