Uniform Distribution Theory **6** (2011), no.1, 143–184

uniform distribution theory

ON ROTH'S ORTHOGONAL FUNCTION METHOD IN DISCREPANCY THEORY

Dmitriy Bilyk

ABSTRACT. Motivated by the recent surge of activity on the subject, we present a brief non-technical survey of numerous classical and new results in discrepancy theory related to Roth's orthogonal function method.

Communicated by Yukio Ohkubo

The goal of this paper is to survey the use of Roth's orthogonal function method in the theory of irregularities of distribution, which, in a broad sense, means applications of orthogonal function decompositions to proving discrepancy estimates. The idea originated in a seminal paper of Klaus Roth [61, 1954] that, according to his own words "started a new theory" [26]. Since then and up to today this approach has been widely exploited to produce new important results in the the area. In the present expository article we trace the method from its origins to recent results (to which the author has made some contribution). It is our intention to keep the exposition concise, simple, and unobscured by the technicalities. We shall concentrate on the heuristics and intuition behind the underlying ideas, introduce classical and novel points of view on the method, as well as connections to other areas of mathematics.

Before we proceed to the main part of the discussion, I would like to emphasize the tremendous influence that the aforementioned paper of Roth [61], entitled "On irregularities of distribution", has had on the development of the field. Even the number of papers with identical or similar titles, that appeared in the subsequent years, attests to its importance: 4 papers by Roth himself (*On irregularities of distribution. I–IV*, [61, 62, 63, 64]), one by H. Davenport (*Note on irregularities of distribution*, [33]), 10 by W. M. Schmidt (*Irregularities of distribution. I–IX*, [65, 66, 67, 68, 69, 70, 71, 72, 73, 74]), 2 by J. Beck (*Note on irregularities of distribution. I–II*, [5, 6]), 4 by W. W. L. Chen

2010 Mathematics Subject Classification: 11K38, 11K06.
Keywords: Discrepancy, star-discrepancy, dyadic interval, Haar function, Besov space, Sobolev space.

(*On irregularities of distribution. I–IV*, [21, 22, 23, 24]), at least 2 by Beck and Chen (*Note on irregularities of distribution. I–II*, [3, 8] and several others with similar, but more specific names), as well as the fundamental monograph on the subject by Beck and Chen, *"Irregularities of distribution"*, [4].

1. Introduction: history, old and new results.

We now begin by briefly discussing the essence and the history of the field. Discrepancy theory at large is concerned with various forms of the following questions: *How well can uniform distribution be approximated by a discrete set of N points? And what are the errors and limitations that necessarily arise in such approximations?* The latter question is the gist of the theory of irregularities of distribution. The subject arose naturally in number theory in connection with the notion of uniformly distributed sequences. A sequence $\omega = \{\omega_n\}_{n=1}^{\infty} \subset [0,1]$ is called uniformly distributed if, for any subinterval $I \subset [0,1]$, the proportion of points ω_n that fall into I approximates its length, i.e.,

$$\lim_{N\to\infty} \frac{\#\{\omega_n \in I : 1 \le n \le N\}}{N} = |I|. \tag{1.1}$$

This quality can be easily quantified using the notion of *discrepancy*:

$$D_N(\omega) = \sup_{I\subset[0,1]} \big|\#\{\omega_n \in I : 1 \le n \le N\} - N \cdot |I|\big|. \tag{1.2}$$

In fact, it is easy to show that ω is uniformly distributed if and only if $D_N(\omega)/N$ tends to zero as $N \to \infty$ (see e.g. [48]). In [30, 1935], van der Corput posed a question whether there exists a sequence ω for which the quantity $D_N(\omega)$ stays bounded as N gets large. In [1, 1945], van Aardenne-Ehrenfest gave a negative answer to this question, which meant that no sequence can be distributed too well; furthermore, in [2] she gave a quantitative version of this statement, showing that $D_N(\omega)$ must be at least as big as $\log\log N/\log\log\log N$ infinitely often. In [61], Roth greatly improved this result by proving that for any sequence ω the inequality

$$D_N(\omega) \ge C\sqrt{\log N} \tag{1.3}$$

holds for infinitely many values of N. These results signified the birth of a new theory.

Roth had equivalently reformulated the problem to the following setting: let

$$\mathcal{P}_N \subset [0,1]^d$$

be a set of N points and consider the discrepancy function

$$D_N(x_1, \ldots, x_d) = \#\{\mathcal{P}_N \cap [0, x_1) \times \cdots \times [0, x_d)\} - N \cdot x_1 \cdots x_d, \tag{1.4}$$

i.e., the difference of the actual and expected number of points of $\mathcal{P}_N$ in the box $[0, x_1) \times \cdots \times [0, x_d)$. In this setup, Roth proved

Theorem 1.1 (Roth). *In all dimensions $d \geq 2$, for any N-point set*

$$\mathcal{P}_N \subset [0,1]^d,$$

one has

$$\sup_{x \in [0,1]^d} \big|D_N(x)\big| \geq C_d \log^{\frac{d-1}{2}} N, \tag{1.5}$$

where C_d is an absolute constant that depends only on the dimension d.

Moreover, Roth demonstrated that, when $d = 2$, this result is equivalent to (1.3). More generally, uniform lower bounds for the discrepancy function of finite point sets (for all values of N) in dimension d are equivalent to lower estimates for the discrepancy of infinite sequences (1.2) (for infinitely many values of N) in dimension $d - 1$. These two settings are sometimes referred to as 'static' (fixed finite point sets) and 'dynamic' (infinite sequences). In these terms, one can say that the dynamic and static problems are equivalent at the cost of one dimension—the relation becomes intuitively clear if one views the sequence index (or time) as an additional dimension. In what follows, we shall work with Roth's, more geometrical, static reformulation.

In fact, instead of working directly with the L^∞ norm of the discrepancy function

$$\|D_N\|_\infty = \sup_{x \in [0,1]^d} \big|D_N(x)\big|$$

(the normalized quantity $\frac{1}{N}\|D_N\|_\infty$ is classically referred to as *the star-discrepancy*; since in this paper we only work with the unnormalized version, we shall sometimes use this name for $\|D_N\|_\infty$), Roth considered a smaller quantity, its L^2 norm $\|D_N\|_2$—inequality (1.5) actually holds with the L^2 norm of D_N on the left hand side. This substitution allows for an introduction of a variety of Hilbert space techniques, including orthogonal decompositions.

Roth's ingenious idea consisted of expanding the discrepancy function D_N in the classical orthogonal Haar basis and considering only the projection onto the span of those Haar functions which are supported on rectangles of volume roughly equal to $\frac{1}{N}$ (heuristically justified by the fact that, for a well distributed set, they contain approximately one point per rectangle). To be even more precise, the size of the rectangles R was chosen so that $|R| \approx \frac{1}{2N}$, ensuring that at least about half of all rectangles are free of points of $\mathcal{P}_N$. The Haar coefficients of D_N, corresponding to these empty rectangles, are then easily computable, which leads directly to the estimate (1.5). This approach strongly resonates with Kolmogorov's method of proving lower error bounds for cubature formulas, see, e.g., [80, Chapter IV].

Roth's result has been extended to other L^p norms, $1 < p < \infty$, by W. Schmidt in [74, 1977], who showed that in all dimensions $d \geq 2$, for all $p \in (1, \infty)$ the inequality

$$\|D_N\|_p \geq C_{d,p} \log^{\frac{d-1}{2}} N, \tag{1.6}$$

holds for some constant $C_{d,p}$ independent of the collection of the points $\mathcal{P}_N$. Schmidt's approach was a direct extension of Roth's method: rather then working with arbitrary integrability exponents p, he considers only those p's for which the dual exponent q is an even integer. This allows one to iterate the orthogonality arguments. Even though it took more than twenty years to extend Roth's L^2 inequality to other L^p spaces, a contemporary harmonic analyst may realize that such an extension can be derived in several lines using Littlewood-Paley inequalities. A more thorough discussion will be provided in § 3.

However, the endpoint case $p = \infty$, corresponding to the *star-discrepancy*, is the most natural and important in the theory. As it turns out, Roth's inequality (1.5) is not sharp for the sup-norm of the discrepancy function. It is perhaps not surprising: intuitively, the discrepancy function is highly irregular and comes close to its maximal values only on small sets. Hence, its average (e.g., L^2 norm) must necessarily be much smaller than its extremal (i.e., L^∞) norm. This heuristics also guides the use of some of the methods that have been exploited in the proofs of the star-discrepancy estimates, such as Riesz products.

In 1972, W. M. Schmidt proved that in dimension $d = 2$ one has the following lower bound:

$$\sup_{x \in [0,1]^d} \big|D_N(x)\big| \geq C \log N, \tag{1.7}$$

which is sharp. Indeed, two-dimensional constructions, for which $\|D_N\|_\infty \leq C \log N$ holds for all N (or, equivalently, one-dimensional sequences ω for which $D_N(\omega) \leq C \log N$ infinitely often), have been known for a long time and go back to the works of Lerch [52, 1904], van der Corput [30, 1935] and others. Several other proofs of inequality (1.7) have been given later [54, 1979], [9, 1982], [39, 1981]. The latter (due to Halász) presents great interest to us as it has been built upon Roth's Haar function method—we will reproduce and analyze the argument in § 4. Incidentally, the title of Halász's article [39] ("On Roth's method in the theory of irregularities of point distributions") almost coincides with the title of the present paper.

Higher dimensional analogs of Schmidt's estimate (1.7), however, turned out to be extremely proof-resistant. For a long time inequality (1.5) remained the best known bound in dimensions three and above. In fact, the first gain over the L^2 estimate was obtained only thirty-five years after Roth's result by Beck [7, 1989], who proved that in dimension $d = 3$, the discrepancy function satisfies

$$\|D_N\|_\infty \geq C \log N \cdot (\log\log N)^{\frac{1}{8}-\varepsilon}. \tag{1.8}$$

Almost twenty years later, in 2008, the author jointly with M. Lacey and A. Vagharshakyan ([11], $d = 3$; [12], $d \geq 4$) obtained the first significant improvement of the L^∞ bound in *all dimensions* $d \geq 3$:

Theorem 1.2 (Bilyk, Lacey, Vagharshakyan). *For all $d \geq 3$, there exists some $\eta = \eta(d) > 0$, such that for all $\mathcal{P}_N \subset [0,1]^d$ with $\#\mathcal{P}_N = N$ we have the estimate:*

$$\|D_N\|_\infty \geq C_d \, (\log N)^{\frac{d-1}{2}+\eta(d)}. \tag{1.9}$$

The exact rate of growth of the star-discrepancy in higher dimensions remains an intriguing open problem; in their book, [4] Beck and Chen named it ''the great open problem" and called it ''excruciatingly difficult". The precise form of the conjecture is a subject of debate among the experts in the field. We start with the form which is directly pertinent to the orthogonal function method.

Conjecture 1.3. *For all $d \geq 3$ and all $\mathcal{P}_N \subset [0,1]^d$ with $\#\mathcal{P}_N = N$ we have the estimate:*

$$\|D_N\|_\infty \geq C_d \, (\log N)^{\frac{d}{2}}. \tag{1.10}$$

This conjecture is motivated by connections of this field to other areas of mathematics and, in particular, by a related conjecture in analysis, Conjecture 4.1, which is known to be sharp (a comprehensive review of this relation will be given in § 4). This means, above all, that this is the best result that one can achieve using Roth's Haar function method.

On the other hand, the best known examples [40, 42] of well distributed sets in higher dimensions have star-discrepancy of the order

$$\|D_N\|_\infty \leq C_d \, (\log N)^{d-1}. \tag{1.11}$$

Numerous constructions of such sets are known and are currently a subject of massive ongoing research, see, e.g. the book [34]. These upper bounds together with the estimates for a "smooth" version of discrepancy (see Temlyakov [85]), provide grounds for the alternative form of the conjecture (which is actually older and more established)

Conjecture 1.4. *For all $d \geq 3$ and all $\mathcal{P}_N \subset [0,1]^d$ with $\#\mathcal{P}_N = N$ we have the estimate:*

$$\|D_N\|_\infty \geq C_d \, (\log N)^{d-1}. \tag{1.12}$$

One can notice that both conjectures coincide with Schmidt's estimate (1.7) when $d = 2$. M. Skriganov has proposed yet another form of the conjecture [76]:

$$\|D_N\|_\infty \geq C_d\,(\log N)^{\frac{d-1}{2}+\frac{d-1}{d}}. \tag{1.13}$$

In contrast to the L^∞ inequalitites, it is well known that in the average (L^2 or L^p) sense Roth's bound (1.3), as well as inequality (1.6), is sharp. This was initially proved by Davenport [33] in two dimensions for $p = 2$, who constructed point distributions with $\|D_N\|_2 \leq C\sqrt{\log N}$. Subsequently, different constructions have been obtained by numerous other authors, including Roth [63, 64], Chen [21], Frolov [37]. It should be noted that most of the optimal constructions in higher dimensions $d \geq 3$ are probabilistic in nature and are obtained as randomizations of some classic low-discrepancy sets. In fact, deterministic examples of sets with $\|D_N\|_p \leq C_{d,p} \log^{\frac{d-1}{2}} N$ have been constructed only in the last decade by Chen and Skriganov [27] ($p = 2$) and Skriganov [75] ($p > 1$). It would be interesting to note that their results are also deeply rooted in Roth's orthogonal function method—they use the orthogonal system of Walsh functions to analyze the discrepancy function and certain features of the argument remind of the ideas that appear in Roth's proof.

The other endpoint of the L^p scale, $p = 1$, is not any less difficult than the star-discrepancy estimate. Indeed, the only information that is available is the two-dimensional inequality due to Halász (also contained in the aforementioned paper [39]), which also makes use of the Roth's orthogonal function method:

$$\|D_N\|_1 \geq C\sqrt{\log N}, \tag{1.14}$$

which essentially states that the L^1 norm of discrepancy behaves roughly like its L^2 norm. It is conjectured that the same similarity continues to hold in higher dimensions.

Conjecture 1.5. *For all $d \geq 3$ and all sets of N points in $[0,1]^d$:*

$$\|D_N\|_1 \geq C_d\,(\log N)^{\frac{d-1}{2}}. \tag{1.15}$$

However, almost no results pertaining to this conjecture have been discovered for $d \geq 3$. The only known relevant fact is that Halasz's $\log^{\frac{1}{2}} N$ bound still holds in higher dimensions, i.e., it is not even known if the exponent increases with dimension. The reader is referred to (4.15)–(4.17) for Halasz's L^1 argument.

The outline of the paper is as follows. In § 2 we discuss Roth's original proof and its variations. The next section, § 3, introduces a powerful tool from harmonic analysis, the Littlewood-Paley theory, which allows one to extend Roth's L^2 inequality to L^p, $1 < p < \infty$ as well as to numerous other norms.

We survey a wide range of results and conjectures on lower discrepancy estimates in various function spaces. Section 4 discusses the most important case: the L^∞ estimates for the discrepancy function. We bring up the closely related *small ball inequality*, demonstrate two-dimensional proofs, and explore connections between the problems as well as links to other fields such as probability and approximation theory. In § 5 we give a short three-page account of very technical recent results of the author, Lacey, and Vagharshakyan which improve lower bounds of the star-discrepancy in all dimensions. We try to make the presentation very intuitive and concentrate on the heuristics and ideas behind the proof. Finally, the last section deals with applications of Roth's method to obtaining upper discrepancy estimates, in particular for the variations of the van der Corput set.

We often make use of the symbol "$\lesssim$": $F \lesssim G$ means that there exists a constant $C > 0$ such that $F \leq CG$. The constant is allowed to depend on the dimension and, perhaps, some other parameters, but never on the number of points N. The relation $F \approx G$ means that $F \lesssim G$ and $G \lesssim F$.

The aim of this survey is really two-fold: to acquaint specialists in discrepancy theory with some of the techniques of harmonic analysis which may be used in this subject, as well as to present the circle of problems in the field of irregularities of distribution to the analysts. Numerous books written on discrepancy theory present Roth's method and related arguments, see [4, 20, 34, 48, 56, 80]; the book [57] studies the relations between the uniform distribution and harmonic analysis, [87] studies the subject from the point of view of function space theory, while [77] specifically investigates the connections between discrepancy and Haar functions. In addition, the survey [29] explores various ideas of Roth in discrepancy theory, including the method discussed here, and [25] stresses the applications of Fourier analysis to the theory of irregularities of distribution. We sincerely hope however that the present paper will provide some novel ideas and useful insights and will be of interest to both novices and specialists in the field.

2. Roth's L^2 proof

We now turn to a more detailed discussion of Roth's method. We shall begin by reproducing the proof of his result, Theorem 1.1, [61], although our exposition will slightly differ from the style of the original paper [61] (the arguments, however, will be identical to Roth's). We shall make use of somewhat more modern notation which is closer in spirit to functional and harmonic analysis. Hopefully, this will allow us to make the idea of the proof more transparent.

We start by defining the Haar basis in in $L^2[0,1]$. Let $\mathbf{1}_I(x)$ stand for the characteristic function of the interval I. Consider the collection of all *dyadic* subintervals of [0,1]:

$$\mathcal{D} = \left\{ I = \left[\frac{m}{2^n}, \frac{m+1}{2^n} \right) : m, n \in \mathbb{Z},\, n \geq 0,\, 0 \leq m < 2^n \right\}. \tag{2.1}$$

Dyadic intervals form a grid, meaning that any two intervals in $\mathcal{D}$ are either disjoint, or one is contained in another. In addition, for every interval $I \in \mathcal{D}$, its left and right halves (we shall denote them by I_l and I_r) are also dyadic. The Haar function corresponding to the interval I is then defined as

$$h_I(x) = -\mathbf{1}_{I_l}(x) + \mathbf{1}_{I_r}(x). \tag{2.2}$$

Notice that in our definition, the Haar functions are normalized to have norm one in L^∞ (their L^2 norm is $\|h_I\|_2 = |I|^{1/2}$). This will cause some of the classical formulas to look a little unusual for those readers who are accustomed to the L^2 normalization.

These functions have been introduced by Haar [38, 1910] and have played an extremely important role in analysis, probability, signal processing etc. They are commonly viewed as the first example of *wavelets*. Their orthogonality, i.e., the relation

$$\langle h_{I'}, h_{I''} \rangle = \int_0^1 h_{I'}(x) \cdot h_{I''}(x)\, dx = 0, \qquad I', I'' \in \mathcal{D},\, I' \neq I'', \tag{2.3}$$

follows easily form the facts that $\mathcal{D}$ is a grid and that if $I' \subsetneq I''$, then I' is contained either in the left or right half of I'', hence $h_{I''}$ is constant on the support of I'. It is well known that the system $\mathcal{H} = \mathbf{1}_{[0,1]} \cup \{h_I : I \in \mathcal{D}\}$ forms an orthogonal basis in $L^2[0,1]$ and an unconditional basis in $L^p[0,1]$, $1 < p < \infty$.

In higher dimensions, we consider the family of *dyadic rectangles* $\mathcal{D}^d = \{R = R_1 \times \cdots \times R_d : R_j \in \mathcal{D}\}$. For a dyadic rectangle R, the Haar function supported by it is defined as a coordinatewise product of the one-dimensional Haar functions:

$$h_R(x_1, \ldots, x_d) = h_{I_1}(x_1) \cdots h_{I_d}(x_d). \tag{2.4}$$

The orthogonality of these functions is easily derived from the one dimensional property. It is also well-known that the 'product' Haar system $\mathcal{H}^d = \{f(x) = f_1(x_1) \cdots f_d(x_d) : f_k \in \mathcal{H}\}$ is an orthonormal basis of $L^2([0,1]^d)$—often referred to as the product basis. We note that this is not the only way to extend wavelet bases to higher dimensions [32], but this multi-parameter approach is the correct tool for the problems at hand, where the dimensions of the underlying rectangles are allowed to vary absolutely independently (e.g., some rectangles may be long and thin, while others may resemble a cube). This is precisely the setting of the product (multi-parameter) harmonic analysis—we shall keep returning to this point throughout the text.

As mentioned in the introduction, Roth realized that the magnitude of the discrepancy function is essentially determined by the part of the Haar decomposition, which corresponds to rectangles of volume $|R| \approx \frac{1}{N}$. Let us choose the number $n \in \mathbb{N}$ so that $2^{n-2} \leq N < 2^{n-1}$, i.e., $n \approx \log_2 N$ (although the precise choice of n is important for the argument), and consider dyadic rectangles of R of volume 2^{-n}. These rectangles come in a variety of shapes. To keep track of them we introduce a collection of vectors with integer coordinates

$$\mathbb{H}_n^d = \{\vec{r} = (r_1, \ldots, r_d) \in \mathbb{Z}_+^d : \|\vec{r}\|_1 = n\}, \tag{2.5}$$

where the ℓ_1 norm is defined as $\|\vec{r}\|_1 = |r_1| + \cdots + |r_d|$. These vectors will specify the shape of the dyadic rectangles in the following sense: for $R \in \mathcal{D}^d$, we say that $R \in \mathcal{D}_{\vec{r}}^d$ if $|R_j| = 2^{-r_j}$ for $j = 1, \ldots, d$. Obviously, if $R \in \mathcal{D}_{\vec{r}}^d$ and $\vec{r} \in \mathbb{H}_n^d$, then $|R| = 2^{-n}$. Besides, it is evident that, for a fixed $\vec{r}$, all the rectangles $R \in \mathcal{D}_{\vec{r}}^d$ are disjoint. It is also straightforward to see that the cardinality

$$\#\mathbb{H}_n^d = \binom{n+d-1}{d-1} \approx n^{d-1}, \tag{2.6}$$

which agrees with the simple logic that we have $d-1$ "free" parameters: the first $d-1$ coordinates can be chosen essentially freely, while the last one would be fixed due to the condition $\|\vec{r}\|_1 = n$ or $|R| = 2^{-n}$.

We now define a function f on $[0,1]^d$ to be an r-function with parameter $\vec{r} \in \mathbb{Z}_+^d$ if f has the form

$$f(x) = \sum_{R \in \mathcal{D}_{\vec{r}}^d} \varepsilon_R h_R(x), \tag{2.7}$$

for some choice of signs $\varepsilon_R = \pm 1$. These functions are generalized Rademacher functions (hence the name)—indeed, setting all the signs $\varepsilon_R = 1$, one obtains the familiar Rademacher function. It is trivial that if f is an r-function, then $f^2 = 1$ and thus $\|f\|_2 = 1$. Such functions play the role of building blocks in numerous discrepancy arguments, therefore their L^2 normalization justifies the choice of the L^∞ normalization for the Haar functions. In addition, the fact that two r-functions corresponding to different vectors $\vec{r}$ are orthogonal readily follows from the orthogonality of the family of Haar functions.

Next, we would like to compute how discrepancy function D_N interacts with the Haar functions in certain cases. Notice that discrepancy function can be written in the form

$$D_N(x) = \sum_{p \in \mathcal{P}_N} \mathbf{1}_{[p,\vec{1}]}(x) - N \cdot x_1 \cdots x_d, \tag{2.8}$$

where $\vec{1} = (1, \dots, 1)$ and $[p, \vec{1}] = [p_1, 1] \times \cdots \times [p_d, 1]$. We shall refer to the first term as the *counting* part and the second as the *volume (area)* or the *linear* part.

It is easy to see that, in one dimension, we have

$$\int \mathbf{1}_{[q,1]}(x) \cdot h_I(x)\, dx = \int_q^1 h_I(x)\, dx = 0$$

unless I contains the point q. This implies that for $p \in [0,1]^d$

$$\int_{[0,1]^d} \mathbf{1}_{[p,\vec{1}]}(x) \cdot h_R(x)\, dx = \prod_{j=1}^{d} \int_{p_j}^1 h_{R_j}(x_j)\, dx_j = 0 \tag{2.9}$$

when $p \notin R$. Assume now that a rectangle $R \in \mathcal{D}^d$ is empty, i.e., does not contain points of $\mathcal{P}_N$. It follows from the previous identity that for such a rectangle, the inner product of the corresponding Haar function with the counting part of the discrepancy function is zero:

$$\left\langle \sum_{p \in \mathcal{P}_N} \mathbf{1}_{[p,\vec{1}]},\, h_R \right\rangle = 0, \tag{2.10}$$

in other words, the inner product $\langle D_N, h_R \rangle$ is determined purely by the linear part of D_N.

It is however a simple exercise to compute the inner product of the linear part with any Haar function:

$$\langle N x_1 \dots x_d, h_R \rangle = N \prod_{j=1}^{d} \big\langle x_j, h_{R_j}(x_j) \big\rangle = N \cdot \frac{|R|^2}{4^d}. \tag{2.11}$$

Hence we have shown that if a rectangle $R \in \mathcal{D}^d$ does not contain points of $\mathcal{P}_N$ in its interior, we have

$$\langle D_N, h_R \rangle = -N |R|^2 4^{-d}. \tag{2.12}$$

These, somewhat mysterious, computations can be explained geometrically (see [74], also [20, Chapter 3]). For simplicity, we shall do it in dimension $d = 2$, but this argument easily extends to higher dimensions. Let $R \subset [0,1]^2$ be an arbitrary dyadic rectangle of dimensions $2h_1 \times 2h_2$ which does not contain any points of $\mathcal{P}_N$ and let $R' \subset R$ be the lower left quarter of R. Notice that, for any point $x = (x_1, x_2) \in R'$, the expression

$$\begin{aligned} D_N\big(x\big) - D_N\big(x + (h_1, 0)\big) + D_N\big(x + (h_1, h_2)\big) - D_N\big(x + (0, h_2)\big) = \\ - N \cdot h_1 h_2 = -N \cdot \frac{|R|}{4}. \end{aligned} \tag{2.13}$$

Indeed, since R is empty, the counting parts will cancel out, and the area parts will yield precisely the area of the rectangle with vertices at the four points in the identity above. Hence, it is easy to see that

$$\int_{R'}\Big(D_N(x) - D_N\big(x+(h_1,0)\big) + D_N\big(x+(h_1,h_2)\big) - D_N\big(x+(0,h_2)\big)\Big)\,dx = \\ - N\cdot\left(\frac{|R|}{4}\right)^2, \quad (2.14)$$

while, on the other hand,

$$\int_{R'}\Big(D_N(x) - D_N\big(x+(h_1,0)\big) + D_N\big(x+(h_1,h_2)\big) - D_N\big(x+(0,h_2)\big)\Big)\,dx = \\ \int_R D_N(x)\cdot h_R(x)\,dx = \langle D_N, h_R\rangle. \quad (2.15)$$

In other words, the inner product of discrepancy with the Haar function supported by an empty rectangle picks up the local discrepancy arising purely as the area of the rectangle.

We are now ready to prove a crucial preliminary lemma.

Lemma 2.1. *Let $\mathcal{P}_N \subset [0,1]^d$ be a distribution of N points and chose $n \in \mathbb{N}$ so that $2^{n-2} \le N < 2^{n-1}$. Then, for any $\vec{r} \in \mathbb{H}_n^d$, there exists an* r*-function $f_{\vec{r}}$ with parameter $\vec{r}$ such that*

$$\langle D_N, f_{\vec{r}}\rangle \ge c_d > 0, \quad (2.16)$$

where the constant c_d depends on the dimension only.

P r o o f. Construct the function $f_{\vec{r}}$ in the following way:

$$f_{\vec{r}} = \sum_{R\in\mathcal{D}^d_{\vec{r}}:\, R\cap\mathcal{P}_N=\emptyset} (-1)\cdot h_R + \sum_{R\in\mathcal{D}^d_{\vec{r}}:\, R\cap\mathcal{P}_N\neq\emptyset} \operatorname{sgn}\big(\langle D_N, h_R\rangle\big)\cdot h_R \quad (2.17)$$

By our choice of n, at least 2^{n-1} of the 2^n rectangles in $\mathcal{D}^d_{\vec{r}}$ must be free of points of $\mathcal{P}_N$. It then follows from (2.10) and (2.11) that

$$\begin{aligned}\langle D_N, f_{\vec{r}}\rangle \ge -\sum_{R\cap\mathcal{P}_N=\emptyset}\langle D_N, h_R\rangle &= \sum_{R\cap\mathcal{P}_N=\emptyset}\langle Nx_1\dots x_d, h_R\rangle \\ &= \sum_{R\cap\mathcal{P}_N=\emptyset} N\cdot\frac{|R|^2}{4^d} \\ &\ge 2^{n-1}\cdot 2^{n-2}\cdot\frac{2^{-2n}}{4^d} = c_d. \quad (2.18)\end{aligned}$$

□

Remark. Roth [61] initially defined the functions $f_{\vec r}$ slightly differently: he set them equal to zero on those dyadic rectangles which do contain points of $\mathcal{P}_N$, i.e., Roth's functions consisted only of the first term of (2.17). While this bears no effect on this argument, it was later realized by Schmidt [74] that in more complex situations it is desirable to have more uniformity in the structure of these building blocks. He simply chose the sign that increases the inner product on non-empty rectangles (the second term in (2.17)). Schmidt's paper, as well as subsequent papers by Halász [39], Beck [7], the author and collaborators [11, 12, 14], make use of the r-functions as defined here (2.7).

Lemma 2.1 produces a rather large collection of *orthogonal*(!) functions such that the projections of D_N onto each of them is big, hence the norm of D_N must be big: this is the punchline of Roth's proof.

Proof of Theorem 1.1. Roth's proof proceeds by duality. Construct the following test function:

$$F = \sum_{\vec r \in \mathbb{H}_n^d} f_{\vec r}\,, \tag{2.19}$$

where $f_{\vec r}$ are the r-functions provided by Lemma 2.1. Orthogonality of $f_{\vec r}$'s yields:

$$\|F\|_2 = \Bigg(\sum_{\vec r \in \mathbb{H}_n^d} \|f_{\vec r}\|_2^2 \Bigg)^{1/2} \; = \; (\#\mathbb{H}_n^d)^{1/2} \approx \; n^{\frac{d-1}{2}}, \tag{2.20}$$

while Lemma 2.1 guarantees that

$$\langle D_N, F\rangle \geq (\#\mathbb{H}_n^d)\cdot c_d \approx n^{d-1}. \tag{2.21}$$

Now Cauchy-Schwarz inequality easily implies that:

$$\|D_N\|_2 \geq \frac{\langle D_N, F\rangle}{\|F\|_2} \gtrsim \; n^{\frac{d-1}{2}} \; \approx \; \big(\log N\big)^{\frac{d-1}{2}}, \tag{2.22}$$

which finishes the proof. $\square$

Second proof of Theorem 1.1. One can formulate a different proof which avoids using duality, although it boils down to the same idea. This proof first appeared in an unpublished manuscript of A. Pollington and was reproduced in [25]. Use the orthogonality of Haar functions, Bessel's inequality and (2.11) to write

$$\begin{aligned}\|D_N\|_2^2 \; &\geq \sum_{|R|=2^{-n},\, R\cap\mathcal{P}_N=\emptyset} \frac{|\langle D_N, h_R\rangle|^2}{|R|} \; = \sum_{\vec r\in\mathbb{H}_n^d} \; \sum_{R\in\mathcal{D}_{\vec r}^d:\, R\cap\mathcal{P}_N=\emptyset} N^2\cdot\frac{2^{-4n}}{2^{-n}\cdot 4^{2d}} \\ &\gtrsim \; (\#\mathbb{H}_n^d)\cdot \; 2^{n-1}\cdot 2^{2n-4}2^{-3n} \; \approx n^{d-1} \; \approx \; \big(\log N\big)^{d-1}. \end{aligned} \tag{2.23}$$

The first line of the above calculation may look a bit odd: this is a consequence of the L^∞ normalization of the Haar functions. $\square$

One can easily extend the first proof to an L^p bound, $1 < p < \infty$, provided that one has the estimate for the L^q norm of the test function F, where q is the dual index to p, i.e., $1/p + 1/q = 1$. Ineed, it will be shown in the next section as a simple consequence of the Littlewood-Paley inequalities that for any $q \in (1, \infty)$ we have the same estimate as for the L^2 norm: $\|F\|_q \approx n^{\frac{d-1}{2}}$, see (3.10) and (3.12). Hence, replacing Cauchy-Schwarz by Hölder's inequality in (2.22), one immediately recovers Schmidt's result [74]:

$$\|D_N\|_p \gtrsim (\log N)^{\frac{d-1}{2}}. \tag{2.24}$$

Schmidt had originally estimated the L^q norms of the function F in the case when $q = 2m$ is an even integer, by using essentially L^2 techniques: squaring out the integrands and analyzing the orthogonality of the obtained terms. We point out also that an analog of the second proof (2.23) can be carried out for the L^p too using the device of the product Littlewood-Paley square function instead of orthogonality. The reader is invited to consult the next section, § 3, for details.

Recently, Hinrichs and Markashin [44] have slightly modified Roth's method to obtain the best known value of the constant C_d in Theorem 1.1: they have noticed that one can extend the summation in (2.23) to include rectangles with larger volume $|R| \geq 2^{-n}$. A careful computation then yields

$$C_2 = \frac{7}{216\sqrt{\log 2}} = 0.038925\ldots \quad \text{and} \quad C_d = \frac{7}{27 \cdot 2^{2d-1}\sqrt{(d-1)!}(\log 2)^{\frac{d-1}{2}}}$$

for $d \geq 3$, where all logarithms are taken to be natural.

3. Littlewood-Paley Theory

While Roth's method in its original form provides sharp information about the behavior of the L^2 norm of the discrepancy function, additional ideas and tools are required in order to extend the result to other function spaces, such as L^p, $1 < p < \infty$. In particular, the L^2 arguments of the previous section made essential use of orthogonality. Therefore, one needs an appropriate substitute for this notion in the case $p \neq 2$. A hands-on approach to this problem has been discovered by Schmidt in [74].

However, harmonic analysis provides a natural tool which allows one to push orthogonality arguments from L^2 to L^p, as well as to more general function spaces. This tool is the so-called Littlewood-Paley theory. In this section, we shall give the necessary definitions, facts, and references relevant to our settings and concentrate on the applications of this theory to the irregularities of distribution.

We start by considering the one-dimensional case. Let f be a measurable function on the interval $[0,1]$. The dyadic (Haar) square function of f is then defined as

$$Sf(x) = \left(\left| \int_0^1 f(t)dt \right|^2 + \sum_{I\in\mathcal{D}} \frac{|\langle f, h_I\rangle|^2}{|I|^2} \mathbf{1}_I(x) \right)^{\frac{1}{2}}$$

$$= \left(\left| \int_0^1 f(t)dt \right|^2 + \sum_{k=0}^{\infty} \left(\sum_{I\in\mathcal{D},\, |I|=2^{-k}} \frac{\langle f, h_I\rangle}{|I|} h_I(x) \right)^2 \right)^{\frac{1}{2}} \tag{3.1}$$

We stress again that the formula may look unusual to a reader familiar with the subject due to the uncommon (L^∞, not L^2) normalization of the Haar functions. To intuitively justify the correctness of this definition, notice that $Sh_I = \mathbf{1}_I$ for any $I \in \mathcal{D}$. In particular, if the function has the form $f = \sum_{I\in\mathcal{D}} a_I h_I$, then its square function is

$$Sf = \left(\sum_{I\in\mathcal{D}} a_I^2 \mathbf{1}_I \right)^{\frac{1}{2}} = \left(\sum_{k=0}^{\infty} \left(\sum_{I\in\mathcal{D}\colon |I|=2^{-k}} a_I h_I \right)^2 \right)^{\frac{1}{2}}. \tag{3.2}$$

Since Haar functions (together with the constant $\mathbf{1}_{[0,1]}$) form an orthogonal basis of $L^2[0,1]$, Parseval's identity immediately implies that

$$\|Sf\|_2 = \|f\|_2. \tag{3.3}$$

A non-trivial generalization of this fact to an equivalence of L^p norms, $1<p<\infty$, is referred to as the Littlewood-Paley inequalities. They can also be viewed as a generalization of the famous Khintchine inequality:

Theorem 3.1 (Littlewood-Paley inequalities, [88]). *For $1 < p < \infty$, there exist constants $B_p > A_p > 0$ such that for every function $f \in L^p[0,1]$ we have*

$$A_p\|Sf\|_p \le \|f\|_p \le B_p\|Sf\|_p. \tag{3.4}$$

The asymptotic behavior of the constants A_p and B_p is known [88] and is very useful in numerous arguments, especially when (3.4) is applied for very high values of p. In particular $B_p \approx \sqrt{p}$ when p is large. Also, a simple duality argument shows that $A_q = B_p^{-1}$, where q is the dual index of p. The reader is invited to consult the following references for more details: [88, 78, 16].

What is very important is that these inequalities continue to hold for the Hilbert space-valued functions (in this case, all the arising integrals are understood as Bochner integrals). This delicate fact allows one to extend the

Littlewood-Paley inequalities to the multi-parameter setting in a fairly straightforward way by successively applying (3.4) in each dimension while treating the other dimensions as vector-valued coefficients [59, 78]. We note that in the general case one would apply the one dimensional Littlewood-Paley inequality d times (once in each coordinate). However, in the setting introduced by Roth's method (where the attention is restricted to dyadic boxes R of fixed volume $|R| = 2^{-n}$) one would apply it only $d-1$ times since this is the number of the free parameters—once the lengths of $d-1$ sides are specified, the last one is determined automatically by the condition $|R| = 2^{-n}$.

Rather then stating the relevant inequalities in full generality (which an interested reader may find in [59, 11]), we illustrate the use of this approach by a simple example, important to the topic of our discussion. Recall that the test function (2.19) in Roth's proof was constructed as

$$F = \sum_{\vec{r}\in\mathbb{H}_n^d} f_{\vec{r}} = \sum_{R:\,|R|=2^{-n}} \varepsilon_R h_R\,,$$

where $\varepsilon_R = \pm 1$. We want to estimate the L^q norm of F. Applying the one-dimensional Littlewood-Paley inequality in the first coordinate x_1, we obtain

$$\|F\|_q = \Bigg\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \Bigg\|_q \le B_q \Bigg\| \Bigg[\sum_{r_1=1}^{n} \Bigg| \sum_{\substack{|R|=2^{-n}\\ |R_1|=2^{-r_1}}} \varepsilon_R h_R \Bigg|^2 \Bigg]^{1/2} \Bigg\|_q . \tag{3.5}$$

In the two-dimensional case for any value of r_1 all the rectangles satisfying the conditions of the innermost summation are disjoint, and for each point x we have:

$$\begin{aligned} \sum_{r_1=1}^{n} \Bigg| \sum_{\substack{|R|=2^{-n}\\ |R_1|=2^{-r_1}}} \varepsilon_R h_R(x) \Bigg|^2 &= \sum_{r_1=1}^{n} \sum_{\substack{|R|=2^{-n}\\ |R_1|=2^{-r_1}}} |\varepsilon_R|^2 \mathbf{1}_R(x) \\ &= \sum_{R\in\mathcal{D}^2,\,|R|=2^{-n}} \mathbf{1}_R(x) = \#\mathbb{H}_n^2 \approx n, \end{aligned} \tag{3.6}$$

since $\varepsilon_R^2 = 1$ and every point is contained in $\#\mathbb{H}_n^2$ dyadic rectangles (one per each shape).

In the case $d \ge 3$, the expression on the right-hand side of (3.5) can be viewed as a Hilbert space-valued function. Indeed, fix all the coordinates except x_2 and define an ℓ^2-valued function

$$F_2(x_2) = \sum_{I\in\mathcal{D}} \left\{ \sum_{\substack{|R|=2^{-n},\, R_2=I\\ |R_1|=2^{-r_1}}} \varepsilon_R \prod_{j\neq 2} h_{R_j}(x_j) \right\}_{r_1=1}^{n} h_I(x_2)\,. \tag{3.7}$$

Then the expression inside the L^q norm on the right hand side of (3.5) is exactly $|F|_{\ell^2}$. Applying the Hilbert space-valued Littlewood-Paley inequality in the second coordinate, we get

$$\begin{aligned} \|F\|_q &= \Big\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \Big\|_q \le B_q \big\| \, |F_2|_{\ell^2} \big\|_q \\ &\le B_q^2 \left\| \left[\sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \left| \sum_{\substack{|R|=2^{-n}\\ |R_j|=2^{-r_j}\,,\ j=1,2}} \varepsilon_R h_R \right|^2 \right]^{1/2} \right\|_q . \end{aligned} \tag{3.8}$$

And if $d=3$, then analog of (3.6) holds, completing the proof in this case. In the case of general d we continue applying the vector-valued Littlewood-Paley inequalities inductively in a similar fashion a total of $d-1$ times to obtain

$$\begin{aligned} \|F\|_q &= \Big\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \Big\|_q \le \cdots \\ \cdots &\le B_q^{d-1} \left\| \left[\sum_{r_1=1}^{n} \cdots \sum_{r_{d-1}=1}^{n} \left| \sum_{\substack{|R|=2^{-n}\\ |R_j|=2^{-r_j}\,,\ j=1,\dots,d-1}} \varepsilon_R h_R \right|^2 \right]^{1/2} \right\|_q . \end{aligned} \tag{3.9}$$

Just as explained in (3.6), in this case all the rectangles in the innermost summation are disjoint and thus

$$\|F\|_q \le B_q^{d-1} \left\| \left[\sum_{R\in\mathcal{D}^d,\, |R|=2^{-n}} |\varepsilon_R|^2 \mathbf{1}_R \right]^{\frac12} \right\|_q = B_q^{d-1} \left(\#\mathbb{H}_n^d\right)^{\frac12} \approx n^{\frac{d-1}{2}}\,. \tag{3.10}$$

Remark. For a function $f = \sum_{R \in \mathcal{D}^d} a_R h_R$ on $[0,1]^d$, the expression

$$S_d f(x) = \left[\sum_{R \in \mathcal{D}^d,\, |R| = 2^{-n}} |a_R|^2 \mathbf{1}_R(x) \right]^{\frac{1}{2}}$$

is called the *product* dyadic square function of f. The product Littlewood-Paley inequalities (whose proof is essentially identical to the argument presented above) state that

$$A_p^d \, \|S_d f\|_p \le \|f\|_p \le B_p^d \, \|S_d f\|_p \,. \tag{3.11}$$

With these inequalities at hand, one can estimate the L^q norm of F in a single line:

$$\begin{aligned} \|F\|_q &= \left\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \right\|_q \approx \|S_d f\|_q \\ &= \left\| \left[\sum_{R \in \mathcal{D}^d,\, |R| = 2^{-n}} |\varepsilon_R|^2 \mathbf{1}_R \right]^{\frac{1}{2}} \right\|_q = \left(\# \mathbb{H}_n^d \right)^{\frac{1}{2}} \approx n^{\frac{d-1}{2}}. \end{aligned} \tag{3.12}$$

We chose to include the illustrative proof of this estimate in order to demonstrate the essence of the product Littlewood-Paley theory. In addition, the argument leading to (3.10) gives a better implicit constant than the general inequalities (B_q^{d-1} versus B_q^d, according to the number of free parameters). While we generally are not concerned with the precise values of constants in this note, the behavior of this particular constant plays an important role in some further estimates, see (3.19).

The proof of of Schmidt's L^p lower bound (1.6) can now be finished immediately. Let q be the dual index of p, i.e. $1/p + 1/q = 1$ and let F be as defined in (2.19). Then, replacing Cauchy-Schwarz with Hölder's inequality in (2.22) and using (3.12), we obtain:

$$\|D_N\|_p \ge \frac{\langle D_N, F \rangle}{\|F\|_q} \gtrsim n^{\frac{d-1}{2}} \approx \big(\log N \big)^{\frac{d-1}{2}}. \tag{3.13}$$

An analog of the second proof (2.23) of Roth's estimate (1.5) can also be carried out easily using the Littlewood-Paley square function. We include it since it provides a foundation for discrepancy estimates in other function spaces. It is particularly useful when one deals with quasi-Banach spaces and is forced to avoid duality arguments. We start with a simple lemma:

Lemma 3.2. *Let $A_k \subset [0,1]^d$, $k \in \mathbb{N}$, satisfy $\mu(A_k) \geq c$, where μ is the Lebesgue measure, then for any $M \in \mathbb{N}$*

$$\mu\left(\left\{x \in [0,1]^d : \sum_{k=1}^{M} \mathbf{1}_{A_k}(x) > \frac{1}{2}cM\right\}\right) > \frac{1}{2}c.$$

P r o o f. Assume this is not true, then we immediately arrive to a contradiction

$$\begin{aligned} cM &\leq \int \sum_{k=1}^{M} \mathbf{1}_{A_k}(x)\, dx \\ &\leq \frac{1}{2}cM \cdot \mu\Bigg(\sum_{k=1}^{M} \mathbf{1}_{A_k} < \frac{1}{2}cM\Bigg) + M \cdot \mu\Bigg(\sum_{k=1}^{M} \mathbf{1}_{A_k} > \frac{1}{2}cM\Bigg) \\ &< \frac{1}{2}cM + M \cdot \frac{1}{2}c = cM\,. \end{aligned}$$

We shall apply the lemma as follows: for each $\vec{r} \in \mathbb{H}_n^d$, let $A_{\vec{r}}$ be the union of rectangles $R \in \mathcal{D}_{\vec{r}}^d$ which do not contain points of $\mathcal{P}_N$. Then $\mu(A_{\vec{r}}) \geq c = \frac{1}{2}$ and $M = \#\mathbb{H}_n^d \approx n^{d-1}$. Let $E \subset [0,1]^d$ be the set of points where at least $M/4$ empty rectangles intersect. By the lemma above, $\mu(E) > \frac{1}{4}$. On this set, using (2.12):

$$S_d D_N(x) = \left[\sum_{R \in \mathcal{D}^d} \frac{\langle D_N, h_R\rangle^2}{|R|^2} \mathbf{1}_R(x)\right]^{\frac{1}{2}} \gtrsim \Big(M \cdot N^2\, 2^{-2n}\Big)^{\frac{1}{2}} \approx n^{\frac{d-1}{2}}. \tag{3.14}$$

Integrating this estimate over E finishes the proof of (1.6)

$$\|D_N\|_p \gtrsim \|S_d D_N\|_p \gtrsim n^{\frac{d-1}{2}} \approx (\log N)^{\frac{d-1}{2}}.$$

$\square$

3.1. Lower bounds in other function spaces

The use of the Littlewood-Paley theory opens the door to considering much wider classes of functions than just the L^p spaces. There has been a splurge of activity in this direction. We shall give a very brief overview of the results and conjectures related to various function spaces. All of the results described below are direct descendants of Theorem 1.1 and Roth's method as every single one of them makes use of the Haar coefficients of the discrepancy function.

In particular, a direct extension of the above argument provides a lower bound of the discrepancy function in product Hardy spaces H^p, $0 < p \leq 1$. These spaces are generalizations of the classical classes introduced by Hardy, see [17, 18]. The norm of these spaces can be characterized using the square function:

$$\|f\|_{H^p} \approx \|S_d f\|_p.$$

We have the following result due to Lacey [49]: for $0 < p \leq 1$,

$$\|\widetilde{D}_N\|_{H^p} \geq C_{d,p}(\log N)^{\frac{d-1}{2}}, \tag{3.15}$$

where $\widetilde{D}_N = \sum_{R\in\mathcal{D}^d} \frac{\langle D_N, h_R\rangle}{|R|} h_R$, i.e. the discrepancy function D_N modified as to have mean zero over every subset of coordinates. The proof of this result is a verbatim repetition of the previous proof—one simply estimates the norm of the square function. As this example clearly illustrates, in harmonic analysis Hardy spaces H^p serve as a natural substitute for L^p spaces when $p \leq 1$. Numerous analytic tools, such as square functions, maximal functions, atomic decompositions [78], allow one to extend the L^p estimates to the H^p setting for $0 < p \leq 1$. The L^p behavior of the discrepancy function for this range of p, however, still remains a mystery. It is conjectured that the L^p norm should obey the same asymptotics in N for *all* values of $p > 0$.

Conjecture 3.3. *For all $p \in (0,1]$ the discrepancy function satisfies the estimate*

$$\|D_N\|_p \geq C_{d,p}(\log N)^{\frac{d-1}{2}}. \tag{3.16}$$

The only currently available information regarding this conjecture is the result of Halász [39] who proved that it indeed holds in dimension $d = 2$ for the L^1 norm:

$$\|D_N\|_1 \geq C\sqrt{\log N}. \tag{3.17}$$

We shall discuss Halász's ingenious method in § 4. Halász was able to extend this inequality to higher dimensions, but only with the same right-hand side. Thus it is not known whether the L^1 bound even grows with the dimension. As to the case $p < 1$, no information whatsoever is available at this time.

In attempts to get close to L^1, Lacey [49] has proved that if one replaces L^1 with the Orlicz space $L(\log L)^{\frac{d-2}{2}}$, then the conjectured bound holds

$$\|D_N\|_{L(\log L)^{(d-2)/2}} \geq C_d(\log N)^{\frac{d-1}{2}}. \tag{3.18}$$

We remark that an adaptation of the proof of Schmidt's L^p bound given in the previous subsection, specifically estimate (3.10), can easily produce a slightly weaker inequality

$$\|D_N\|_{L(\log L)^{(d-1)/2}} \geq C_d(\log N)^{\frac{d-1}{2}} \tag{3.19}$$

Indeed, let F once again be as defined in (2.19). It is well known that (see, e.g., [55]) the dual of $L(\log L)^{(d-1)/2}$ is the exponential Orlicz space $\exp(L^{2/(d-1)})$. Hence we need to estimate the norm of F in this space. It is also a standard fact that

$$\|F\|_{\exp(L^\alpha)} \approx \sup_{q>1} q^{-1/\alpha}\|F\|_q \tag{3.20}$$

(this can be established using Taylor series for e^x and Stirling's formula). We recall that the constant arising in the Littlewood-Paley inequalities (3.4) is $B_q \approx \sqrt{q}$ for large q and the implicit constant in (3.10) is B_q^{d-1}. Thus we obtain

$$\begin{aligned} \|F\|_{\exp(L^{2/(d-1)})} &\lesssim \sup_{q>1} q^{-(d-1)/2} \cdot B_q^{d-1} n^{\frac{d-1}{2}} \\ &\approx \sup_{q>1} q^{-(d-1)/2} \cdot q^{(d-1)/2} n^{\frac{d-1}{2}} = n^{\frac{d-1}{2}}, \end{aligned} \tag{3.21}$$

and (3.19) immediately follows by duality. These estimates are similar in spirit to the famous Chang-Wilson-Wolff inequality [19].

In a different vein, Triebel has recently studied the behavior of the discrepancy in Besov spaces [86, 87]. He proves, among other things, that

$$\begin{gathered} \|D_N\|_{S^r_{pq}B([0,1]^d)} \geq C_{d,p,q,r}\, N^r \left(\log N\right)^{\frac{d-1}{q}}, \\ 1 < p,\, q < \infty, \qquad \frac{1}{p} - 1 < r < \frac{1}{p}. \end{gathered} \tag{3.22}$$

Here the space $S^r_{pq}B([0,1]^d)$ is the Besov space with dominating mixed smoothness. The exact definition of this class (which can be stated in terms very reminiscent of the Littlewood-Paley square function Sf) is technical and would take our discussion far afield. We would only mention that the index p represents integrability, r measures smoothness, and q is a certain correction index. In particular, the case $q = 2$ corresponds to the well-known Sobolev spaces which, roughly speaking, consist of functions, with r^{th} mixed derivative in L^p. Furthermore, when $r = 0$, one recovers Schmidt's L^p estimates. At least in dimension $d = 2$, the estimates (3.1) are sharp [43], see § 6. For more details, the reader is directed towards Triebel's recent book [87] concentrating on discrepancy and numerical integration in this context as well as to his numerous other famous books for a comprehensive treatise of the theory of function spaces in general.

The recent work of Ou [58] deals with the growth of the discrepancy function in weighted L^p spaces. A non-negative measurable function ω on $[0,1]^d$ is called an A_p *(dyadic product) weight* if

$$\sup_{R\in\mathcal{D}^d} \left(\int_R \omega\right)\left(\int_R \omega^{-\frac{1}{p-1}}\right)^{p-1} < \infty. \tag{3.23}$$

The space $L^p(\omega)$ is then defined as the L^p space with respect to the measure $\omega(x)\,dx$. The A_p weights play a tremendously important role in harmonic analysis: they give the largest reasonable class of measures such that the standard boundedness properties of classical operators (such as maximal functions, singular integrals, square functions) continue to hold in L^p spaces with respect to these measures. By an adaptation of the square function argument (3.14), Ou was able to show that

$$\|D_N\|_{L^p(\omega)} \geq C_{d,p,\omega} (\log N)^{\frac{d-1}{2}}, \tag{3.24}$$

i.e., the behavior in weighted L^p spaces is essentially the same as in their Lebesgue-measure prototypes.

Approaching the other end of the L^p scale in attempts to understand the precise nature of the kink that occurs at the passage from the average (L^p) to the maximum (L^∞) norm, Bilyk, Lacey, Parissis, and Vagharshakyan [14] computed the lower bounds of the discrepancy function in spaces which are "close" to L^∞. One such space is the product dyadic BMO (which stands for *bouded mean oscillation*). This space, introduced by Chang and Fefferman [17], is a proper generalization of the classical BMO space to the multiparameter setting. In particular, the famous H^1 – BMO duality is preserved. Just as H^1 often serves as a natural substitute for L^1, in many problems of harmonic analysis BMO naturally replaces L^∞. However, Bilyk, Lacey, Parissis, and Vagharshakyan showed that in this case the BMO norm behaves like L^p norms rather then L^∞:

$$\|D_N\|_{\mathrm{BMO}} \geq C_d (\log N)^{\frac{d-1}{2}}. \tag{3.25}$$

In fact, this estimate is not hard to obtain with the help of the same test function F (2.19) that we have used several times already—all we have to do is to estimate its dual (H^1) norm. Just as in (3.12):

$$\begin{aligned} \|F\|_{H^1} \approx \|SF\|_1 &= \left\| \left[\sum_{R \in \mathcal{D}^d,\, |R| = 2^{-n}} |\varepsilon_R|^2 \mathbf{1}_R \right]^{\frac{1}{2}} \right\|_1 \\ &= \left(\#\mathbb{H}_n^d \right)^{\frac{1}{2}} \approx n^{\frac{d-1}{2}}, \end{aligned} \tag{3.26}$$

which immediately yields the result. In addition, the authors prove lower bounds in the aforementioned exponential Orlicz spaces. These spaces $\exp(L^\alpha)$ serve as an intermediate scale between the L^p spaces, $p < \infty$, and L^∞. The following estimate is contained in [14]: *in dimension $d = 2$ for all $2 \leq \alpha < \infty$ we have*

$$\|D_N\|_{\exp(L^\alpha)} \geq C (\log N)^{1 - \frac{1}{\alpha}}. \tag{3.27}$$

We note that this inequality can be viewed as a smooth interpolation of lower bounds between L^p and L^∞. Indeed, when $\alpha = 2$ (the subgaussian case $\exp(L^2)$), the estimate is $\sqrt{\log N}$—the same as in L^2. On the other hand, as α approaches infinity, the right hand side approaches the L^∞ bound—$\log N$. The proof of this estimate closely resembles Halász's proof of the L^∞ bound (see (4.11) below), with the obvious modification that the test function has to be estimated in the dual space $L(\log L)^{1/\alpha}$. Hence the same problems and obstacles that arise when dealing with star-discrepancy prevent straightforward extensions of this estimate to higher dimensions. We finish this discussion by mentioning that both of these estimates were shown to be sharp, see § 6.

4. The star-discrepancy (L^∞) lower bounds and the small ball inequality

We now turn our attention to the most important case: L^∞ bounds of the discrepancy function. As explained in the introduction, when the set $\mathcal{P}_N$ is distributed rather well, its discrepancy comes close to its maximal values only on a thin set, while staying relatively small on most of $[0,1]^d$. Therefore the extremal L^∞ norm of this function has to be much larger than the averaging L^2 norm. This heuristic was first confirmed by Schmidt [71] who proved

$$\|D_N\|_\infty \geq C \log N. \tag{4.1}$$

Other proofs of this inequality have been later given by Liardet [54, 1979], Béjian [9, 1982] (who produced the best known value of the constant $C = 0.06$), and Halász [39, 1981]. The proof of Halász is the most relevant to the topic of the present survey as it relies on Roth's orthogonal function idea and takes it to a new level. However, before we proceed to Halász's proof of Schmidt's lower bound, we shall discuss another related inequality.

The *small ball inequality*, which arises naturally in probability and approximation, besides being important and significant in its own right, also serves as a model for the lower bounds for the star-discrepancy (1.10). This inequality is concerned with the lower estimates of the supremum norm of linear combinations of multivariate Haar functions supported by dyadic boxes of fixed volume (we call such sums 'hyperbolic') and can be viewed as a reverse triangle inequality. It is linked to the discrepancy function through Roth's orthogonal function method. Although no formal connections are known, most arguments designed for this inequality can be transferred to the discrepancy setting. We now state the conjecture:

Conjecture 4.1 (The Small Ball Conjecture). *In dimensions $d \geq 2$, for any choice of the coefficients α_R one has the following inequality:*

$$n^{\frac{d-2}{2}} \left\| \sum_{R \in \mathcal{D}^d \colon |R| = 2^{-n}} \alpha_R h_R \right\|_\infty \gtrsim 2^{-n} \sum_{R \colon |R| = 2^{-n}} |\alpha_R|. \tag{4.2}$$

The point of interest in this conjecture is the precise exponent of n on the left-hand side. If one replaces $n^{(d-2)/2}$ by $n^{(d-1)/2}$, this inequality becomes almost trivial, and, in fact, holds for the L^2 norm. Indeed, using the orthogonality of Haar functions and keeping in mind that $\|h_R\|_2 = |R|^{1/2}$, we obtain

$$\begin{aligned} \Bigg\| \sum_{R\in\mathcal{D}^d:\, |R|=2^{-n}} \alpha_R h_R \Bigg\|_2 &= \Bigg(\sum_{|R|=2^{-n}} |\alpha_R|^2 2^{-n} \Bigg)^{\frac{1}{2}} \\ &\gtrsim \frac{\sum_{|R|=2^{-n}} |\alpha_R| 2^{-n/2}}{\left(n^{d-1}2^n\right)^{\frac{1}{2}}} \\ &= n^{-\frac{d-1}{2}} \cdot 2^{-n} \sum_{|R|=2^{-n}} |\alpha_R|, \end{aligned} \tag{4.3}$$

where in the last line we have used the Cauchy-Schwarz inequality and the fact that the number of terms in the sum is of the order $n^{d-1}2^n$. The presence of the quantity $d-1$ in this context is absolutely natural, as it is, in fact, the number of free parameters (dictated by the condition $|R| = 2^{-n}$). The passage to $d-2$ for the L^∞ norm requires a much deeper analysis and brings out a number of complications.

This result and the conjecture should be compared to Roth's L^2 discrepancy estimate (1.5) and Conjecture 1.3. The computation just presented is very close the proof (2.23) of (1.5). In fact, the resemblance becomes even more striking if one restricts the attention to the case when all the coefficients $\alpha_R = \pm 1$. In this case $2^{-n}\sum_{|R|=2^{-n}} |\alpha_R| \approx n^{d-1}$ and the L^2 estimate (4.3) becomes

$$\Bigg\| \sum_{R\in\mathcal{D}^d:\, |R|=2^{-n}} \alpha_R h_R \Bigg\|_2 \gtrsim n^{\frac{d-1}{2}}, \tag{4.4}$$

while the conjectured L^∞ inequality 4.2 for $\alpha = \pm 1$ turns into

$$\Bigg\| \sum_{R\in\mathcal{D}^d:\, |R|=2^{-n}} \alpha_R h_R \Bigg\|_\infty \gtrsim n^{\frac{d}{2}}. \tag{4.5}$$

Recalling that n in Roth's argument was chosen to be approximately $\log_2 N$, one immediately sees the similarity of these inequalities to (1.5) and (1.10).

Choosing α_R's to be either independent Gaussian random variables or independent random signs $\alpha_R = \pm 1$ verifies that this conjecture is sharp, see, e.g., [11]. This provides one of the reasons to believe that the correct estimate for the star-discrepancy should be Conjecture 1.3: $\|D_N\|_\infty \gtrsim (\log N)^{d/2}$.

We shall now illustrate the connection between this inequality and the discrepancy estimates.

The Small Ball Conjecture has been proved in $d = 2$ by M. Talagrand [79] in 1994. In 1995, V. Temlyakov [82] (see also [83, 84]) has given another, very elegant proof of this inequality in two dimensions, which closely resembled the argument of Halász [39] for (1.7). We shall present Temlyakov's proof first as it is somewhat "cleaner" and avoids technicalities. Then we shall explain which adjustments need to be made in order translate this argument to a proof of Schmidt's estimate.

The proof is based on Riesz products. An important feature of the two-dimensional case is the following *product rule*: if R, $R' \in \mathcal{D}^2$ are not disjoint, $R \neq R'$, and $|R| = |R'|$, then

$$h_R \cdot h_{R'} = \pm h_{R \cap R'}, \tag{4.6}$$

i.e., the product of two Haar functions is again Haar. The proof of this fact is straightforward. Unfortunately, this rule does not hold in higher dimensions. Indeed, for $d \geq 3$ one can have two different boxes of the same volume which coincide in one of the coordinates, say $R_1 = R_1'$. Then, $h_{R_1} \cdot h_{R_1'} = h_{R_1}^2 = \mathbf{1}_{R_1}$– –we lose orthogonality in the first coordinate. The fact that the product rule fails in higher dimensions is a major obstruction on the path to solving the conjecture.

P r o o f o f t h e s m a l l b a l l c o n j e c t u r e i n d i m e n s i o n $d = 2$. For each $k = 0, \ldots, n$ consider the r-functions $f_k = \sum_{|R|=2^{-n},\, |R_1|=2^{-k}} \operatorname{sgn}(\alpha_R) h_R$. Obviously, in two dimensions, the conditions $|R| = 2^{-n}$ and $|R_1| = 2^{-k}$ uniquely define the shape of a dyadic rectangle. We are now ready to construct the test function as a Riesz product:

$$\Psi := \prod_{k=1}^{n} \Big(1 + f_k \Big). \tag{4.7}$$

First of all, notice that Ψ is non-negative. Indeed, since f_k's only take the values ± 1, each factor above is equal to either 0 or 2. Thus, we can say even more than $\Psi \geq 0$: the only possible values of Ψ are 0 and 2^{n+1}. Next, we observe that $\int \Psi(x)dx = 1$. This can be explained as follows. Expand the product in (4.7). The leading term is equal to 1. All the other terms are products of Haar functions; therefore, by the product rule, they themselves are Haar functions and have integral zero. So, Ψ is a non-negative function with integral 1. In other words, it has L^1 norm 1: $\|\Psi\|_1 = 1$.

A similar argument applies to the inner product of $\sum_{|R|=2^{-n}} \alpha_R h_R$ and Ψ. Multiplying out the product in (4.7) and using the product rule, one can see that

$$\Psi = 1 + \sum_{R\in\mathcal{D}^d:\,|R|=2^{-n}} \operatorname{sgn}(\alpha_R) h_R \,+\, \Psi_{>n}, \tag{4.8}$$

where $\Psi_{>n}$ is a linear combination of Haar functions supported by rectangles of area less than 2^{-n}. The first and the third term are orthogonal to $\sum_{|R|=2^{-n}} \alpha_R h_R$. Hence, using Hölder's inequality,

$$\begin{aligned}
\Bigg\| \sum_{R\in\mathcal{D}^d:\,|R|=2^{-n}} \alpha_R h_R \Bigg\|_\infty &\geq \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \Psi \Big\rangle \\
&= \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \sum_{|R|=2^{-n}} \operatorname{sgn}(\alpha_R) h_R \Big\rangle \qquad (4.9)\\
&= \sum_{R\in\mathcal{D}^d:\,|R|=2^{-n}} \alpha_R \cdot \operatorname{sgn}(\alpha_R) \cdot \|h_R\|_2^2 \\
&= 2^{-n} \cdot \sum_{|R|=2^{-n}} |\alpha_R|, \qquad (4.10)
\end{aligned}$$

and we are done (notice that for $d = 2$ we have $n^{\frac{d-2}{2}} = 1$). □

Proof of Schmidt's lower bound for the star-discrepancy. We now explain how the same idea can be used to prove a discrepancy estimate. In place of the r-functions f_k used above, we shall utilize the r-functions $f_k = \sum_{|R|=2^{-n}} \varepsilon_R h_R$ such that $\langle D_N, f_k\rangle \geq c$, which were used in Roth's proof (2.22) of the L^2 estimate (1.5) and whose existence is guaranteed by Lemma 2.1. The test function is then constructed in a fashion very similar to (4.7):

$$\Phi := \prod_{k=0}^{n} \Big(1 + \gamma f_k\Big) - 1 = \gamma \sum_{k=0}^{n} f_k \,+\, \Phi_{>n}, \tag{4.11}$$

where $\gamma > 0$ is a small constant, and $\Phi_{>n}$ is in the span of Haar functions with support of area less than 2^{-n}. In complete analogy with the previous proof, we have that $\|\Phi\|_1 \leq 2$. Also,

$$\Big\langle D_N, \gamma \sum_{k=0}^{n} f_k \Big\rangle \geq\, c\gamma(n+1) \geq\, C'\gamma \log N. \tag{4.12}$$

Up to this point the proof repeated the proof of the two-dimensional small ball conjecture verbatim. In this regard, one can view the small ball inequality as the linear part of the star-discrepancy estimate. Notice that subtracting 1 in the definition of Φ eliminated the need to estimate the "constant" term $\int D_N(x)dx$.

All that remains is to show that the higher-order terms $\Phi_{>n}$ yield a smaller input. This can be done by "brute force". We first prove an auxiliary lemma valid in every dimension. This lemma is closely related to Lemma 2.1

Lemma 4.2. *Let $f_{\vec{s}}$ be any* r*-function with parameter $\vec{s}$. Denote $s = \|\vec{s}\|_1$. Then, for some constant $\beta_d > 0$,*

$$\langle D_N, f_{\vec{s}} \rangle \le \beta_d N 2^{-s}. \tag{4.13}$$

P r o o f. It follows from (2.11), that the counting part of D_N satisfies

$$|\langle N x_1 \cdot x_d, f_{\vec{s}} \rangle| \lesssim 2^s \cdot N 2^{-2s} = N 2^{-s}.$$

As to the counting part, it follows from the proof of Lemma 2.1 that $\mathbf{1}_{[p,\vec{1}]}$ is orthogonal to the functions h_R for all $R \in \mathcal{D}^d_{\vec{s}}$ except for the rectangle R which contains the point p. It is easy to check that $\langle \mathbf{1}_{[p,\vec{1}]}, f_{\vec{s}} \rangle = \langle \mathbf{1}_{[p,\vec{1}]}, h_R \rangle \lesssim |R| = 2^{-s}$. The estimate for the counting part of D_N then follows by summing over all the points of $\mathcal{P}_N$. □

We now estimate the higher order terms in $\langle D_N, \Phi \rangle$. Write

$$\Phi_{>n} = F_2 + F_3 + \cdots + F_n,$$

where

$$F_k = \gamma^k \sum_{0 \le j_1 < j_2 < \cdots < j_k \le n} f_{j_1} \cdot f_{j_2} \ldots f_{j_k}.$$

Notice that, due to the product rule, the product $f_{j_1} \cdot f_{j_2} \ldots f_{j_k}$ is an r-function with parameter $\vec{s} = (n - j_1, j_k)$, so $s = n - j_1 + j_k$. We reorganize he sum according to the parameter s, $n+1 \le s \le 2n$. To obtain a term which yields an r-function corresponding to a fixed value of s, we need to have $j_k = j_1 + s - n \le n$. This can be done in $2n - s + 1$ ways ($j_1 = 0, \ldots, 2n - s$). For each such choice of j_1 and j_k we can choose the "intermediate" $k-2$ values in $\binom{s-n-1}{k-2}$ ways. Notice that we must have $2 \le k \le s - n + 1$. We obtain

$$\begin{aligned}
\langle D_N, \Phi_{>n} \rangle &= \sum_{k=2}^{n} \langle D_N, F_k \rangle = \sum_{s=n+1}^{2n} (2n - s + 1) \sum_{k=2}^{s-n+1} \binom{s-n-1}{k-2} \cdot \gamma^k \cdot \beta_2 N 2^{-s} \\
&\le \beta_2 n \sum_{s=n+1}^{2n} \gamma^2 (1+\gamma)^{s-n-1} \, N \, 2^{-s} \le \frac{1}{4} \beta_2 \gamma^2 n \sum_{s=n+1}^{\infty} \left(\frac{1+\gamma}{2} \right)^{s-n-1} \\
&= \frac{\gamma^2 \beta_2}{2(1-\gamma)} n,
\end{aligned}$$

where we used that $N \le 2^{n-1}$. Since $n \le \log_2 N + 2$, by making γ very small we can assure that this quantity is less than $\frac{1}{2} C' \gamma \log N$, a half of (4.12). We finally obtain that:

$$\|D_N\|_\infty \geq \frac{1}{2}\langle D_N, \Phi\rangle \geq \frac{1}{2}\Big(C'\gamma \log N - \frac{1}{2}C'\gamma \log N\Big) \gtrsim \log N, \tag{4.14}$$

which finishes the proof of Schmidt's bound. □

We would like to point out that it is not surprising that the Riesz product approach is successful in these problems. As discussed earlier, the extremal values of the discrepancy function (as well as of hyperbolic Haar sums) are achieved on a very thin set. Riesz products are known to capture such sets extremely well. In fact, we can see that Temlyakov's test function $\Psi = 2^{n+1}\mathbf{1}_E$, where E is the set on which all the functions f_k are positive, and in particular the L^∞ norm is attained. We shall make a further remark about the structure of this set E in § 6.

To reinforce the potency of the powerful blend of Roth's method and the Riesz product techniques, we describe the proof of the L^1 lower bound (1.14) for the discrepancy function contained in the same fascinating paper by Halász [39] (while the L^∞ bound was already known, this result was completely new at the time). This argument introduces another brilliant idea: using complex numbers. The test function is constructed as follows:

$$\Gamma := \prod_{k=0}^{n}\Big(1 + \frac{i\gamma}{\sqrt{\log N}} f_k\Big) - 1 = \frac{i\gamma}{\sqrt{\log N}}\sum_{k=0}^{n} f_k \; + \Gamma_{>n}\,, \tag{4.15}$$

where a small constant $\gamma > 0$ and the "-1" in the end play the same role as in the previous argument, and $\Gamma_{>n}$ is the sum of the higher-order terms. Then one can see that

$$\|\Gamma\|_\infty \leq \Big(1 + \frac{\gamma^2}{\log N}\Big)^{\frac{n}{2}} + 1 \leq e^{\gamma^2/2} + 1 \lesssim 1. \tag{4.16}$$

Just as before, one can show that the input of $\Gamma_{>n}$ will be small provided that γ is small enough. Hence,

$$\|D_N\|_1 \gtrsim |\langle D_N, \Gamma\rangle| \gtrsim \frac{\gamma}{\sqrt{\log N}}\langle D_N, \sum_{k=0}^{n} f_k\rangle \gtrsim \frac{n+1}{\sqrt{\log N}} \approx \sqrt{\log N}. \tag{4.17}$$

The absolute efficiency of Riesz products in the two-dimensional case of these problems is justified by the fact that the condition $|R| = 2^{-n}$ effectively leaves only one free parameter (e.g., the value of $|R_1|$ defines the shape of the rectangle) and creates lacunarity ($|R_1| = 2^{-k}$, $k = 0, 1, \ldots, n$). Historically, Riesz products were specifically designed to work in such settings (lacunary Fourier series, see, e.g., [89], [60, 1918]). From the probabilistic point of view, Riesz products work best when the factors behave similarly to independent random variables, which relates perfectly to our problems for $d = 2$, since the functions f_k actually are independent random variables. The failure of the product rule explains the loss of independence in higher dimensions. This approach towards Conjecture 4.1 is taken in [15].

While the failure of the product rule or lack of independence are huge obstacles to the Riesz product method in higher dimensions, they are not intrinsic to our problems. However, there are direct indications that the small ball inequality is much more difficult and delicate in dimensions $d \geq 3$ than in $d = 2$. Consider the signed ($\alpha_R = \pm 1$) case, see (4.5). In this case, at every point $x \in [0,1]^d$ the sum on the left-hand side has $\#\mathbb{H}_n^d \approx n^{d-1}$ terms, while the right-hand side of the inequality is $n^{d/2}$. In dimension $d = 2$, these two numbers are equal, which means that the L^∞ norm is achieved at those points where essentially all the terms have the same sign (the function Ψ finds precisely those points). In dimensions $d \geq 3$ on the other hand, n^{d-1} is much greater than $n^{d/2}$, while we know that the conjecture is sharp. This means that for certain choices of coefficients, very subtle cancellation will happen at all points of the cube, where even in the worst case one sign will outweigh the other by a very small fraction, $\frac{n^{d/2}}{n^{d-1}}$, of all terms. (Of course, in some specific cases, say $\alpha_R = 1$ for all R, we shall have n^{d-1} as the lower bound.)

An alternative viewpoint stems from the close examination of the structure of the two-dimensional Riesz products Ψ. Consider again the signed case $\alpha_R = \pm 1$ and denote $H_n = \sum_{|R|=2^{-n}} \alpha_R h_R$. It can be shown that $\|H_n\|_1 \approx \|H_n\|_2 \approx n^{1/2}$. Indeed, Hölder's inequality implies that

$$\|H_n\|_2 \leq \|H_n\|_1^{1/3} \cdot \|H_n\|_4^{2/3}.$$

It is easy to see that $\|H_n\|_2 \approx \|H_n\|_4 \approx n^{(d-1)/2} = n^{1/2}$ (the latter computation for the L^4 norm is carried out using the Littlewood-Paley inequalities and is almost identical to (3.12)). The estimate for the L^1 norm of H_n then follows. Equality (4.8), on the other hand, implies that the L^1 norm of $H_n - (-\Psi_{>n})$ is at most $1 + \|\Psi\|_1 = 2$, i.e., H_n, a hyperbolic sum of Haars of order n, can be well approximated in the L^1 norm by a linear combination of Haar functions of higher order. In fact, the Small Ball Conjecture 4.1 would follow if we can prove that for any choice of $\alpha_R = \pm 1$ we have

$$\operatorname{dist}_{L^1}\left(\sum_{R:\, |R|=2^{-n}} \alpha_R h_R,\, H_{>n} \right) \lesssim n^{\frac{d-2}{2}},$$

where $H_{>n}$ is the span of Haar functions supported by rectangles of size $|R| < 2^{-n}$.

Before proceeding to a discussion of the recent progress in the multidimensional case, we would like to briefly explain the connections of Conjecture 4.1 to other areas of mathematics. While the connection of the small ball conjecture to discrepancy function is indirect, it does have important formal implications in probability and approximation theory.

Approximation theory: Entropy of mixed smoothness classes 1. Let $MW^p([0,1]^d)$ be the space of functions on $[0,1]^d$ with mixed derivative $\frac{\partial^d f}{\partial x_1 \partial x_2 \dots \partial x_d}$ in L^p and consider its unit ball $B(MW^p)$. It is compact in the L^∞ metric and its compactness may be quantified through the device of *covering numbers*. Define $N(\varepsilon, p, d)$ to be the least number N of L^∞ balls of radius ε needed to cover the unit ball $B(MW^p)$. The task at hand is to determine the correct order of growth of these numbers as $\varepsilon \downarrow 0$.

Conjecture 4.3. *For $d \geq 2$, we have $\log N(\varepsilon, 2, d) \simeq \varepsilon^{-1}(\log 1/\varepsilon)^{d-1/2}$, as $\varepsilon \downarrow 0$.*

The case $d = 2$ follows from the work of Talagrand [79], and the upper bound is known in full generality [35]. It is well known [81] that inequalities akin to the small ball conjecture (4.2) imply lower bounds on the covering numbers.

Probability. The small ball problem for the Brownian sheet is concerned with finding the exact behavior of the small deviation probability

$$\mathbb{P}\left(\|B\|_{C([0,1]^d)} < \varepsilon\right),$$

where B is the Brownian sheet, i.e., a centered multiparameter Gaussian process characterized by the covariance relation $\mathbb{E}X_s \cdot X_t = \prod_{j=1}^d \min(s_j, t_j)$.

Kuelbs and Li [47] have discovered a tight connection between the small ball probabilities and the properties of the reproducing kernel Hilbert space, which in the case of the Brownian sheet is $WM^2([0,1]^d)$. Their result, applied to the setting of the Brownian sheet in [35], yields an equivalent conjecture:

Conjecture 4.4. *In dimensions $d \geq 2$, for the Brownian sheet B we have*

$$-\log \mathbb{P}\left(\|B\|_{C([0,1]^d)} < \varepsilon\right) \simeq \varepsilon^{-2}(\log 1/\varepsilon)^{2d-1}, \quad \varepsilon \downarrow 0.$$

In $d \geq 3$, the upper bounds are established, see [35]. The lower bound for $d = 2$ has been obtained by Talagrand [79] using (4.2). The idea lies in employing an orthogonal decomposition of B with coefficients being independent standard Gaussians. Inequality (4.2) then allows one to pass from B to sums of absolute values of standard Gaussian random variables, which is a familiar object in probability theory. A detailed account of small ball probabilities for Gaussian processes can be found in [53]. It is worth mentioning that Conjecture 4.4 explains the nomenclature *small ball inequality.*

5. Higher dimensions

For a long time there have been virtually no improvements over the L^2 bound neither in the small ball conjecture, nor in the star-discrepancy bound. In the seminal 1989 paper on discrepancy [7], J. Beck gains a factor of $(\log\log N)^{\frac{1}{8}-\varepsilon}$ over Roth's L^2 bound. A corresponding logarithmic improvement for the small ball inequality can also be extracted from his argument. In 2008, largely building upon Beck's work and enhancing it with new ideas and methods, the author, M. Lacey, and A. Vagharshakyan [11], [12], obtained the first significant improvement over the 'trivial' estimate in all dimensions greater than two:

Theorem 5.1. *In all dimensions $d \geq 3$ there exists $\eta(d) > 0$ such that for all choices of coefficients we have the inequality:*

$$n^{\frac{d-1}{2}-\eta(d)} \Bigg\| \sum_{R:\,|R|=2^{-n}} \alpha_R h_R \Bigg\|_\infty \gtrsim 2^{-n} \sum_{R:\,|R|=2^{-n}} |\alpha_R|. \tag{5.1}$$

P r o o f. The proof of this result was then modified to the discrepancy framework (in a way analogous to the one described in the previous section) to obtain (1.9). The inequality (5.1) also directly translates into improved lower bounds in both Conjectures 4.3 and 4.4.

Since complete technical details of the proof of (5.1), which can be found in [11, 12, 13] as well as Lacey's notes on the subject [50], would take up more space than the rest of this survey, we shall simply present the main ideas of the argument and the heuristics behind them. Then an interested reader can follow the complete proof in the listed references.

Following the idea of Beck, the test function is constructed as a "short" Riesz product. For $\vec{r} \in \mathbb{H}_n^d$, we consider the r-functions $f_{\vec{r}} = \sum_{R\in\mathcal{D}^d_{\vec{r}}} \operatorname{sgn}(\alpha_R) h_R$. Let q be an integer such that $q \approx an^\varepsilon$ for small constants a, $\varepsilon > 0$. Divide the set $\{0, 1, \dots, n\}$ into q disjoint (almost) equal intervals of length about n/q: I_1, $I_2, \dots, I_q$ numbered in increasing order. Let $\mathbb{A}_j := \{\vec{r} \in \mathbb{H}_n^d \mid r_1 \in I_j\}$. The cardinality of $\mathbb{A}_j$ is then $\#\mathbb{A}_j \approx n^{d-1}/q$. Indeed, the first coordinate r_1 can be chosen in n/q ways, the next $d-2$—roughly in n ways each, and the last one is fixed due to the condition $\|\vec{r}\|_1 = n$. We construct the function $F_j = \sum_{\vec{r}\in\mathbb{A}_j} f_{\vec{r}}$. Due to orthogonality, $\|F_j\|_2 \approx \#\mathbb{A}_j \approx n^{(d-1)/2}/\sqrt{q}$. We now introduce the "false" L^2 normalization: $\widetilde{\rho} = aq^{1/4} n^{-(d-1)/2}$. We are now ready to define the Riesz product

$$\Psi := \prod_{j=1}^{q} (1 + \widetilde{\rho} F_j). \tag{5.2}$$

Let us explain the effects that this construction creates and compare it to the two-dimensional Temlyakov's test function (4.7). First of all, the grouping of r-functions by the values of the first coordinate is reminiscent of the construction in (4.7). Here, rather than specifying the value of $|R_1|$, we indicate the range of values that it may take. This idea allows us to preserve some lacunarity in the Riesz product while keeping its size under control. In particular, if $i < j$, then the Haar functions involved in F_j, in the first coordinate, are supported on intervals strictly smaller than those that support the Haar functions in F_i. It follows that for any $k \le q$ and $1 \le j_1 < j_2 < \cdots < j_k \le q$

$$\int_{[0,1]^d} F_{j_1}(x) \cdots F_{j_k}(x) = 0, \tag{5.3}$$

since the integral in the first coordinate is already zero (all the Haar functions are distinct). In particular,

$$\int_{[0,1]^d} \Psi(x)dx = 1, \tag{5.4}$$

as (5.3) implies that all the higher order terms have mean zero. By comparison, Beck's [7] construction of the short Riesz product was probabilistic, which made it much more difficult to collect definitive information about the interaction of different factors in the product.

Secondly, recall that the Riesz product in (4.7) was non-negative allowing one to replace the L^1 norm with the integral which is much easier to compute. While in our case positivity everywhere is too much to hope for, it can be shown that the product is positive with large probability. The "false" L^2 normalization $\widetilde{\rho}$ makes the L^2 norm of $\widetilde{\rho}F_j$ small: $\|\widetilde{\rho}F_j\|_2 \approx q^{-1/4} \approx n^{-\varepsilon/4} \ll 1$. Thus $(1 + \widetilde{\rho}F_j)$ is positive on a set of large measure, therefore, so is the product (5.2). This heuristic is quantified in (5.9).

However, we cannot take Ψ to be our test function since we do not know exactly how it interacts with $\sum_{|R|=2^{-n}} \alpha_R h_R$. As explained in the remarks after the product rule (4.6), problems arise when the rectangles supporting the Haar functions coincide in one of the coordinates, in other words, when for two vectors $\vec{r}$, $\vec{s} \in \mathbb{H}_n^d$ and for some $k = 1, \ldots, d$, we have $r_k = s_k$. We say that a *coincidence* occurs in this situation. We say that a collection of vectors $\{\vec{r}_j\}_{j=1}^m \subset \mathbb{H}_n^d$ is *strongly distinct* if no coincidences occur between the elements of the collection, i.e., for all $1 \le i, j \le m$, $1 \le k \le d$, we have $r_{i,k} \ne r_{j,k}$. We can then write

$$\Psi = 1 + \Psi^{sd} + \Psi^{\neg sd}, \tag{5.5}$$

where

$$\Psi^{sd} = \sum_{k=1}^{q} \widetilde{\rho}^k \sum_{1 \le j_1 < j_2 < \cdot < j_k \le q} \left(\widetilde{\sum f_{\vec{r}_{j_1}} \cdots f_{\vec{r}_{j_k}}} \right), \tag{5.6}$$

and the tilde above the innermost sum indicates that the sum is extended over all collections of vectors $\{\vec{r}_{j_t} \in \mathbb{A}_{j_t} : t = 1, \dots, k\}$ which are strongly distinct. To put it simpler, $\Psi^{\neg sd}$ consists of the terms that involve coincidences, and Ψ^{sd}—of the ones that do not.

The function Ψ^{sd} is then taken to be the test function. Since all the coincidences are eliminated, the product rule (4.6) is applicable and an argument similar to (4.9)–(4.10) can be carried out, provided we can show that $\|\Psi^{sd}\|_1 \leq 1$.

An enormous part of of the proof in [11, 12] is devoted to the study of analytic and combinatorial aspects of the coincidences, i.e.,the behavior of $\Psi^{\neg sd}$. An important starting point is the following non-trivial lemma, which as a tribute to J. Beck's ideas we call the *Beck gain*:

Lemma 5.2 (Beck Gain). *For every $p \geq 2$ we have the following inequality*

$$\Bigg\| \sum_{\substack{\vec{r} \neq \vec{s} \in \mathbb{H}_n^d \\ r_1 = s_1}} f_{\vec{r}} \cdot f_{\vec{s}} \Bigg\|_p \lesssim p^{\frac{2d-1}{2}}\, n^{\frac{2d-3}{2}}. \tag{5.7}$$

The important point of this lemma is the precise power of n in the estimate. Let us explain that the exponent $\frac{2d-3}{2}$ is very natural. Indeed, d-dimensional vectors $\vec{r}$ and $\vec{s}$ have d parameters each. The condition $\|\vec{r}\|_1 = \|\vec{s}\|_1 = n$ eliminates one free parameter in each vector. Additionally, the coincidence $r_1 = s_1$ freezes one more parameter. Hence, the total number of free parameters in the sum is $2d - 3$ and each can take roughly n values. Thus the total number of terms is of the order of n^{2d-3} and (5.7) essentially says that they behave as if they were orthogonal. The power of p doesn't seem to be sharp (perhaps, $\frac{2d-3}{2}$ should also be the correct exponent of p), but it is important for further estimates that this dependence is polynomial in p.

One can also view it in the following way. It is not hard to show using Littlewood-Paley inequalities that $\| \sum_{\vec{r} \neq \vec{s} \in \mathbb{H}_n^d} f_{\vec{r}} \cdot f_{\vec{s}} \|_p \lesssim n^{d-1}$. Therefore, by imposing the condition $r_1 = s_1$ one *gains* $\sqrt{n}$ in the estimate. This lemma, albeit in a weaker form (just for $p = 2$ and with a larger power of n) appeared in the aforementioned paper of Beck [7]. In his argument, in order to compute the L^2 norm, Beck expands the square of the sum and notices that the integral of each term is zero unless there is a coincidence in each coordinate. Careful combinatorial analysis of these coincidences then produces the desired inequality. The extension and generalization obtained in [11, 12] is achieved by replacing the process of expanding the square by the applications of the Littlewood-Paley square function (3.2), which is a natural substitution in harmonic analysis, when one wants to pass from L^2 to L^p, $p \neq 2$. Every application of the Littlewood-Paley inequality (3.4) yields a constant $B_p \approx \sqrt{p}$, see [11, Lemma 8.2], [12, Lemma 5.2] for complete details. The lemma was initally proved in $d = 3$ [11] and then extended to $d \geq 3$ [12] by a tricky induction argument.

Furthermore, one needs to analyze more complicated instances of coincidences which arise in $\Psi^{\neg sd}$. Their high combinatorial complexity in large dimensions aggravates the difficulty of the problem. Further success of the Riesz product method requires inequalities of the type:

$$\left\| \sum f_{\vec{r_1}} \cdots f_{\vec{r_k}} \right\|_p \lesssim p^{\alpha M} n^{\frac{M}{2}}, \tag{5.8}$$

where the sum is extended over all k-tuples $\vec{r_1}, \ldots, \vec{r_k}$ with a specified configuration of coincidences and M is the number of free parameters imposed by this configuration; $\alpha > 0$ is a constant which is conjectured to be $\frac{1}{2}$. Estimates of this type suggest that free parameters behave orthogonally even for longer coincidences. When $k = 2$ (the graph describing the coincidence consists of two vertices and one edge), estimate (5.8) turns precisely into inequality (5.7) of Lemma 5.7. In [12] a partial result with a larger power of n is obtained for $k > 2$. Roughly speaking, these inequalities are proved by choosing a large *matching* (disjoint collection of edges) in the associated graph and applying the Beck gain (5.7) to each simple coincidence, see [12, Theorem 8.3].

In the end, one arrives to the following estimates (see Lemma 4.8 in [12]):

Lemma 5.3. *We have these estimates:*

$$\begin{aligned} \mu\big(\{\Psi < 0\}\big) &\lesssim \exp(-A\sqrt{q})\,; && (5.9)\\ \|\Psi\|_2 &\lesssim \exp(a'\sqrt{q})\,; && (5.10)\\ \int \Psi(x)dx &= 1\,; && (5.11)\\ \|\Psi\|_1 &\lesssim 1\,; && (5.12)\\ \|\Psi^{\neg sd}\|_1 &\lesssim 1\,; && (5.13)\\ \|\Psi^{sd}\|_1 &\lesssim 1\,, && (5.14) \end{aligned}$$

where $0 < a' < 1$ *is a small constant,* $A > 1$ *is a large constant, and* μ *is the Lebesgue measure.*

Some remarks are in order. The first two inequalities rely on Beck gain (5.7). Inequality (5.9) is a quantification of the fact discussed earlier that, due to the false L^2 normalization $\widetilde{\rho}$, Ψ is negative on a very small set (a weaker version can be proved without referring to Beck gain). The L^2 bound (5.10) can be explained heuristically: if F_j's were independent random variables, we would immediately obtain (5.10):

$$\int \prod_{j=1}^{q} (1+\widetilde{\rho} F_j)^2\, dx = \prod_{j=1}^{q} \int (1+\widetilde{\rho} F_j)^2\, dx \le \prod_{j=1}^{q} (1+\widetilde{\rho}^2 \|F_j\|_2^2)$$
$$\le \left(1+\frac{a^2}{\sqrt{q}}\right)^q \le e^{a^2\sqrt{q}}.$$

While they are not independent, one applies a conditional expectation argument and Beck gain (5.7) (the lack of independence is the result of coincidences). Equality (5.11) has already been explained, see (5.4), and estimate (5.12) easily follows from (5.9)–(5.11) using Cauchy-Schwarz inequality:

$$\begin{aligned}\|\Psi\|_1 = \int \Psi(x)dx - 2\int_{\{\Psi<0\}} \Psi(x)dx &\le 1 + 2\mu\big(\{\Psi<0\}\big)^{1/2} \cdot \|\Psi\|_2 \\ &\lesssim 1 + \exp(-A\sqrt{q}/2 + a'\sqrt{q}) \ \lesssim\ 1.\end{aligned}$$

Estimate (5.13) is the result of the analysis of coincidences and (5.14) is implied by the previous two since $\Psi^{sd} = \Psi - 1 - \Psi^{\neg sd}$.

Finally, we obtain, as in (4.9)-(4.10):

$$\begin{aligned}\Bigg\| \sum_{R\in\mathcal{D}^d:\,|R|=2^{-n}} \alpha_R h_R \Bigg\|_\infty \ &\gtrsim\ \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \Psi^{sd} \Big\rangle \\ &=\ \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \widetilde{\rho} \sum_{|R|=2^{-n}} \operatorname{sgn}(\alpha_R) h_R \Big\rangle \\ &=\ \widetilde{\rho} \sum_{R\in\mathcal{D}^d:\,|R|=2^{-n}} \alpha_R \cdot \operatorname{sgn}(\alpha_R) \cdot \|h_R\|_2^2 \\ &\approx\ n^{-\frac{d-1}{2}+\frac{\varepsilon}{4}}\, 2^{-n} \cdot \sum_{|R|=2^{-n}} |\alpha_R|, \end{aligned} \tag{5.15}$$

so, (5.1) holds with $\eta = \varepsilon/4$. □

To finish this discussion, we mention that the *signed* small ball inequality, i.e., a version with $\alpha_R = \pm 1$ for each R, see (4.5) may be viewed as a toy model of Conjecture 4.1. It avoids numerous technicalities, while preserving most of the complications arising from the combinatorial complexity of higher dimensional dyadic boxes. In [13], the same authors came up with a significant simplification of the arguments in [11, 12] for the signed case (in fact, it only required the simplest estimate for coincidences (5.7), and not the more complicated (5.8)). It yielded the bound $\big\| \sum_{|R|=2^{-n}} \alpha_R h_R \big\|_\infty \gtrsim n^{\frac{d-1}{2}+\eta}$ for $\alpha_R = \pm 1$ in all dimensions and allowed them to obtain an explicit value of the gain $\eta(d) = \frac{1}{8d} - \varepsilon$.

In [15], Bilyk, Lacey, Parissis, and Vagharshakayan initiated a different approach to the proof of the signed small ball inequality. While in dimension $d = 2$, the small ball inequality (4.2) can be proved as a consequence of the independence of the random variables $f_{\vec{r}}$, in higher dimensions due to coincidences independence no longer holds. This shortcoming can by partially compensated for by delicate conditional expectation arguments. The proof of the three-dimensional inequality in [15] yields the best currently known gain: $\eta(3) = \frac{1}{8}$. Unfortunately, at this time it is not clear how to transfer this method to the discrepancy setting or extend it to higher dimensions.

6. Low discrepancy distributions: the van der Corput set

All the previous sections of this survey were concerned with the proofs of various lower bounds for discrepancy. In the last section we would like to illustrate how Roth's idea of incorporating dyadic harmonic analysis into discrepancy theory helps in proving some upper discrepancy estimates. We recall a very standard construction, the so-called "digit-reversing" van der Corput set [31] (also known as the Hammersley point set):

$$\mathcal{V}_n = \big\{ (0.x_1 x_2 \dots x_n, 0.x_n x_{n-1} \dots x_1) : \; x_k = 0,\, 1;\, k = 1, \dots, n \big\}, \tag{6.1}$$

where the coordinates are written as binary fractions. This set has 2^n points and its star-discrepancy is optimal in the order of magnitude: $\|D_{\mathcal{V}_n}\|_\infty \le n + 1 \approx \log N$. The crucial property of this set, which allows one to deduce such a favorable discrepancy bound is the fact that it forms a binary net: any dyadic rectangle R of area $|R| = 2^{-n}$ contains precisely one point of $\mathcal{V}_n$, and hence the discrepancy of $\mathcal{V}_n$ with respect to such rectangles is zero.

Various norms of the discrepancy function of variations of this set have been studied by many authors: [10],[14], [28], [31], [36], [41], [43], [45], [51], [62] to name just a few. It is well known that, while $\mathcal{V}_n$ has optimal star-discrepancy, its L^2 discrepancy is also of the order $\log N$ as opposed to the optimal $\sqrt{\log N}$. The problem actually lies in the fact that $\int D_{\mathcal{V}_n}(x)dx = \frac{n}{8} + \overline{o}\,(n) \approx \log N$ as observed in [41, 28, 10, 43]. In other words, in any reasonable orthogonal decomposition of $D_{\mathcal{V}_n}$ the zero-order coefficient is already too big. There are several standard remedies which allow one to overcome this problem.

i) Random shifts: Roth [62, 64] showed probabilistically that there exists a shift of $\mathcal{V}_n$ modulo 1 which achieves optimal L^2 discrepancy; a deterministic example was constructed in [10].

ii) Symmetrization: this idea was introduced by Davenport [33] to construct the first example of a set with optimal order of L^2 discrepancy, see also [28].

iii) Digit shifts: this idea that goes back to [41],[22] works extremely well for the van der Corput set. We alter $\mathcal{V}_n$ as follows:

$$\mathcal{V}_n^\sigma = \Big\{ \big(0.x_1x_2\dots x_n, 0.(x_n \oplus \sigma_n)(x_{n-1}\oplus\sigma_{n-1})\dots(x_1\oplus\sigma_1)\big) : \\ x_k = 0,\, 1;\, k = 1,\dots,n \Big\}, \quad (6.2)$$

where $\sigma = (\sigma_k)_{k=1}^n \in \{0,1\}^n$ is fixed and $\oplus$ denotes addition modulo 2, i.e., after flipping the digits we also change some of them (those for which $\sigma_k = 1$).

This procedure has been thoroughly studied for the van der Corput set: it is known to improve its distributional quality [45, 36]. In particular, when approximately half of the digits are shifted, i.e., $\sum \sigma_k \approx \frac{n}{2}$, this set has optimal L^2 discrepancy [41, 46, 14].

The nice dyadic structure of this set makes it perfectly amenable to the methods of harmonic analysis. For example, in [10] it is analyzed using Fourier series, in [51, 28] (see also the book [34]) the authors exploit Walsh functions, while the estimates in [14, 43] are based on Haar coefficients. We will concentrate on the latter results since they directly relate to Roth's method and complement previously discussed lower bounds.

The author, Lacey, Parissis, and Vagharshakyan have shown that the BMO (3.25) and $\exp(L^\alpha)$ (3.27) lower estimates in dimension $d = 2$ are sharp. In particular, for the digit-shifted van der Corput set $\mathcal{V}_n^\sigma$ with $\sum \sigma_k \approx \frac{n}{2}$ we have $\|D_{\mathcal{V}_n^\sigma}\|_{\exp(L^\alpha)} \lesssim (\log N)^{1-\frac{1}{\alpha}}$, $\alpha \geq 2$, while $\|D_{\mathcal{V}_n}\|_{\mathrm{BMO}} \lesssim \sqrt{\log N}$ for the standard van der Corput set. These inequalities were based on estimates of the Haar coefficients of the discrepancy function, namely

$$\Big|\big\langle D_{\mathcal{V}_n^\sigma}, h_R \big\rangle\Big| \lesssim \min\big\{1/N,\, |R|\big\}.$$

While this estimate for small rectangles is straightforward (the counting and linear part can be bounded separately), coefficients corresponding to large rectangles involve subtle cancellations suggested by the structure and self-similarities of $\mathcal{V}_n^\sigma$. The BMO and $\exp(L^\alpha)$ can the be obtained by applying arguments of Littlewood-Paley type. Almost simultaneously, Hinrichs [43] estimated the

Besov norm of the same digit-shifted van der Corput set using a very similar method. In fact, he went much further and computed all the Haar coefficients of $D_{\mathcal{V}_n^\sigma}$ exactly. This led to showing that the lower Besov space estimate (3.1) of Triebel [86] is sharp in $d = 2$, more precisely $\|D_{\mathcal{V}_n^\sigma}\|_{S^r_{pq}B([0,1]^d)} \lesssim N^r (\log N)^{\frac{1}{q}}$ for $1 \le p,\, q \le \infty$, $0 \le r < \frac{1}{p}$.

We close our exposition with an amusing observation which pinpoints yet another connection between the small ball inequality (4.2) and discrepancy. Consider the two-dimensional case and assume that all the coefficients α_R are non-negative. Recall Temlyakov's test function (4.7):

$$\Psi = \prod_{k=1}^{n} (1 + f_k),$$

where, since $\operatorname{sgn}(\alpha_R) = +1$, the r-functions $f_k = \sum_{|R|=2^{-n},\, |R_1|=2^{-k}} h_R$ are actually Rademacher functions. As discussed in § 4 right after (4.14), $\Psi = 2^{n+1}\mathbf{1}_E$, where $E = \{x \in [0,1]^2 :\, f_k(x) = +1,\, k = 0, 1, \dots, n\}$.

We shall describe the geometry of the set E. Evidently, it consists of 2^{n+1} dyadic squares of area $2^{-2(n+1)}$. We characterize the locations of the lower left corners of these squares. If $t \in [0,1]$ and a dyadic interval I of length 2^{-k} contains t, then $h_I(t) = -1$ if the $(k+1)^{st}$ binary digit of t is 0, and $h_I(t) = 1$ if it is 1. Thus $f_k(x_1, x_2) = +1$ exactly when the $(k+1)^{st}$ digit of x_1 and the $(n-k+1)^{st}$ digit of x_2 are the same. Therefore, $(x_1, x_2) \in E$ when this holds for all $k = 0, 1, \dots, n$, i.e., the first $n+1$ binary digits of x_2 are formed as the reversed sequence of the first $n+1$ digits of x_1—but this is precisely the definition of the van der Corput set $\mathcal{V}_{n+1}$! So,

$$E = \mathcal{V}_{n+1} + \left[0, 2^{-(n+1)}\right) \times \left[0, 2^{-(n+1)}\right).$$

REFERENCES

[1] VAN AARDENNE-EHRENFEST, T.: *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] VAN AARDENNE-EHRENFEST, T.: *On the impossibility of a just distribution*, Nederl. Akad. Wetensch., Proc. **52** (1949), 734–739; Indag. Math. (N.S.) **11** (1949), 264–269.

[3] BECK, J. – CHEN, W. W. L.: *Note on irregularities of distribution*, Mathematika **33** (1986), no. 1, 148–163.

[4] BECK, J. – CHEN, W. W. L.: *Irregularities of Distribution*, *Cambridge Tracts in Mathematics* **89**, Cambridge University Press, Cambridge, 1987.

[5] BECK, J.: *Irregularities of distribution. I*, Acta Math. **159** (1987), no. 1–2, 1–49.

[6] BECK, JÓZSEF: *Irregularities of distribution. II*, Proc. London Math. Soc. (3) **56** (1988), no. 1, 1–50.

[7] articleBECK, JÓZSEF: *A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution*, Compos. Math. **72** (1989), no. 3, 269–339.

[8] BECK, J. – CHEN, W. W. L.: *Note on irregularities of distribution. II*, Proc. London Math. Soc. (3) **61** (1990, no. 2, 251–272.

[9] BÉJIAN, R.: *Minoration de la discrépance d'une suite quelconque sur T*, Acta Arith. **41** (1982), no. 2, 185–202.

[10] BILYK, D.: *Cyclic shifts of the van der Corput set*, Proc. Amer. Math. Soc. **137** (2009), no. 8, 2591–2600.

[11] BILYK, D. – LACEY, M.: *On the small ball inequality in three dimensions*, Duke Math. J. **143** (2008), no. 1, 81–115.

[12] BILYK, D. – LACEY, M. – VAGHARSHAKYAN, A.: *On the small ball inequality in all dimensions*, J. Funct. Anal. **254** (2008), no. 9, 2470–2502.

[13] BILYK, D. – LACEY, M. – VAGHARSHAKYAN, A.: *On the signed small ball inequality*, Online J. Anal. Comb. **3**, (2008), Art. 6, 7 pp.

[14] BILYK, D. – LACEY, M. – PARISSIS, I. – VAGHARSHAKYAN, A.: *Exponential squared integrability of the discrepancy function in two dimensions*, Mathematika **55** (2009), no. 1–2, 1–27.

[15] BILYK, D. – LACEY, M. – PARISSIS, I. – VAGHARSHAKYAN, A.: *A three-dimensional signed small ball inequality*, Walter Philipp memorial volume (2010) (to appear).

[16] BURKHOLDER, D. L.: *Sharp inequalities for martingales and stochastic integrals*, Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987). Astérisque (1988) **157–158**, 75–94.

[17] CHANG, S.-Y. A. – FEFFERMAN, R.: *A continuous version of duality of H^1 with BMO on the bidisc*, Ann. Math. (2), **112** (1980), no. 1, 179–201.

[18] CHANG, S.-Y. A. – FEFFERMAN, R.: *Some recent developments in Fourier analysis and H^p-theory on product domains*, Bull. Amer. Math. Soc. (N.S.) **12** (1985), no. 1, 1–43.

[19] CHANG, S.-Y. A. – WILSON, J. M. – WOLFF, T. H.: *Some weighted norm inequalities concerning the Schrödinger operators*, Comment. Math. Helv. **60** (1985), no. 2, 217–246.

[20] CHAZELLE, B.: *The Discrepancy Method. Randomness and Complexity*, Cambridge University Press, Cambridge, 2000.

[21] CHEN, W. W. L.: *On irregularities of distribution*, Mathematika **27** (1981), no. 2, 153–170.

[22] CHEN, W. W. L.: *On irregularities of distribution. II*, Quart. J. Math. Oxford Ser. (2) **34** (1983), no. 135, 257–279.

[23] CHEN, W. W. L.: *On irregularities of distribution. III*, J. Austral. Math. Soc. Ser. A **60** (1996), no. 2, 228–244.
[24] CHEN, W. W. L.: *On irregularities of distribution. IV*, J. Number Theory **80**, (2000), no. 1, 44–59.
[25] CHEN, W. W. L.: *Fourier techniques in the theory of irregularities of point distribution*, in:*Fourier Analysis and Convexity*, *Appl. Numer. Harmon. Anal.*, Birkhäuser Boston, Boston, MA, 2004, pp. 59–82.
[26] CHEN, W. W. L.: *private communication*,AIM, Palo Alto, CA, 2008.
[27] CHEN, W. W. L. – SKRIGANOV, M.M.: *Explicit constructions in the classical mean squares problem in the irregularities of point distribution*, J. Reine Angew. Math. **545** (2002), 67–95.
[28] CHEN, W. W. L. – SKRIGANOV, M.M.: *Davenport's theorem in the theory of irregularities of point distribution*, English, with English and Russian summaries, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **269** (2000); Vopr. Kvant. Teor. Polya i Stat. Fiz. **16**, 339–353, 371; translation in J. Math. Sci. (N. Y.) **115** (2003), no. 1, 2076–2084.
[29] CHEN, W. – TRAVAGLINI, G.: *Some of Roth's ideas in discrepancy theory*, in: *Analytic Number Theory*, Cambridge Univ. Press, Cambridge, 2009, pp. 150–163.
[30] VAN DER CORPUT, J. G.: *Verteilungsfunktionen I*, Proc. Akad. Wetensch. Amsterdam, **38** (1935), 813–821.
[31] VAN DER CORPUT, J. G.: *Verteilungsfunktionen II*, Akad. Wetensch. Amsterdam, Proc. **38** (1935), 1058–1068.
[32] DAUBECHIES, INGRID: *Ten Lectures on Wavelets*, in: *CBMS-NSF Regional Conference Series in Applied Mathematics* **61**, Society for Industrial and Applied Mathematics (SIAM) , Philadelphia, PA, 1992.
[33] DAVENPORT, H.: *Note on irregularities of distribution*, Mathematika, **3**, (1956), 131–135.
[34] DICK, J. – PILLICHSHAMMER, F.: *Digital Nets and Sequences. Discrepancy theory and quasi-Monte Carlo integration*, Cambridge University Press, Cambridge, 2010.
[35] DUNKER, T. – KÜHN, T. – LIFSHITS, M. – LINDE, W.: *Metric entropy of the integration operator and small ball probabilities for the Brownian sheet*, English, with English and French summaries, C. R. Acad. Sci. Paris Sér. I Math. **326** (1998), no. 3, 347–352.
[36] FAURE, H. – PILLICHSHAMMER, F.: *L_p discrepancy of generalized two-dimensional Hammersley point sets*, Monatsh. Math. **158** (2009), no. 1, 31–61.
[37] *Upper bounds on the discrepancy in L_p*, $2 \leq p < \infty$, Dokl. Akad. Nauk USSR **252** (1980), 805–807.
[38] HAAR, A.: *Zur Theorie der orthogonalen Funktionensysteme*, Math. Annal. **69** (1910), 331–371.
[39] HALÁSZ, G.: *On Roth's method in the theory of irregularities of point distributions*, in: *Recent Progress in Analytic Number Theory*, **2** (Durham, 1979), Academic Press, London, 1981, pp. 79–94.

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

[41] HALTON, J.H. – ZAREMBA, S.C.: *The extreme and L^2 discrepancies of some plane sets*, Monatsh. Math. **73** (1969), 316–328.

[42] HAMMERSLEY, J.M.: *Monte Carlo methods for solving multivariable problems*, Ann. New York Acad. Sci **86** (1960), 844–874.

[43] HINRICHS, A.: , *Discrepancy of Hammersley points in Besov spaces of dominating mixed smoothness*, Math. Nachr. **283** (2010), no. 3, 478–488.

[44] HINRICHS, A. – MARKASHIN, L.: *On lower bounds for the L_2 discrepancy*, J. Complexity (2010) (to appear).

[45] KRITZER, P.: *On some remarkable properties of the two-dimensional Hammersley point set in base 2*, (English, with English and French summaries), J. Théor. Nombres Bordeaux **18** (2006), no. 1, 203–221.

[46] KRITZER, P. – PILLICHSHAMMER, F.: , *An exact formula for the L_2 discrepancy of the shifted Hammersley point set*, Unif. Distrib. Theory **1** (2006), no. 1, 1–13 (electronic).

[47] KUELBS, J. – LI, W. V.: *Metric entropy and the small ball problem for Gaussian measures*, J. Funct. Anal. **116** (1993), no. 1, 133–157.

[48] KUIPERS, L. – NIEDERREITER, H.: *Uniform Distribution of Sequences*, *Pure and Applied Mathematics*, Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.

[49] LACEY, M.: *On the discrepancy function in arbitrary dimension, close to L^1*, Anal. Math. **34** (2008), no. 2, 119–136.

[50] LACEY, M.: *Small Ball and Discrepancy Inequalities*, (2008), available on arxiv.org.

[51] LARCHER, G. – PILLICHSHAMMER, F.: *Walsh series analysis of the L_2--discrepancy of symmetrisized point sets*, Monatsh. Math. **132** (2001), no. 1, 1–18.

[52] LERCH, M.: *Question 1547*, L'Intermediaire Math. **11** (1904), 144–145.

[53] LI, W. V. – SHAO, Q.-M.: *Gaussian processes: inequalities, small ball probabilities and applications*, in: *Stochastic processes: theory and methods*, *Handbook of Statist.* **19**, North-Holland, Amsterdam, 2001, pp. 533–597.

[54] LIARDET, P.: *Discrépance sur le cercle.*, Primaths. I, Univ. Marseille, 1979

[55] LINDENSTRAUSS, J. – TZAFRIRI, L.: *Classical Banach Spaces. I. Sequence spaces Ergebnisse der Mathematik und ihrer Grenzgebiete*, **92**, Springer-Verlag, Berlin, 1977.

[56] MATOUŠEK, J.: *Geometric Discrepancy*, An illustrated guide, *Algorithms and Combinatorics* **18**, Springer-Verlag, Berlin, 1999.

[57] MONTGOMERY, H. L.: *Ten Lectures On the Interface Between Analytic Number Theory and Harmonic Analysis*, in: *CBMS Regional Conference Series in Mathematics* **84**, Amer. Math. Soc., Providence, 1994.

[58] OU, W.: *Irregularity of distributions and multiparameter A_p weights*, Unif. Distrib. Theory **5** (2010), no. 2, 131–139.

[59] PIPHER, J.: *Bounded double square functions*, English, with French summary, Ann. Inst. Fourier (Grenoble) **36** (1986), no. 2, 69–82.

[60] RIESZ, F.: *Über die Fourierkoeffizienten einer stetigen Funktion von beschränkter Schwankung*, Math. Z. **2** (1918), no. 3–4, 312–315.

[61] ROTH, K. F.: *On irregularities of distribution*, Mathematika **1** (1954), 73–79.

[62] ROTH, K. F.: *On irregularities of distribution. II*, Comm. Pure Appl. Math. **29** (1976), no. 6, 739–744.

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

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

[65] SCHMIDT, W. M.: *Irregularities of distribution*, Quart. J. Math. Oxford Ser. (2) **19** (1968), 181–191.

[66] SCHMIDT, W. M.: *Irregularities of distribution. II*, Trans. Amer. Math. Soc. **136** (1969), 347–360.

[67] SCHMIDT, W. M.: *Irregularities of distribution. III*, Pacific J. Math. **29** (1969), 225–234.

[68] SCHMIDT, W. M.: *Irregularities of distribution. IV*, Invent. Math. **7** (1969), 55–82.

[69] SCHMIDT, W. M.: *Irregularities of distribution. V*, Proc. Amer. Math. Soc., **25**, (1970), 608–614.

[70] SCHMIDT, W. M.: *Irregularities of distribution. VI*, Compositio Math. **24** (1972), 63–74.

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

[72] SCHMIDT, W. M.: *Irregularities of distribution. VIII*, Trans. Amer. Math. Soc. **198**, (1974), 1–22.

[73] SCHMIDT, W. M.: *Irregularities of distribution. IX*, in: *Collection of Articles in Memory of Juriĭ Vladimirovič Linnik*, Acta Arith. **27** (1975), pp. 385–396.

[74] SCHMIDT, W. M.: *Irregularities of distribution. X*, in: *Number Theory and Algebra*, Academic Press, New York, 1977, pp. 311–329.

[75] SKRIGANOV, M. M.: *Harmonic analysis on totally disconnected groups and irregularities of point distributions.*, J. Reine Angew. **600**, (2006), 25–49.

[76] SKRIGANOV, M. M.: *private communication*, AIM, Palo Alto, CA, (2008).

[77] SOBOL', I.M.: *Multidimensional Quadrature Formulas and Haar Functions*, (Russian), Izdat. Nauka, Moscow, 1969.

[78] STEIN, ELIAS M.: *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, *Princeton Mathematical Series* **43**, (With the assistance of Timothy S. Murphy); *Monographs in Harmonic Analysis, III*, Princeton University Press, Princeton, NJ, 1993.

[79] TALAGRAND, M.: *The small ball problem for the Brownian sheet*, Ann. Probab. **22** (1994), no. 3, 1331–1354.

[80] TEMLYAKOV, V. N.: *Approximation of Periodic Functions*, Nova Science, New York, 1994.

[81] TEMLYAKOV, V. N.: *Approximation of Functions with Bounded Mixed Derivative*, in: *Proc. Steklov Inst. Math.* **178** (1989), no. 1,

[82] TEMLYAKOV, V. N.: *Some Inequalities for Multivariate Haar Polynomials*, East Journal on Approximations **1** (1995), no. 1, 61–72.

[83] TEMLYAKOV, V. N.: *An inequality for trigonometric polynomials and its application for estimating the entropy numbers*, J. Complexity, **11** (1995), no. 2, 293–307.

[84] TEMLYAKOV, V. N.: *An Inequality for Trigonometric Polynomials and its Application for Estimating the Kolmogorov Widths*, East J. Approx. **2**, (1996), 253-262.

[85] TEMLYAKOV, V. N.: *Cubature formulas, discrepancy, and nonlinear approximation*, J. Complexity **19**, no. 3, (2003), 352–391.

[86] TRIEBEL, H.: *Numerical integration and discrepancy, a new approach*, Math. Nachr. **283** (2010), no. 1, 139–159.

[87] TRIEBEL, H.: *Bases in Function Spaces, Sampling, Discrepancy, Numerical Integration*, European Math. Soc. Publishing House, Zürich (to appear).

[88] WANG, GANG: *Sharp square-function inequalities for conditionally symmetric martingales*, Trans. Amer. Math. Soc. **328** (1991), no. 1, 393–419.

[89] ZYGMUND, A.: *Trigonometric Series. Vol. I, II, Cambridge Mathematical Library*, 3rd. edition, (With a foreword by Robert A. Fefferman), Cambridge University Press, Cambridge, 2002.

Received February 21, 2011
Accepted April 4, 2011

Dmitriy Bilyk
Department of Mathematics
University of South Carolina
LeConte College, office 300D
1523 Greene Street
Columbia, SC 29208
U.S.A
E-mail: bilyk@math.sc.edu