IRREGULARITIES OF POINT DISTRIBUTION

WILLIAM CHEN

Survey given at the Workshop on
Small Ball Inequalities in Analysis, Probability Theory, and Irregularities of Distribution,
American Institute of Mathematics,
8 December 2008

1. Brief History

The theory of irregularities of point distribution can be thought of as a quantitative version of the theory of uniform distribution, and is commonly thought to have originated from the work of van der Corput [13, 14] in 1935.

Conjecture (van der Corput)**.** *Suppose that $s_1, s_2, s_3, \ldots$ is infinite sequence in $[0,1]$. Then given any positive real number κ, there exist a positive integer n and subintervals $I, J \subset [0,1]$ of equal length such that*

$$\#(\{s_1, \ldots, s_n\} \cap I) - \#(\{s_1, \ldots, s_n\} \cap J) > \kappa.$$

Here, think of κ as a prescribed error. The conjecture then says that any such prescribed error can be exceeded by a suitable truncation of the infinite sequence. In other words, no infinite sequence in $[0,1]$ can be too evenly distributed.

The conjecture was proved in 1945 by van Aardenne-Ehrenfest [1]. In 1949, she established the following quantitative version of the result.

Theorem ([2])**.** *Suppose that $s_1, \ldots, s_N \in [0,1]$, where the integer N is sufficiently large. Then there exist $n \in \{1, \ldots, N\}$ and $\alpha \in [0,1]$ such that*

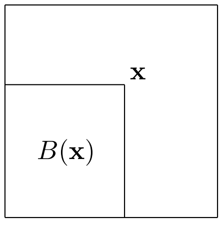
$$|\#(\{s_1, \ldots, s_n\} \cap [0, \alpha]) - n\alpha| \gg g(N), \tag{1}$$

where the function $g(N)$ can be taken to be

$$g(N) = \frac{\log \log N}{\log \log \log N}.$$

In 1954, Roth studied the problem in yet another formulation.

Theorem ([18])**.** *Suppose that $\mathcal{P}$ is a distribution of N points in the unit square $[0,1)^2$, where the integer N is sufficiently large. For every $\mathbf{x} = (x_1, x_2) \in [0,1]^2$, consider the rectangle $B(\mathbf{x}) = [0, x_1) \times [0, x_2)$*



and the discrepancy function $D[\mathcal{P}; B(\mathbf{x})] = \#(\mathcal{P} \cap B(\mathbf{x})) - N x_1 x_2$. Then

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

where the function $f(N)$ can be taken to be $f(N) = (\log N)^{\frac{1}{2}}$.

Furthermore, Roth showed that these two formulations are equivalent. More precisely, if the inequality (1) holds for some function $g(N)$, then the inequality (2) holds for any function $f(N) \asymp g(N)$. Conversely, if the inequality (2) holds for some function $f(N)$, then the inequality (1) holds for any function $g(N) \asymp f(N)$.

Note also that Roth's formulation gives us symmetry between the variables, as x_1 and x_2 now perform the roles of n (more precisely n/N) and α. It also introduces geometry into the subject, and we shall exploit this later in Section 3.

2. The Classical Problem

Roth's formulation immediately lends itself to generalization to higher dimensions. Let the dimension $K \geqslant 2$ be fixed, and let $\mathcal{P}$ be a distribution of N points in the unit cube $[0,1)^K$. For every $\mathbf{x} = (x_1, \ldots, x_K) \in [0,1]^K$, consider the rectangular box $B(\mathbf{x}) = [0, x_1) \times \ldots \times [0, x_K)$ and the discrepancy function

$$D[\mathcal{P}; B(\mathbf{x})] = \#(\mathcal{P} \cap B(\mathbf{x})) - N x_1 \ldots x_K.$$

Let

$$\|D[\mathcal{P}]\|_\infty = \sup_{\mathbf{x} \in [0,1]^K} |D[\mathcal{P}; B(\mathbf{x})]|,$$

and for any fixed real number $W \in (0, \infty)$, let

$$\|D[\mathcal{P}]\|_W = \left(\int_{[0,1]^K} |D[\mathcal{P}; B(\mathbf{x})]|^W \, \mathrm{d}\mathbf{x} \right)^{\frac{1}{W}}.$$

We want to say that all distributions are bad, and that some distributions have reasonably small badness. Hence we are interested in obtaining lower and upper bounds to the quantities

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \qquad \text{and} \qquad \inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_W.$$

We first state the "old" estimates that apply to all dimensions.

The first lower bound was obtained in 1954 by Roth, who also established in 1980 the corresponding upper bound.

Theorem ([18, 20])**.** *We have*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_2 \asymp_K (\log N)^{\frac{K-1}{2}}.$$

Corollary. *We have*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \gg_K (\log N)^{\frac{K-1}{2}}. \tag{3}$$

This corollary is not best possible. Furthermore, the corresponding upper bound currently known, due to Halton, appears to be rather weak – indeed, the proof is rather easy. The case $K = 2$ was established in 1904 by Lerch [17].

Theorem ([16])**.** *We have*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \ll_K (\log N)^{K-1}. \tag{4}$$

A lower bound in the case $W > 1$ was obtained by Schmidt in 1977, and this was shown to be best possible by Chen in 1980.

Theorem ([23, 10])**.** *For every fixed real number $W > 1$, we have*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_W \asymp_{K,W} (\log N)^{\frac{K-1}{2}}.$$

In summary, the cases when $1 < W < \infty$ are rather well understood.

For the case $W = 1$, the following result by Halász is sharp when $K = 2$.

Theorem ([15])**.** *We have*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_1 \gg_K (\log N)^{\frac{1}{2}}.$$

We next turn our attention to the special case $K = 2$, where a lower bound superior to that given in (3) was established in 1972 by Schmidt, who studied the problem in the earlier fomulation by van Aardenne-Ehrenfest and showed that the inequality (1) holds with $g(N) = \log N$. In view of the upper bound (4) given by Lerch, we have the following result.

Theorem ([21])**.** *For the special case $K = 2$, we have*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \asymp \log N.$$

In 1981, Halász [15] obtained an alternative proof of the lower bound, using a variation of Roth's technique in [18].

In summary, the special case $K = 2$ is rather well understood.

When the dimension $K > 2$, we can summarize our results stated thus far as follows. We have

$$(\log N)^{\frac{K-1}{2}} \ll_K \inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \ll_K (\log N)^{K-1},$$

and

$$(\log N)^{\frac{1}{2}} \ll_K \inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_1 \ll_K (\log N)^{\frac{K-1}{2}}.$$

For the special case $K = 3$, we have the following improvement due to Beck in 1989.

Theorem ([6])**.** *For the special case $K = 3$, we have, for any fixed $\delta > 0$,*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \gg_\delta (\log N)(\log\log N)^{\frac{1}{8}-\delta}.$$

This was improved very recently by Bilyk and Lacey.

Theorem ([8])**.** *For the special case $K = 3$, there exists $\eta > 0$ such that*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \gg (\log N)^{1+\eta}.$$

The ideas were extended to all dimensions by Bilyk, Lacey, and Vagharshakyan.

Theorem ([9])**.** *For every dimension $K \geqslant 3$, there exists $\eta(K) > 0$ such that*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \gg_K (\log N)^{\frac{K-1}{2}+\eta(K)}.$$

This last result is currently the best lower bound known for any dimension $K \geqslant 3$. The open question here is whether $\eta(K) = \frac{1}{2}$ is admissible, like when $K = 2$.

3. Generalization of the Classical Problem

Recall our last comment in Section 1, that Roth[1] brought geometry into the subject, and provided an opportunity to generalize the classical problem. This was first undertaken by Schmidt who in the late 1960s and early 1970s wrote a series of beautiful papers on the subject, and his effort was followed by Beck in the 1980s. The questions mostly take the following line.

Let $\mathcal{P}$ be a distribution of N points in torus $[0, 1)^K$, where the dimension $K \geqslant 2$ is fixed and the natural number N is sufficiently large. Let $\mathcal{A}$ be an infinite collection

[1]In a recent conversation I had with Roth, he offered the comment that he rated this above all his other mathematical achievements. Not only did he solve a difficult and beautiful problem, but he pushed the subject in a direction that opened it up to many interesting and deep questions, and provided others with the impetus to pursue this very intriguing area of research.

of measurable subsets $A \subset [0,1)^K$, endowed with an integral-geometric measure normalized such that the total measure is equal to 1. For any such measurable set $A \in \mathcal{A}$, consider the discrepancy function

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

Let

$$\|D[\mathcal{P}]\|_\infty = \sup_{A \in \mathcal{A}} |D[\mathcal{P};A]|,$$

and for any fixed real number $W \in (0,\infty)$, let

$$\|D[\mathcal{P}]\|_W = \left(\int_{\mathcal{A}} |D[\mathcal{P};A]|^W \, \mathrm{d}A \right)^{\frac{1}{W}}.$$

We now study the quantities

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \qquad \text{and} \qquad \inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_W.$$

4. Blowing Up the Trivial Error

Most lower bound arguments make use of the "trivial errors", always small, that occur frequently and seek ways to combine them in efficient ways to obtain a large discrepancy. We shall give a quick illustration by discussing[2] Schmidt's chocolate cake theorem.

Let $\mathcal{P}$ be a distribution of N points in $[0,1)^2$. Consider the collection $\mathcal{A}$ of all convex subsets of $[0,1]^2$. For any convex set $A \in \mathcal{A}$, consider the discrepancy function $D[\mathcal{P};A] = \#(\mathcal{P} \cap A) - N\mu(A)$, and let

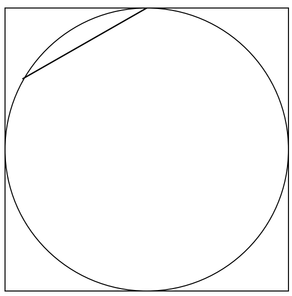
$$\|D[\mathcal{P}]\|_\infty = \sup_{A \in \mathcal{A}} |D[\mathcal{P};A]|.$$

Theorem ([22])**.** *We have*

$$\inf_{\#\mathcal{P}=N} \|D[\mathcal{P}]\|_\infty \gg N^{\frac{1}{3}}.$$

We comment that the exponent $\frac{1}{3}$ is best possible, as shown by Beck [5].

Proof. Imagine that $[0,1]^2$ is a square plate and A is a circular cake of diameter 1. The plate and cake are decorated with a collection $\mathcal{P}$ of N chocolates.



The picture above also shows a disc-segment of area $\frac{1}{2N}$. Elementary calculation shows that we can mark out $\gg N^{\frac{1}{3}}$ non-overlapping disc-segments of this kind, and each should contain precisely half a chocolate, but of course each contains none or at least one.

[2]Schmidt proved the theorem. I added the chocolate interpretation – Schmidt *loves* chocolates.

Now let $B_1, \ldots, B_s$ denote those disc-segments that have no chocolates, and let $C_1, \ldots, C_t$ denote those disc-segments with chocolates. Clearly $s + t \gg N^{\frac{1}{3}}$. Furthermore, we have the trivial errors

$$D[\mathcal{P}; B_i] = -\tfrac{1}{2} \qquad \text{and} \qquad D[\mathcal{P}; C_j] \geqslant \tfrac{1}{2}.$$

Observe that no matter how many disc-segments we cut away, the remainder of the cake is convex. Someone who is worried about being found out stealing small disc-segments of the cake may choose to steal disc-segments with no chocolates, so that

$$D[\mathcal{P}; A \setminus (\bigcup B_i)] = D[\mathcal{P}; A] + \tfrac{1}{2}s.$$

Those that are more daring will choose to steal the disc-segments with the chocolates, so that

$$D[\mathcal{P}; A \setminus (\bigcup C_j)] \leqslant D[\mathcal{P}; A] - \tfrac{1}{2}t.$$

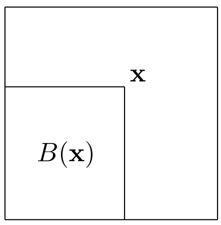
Combining these two estimates, we obtain

$$D[\mathcal{P}; A \setminus (\bigcup B_i)] - D[\mathcal{P}; A \setminus (\bigcup C_j)] \gg N^{\frac{1}{3}},$$

so that either

$$|D[\mathcal{P}; A \setminus (\bigcup B_i)]| \gg N^{\frac{1}{3}} \qquad \text{or} \qquad |D[\mathcal{P}; A \setminus (\bigcup C_j)]| \gg N^{\frac{1}{3}}. \qquad \square$$

Let us now return to Roth's problem in dimension $K = 2$. Suppose that $\mathcal{P}$ is a distribution of N points in the unit square $[0,1)^2$, where the integer N is sufficiently large. For every $\mathbf{x} = (x_1, x_2) \in [0,1]^2$, consider the rectangle $B(\mathbf{x}) = [0, x_1) \times [0, x_2)$ and the discrepancy function $D[\mathcal{P}; B(\mathbf{x})] = \#(\mathcal{P} \cap B(\mathbf{x})) - N x_1 x_2$.



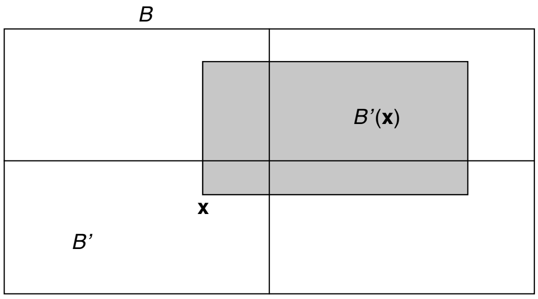
Since the point set $\mathcal{P}$ is arbitrary, we have no precise information on these points, and so it is hard to extract discrepancy near these points. On the other hand, note that parts of $[0,1]^2$ are short of points of $\mathcal{P}$, giving rise to "trivial discrepancies", so we try to exploit these.

Suppose that $\mathcal{P}$ has N points. If we partition the unit square $[0,1]^2$ into more than $2N$ subsets, then at least half of these subsets are devoid of points of $\mathcal{P}$. More precisely, choose n to satisfy $2N \leqslant 2^n < 4N$, and partition $[0,1]^2$ into similar rectangles of area 2^{-n}. Then at least half of these rectangles contain no points of $\mathcal{P}$. We shall extract discrepancy from such "empty" rectangles.

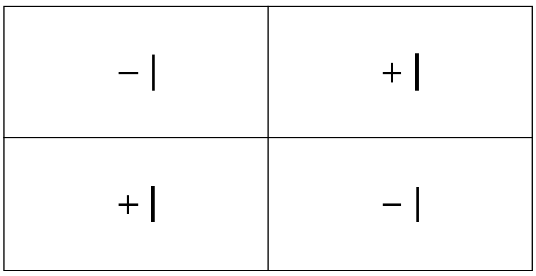
A typical rectangle of area $2^{-n} = 2^{-r_1} \times 2^{-r_2}$ is of the form

$$B = \prod_{j=1}^{2} [m_j 2^{-r_j}, (m_j + 1) 2^{-r_j}). \tag{5}$$

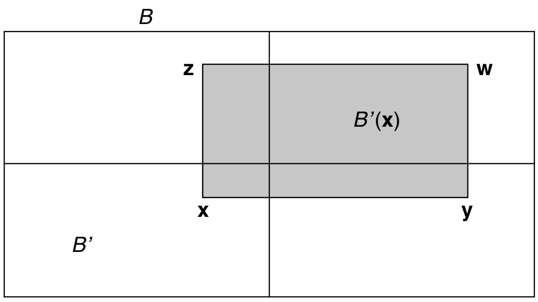
Let B' denote the bottom-left quarter of B. For any $\mathbf{x} \in B'$, consider the rectangle $B'(\mathbf{x})$ of area 2^{-n-2} and bottom left vertex $\mathbf{x}$ as shown.



Suppose that B contains no point of $\mathcal{P}$. Then $B'(\mathbf{x})$ contains no point of $\mathcal{P}$, and so has trivial discrepancy $-N2^{-n-2}$. A device to pick up this trivial discrepancy is provided by the Rademacher function $R_{r_1,r_2}(\mathbf{x})$ defined locally on B as follows.



For any $\mathbf{x} \in B'$, let $\mathbf{y}, \mathbf{z}, \mathbf{w}$ denote the other vertices of $B'(\mathbf{x})$ as shown.



Writing $D(\mathbf{x})$ as an abbreviation for $D[\mathcal{P}; B(\mathbf{x})]$, we have

$$\begin{aligned}\int_B D[\mathcal{P}; B(\mathbf{x})] R_{r_1,r_2}(\mathbf{x})\,\mathrm{d}\mathbf{x} &= \int_B D(\mathbf{x}) R_{r_1,r_2}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\ &= \int_{B'} (D(\mathbf{x}) - D(\mathbf{y}) - D(\mathbf{z}) + D(\mathbf{w}))\,\mathrm{d}\mathbf{x} \\ &= \int_{B'} D[\mathcal{P}; B'(\mathbf{x})]\,\mathrm{d}\mathbf{x} = -N2^{-2n-4}.\end{aligned}$$

To avoid those rectangles B with points of $\mathcal{P}$ undoing what we have achieved, we choose to kill off the effects of those rectangles. Accordingly, we define an auxiliary function on $[0,1]^2$ as follows. Let B be any rectangle of area $2^{-n} = 2^{-r_1} \times 2^{-r_2}$ and of the form (5). Write

$$f_{r_1,r_2}(\mathbf{x}) = \begin{cases} -R_{r_1,r_2}(\mathbf{x}) & \text{if } B \cap \mathcal{P} = \emptyset, \\ 0 & \text{if } B \cap \mathcal{P} \neq \emptyset. \end{cases} \tag{6}$$

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 $2^{-r_1}\times 2^{-r_2}$ rectangles, and noting that at least 2^{n-1} rectangles B satisfy $B\cap\mathcal{P}=\emptyset$, we conclude that

$$\int_{[0,1]^2} D(\mathbf{x}) f_{r_1,r_2}(\mathbf{x})\,\mathrm{d}\mathbf{x} = N2^{-2n-4}\#\{B : B\cap\mathcal{P}=\emptyset\} \gg 1. \tag{7}$$

There are $n+1$ choices of $r_1, r_2 \geqslant 0$ with $r_1+r_2=n$. Accordingly, we consider the auxiliary function

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

Then it follows from (7) that

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

The Cauchy–Schwarz inequality gives

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

We therefore need

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

but this follows easily from the orthogonality condition

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

Thus

$$\int_{[0,1]^2} |D(\mathbf{x})|^2\,\mathrm{d}\mathbf{x} \gg \log N \tag{8}$$

as required.

We comment that this argument can be generalized to arbitrary dimensions with no extra difficulties.

For dimension $K=2$, Halász devised an ingenious variation of Roth's technique to show that

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

Consider the auxiliary function

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

where $\alpha\in(0,\frac{1}{2})$ will be chosen suitably and fixed later, and where the functions $f_{r_1,r_2}(\mathbf{x})$ are precisely those studied by Roth and given in (6). Clearly

$$\left|\int_{[0,1]^2} D(\mathbf{x})H(\mathbf{x})\,\mathrm{d}\mathbf{x}\right| \leqslant \left(\sup_{\mathbf{x}\in[0,1]^2} |D(\mathbf{x})|\right) \int_{[0,1]^2} |H(\mathbf{x})|\,\mathrm{d}\mathbf{x},$$

and it is relatively easy to show that

$$\int_{[0,1]^2} |H(\mathbf{x})|\,\mathrm{d}\mathbf{x} \leqslant 2,$$

so it remains to show that

$$\left|\int_{[0,1]^2} D(\mathbf{x})H(\mathbf{x})\,\mathrm{d}\mathbf{x}\right| \gg n.$$

But then note that $H(\mathbf{x}) = \alpha F(\mathbf{x}) + \alpha^2 \ldots$, and so

$$\int_{[0,1]^2} D(\mathbf{x})H(\mathbf{x})\,\mathrm{d}\mathbf{x} = \alpha \int_{[0,1]^2} D(\mathbf{x})F(\mathbf{x})\,\mathrm{d}\mathbf{x} + \alpha^2 \ldots.$$

A reasonably crude argument then gives

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

so a suitable choice of α does the trick.

Unfortunately we cannot handle the corresponding ... satisfactorily when $K > 2$. We refer the reader to the discussions of Bilyk–Lacey–Vagharshakyan.

5. Use of Fourier Transforms

In the classical problem, we consider aligned rectangular boxes. In more general problems, we consider other sets of geometric objects. The question arises as to whether we can relate discrepancy with their geometry and, if so, determine such relationship.

Note first of all that any discrepancy function contains information on geometry as well as measure, the former through the characteristic function of the geometric objects and the latter through the difference between the discrete counting measure of the points of $\mathcal{P}$ and the continuous measure arising from volume.

Let the dimension $K \geqslant 2$ be fixed. Let A be a set of finite volume in euclidean space $\mathbb{R}^K$, let $\mathcal{P}$ be a distribution of N points in the unit cube $[0,1]^K$, and consider the discrepancy function

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

For a translate $A + \mathbf{x}$ of A, we have the discrepancy function

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

if we assume, for convenience, that A is symmetric across the origin. This shows that under translation, the discrepancy function is a convolution of the form

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

Here χ_A is purely geometric, whereas $\mathrm{d}Z_0 - N\mathrm{d}\mu_0$ is purely measure-theoretic, but they are held together by a convolution.

In 1964, in his work on integer sequences relative to long arithmetic progressions, Roth established his famous $\frac{1}{4}$-theorem. This represented the first successful use of Fourier transform in irregularities of point distribution, and became the catalyst that propelled Beck to arguably the most fascinating results in the subject.

Passing over to Fourier transforms, the convolution (9) becomes

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

and this provides a separation of geometry and measure.

For lower bounds which need to be valid for arbitrary distributions, we have very little useful information on the measure, so we need to concentrate on $\widehat{\chi_A}$ 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. We search for trivial discrepancy, and seek ways to blow them up.

For upper bounds, we attempt to construct specific point sets $\mathcal{P}$, so we have good information on them, and this gives better control over $\widehat{\mathrm{d}Z_0 - N\mathrm{d}\mu_0}$.

Let the dimension $K \geqslant 2$ be fixed, and consider the unit cube $[0,1]^K$, treated as a torus. Suppose that $A \subseteq [0,1]^K$ is a compact and convex set that also satisfies a technical condition which we shall not discuss here. We now subject the set A to contractions $\lambda \in [0,1]$, proper orthogonal transformations $\tau \in \mathcal{T}$, and translations $\mathbf{x} \in [0,1]^K$ to obtain similar copies $A(\lambda,\tau,\mathbf{x})$, and study the discrepancy function $D[\mathcal{P};A(\lambda,\tau,\mathbf{x})]$. In 1987, Beck obtained the following amazing result which shows that the discrepancy is heavily dependent on the boundary surface of A.

Theorem ([3]). *For every distribution $\mathcal{P}$ of N points in $[0,1]^K$, we have*

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

This result is sharp, as demonstrated by Beck and Chen in 1990.

Theorem ([7]). *There exists a distribution $\mathcal{P}$ of N points in $[0,1]^K$ such that*

$$\int_{[0,1]^K}\int_{\mathcal{T}}\int_0^1 |D[\mathcal{P};A(\lambda,\tau,\mathbf{x})]|^2 \,\mathrm{d}\lambda\,\mathrm{d}\tau\,\mathrm{d}\mathbf{x} \ll_A N^{1-\frac{1}{K}}.$$

Next, let us consider the unit square $[0,1]^2$, treated as a torus. Suppose that $A \subseteq [0,1]^2$ is a compact and convex set that again satisfies a technical condition that we shall not discuss here. We now subject the set A to contractions $\lambda \in [0,1]$ and translations $\mathbf{x} \in [0,1]^2$, but no proper orthogonal transformations, to obtain homothetic copies $A(\lambda,\mathbf{x})$, and study the discrepancy function $D[\mathcal{P};A(\lambda,\mathbf{x})]$. In 1988, Beck obtained the following equally amazing result which we shall not state precisely but which shows that the discrepancy is heavily dependent on our ability to approximate the set A by inscribed polygons.

Theorem ([4]). *For every distribution $\mathcal{P}$ of N points in $[0,1]^2$, we have*

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

where $\xi_N(A)$ is some measure of approximability of A by inscribed polygons.

For the factor $\log N$, compare this estimate with the classical lower bound (8). Also, we should mention that $\xi_N(A)$ is a constant if A is a convex polygon and a power of N if A is a circular disc.

6. Large Errors – Probabilistic Techniques

The case $K = 2$ of (10) gives the estimate

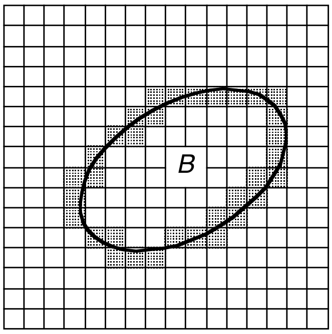
$$\int_{[0,1]^2}\int_{\mathcal{T}}\int_0^1 |D[\mathcal{P};A(\lambda,\tau,\mathbf{x})]|^2 \,\mathrm{d}\lambda\,\mathrm{d}\tau\,\mathrm{d}\mathbf{x} \gg_A N^{\frac{1}{2}},$$

which in turn leads to the estimate

$$\sup_{\substack{\lambda\in[0,1]\\ \tau\in\mathcal{T}\\ \mathbf{x}\in[0,1]^2}} |D[\mathcal{P};A(\lambda,\tau,\mathbf{x})]| \gg_A N^{\frac{1}{4}}.$$

In this section, we shall briefly illustrate the probabilistic techniques that show that the exponent $\frac{1}{4}$ is best possible. We shall also understand that the discrepancy is heavily dependent on the boundary curve of A.

Assume that $N = M^2$ for simplicity, split the unit square $[0,1]^2$ into M^2 little squares in the usual way, and place one point in each little square. Let B be a similar copy of A obtained under the contraction, proper orthogonal transformation and translation described earlier.



Suppose that S is one of the M^2 little squares. We shall examine its contribution to the discrepancy function

$$D[\mathcal{P}; B] = \#(\mathcal{P} \cap B) - N\mu(B).$$

If $S \subseteq B$, then S contributes 1 to $\#(\mathcal{P} \cap B)$ and again 1 to $N\mu(B)$, so makes a net contribution of zero. If $S \cap B = \emptyset$, then S contributes 0 to $\#(\mathcal{P} \cap B)$ and again 0 to $N\mu(B)$, so also makes a net contribution of zero. Thus the only contribution to the discrepancy comes from those little squares S where $S \cap \partial B \neq \emptyset$. This is where the boundary curve[3] comes to the party. It is very easy to see that $\#\{S : S \cap \partial B \neq \emptyset\} \ll_A M$, and that $|D[\mathcal{P}; S]| \leqslant 1$ for every little square S. We thus immediately obtain the trivial estimate

$$|D[\mathcal{P}; B]| \leqslant \sum_{S \cap \partial B \neq \emptyset} |D[\mathcal{P}; S]| \ll_A M = N^{\frac{1}{2}}.$$

Next, we bring in randomness and independence. We now assume that the point in each little square S is uniformly distributed within S, and independently of other points. Let $\tilde{\mathbf{p}}_S$ denote the random point in S, and let $\tilde{\mathcal{P}}$ denote the collection of all $N = M^2$ random points. Write

$$\psi_S = \begin{cases} 1 & \text{if } \tilde{\mathbf{p}}_S \in B, \\ 0 & \text{if } \tilde{\mathbf{p}}_S \notin B, \end{cases}$$

and let $D[\tilde{\mathcal{P}}; S] = \psi_S - \mathbb{E}\psi_S$. Then

$$D[\tilde{\mathcal{P}}; B] = \sum_{S \cap \partial B \neq \emptyset} (\psi_S - \mathbb{E}\psi_S).$$

We now use large deviation in probability theory, results such as Bernstein–Chernoff or Hoeffding. To cut a long story short, the estimate $N^{\frac{1}{2}}$ then becomes something like $N^{\frac{1}{4}}(\log N)^{\frac{1}{2}}$ – square root of the trivial estimate together with a small price for using probability theory.

7. The Classical Problem – Upper Bounds

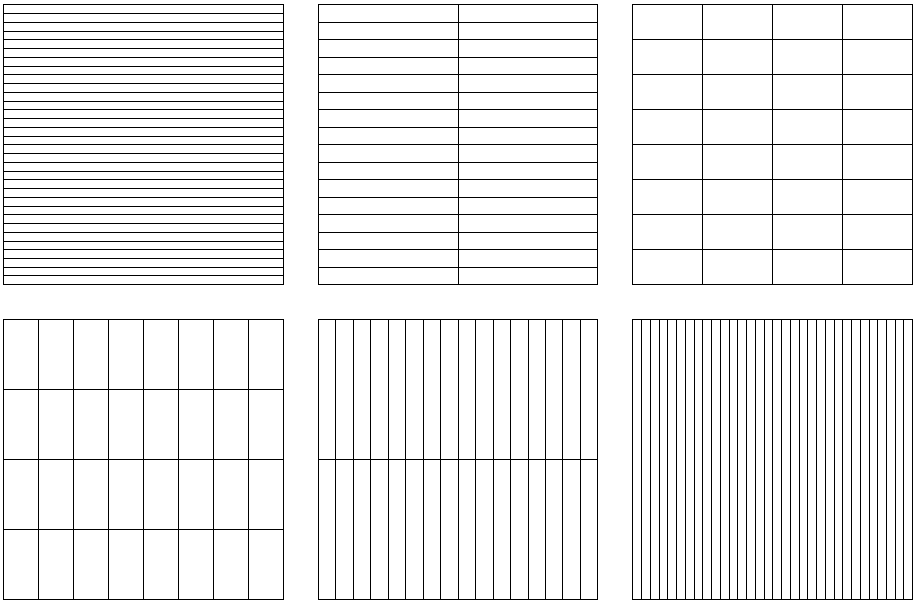
To complete this very brief survey, we return to the classical problem of rectangles in the unit square as studied by Roth at the beginning. Some of the point constructions have been known for quite a while, but there has been a fair amount of more recent work that has changed and improved our understanding of the problems. Here we concentrate on the van der Corput point sets which have a rather rich literature.

[3] In higher dimensions, this becomes the boundary surface.

First of all, we shall introduce the van der Corput set $\mathcal{P}_h$ of 2^h points in the unit square $[0,1)^2$.

Consider the following question. Let h be a fixed positive integer. Does there exist a set of 2^h points such that whenever we partition $[0,1)^2$ into congruent rectanges of size $2^{-h_1} \times 2^{-h_2}$, where $0 \leqslant h_1, h_2 \leqslant h$ and $h_1 + h_2 = h$, each rectangle contains precisely one point of the set? For convenience, let us assume that all rectangles include the left and bottom edges and exclude the right and top edges.

For instance, if $h = 5$, then we shall partition $[0,1)^2$ into 32 congruent rectangles in six different ways. In each of them, we need each rectangle to contain precisely one point of the same set.

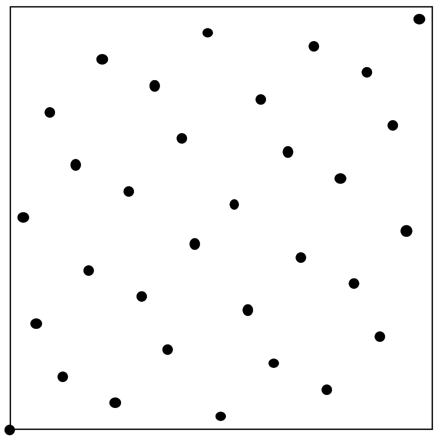


We can answer the question in the affirmative by giving the example of the van der Corput set $\mathcal{P}_h$ of 2^h points in the unit square $[0,1)^2$. This set is best described by using dyadic expansion, and we write

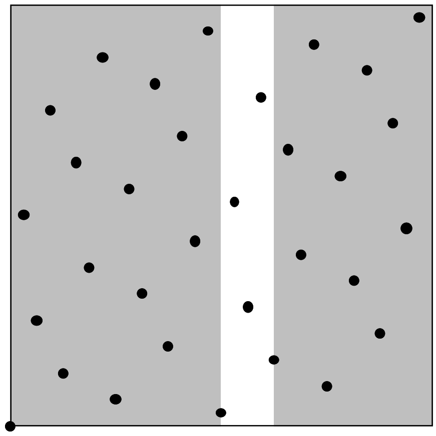
$$\mathcal{P}_h = \{(0.a_1 \dots a_h, 0.a_h \dots a_1) : a_1, \dots, a_h \in \{0,1\}\}.$$

Note that the digits of the second coordinate are in reverse order from the digits of the first coordinate.

The 32 points of $\mathcal{P}_5$ are distributed as below.



The van der Corput sets $\mathcal{P}_h$ have nice periodicity properties. To illustrate what we mean, let us look at the case of $\mathcal{P}_5$ again.



Suppose that we consider only those points whose first coordinate lie in the dyadic interval $[4\times 2^{-3}, 5\times 2^{-3})$. In the picture above, there are precisely four such points. Note that the second coordinate has period 2^{-2}.

This periodicity property opens the door to classical Fourier series.

Consider a rectangle of the form $B(x_1,x_2) = [0,x_1) \times [0,x_2)$. For simplicity, let us assume that x_1 is an integer multiple of 2^{-h}. Write $x_1 = 0.a_1 \ldots a_h$, and keep it fixed. Then

$$[0,x_1) = \bigcup_{\substack{i=1\\a_i=1}}^{h} [0.a_1 \ldots a_{i-1}, 0.a_1 \ldots a_i),$$

and it can be shown that

$$\begin{aligned} D[\mathcal{P}_h; B(x_1,x_2)] &= \sum_{\substack{i=1\\a_i=1}}^{h} D[\mathcal{P}_h; [0.a_1 \ldots a_{i-1}, 0.a_1 \ldots a_i) \times [0,x_2)] \\ &= \sum_{\substack{i=1\\a_i=1}}^{h} \left(\alpha_i - \psi\left(\frac{x_2+\beta_i}{2^{i-h}}\right)\right), \end{aligned} \tag{11}$$

where α_i, β_i are constants and where $\psi(z) = z - [z] - \frac{1}{2}$ is the saw-tooth function. Note that the summand is periodic in the variable x_2 with period 2^{i-h}.

We note at once that $|D[\mathcal{P}_h; B(x_1,x_2)]| \ll h$, so that $\mathcal{P}_h$ gives the best L^∞ upper bound in the classical problem in dimension $K=2$.

On the other hand, it is easy to see that $|D[\mathcal{P}_h; B(x_1,x_2)]|^2$ contains the term

$$\sum_{\substack{i,j=1\\a_i=a_j=1}}^{h} \alpha_i \alpha_j,$$

caused by the restriction[4] that all the rectangles $B(x_1,x_2)$ under consideration share a common lower left vertex, namely the origin, and this leads to the estimate

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

[4]Those people who wish to have an easy life and avoid difficulties would argue to the death that this is an unnatural restriction. However, overcoming this handicap calls for much deeper understanding of the questions. Weyl, Roth, Schmidt, Beck and Sobolev cannot all be wrong!

and so $\mathcal{P}_h$ does *not* lead to the best L^2 upper bound.

To overcome this difficulty, Roth procceded to translate the points in $\mathcal{P}_h$ in the x_2-direction to obtain a translated point set $\mathcal{P}_h(t)$. Keeping x_2 as well as x_1 fixed, we can show that

$$D[\mathcal{P}_h(t);B(x_1,x_2)] = \sum_{\substack{i=1\\a_i=1}}^{h} \left(\psi\left(\frac{z_i+t}{2^{i-h}}\right) - \psi\left(\frac{w_i+t}{2^{i-h}}\right)\right),$$

where z_i, w_i are constants. This is a sum of quasi-orthogonal functions in the translation variable t, and one can show that

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

We now integrate trivially over $\mathbf{x}=(x_1,x_2)$, and conclude that there exists t^* such that

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

Note that $N=2^h$.

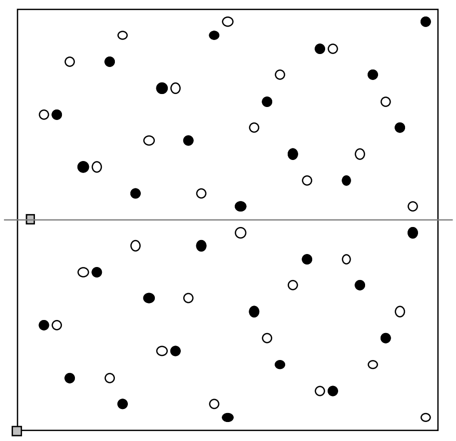
Another way of overcoming this difficulty is to use the reflection principle of Davenport. We reflect the set $\mathcal{P}_h$ across the axis $x_2=\frac{1}{2}$ to obtain the set

$$\mathcal{P}_h^* = \{(p_1, 1-p_2) : (p_1,p_2) \in \mathcal{P}_h\}.$$

Then one can show that

$$D[\mathcal{P}_h^*;B(x_1,x_2)] = \sum_{\substack{i=1\\a_i=1}}^{h} \left(-\alpha_i - \psi\left(\frac{x_2+\gamma_i}{2^{i-h}}\right)\right),$$

where α_i, γ_i are constants and α_i is exactly the same as in (11).



Then

$$D[\mathcal{P}_h \cup \mathcal{P}_h^*;B(x_1,x_2)] = -\sum_{\substack{i=1\\a_i=1}}^{h} \left(\psi\left(\frac{x_2+\beta_i}{2^{i-h}}\right) + \psi\left(\frac{x_2+\gamma_i}{2^{i-h}}\right)\right),$$

a sum of quasi-orthogonal functions in the variable x_2, and one can show that

$$\int_{[0,1]} |D[\mathcal{P}_h;B(x_1,x_2)]|^2 \,\mathrm{d}x_2 \ll h.$$

We now integrate trivially over x_1.

Recall that

$$\mathcal{P}_h = \{(0.a_1 \ldots a_h, 0.a_h \ldots a_1) : a_1, \ldots, a_h \in \{0,1\}\},$$

where the digits of the second coordinate are in reverse order from the digits of the first coordinate. Clearly $\mathcal{P}_h$ is isomorphic to the group $\mathbb{Z}_2^h$. We can then make use of the group characters which are the Walsh functions. This opens the door to Fourier–Walsh analysis.

We can further consider the analogues of the van der Corput point sets base p for some prime p. The van der Corput set $\mathcal{P}_h$ of p^h points is then best described by using p-adic expansion, and we write

$$\mathcal{P}_h = \{(0.a_1 \ldots a_h, 0.a_h \ldots a_1) : a_1, \ldots, a_h \in \{0, 1, \ldots, p-1\}\}.$$

Note again that the digits of the second coordinate are in reverse order from the digits of the first coordinate. In this case, $\mathcal{P}_h$ is isomorphic to the group $\mathbb{Z}_p^h$. The group characters are now the base p Walsh functions. This opens the door to base p Fourier–Walsh analysis.

Finally, we mention digit shifts which give rise to points of the form

$$(0.(a_1 \oplus u_1) \ldots (a_h \oplus u_h), 0.(a_h \oplus v_h) \ldots (a_1 \oplus v_1)),$$

where $(u_1, \ldots, u_h, v_1, \ldots, v_h) \in \mathbb{Z}_p^{2h}$. The effect of digit shifts can be analyzed by base p Fourier–Walsh analysis.

A few years ago, explicit point sets with best L^2 upper bounds were successfully constructed by Chen and Skriganov [11], through the use of base p Fourier–Walsh analysis and other ideas. More recently, Chen and Skriganov [12] obtained much deeper understanding of the relationship between explicit constructions, digit shifts and orthogonality, again through the use of base p Fourier–Walsh analysis.

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. On the discrepancy of convex plane sets. *Monatsh. Math.*, 105 (1988), 91–106.

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

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

[11] 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.

[12] 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).

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

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

[15] 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).

[16] J.H. Halton. On the efficiency of certain quasirandom sequences of points in evaluating multidimensional integrals. *Numer. Math.*, 2 (1960), 84–90.
[17] M. Lerch. Question 1547. *L'Intermediaire Math.*, 11 (1904), 144–145.
[18] K.F. Roth. On irregularities of distribution. *Mathematika*, 1 (1954), 73–79.
[19] K.F. Roth. Remark concerning integer sequences, *Acta Arith.*, 9 (1964), 257–260.
[20] K.F. Roth. On irregularities of distribution IV. *Acta Arith.*, 37 (1980), 67–75.
[21] W.M. Schmidt. Irregularities of distribution VII. *Acta Arith.*, 21 (1972), 45–50.
[22] W.M. Schmidt. Irregularities of distribution IX. *Acta Arith.*, 27 (1975), 386–396.
[23] W.M. Schmidt. Irregularities of distribution X. *Number Theory and Algebra* (H. Zassenhaus, ed.), pp. 311–329 (Academic Press, 1977).

Department of Mathematics, Macquarie University, Sydney, NSW 2109, Australia
E-mail address: wchen@maths.mq.edu.au