Roth's Orthogonal Function Method in Discrepancy Theory and Some New Connections

Dmitriy Bilyk

Abstract In this survey we give a comprehensive, but gentle introduction to the circle of questions surrounding the classical problems of discrepancy theory, unified by the same approach originated in the work of Klaus Roth [85] and based on multiparameter Haar (or other orthogonal) function expansions. Traditionally, the most important estimates of the discrepancy function were obtained using variations of this method. However, despite a large amount of work in this direction, the most important questions in the subject remain wide open, even at the level of conjectures. The area, as well as the method, has enjoyed an outburst of activity due to the recent breakthrough improvement of the higher-dimensional discrepancy bounds and the revealed important connections between this subject and harmonic analysis, probability (small deviation of the Brownian motion), and approximation theory (metric entropy of spaces with mixed smoothness). Without assuming any prior knowledge of the subject, we present the history and different manifestations of the method, its applications to related problems in various fields, and a detailed and intuitive outline of the latest higher-dimensional discrepancy estimate.

1 Introduction

The subject and the structure of the present chapter is slightly unconventional. Instead of building the exposition around the results from one area, united by a common topic, we concentrate on problems from different fields which all share a common method.

The starting point of our discussion is one of the earliest results in discrepancy theory, Roth's 1954 L^2 bound of the discrepancy function in dimensions $d \geq 2$ [85], as well as the tool employed to obtain this bound, which later evolved into a pow-

Dmitriy Bilyk
Department of Mathematics, University of South Carolina, Columbia, SC 29208 USA e-mail: bilyk@math.sc.edu

erful orthogonal function method in discrepancy theory. We provide an extensive overview of numerous results in the subject of irregularities of distribution, whose proofs are based on this method, from the creation of the field to the latest achievements.

In order to highlight the universality of the method, we shall bring out and emphasize analogies and connections of discrepancy theory and Roth's method to problems in a number of different fields, which include numerical integration (errors of cubature formulas), harmonic analysis (the small ball inequality), probability (small deviations of multiparameter Gaussian processes), approximation theory (metric entropy of spaces with dominating mixed smoothness). While some of these problems are related by direct implications, others are linked only by the method of proof, and perhaps new relations are yet to be discovered.

We also present a very detailed and perceptive account of the proof of one of the most recent important developments in the theory, the improved L^∞ bounds of the discrepancy function, and the corresponding improvements in other areas. We focus on the heuristics and the general strategy of the proof, and thoroughly explain the idea of every step of this involved argument, while skipping some of the technicalities, which could have almost doubled the size of this chapter.

We hope that the content of the volume will be of interest to experts and novices alike and will reveal the omnipotence of Roth's method and the fascinating relations between discrepancy theory and other areas of mathematics. We have made every effort to make our exposition clear, intuitive, and essentially self-contained, requiring familiarity only with the most basic concepts of the underlying fields.

1.1 The history and development of the field

Geometric *discrepancy theory* seeks answers to various forms of the following questions: *How accurately can one approximate a uniform distribution by a finite discrete set of points? And what are the errors and limitations that necessarily arise in such approximations?* The subject naturally grew out of the notion of uniform distribution in number theory. 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.

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

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

$$D_N(\omega) = \sup_{I\subset[0,1]} \big|\#\{\omega_n \in I:\, 1 \le n \le N\} - N\cdot|I|\big|, \tag{2}$$

where I is an interval. In fact, it is not hard 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. [64]).

In [38, 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. More precisely, he mildly conjectured that the answer is "No" by stating that he is unaware of such sequences. Indeed, in [1, 1945], [2], van Aardenne-Ehrenfest gave a negative answer to this question, which meant that no sequence can be distributed too well. This result is widely regarded as a predecessor of the theory of *irregularities of distribution*.

This area was turned into a real theory with precise quantitative estimates and conjectures by Roth, who in particular, see [85], greatly improved van Aardenne-Ehrenfest's result by demonstrating that for any sequence ω the inequality

$$D_N(\omega) \geq C\sqrt{\log N} \tag{3}$$

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

Roth in fact worked on the following, more geometrical version of the problem. Let $\mathscr{P}_N \subset [0,1]^d$ be a set of N points and consider the discrepancy function

$$D_N(x_1,\ldots,x_d) = \#\{\mathscr{P}_N \cap [0,x_1)\times\cdots\times[0,x_d)\} - N\cdot x_1\cdot\cdots\cdot x_d, \tag{4}$$

i.e. the difference of the actual and expected number of points of $\mathscr{P}_N$ in the box $[0,x_1)\times\cdots\times[0,x_d)$. Notice that, in contrast to some of the standard references, we are working with the unnormalized version of the discrepancy function, i.e. we do not divide this difference by N as it is often done. Obviously, the most natural norm of this function is the L^∞ norm, i.e. the supremum of $|D_N(x)|$over $x\in[0,1]^d$, often referred to as *the star-discrepancy*. In fact the term *star-discrepancy* is reserved for the sup-norm of the normalized discrepancy function, i.e. $\frac{1}{N}\|D_N\|_\infty$, however since we only use the unnormalized version in this text, we shall abuse the language and apply this term to $\|D_N\|_\infty$.

Instead of directly estimating the L^∞ norm of the discrepancy function $\|D_N\|_\infty = \sup_{x\in[0,1]^d} \big|D_N(x)\big|$, Roth considered a smaller quantity, namely its L^2 norm $\|D_N\|_2$. This substitution allowed for an introduction of a variety of Hilbert space techniques, including orthogonal decompositions. In this setting Roth proved

Theorem 1 (Roth, 1954, [85]). *In all dimensions $d\geq 2$, for any N-point set $\mathscr{P}_N \subset [0,1]^d$, one has*

$$\big\|D_N\big\|_2 \geq C_d \log^{\frac{d-1}{2}} N, \tag{5}$$

where C_d is an absolute constant that depends only on the dimension d. This in particular implies that

$$\sup_{x\in[0,1]^d} \big|D_N(x)\big| \geq C_d \log^{\frac{d-1}{2}} N. \tag{6}$$

It was also shown that, when $d=2$, inequality (6) is equivalent to (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 (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 index of the sequence (or time) as an additional dimension. In this text, we adopt the former geometrical, 'static' formulation of the problems.

According to Roth's own words, these results "started a new theory" [34]. The paper [85] in which it was presented, entitled "On irregularities of distribution", has had a tremendous influence on the further development of the field. Even the number of papers with identical or similar titles, that appeared in subsequent years, attests to its importance: 4 papers by Roth himself (*On irregularities of distribution. I-IV*, [85, 86, 87, 88]), one by H. Davenport (*Note on irregularities of distribution*, [42]), 10 by W. M. Schmidt (*Irregularities of distribution. I-X*, [90, 91, 92, 93, 94, 95, 96, 97, 98, 99]), 2 by J. Beck (*Note on irregularities of distribution. I-II*, [6, 7]), 4 by W. W. L. Chen (*On irregularities of distribution. I-IV*, [29, 30, 31, 32]), at least 2 by Beck and Chen (*Note on irregularities of distribution. I-II*, [10, 12] 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"*, [11].

The technique proposed in the aforementioned paper was no less important than the results themselves. Roth was the first to apply the expansion of the discrepancy function D_N in the classical orthogonal Haar basis. Furthermore, he realized that in order to obtain good estimates of $\|D_N\|_2$ it suffices to consider just its projection onto the span of those Haar functions which are supported on dyadic rectangles of volume roughly equal to $\frac{1}{N}$. This is heuristically justified by the fact that, for a well distributed set, each such rectangle contains approximately one point. To be even more precise, the size of the rectangles R was chosen so that $|R| \approx \frac{1}{2N}$, ensuring that about half of all rectangles are free of points of $\mathscr{P}_N$. The Haar coefficients of D_N, corresponding to these empty rectangles, are then easy to compute, which leads directly to the estimate (5). This idea is the main theme of §2. Roth's approach strongly resonates with Kolmogorov's method of proving lower error bounds for cubature formulas, see e.g. [109, Chapter IV]. We shall discuss these ideas in more detail in §2.3.

A famous quote attributed to G. Polya [83] says,

> What is the difference between method and device? A method is a device which you used twice.

In agreement with this statement, over the past years Roth's clever device has indeed evolved into a powerful and versatile method: it has been applied an enormous number of times to various problems and questions in discrepancy theory and other areas. Our survey is abundant in such applications: discrepancy estimates in other function spaces §3.4, estimates of the star-discrepancy §4.4, §5, the small ball inequality §4.3, §5, constructions of low-discrepancy distrbutions §6.

Roth's L^2 result has been extended to other L^p norms, $1 < p < \infty$, only significantly later by W. Schmidt in [99, 1977], who showed that in all dimensions $d \geq 2$, for all $p \in (1,\infty)$ the inequality

$$\|D_N\|_p \geq C_{d,p} \log^{\frac{d-1}{2}} N, \tag{7}$$

holds for some constant $C_{d,p}$ independent of the collection of points $\mathscr{P}_N$. 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 just a couple of lines using Littlewood–Paley inequalities. A comprehensive discussion will be provided in §3.

While the case $1 < p < \infty$ is thoroughly understood, the endpoint case $p = \infty$, i.e. the *star-discrepancy*, is much more mysterious, despite the fact that it is most natural and important in the theory as it describes the worst possible discrepancy. It turns out that Roth's inequality (6) 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 extremal (i.e. L^∞) norm must necessarily be much larger than its average (e.g. L^2) 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:

$$\sup_{x\in[0,1]^d} \big|D_N(x)\big| \geq C\log N, \tag{8}$$

which is known to be 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 [70, 1904], van der Corput [38, 1935] and others, see e.g. §6.

Several other proofs of Schmidt's inequality (8) have been given later [72, 1979], [13, 1982], [52, 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 §4.4. Incidentally, the title of Halász's article [52] ("On Roth's method in the theory of irregularities of point distributions") almost coincides with the title of this chapter.

Higher dimensional analogs of Schmidt's estimate (8), however, turned out to be extremely proof-resistant. For a long time inequality (6) 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 [8, 1989], who proved that in dimension $d = 3$, discrepancy function satisfies

$$\|D_N\|_\infty \geq C\log N \cdot (\log\log N)^{\frac{1}{8}-\varepsilon}. \tag{9}$$

Almost twenty years later, in 2008, the author jointly with M. Lacey and A. Vagharshakyan ([17], $d = 3$; [18], $d \geq 4$) obtained the first significant improvement of the L^∞ bound in *all dimensions* $d \geq 3$:

Theorem 2 (Bilyk, Lacey, Vagharshakyan). *For all* $d \geq 3$*, there exists some* $\eta = \eta(d) > 0$*, such that for all* $\mathscr{P}_N \subset [0,1]^d$ *with* $\#\mathscr{P}_N = N$ *we have the estimate:*

$$\|D_N\|_\infty \geq C_d \, (\log N)^{\frac{d-1}{2}+\eta} \, . \tag{10}$$

The exact rate of growth of the star-discrepancy in higher dimensions remains an intriguing question; in their book [11], Beck and Chen named it ‘‘the great open problem” and called it ‘‘excruciatingly difficult”.

Even the precise form of the conjecture is a subject of ongoing debate among the experts in the field. The opinions are largely divided between two possible formulations of this conjecture. We start with the form which is directly pertinent to the orthogonal function method.

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

$$\|D_N\|_\infty \geq C_d \, (\log N)^{\frac{d}{2}} \, . \tag{11}$$

This conjecture is motivated by connections of this field to other areas of mathematics and, in particular, by a related conjecture in analysis, the *small ball conjecture* (111), which is known to be sharp, see §4.2. Unfortunately, this relation is not direct – it is not known whether the validity of the small ball conjecture implies the discrepancy estimate (11), the similarity lies just in the methods of proof . But, at the very least, this connection suggests that Conjecture 1 is the best result that one can achieve using Roth’s Haar function method.

On the other hand, the best known examples [53, 55] of well distributed sets in higher dimensions have star-discrepancy of the order

$$\|D_N\|_\infty \leq C_d \, (\log N)^{d-1} \, . \tag{12}$$

Numerous constructions of such sets are known and are currently a subject of massive ongoing research, see e.g. the book [43]. These upper bounds together with the estimates for a “smooth” version of discrepancy (see Temlyakov [114]), provide grounds for an alternative form of the conjecture (which is actually older and more established)

Conjecture 2. For all $d \geq 3$ and all $\mathscr{P}_N \subset [0,1]^d$ with $\#\mathscr{P}_N = N$ we have the estimate:

$$\|D_N\|_\infty \geq C_d \, (\log N)^{d-1} \, . \tag{13}$$

One can notice that both conjectures coincide with Schmidt’s estimate (8) when $d = 2$. Skriganov has proposed yet another form of the conjecture [103]:

$$\|D_N\|_\infty \geq C_d \, (\log N)^{\frac{d-1}{2}+\frac{d-1}{d}} \, , \tag{14}$$

which is exact both in $d = 1$ and $d = 2$.

In contrast to the L^∞ inequalitites, it is well known that in the average (L^2 or L^p) sense Roth's bound (3), as well as inequality (7), is sharp. This was initially proved by Davenport [42] in two dimensions for $p = 2$, who constructed point distributions with $||D_N\|_2 \le C\sqrt{\log N}$. Subsequently, different constructions have been obtained by numerous other authors, including Roth [87, 88], Chen [29], Frolov [48]. It should be noted that most of the optimal constructions in higher dimensions $d \ge 3$ are probabilistic in nature and are obtained as randomizations of some classic low-discrepancy sets. In fact, deterministic examples of sets with $\|D_N\|_p \le C_{d,p} \log^{\frac{d-1}{2}} N$ have been constructed only in the last decade by Chen and Skriganov [35, 36] ($p = 2$) and Skriganov [102] ($p > 1$). 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 one of the ideas that appear in Roth's proof.

The other endpoint of the L^p scale, $p = 1$, is not any less (and perhaps even more) difficult than the star-discrepancy estimates. The only information that is available is the two-dimensional inequality (proved in the aforementioned paper of Halász [52]), which also makes use of Roth's orthogonal function method:

$$\|D_N\|_1 \ge C\sqrt{\log N}, \tag{15}$$

This means 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 3. For all $d \ge 3$ and all sets of N points in $[0,1]^d$:

$$\|D_N\|_1 \ge C_d \left(\log N\right)^{\frac{d-1}{2}}. \tag{16}$$

However, almost no results pertaining to this conjecture have been discovered for $d \ge 3$. The only known relevant fact is that $\sqrt{\log N}$ bound still holds in higher dimensions, i.e. it is not even known if the exponent increases with dimension. The reader is referred to §4.5 for Halász's L^1 argument.

1.2 Preliminary discussion

While the main subject of this chapter is Roth's method in discrepancy theory, we are also equally concentrated on its applications and relations to a wide array of problems extending to topics well beyond discrepancy. One of our principal intentions is to stress the connections between different areas of mathematics and accentuate the use of the methods of harmonic analysis in discrepancy and related fields. Having aimed to cover such a broad range of topics, we left ourselves with little chance to make the exposition very detailed and full of technicalities. Instead, we decided to focus on the set of ideas, connections, arguments, and conjectures that permeate discrepancy theory and several other subjects.

We assume only very basic prior knowledge of any of the underlying fields, introducing and explaining the new concepts as they appear in the text, discussing basic properties, and providing ample references. In particular, we believe this chapter to be a very suitable reading for graduate students as well as for mathematicians of various backgrounds interested in discrepancy or any of the discussed areas. In an effort to make our exposition reader-friendly and accessible, we often sacrifice generality, and sometimes even rigor, in favor of making the presentation more intuitive, providing simpler and more transparent arguments, or explaining the heuristics and ideas behind the proof. The reader however should *not* get the impression that this chapter is void of mathematical content. In fact, a great number of results are meticulously proved in the text and numerous computations, which could have been skipped in a technical research paper, are carried out in full detail.

1.2.1 A brief outline of the chapter

Even though our exposition consists of several distinct sections which sometimes deal with seemingly unrelated subjects, every section naturally continues and interlaces with the discussion of the previous ones. In the next several paragraphs we give a brief 'sneak preview' of the content of this chapter.

- In §2 we introduce the reader to the main ideas of Roth's L^2 method. We start with the necessary definitions and background information on Haar functions and product orthogonal bases and then proceed to explain a general principle behind Roth's argument. We then give the proof of the L^2 discrepancy bound, Theorem 1. We present Roth's original proof which relies on duality and the Cachy–Schwarz inequality, as well as a slightly different argument which makes use of orthogonality and Bessel's inequality directly. In the end of §2 we turn to Kolmogorov's method of obtaining lower bounds for errors of cubature formulas on various function classes governed by the behavior of the mixed derivative. The method is based on the same idea as Roth's method in discrepancy theory and provides an important connection between these two intimately related areas.

- Extensions of Theorem 1 even to L^p spaces with $p \neq 2$ turned out to be somewhat delicate and not immediate. However, harmonic analysis provides means to make these extensions almost automatic. This instrument, the Littlewood–Paley inequalities, is the subject of §3. The Littlewood–Paley serves as a natural substitute for orthogonality in non-Hilbert spaces, e.g. L^p. In §3 we discuss the relevant version of this theory – the dyadic Littlewood–Paley inequalities, starting with the one-dimensional case and then moving forward to the multiparameter setting. We also discuss the connections of this topic to objects in probability theory such as the famous Khintchine inequality and the martingale difference square function. Unfortunately, unlike many other methods of harmonic analysis, Littlewood–Paley theory has not yet become a "household name" among experts in various fields outside analysis. It is our sincere hope that our exposition will further publicize and popularize this powerful method.

- Next, we demonstrate how these tools can be used to extend Roth's L^2 discrepancy estimate to L^p essentially in one line. Further, a large portion of §3 is devoted to the discussion of discrepancy estimates analogous to Theorem 1 in various function spaces, such as Hardy, Besov, BMO, weighted L^p, and exponential Orlicz spaces. All of these results, in one way or another, take their roots in Roth's method and the Littlewood–Paley (or similar in spirit) inequalities.

- In §4 we turn to arguably the most important problem of discrepancy theory – sharp estimates of the star-discrepancy (L^∞ norm of the discrepancy function). We introduce the *small ball inequality* – a purely analytic inequality which is concerned with lower bounds of the supremum norm of sums of Haar functions supported by rectangles of fixed size. The very structure of these sums suggests certain connections with Roth's method in discrepancy. And indeed, even though it is not known if one problem directly implies the other, there are numerous similarities in the known methods of proof and the small ball inequality may be viewed as a linear model of the star-discrepancy method. We state the small ball conjecture and discuss known results and its sharpness, which indirectly bears some effect on the sharpness of the relevant discrepancy conjectures.

- In §4.3 we present a beautiful proof of the small ball conjecture in dimension $d = 2$. We then proceed to demonstrate an amazingly similar proof of Schmidt's lower bound (8) for the star-discrepancy in $d = 2$ as well as a proof of the L^1 discrepancy bound (15). All three proofs are based on an ingenious method known as the Riesz product. To reinforce the connections of these problems with the classical problems of analysis, in §4.6 we briefly discuss the area in which Riesz product historically first appeared – lacunary Fourier series. We give a proof of Sidon's theorem whose statement, as well as the argument used to prove it, resemble both the small ball inequality and the discrepancy estimates in great detail and perhaps have inspired their respective proofs.

- The small ball inequality turns out to be connected to other areas of mathematics besides discrepancy – in particular, approximation theory and probability. In §§4.7-4.8 we describe the relevant problems: the small deviation probabilities for the Brownian sheet and the metric entropy of function classes with mixed smoothness. We demonstrate that the small ball inequality directly implies lower bounds in both of these problems and hence indirectly ties them to discrepancy.

- A substantial part of this chapter, §5, focuses on the important recent developments in the subject, namely the first significant improvement in the small ball inequality and the L^∞ discrepancy estimates in all dimensions $d \geq 3$. We thoroughly discuss the main steps and ingredients of the proof, intuitively explain many parts of the argument and pinpoint the arising difficulties without going too deep into the technical details. This approach, in our opinion, will allow one to comprehend the ‘big picture’ and the strategy of the proof. An interested reader will then be well-equipped and prepared to fill in the complicated technicalities by consulting the provided references.

- Finally, in §6 our attention makes a 180-degree turn from lower bounds to constructions of well-distributed point sets and upper discrepancy estimates. We introduce one of the most famous low-discrepancy distributions in two dimensions – the van der Corput digit reversing set, whose binary structure makes it a perfect fit for the tools of dyadic analysis and Roth’s method. We describe certain modifications of this set, which achieve the optimal order of discrepancy in various function spaces, in particular, demonstrating the sharpness of some of the results in §3.4.

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 [11, 27, 43, 64, 75, 109]; the book [77] studies the relations between uniform distribution and harmonic analysis, [116] views the subject through the lens of the function space theory, while [104] specifically investigates the connections between discrepancy and Haar functions. In addition, the survey [37] explores various ideas of Roth in discrepancy theory, including the method discussed here. Finally, [66] and [16] are very similar in spirit to this chapter; however, the survey [16] is much more concise than the present text, and the set of notes [66] focuses primarily on the underlying harmonic analysis. We have tried to make the to presentation accessible to a wide audience, rather than experts in one particular area, yet at the same time inclusive, embracing and accentuating the connections between numerous topics. We sincerely hope that, despite a vast amount of literature on the subject, this chapter will provide some novel ideas and useful insights and will be of interest to both novices and specialists in the field.

1.2.2 Some other problems related to Roth’s method

Unfortunately, there are still a number of topics that either grew directly out of Roth’s method or are tangentially, but strongly correlated with it, which we will not be touching upon in this survey, since these discussions would have taken us very far afield. They include, in particular, Beck’s beautiful lower bound on the growth of polynomials with roots on the unit circle [9]. By an argument, very similar to the Halász’s proof of the two-dimensional star-discrepancy estimate, Beck showed that there exists a constant $\delta > 0$ such that for any infinite sequence $\{z_n\}_{n=1}^{\infty}$ of unimodular complex numbers and polynomials $P_N(z) = \prod_{n=1}^{N}(z - z_n)$ the bound

$$\sup_{|z|\le 1} |P_N(z)| > N^{\delta} \tag{17}$$

holds for infinitely many values of N, thus giving a negative answer to a question of Erdős.

Another problem considered by Beck and Roth deals with the so-called combinatorial discrepancy, a natural companion of the geometric discrepancy. Let the function $\lambda : \mathscr{P}_N \to \{\pm 1\}$ represent a "red-blue" coloring of an N-point set $\mathscr{P}_N \subset [0,1]^d$. The *combinatorial discrepancy* of $\mathscr{P}_N$ with respect to a family of sets $\mathscr{B}$ is defined as $T(\mathscr{P}_N) = \inf_{\lambda} \sup_{B \in \mathscr{B}} \Big| \sum_{p \in \mathscr{P}_N \cap B} \lambda(p) \Big|$, i.e. the minimization of the largest disbalance of colors in sets from $\mathscr{B}$ over all possible colorings. In [5], Beck discovered that, when $\mathscr{B}$ is the family of axis-parallel boxes, the quantity $T(N) := \sup_{\mathscr{P}_N} T(\mathscr{P}_N)$ is tightly related to the discrepancy function estimates. In particular, in $d = 2$ one has $T(N) \gtrsim \log N$. In [89] Roth has extended this to real-valued functions λ (continuous coloring) showing that

$$T(N) \gtrsim \frac{(\log N)}{N} \sum_{p \in \mathscr{P}_N} |\lambda(p)|. \tag{18}$$

Roth's argument relied on Haar expansions and Riesz product and almost repeated the proof of the L^∞ discrepancy bound in dimension two with an addition of some new ideas. Recent progress (10) on the discrepancy function directly yields an analogous improvement in the "red-blue" case for $d \geq 3$ and can be adjusted to provide a similar estimate for "continious" colorings in dimension $d = 3$.

There are numerous other examples. Chazelle [28] has applied a discrete version of Roth's orthogonal function method to a problem in computational geometry, obtaining a lower bound for the complexity of orthogonal range searching. The Riesz product techniques, similar to Halász's, have been used in approximation theory for a long time to obtain Bernstein-type inequalities, estimates for entropy numbers and Kolmogorov widths of certain function classes, see e.g. [111, 112, 113]. We shall only briefly discuss some of these connections in §4.8.

This diverse set of topics shows the universality and ubiquitousness of the method and ideas under discussion.

1.2.3 Notation and conventions

Before we proceed to the mathematical part of the text, we would like to explain some of the notation and conventions that we shall be actively using. Since many different constants arise in our discussion, 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 relation $F \approx G$ means that $F \lesssim G$ and $G \lesssim F$. The implicit constants in such inequalities will be allowed to depend on the dimension and, perhaps, some other parameters, but never on the number of points N.

In other words, in this survey we are interested in the asymptotic behavior of the discrepancy when the dimension is fixed and the number of points increases. Therefore, such effects as the *curse of dimensionality* do not come into play. Finding optimal estimates as the dimension goes to infinity is a separate, very interesting and important subject, see e.g. [56]. While one may argue that these questions are sometimes more useful for applications, we firmly insist that the questions discussed here, which go back to van der Corput, van Aardenne-Ehrenfest, and Roth, are at least equally as important, especially considering the fact that in such natural (and low!) dimensions as, say, 3 or 4 the exact rate of growth of discrepancy is far from being understood and the relative gap between the lower and upper estimates is quite unsatisfactory.

Throughout the text several variables will have robustly reserved meanings. The dimension will always be denoted by d. Capital N will always stand for the number of points, while n will represent the scale and will usually satisfy $n \approx \log N$. Unless otherwise specified, all logarithms are taken to be natural, although this is not so important since we are not keeping track of the constants. The discrepancy function of an N-point set $\mathscr{P}_N \subset [0,1]^d$ will be denoted either by $D_{\mathscr{P}_N}$ or, more often, if this creates no ambiguity, simply by D_N. Recall that, unlike a number of standard references, we are considering the unnormalized version of discrepancy, i.e. we do not divide by N in the definition (4). The term *star-discrepancy* refers to the L^∞ norm of D_N.

For a set $A \subset \mathbb{R}^d$, its Lebesgue measure is denoted either by $|A|$ or by $\mu(A)$. For a finite set F, we use $\#F$ to denote its cardinality – the number of elements of F. Whenever we have to resort to probabilistic concepts, $\mathbb{P}$ will stand for probability and $\mathbb{E}$ for expectation.

2 Roth's orthogonal function method and the L^2 discrepancy

Before we begin a detailed discussion of Roth's method, we need to introduce and define its main tool, Haar functions. We shall then explain Roth's main idea and proceed to reproduce his original proof of Theorem 1, although our exposition will slightly differ from the style of the original paper [85] (the argument, 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. Along the way, we shall try to look at the argument at different angles and to find motivation behind some of the steps of the proof.

2.1 Haar functions and Roth's principle

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]:

$$\mathscr{D} = \left\{ I = \left[\frac{m}{2^n}, \frac{m+1}{2^n}\right) : m,n \in \mathbb{Z}, n \geq 0, 0 \leq m < 2^n \right\}. \tag{19}$$

Dyadic intervals form a *grid*, meaning that any two intervals in $\mathscr{D}$ are either disjoint, or one is contained in another. In addition, for every interval $I \in \mathscr{D}$, 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

$$h_I(x) = -\mathbf{1}_{I_l}(x) + \mathbf{1}_{I_r}(x). \tag{20}$$

Notice that in our definition Haar functions are normalized to have unit norm 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 to those readers who are accustomed to the L^2 normalization.

These functions have been introduced by Haar [51, 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

$$\langle h_{I'}, h_{I''} \rangle = \int_0^1 h_{I'}(x) \cdot h_{I''}(x)\, dx = 0, \qquad I', I'' \in \mathscr{D}, I' \neq I'', \tag{21}$$

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

In order to simplify the notation and make it more uniform, we shall sometimes employ the following trick. Denote by $\mathscr{D}_* = \mathscr{D} \cup \{[-1,1]\}$ the dyadic grid on $[0,1]$ with the interval $[-1,1]$ added to it. Then the family $\mathscr{H} = \{h_I\}_{I \in \mathscr{D}_*}$ forms an orthogonal basis of $L^2([0,1])$. In other words, the constant function on $[0,1]$ can be viewed as a Haar function of order -1.

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

$$h_R(x_1, \ldots, x_d) = h_{I_1}(x_1) \cdot \ldots \cdot h_{I_d}(x_d). \tag{22}$$

The orthogonality of these functions is easily derived from the one dimensional property. It is also well known that the 'product' Haar system $\mathscr{H}^d = \{f(x) = f_1(x_1) \cdot \ldots \cdot f_d(x_d) : f_k \in \mathscr{H}\}$ is an orthogonal basis of $L^2([0,1]^d)$ – often referred to as the *product Haar basis*. The construction of product bases starting from a one-dimensional orthogonal basis is also valid for more general systems of orthogonal functions. In view of the previous remark, one can write $\mathscr{H}^d = \{h_R\}_{R \in \mathscr{D}_*^d}$, although most of the times we shall restrict our attention to rectangles in $\mathscr{D}^d$. Thus, every function $f \in L^2([0,1]^d)$ can be written as

$$f = \sum_{R \in \mathscr{D}_*^d} \frac{\langle f, h_R \rangle}{|R|} h_R, \tag{23}$$

where the series converges in L^2. If this expression seems slightly uncoventional, this is a result of the L^∞ normalization of h_R. We note that this is not the only way to extend wavelet bases to higher dimensions [41], but this multiparameter 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 (multiparameter) harmonic analysis – we shall keep returning to this point throughout the text.

One of the numerous important contributions of Klaus Roth to discrepancy theory is the idea of using orthogonal function (in particular, Haar) decompositions in order to obtain discrepancy estimates. This idea was introduced already in his first paper on irregularities of distribution [85]. Even though Haar functions have been introduced almost simultaneously to some questions connected with uniform distribution theory and numerical integration (see Sobol's book [104]), their power for discrepancy estimates only became apparent with Roth's proof of the lower bound for the L^2 bound of the discrepancy function.

In addition to introducing a new tool to the field, Roth has clearly demonstrated a proper way to use it. An orthogonal expansion may be of very little use to us, unless we know how to extract information from it and which coefficients play the most important role. The method of proof of the L^2 bound (5) unambiguously suggests where one should look for the most relevant input of the decomposition to the intrinsic features of the discrepancy function. Further success of this approach in various discrepancy setting and connections to other areas and problems, described throughout this chapter, validates the correctness of the idea and turns it into a method. We formulate it here as a general principle.

Roth's principle: The behavior of the discrepancy function is essentially defined by its projection onto the span of Haar functions h_R supported by rectangles of volume $|R| \approx \frac{1}{N}$, i.e.

$$D_N \sim \sum_{R \in \mathscr{D}^d : \, |R| \approx \frac{1}{N}} \frac{\langle D_N, h_R \rangle}{|R|} h_R. \tag{24}$$

In the formulation of this principle, we interpret the symbols '$\sim$' and '$\approx$' very loosely and broadly. This principle as such should not be viewed as a rigorous mathematical statement. It is rather a circle of ideas and a heuristic approach. In this chapter we shall see many manifestations of this principle in discrepancy theory (both for upper and lower estimates) and will draw parallels with similar methods and ideas in other fields, such as approximation theory, probability, and harmonic analysis.

An intuitive explanation of this principle, perhaps, lies in the fact that, for 'nice' distributions of points $\mathscr{P}_N$, any dyadic rectangle of area $|R| \approx \frac{1}{N}$ would contain roughly one point (or the number of empty rectangles is comparable to the number of of points). At fine scales, the boxes are too small and most of the time they contain no points of $\mathscr{P}_N$ and hence do not carry much information about the discrepancy. While rectangles that are too big (coarse scales) incorporate too much cancellation: the discrepancy of $[0,1]^d$, for example, is always zero. (We should note that large rectangles, however, often give important additional information, see e.g. [58]). Therefore, the intermediate scales are the most important ones. Of course, this justification is too naive and simplistic and does not provide a complete picture. Some details will become more clear after discussing the proof of (5) which we turn to now.

2.2 The proof of the L^2 discrepancy estimate

As promised we shall now reconstruct Roth's original proof of the L^2 estimate (5). Following the general lines of Roth's principle (24), we consider dyadic rectangles $R \in \mathscr{D}^d$ of volume $|R| = 2^{-n} \approx \frac{1}{N}$. To be more exact, let us choose the number $n \in \mathbb{N}$ so that

$$2^{n-2} \le N < 2^{n-1}, \tag{25}$$

i.e. $n \approx \log_2 N$ (although the precise choice of n is important for the argument).

These rectangles come in a variety of shapes, especially in higher dimension. This fact dramatically increases the combinatorial complexity of the related problems. To keep track of these rectangles we introduce a special bookkeeping device – a collection of vectors with non-negative integer coordinates

$$\mathbb{H}_n^d = \{\mathbf{r} = (r_1, \dots, r_d) \in \mathbb{Z}_+^d : \|\mathbf{r}\|_1 = n\}, \tag{26}$$

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

$$\#\mathbb{H}_n^d = \binom{n+d-1}{d-1} \approx n^{d-1}, \tag{27}$$

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 $\|\mathbf{r}\|_1 = n$ or $|R| = 2^{-n}$.

We shall say that a function f on $[0,1]^d$ is an r-function with parameter $\mathbf{r} \in \mathbb{Z}_+^d$ if f is of the form

$$f(x) = \sum_{R \in \mathscr{D}_\mathbf{r}^d} \varepsilon_R h_R(x), \tag{28}$$

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 functions. 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 $\mathbf{r}$ are orthogonal readily follows from the orthogonality of the family of Haar functions.

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

$$D_N(x) = \sum_{p\in\mathscr{P}_N} \mathbf{1}_{[p,\mathbf{1}]}(x) - N\cdot x_1\cdot\dots\cdot x_d, \tag{29}$$

where $\mathbf{1} = (1,\dots,1)$ and $[p,\mathbf{1}] = [p_1,1]\times\dots\times[p_d,1]$. 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

$$\int \mathbf{1}_{[q,1]}(x)\cdot h_I(x)\,dx = \int_q^1 h_I(x)\,dx = 0 \tag{30}$$

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

$$\int_{[0,1]^d} \mathbf{1}_{[p,\mathbf{1}]}(x)\cdot h_R(x)\,dx = \prod_{j=1}^d \int_{p_j}^1 h_{R_j}(x_j)\,dx_j = 0 \tag{31}$$

when $p\notin R$. Assume now that a rectangle $R\in\mathscr{D}^d$ is empty, i.e. does not contain points of $\mathscr{P}_N$. 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:

$$\left\langle \sum_{p\in\mathscr{P}_N} \mathbf{1}_{[p,\mathbf{1}]}, h_R \right\rangle = 0. \tag{32}$$

In other words, if R is free of points of $\mathscr{P}_N$, 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:

$$\langle Nx_1\dots x_d, h_R\rangle = N\prod_{j=1}^d \langle x_j, h_{R_j}(x_j)\rangle = N\cdot\frac{|R|^2}{4^d}. \tag{33}$$

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

$$\langle D_N, h_R\rangle = -N|R|^2 4^{-d}. \tag{34}$$

These, somewhat mysterious, computations can be explained geometrically (see [99], also [27, 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 $\mathscr{P}_N$ 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

$$D_N\big(x\big) - D_N\big(x + (h_1, 0)\big) + D_N\big(x + (h_1, h_2)\big) - D_N\big(x + (0, h_2)\big) = -N \cdot h_1 h_2 = -N \cdot \frac{|R|}{4}. \tag{35}$$

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

$$\int_{R'} \Big(D_N\big(x\big) - D_N\big(x + (h_1, 0)\big) + D_N\big(x + (h_1, h_2)\big) - D_N\big(x + (0, h_2)\big) \Big)\, dx = -N \cdot \frac{|R|}{4} \cdot |R'| = -N \cdot \frac{|R|^2}{4^2}, \tag{36}$$

while, on the other hand,

$$\int_{R'} \Big(D_N\big(x\big) - D_N\big(x + (h_1, 0)\big) + D_N\big(x + (h_1, h_2)\big) - D_N\big(x + (0, h_2)\big) \Big)\, dx = \int_R D_N(x) \cdot h_R(x)\, dx = \langle D_N, h_R \rangle. \tag{37}$$

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

We are now ready to prove a crucial preliminary lemma.

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

$$\langle D_N, f_{\mathbf{r}} \rangle \geq c_d > 0, \tag{38}$$

where the constant c_d depends on the dimension only.

Proof. Construct the function $f_{\mathbf{r}}$ in the following way:

$$f_{\mathbf{r}} = \sum_{R \in \mathscr{D}_{\mathbf{r}}^d : R \cap \mathscr{P}_N = \emptyset} (-1) \cdot h_R + \sum_{R \in \mathscr{D}_{\mathbf{r}}^d : R \cap \mathscr{P}_N \neq \emptyset} \operatorname{sgn}\big(\langle D_N, h_R \rangle\big) \cdot h_R \tag{39}$$

By our choice of n (25), at least 2^{n-1} of the 2^n rectangles in $\mathscr{D}_{\mathbf{r}}^d$ must be free of points of $\mathscr{P}_N$. It then follows from (32) and (33) that

$$\langle D_N, f_{\mathbf{r}} \rangle \geq - \sum_{R \cap \mathscr{P}_N = \emptyset} \langle D_N, h_R \rangle = \sum_{R \cap \mathscr{P}_N = \emptyset} \langle N x_1 \dots x_d, h_R \rangle \tag{40}$$

$$= \sum_{R \cap \mathscr{P}_N = \emptyset} N \cdot \frac{|R|^2}{4^d} \geq 2^{n-1} \cdot 2^{n-2} \cdot \frac{2^{-2n}}{4^d} = c_d.$$

Remark. Roth [85] initially defined the functions $f_{\mathbf{r}}$ slightly differently: he set them equal to zero on those dyadic rectangles which *do* contain points of $\mathscr{P}_N$, i.e. Roth's functions consisted only of the first term of (39). While this bears no effect on this argument, it was later realized by Schmidt [99] 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 (39)). Schmidt's paper, as well as subsequent papers by Halász [52], Beck [8], the author of this chapter and collaborators [17, 18, 19, 20], make use of the r-functions as defined here (28). As we shall see in §5.5, in certain cases this definition brings substantial simplifications, whereas allowing even a small number of zeros in the definition may significantly complicate matters.

We are now completely equipped to prove Roth's theorem. Lemma 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. Roth's original proof made use of duality. Let us construct the following test function:

$$F = \sum_{\mathbf{r} \in \mathbb{H}_n^d} f_{\mathbf{r}}, \tag{41}$$

where $f_{\mathbf{r}}$ are the r-functions provided by Lemma 1. Orthogonality of $f_{\mathbf{r}}$'s yields:

$$\|F\|_2 = \left(\sum_{\mathbf{r} \in \mathbb{H}_n^d} \|f_{\mathbf{r}}\|_2^2 \right)^{1/2} = (\#\mathbb{H}_n^d)^{1/2} \approx n^{\frac{d-1}{2}}, \tag{42}$$

while Lemma 1 guarantees that

$$\langle D_N, F \rangle \geq (\#\mathbb{H}_n^d) \cdot c_d \approx n^{d-1}. \tag{43}$$

Now Cauchy–Schwarz inequality easily implies that:

$$\|D_N\|_2 \geq \frac{\langle D_N, F \rangle}{\|F\|_2} \gtrsim n^{\frac{d-1}{2}} \approx \big(\log N\big)^{\frac{d-1}{2}}, \tag{44}$$

which finishes the proof. □

As one can see, the role of the building blocks is played by the generalized Rademacher functions $f_{\mathbf{r}}$, which we shall observe again in many future arguments. Therefore it is naturally convenient that they are normalized both in L^∞ and in L^2.

One, of course, does not have to use duality to obtain this inequality: we could use orthogonality directly. This proof initially appeared in an unpublished manuscript of A. Pollington and its analogs are often useful when one wants to prove estimates in quasi-Banach spaces and is thus forced to avoid duality arguments, see e.g. (95). For the sake of completeness, we also include this variation of the proof.

Second proof of Theorem 1. The proof is based on the same idea. Let n be chosen as in (25). We use orthogonality, Bessel's inequality and (33) to write

$$\begin{aligned} \|D_N\|_2^2 &\geq \sum_{|R|=2^{-n},\, R\cap\mathscr{P}_N=\emptyset} \frac{|\langle D_N, h_R\rangle|^2}{|R|} = \sum_{\mathbf{r}\in\mathbb{H}_n^d} \sum_{R\in\mathscr{D}_{\mathbf{r}}^d:\, R\cap\mathscr{P}_N=\emptyset} N^2 \cdot \frac{2^{-4n}}{2^{-n}\cdot 4^{2d}} \\ &\gtrsim (\#\mathbb{H}_n^d)\cdot 2^{n-1}\cdot 2^{2n-4} 2^{-3n} \approx n^{d-1} \approx \big(\log N\big)^{d-1}. \end{aligned} \tag{45}$$

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

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. $1/p + 1/q = 1$. Indeed, 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 (85). Hence, replacing Cauchy–Schwarz by Hölder's inequality in (44), one immediately recovers Schmidt's result:

$$\|D_N\|_p \gtrsim (\log N)^{\frac{d-1}{2}}. \tag{46}$$

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 that an analog of the second proof (45) can be carried out also in L^p using the device of the product Littlewood–Paley square function instead of orthogonality. The reader is invited to proceed to the next section, §3, for details.

Recently, Hinrichs and Markhasin [58] have slightly modified Roth's method to obtain the best known value of the constant C_d in Theorem 1. Their idea is quite clever and simple. They have noticed that one can extend the summation in (45) to also include finer scales, i.e. rectangles with smaller volume $|R| \leq 2^{-n}$. A careful computation then yields $C_2 = 0.0327633...$ and $C_d = \frac{1}{\sqrt{21}\cdot 2^{2d-1}\sqrt{(d-1)!}(\log 2)^{\frac{d-1}{2}}}$ for $d \geq 3$, where all logarithms are taken to be natural.

2.3 Lower bounds for cubature formulas on function classes: Kolmogorov's method

Before finishing the discussion of Roth's proof, we would like to highlight its striking similarity to some arguments in the closely related field of numerical integration: namely, Kolmogorov's proof of the lower estimate for the error of cubature formulas in the class $MW_r^p([0,1]^d)$ of functions whose r^{th} mixed derivative has L^p norm at most one. For the purposes of our introductory exposition, we shall define these spaces in a naive fashion. For more general settings and a wider array of numerical integration results and their relations to discrepancy theory, the reader is referred to e.g. [109, 114, 43, 79].

Define the integration operator $(\mathscr{T}_d f)(x_1,\dots,x_d) := \int_0^{x_1}\dots\int_0^{x_d} f(y)\,dy_1\dots dy_d$. For $p\geq 1$ and an integer $r\geq 1$, define the space $MW_r^p([0,1]^d) = (\mathscr{T}_d)^r(L^p([0,1]^d))$, i.e. the image of L^p under the action of an r-fold composition of the integration operators. Let $B(L^p)$ be the unit ball of L^p and $B(MW_r^p) = (\mathscr{T}_d)^r(B(L^p))$ be its image, i.e. the unit ball of MW_r^p or the set of functions whose r^{th} mixed derivative has L^p norm at most one. We shall encounter these classes again in §4.8.

The field of numerical integration is concerned with approximate computations of integrals and evaluations of the arising errors. Let $\mathscr{F}$ be a class of functions on $[0,1]^d$ and $\mathscr{P}_N\subset[0,1]^d$ be a set of N points. For an arbitrary function f on $[0,1]^d$, define the cubature formula associated to $\mathscr{P}_N$ as

$$\Lambda(f,\mathscr{P}_N) = \frac{1}{N}\sum_{p\in\mathscr{P}_N} f(p). \tag{47}$$

Denote by $\Lambda_N(\mathscr{F},\mathscr{P}_N)$ the supremum of the errors of this cubature formula over the class $\mathscr{F}$:

$$\Lambda_N(\mathscr{F},\mathscr{P}_N) := \sup_{f\in F}\left|\Lambda(f,\mathscr{P}_N) - \int\dots\int_{[0,1]^d} f(x)dx_1...dx_d\right|. \tag{48}$$

The infimum of this quantity over all choices of the point set $\mathscr{P}_N$ is the optimal error of the N-point cubature formulas on the class $\mathscr{F}$:

$$\delta_N(\mathscr{F}) := \inf_{\mathscr{P}_N:\#\mathscr{P}_N=N}\Lambda_N(\mathscr{F},\mathscr{P}_N). \tag{49}$$

Notice that the star-discrepancy, $\|D_N\|_\infty$, is equal to $N\cdot\Lambda_N(\mathscr{F},\mathscr{P}_N)$, where $\mathscr{F}$ is the class of characteristic functions of rectangles $[\mathbf{0},x)$. This is only the most trivial of the vast and numerous connections between numerical integrations and discrepancy theory. We recommend, for example, the book [43] for a very friendly introduction to the relations between these fields. Also, in the present book, the chapter by E. Novak and H. Woźniakowski is devoted to the discussion of discrepancy and integration.

We shall also consider the space of functions whose r^{th} mixed derivative satisfies the product Hölder condition. Recall that a univariate function f is called Hölder if the condition $|\Delta_t f(x)|\lesssim|t|$ for the difference operator $\Delta_t f(x) = f(x+t)-f(x)$ holds for all x. The multiparameter nature of the problems under consideration dictates that rather than using the standard generalization of this concept, we use the product version, where the difference operator is applied iteratively in each coordinate. For a vector $\mathbf{t}=(t_1,...,t_d)$, $t_j>0$, and a function f on $[0,1]^d$, define $\Delta_{\mathbf{t}} f(x) = \Delta_{t_d}^{x_d}(...\Delta_{t_1}^{x_1} f)...)(x)$, where the superscript indicates the variable in which the difference operator is being applied. We denote by $H([0,1]^d)$ the class of product

Hölder functions – those functions for which

$$\|\Delta_{\mathbf{t}} f\|_\infty \le C|t_1|\cdot\ldots\cdot|t_d|, \tag{50}$$

and let $B(H([0,1]^d))$ be the unit ball of this space, i.e. functions which are product Hölder with constant one: $\|\Delta_{\mathbf{t}} f\|_\infty \le |t_1|\cdot\ldots\cdot|t_d|$. Furthermore, denote by $B(MH_r([0,1]^d)) = (\mathscr{T}_d)^r(B(H([0,1]^d)))$ the class of functions whose r^{th} mixed derivative has Hölder norm one.

It is not hard to check that for a smooth function f we have

$$\Delta_{\mathbf{t}} f(x) = \int_{x_1}^{x_1+t_1}\cdots\int_{x_d}^{x_d+t_d} \frac{\partial^d f(y)}{\partial x_1\ldots\partial x_d}dy, \tag{51}$$

while it is also clear that $|\Delta_{\mathbf{t}} f(x)| \le 2^d\|f\|_\infty$. Hence

$$|\Delta_{\mathbf{t}} f(x)| \lesssim \min\left\{\left\|\frac{\partial^d f}{\partial x_1\ldots\partial x_d}\right\|_\infty \prod_{j=1}^d |t_j|,\ 2^d\|f\|_\infty\right\}. \tag{52}$$

We shall now demonstrate a method of proof of the lower bounds for the optimal integration errors $\delta_N(\mathscr{F})$ for some function classes. This method, which was invented by Kolmogorov, resembles Roth's method in discrepancy theory to a great extent. We shall prove the following theorem by means of an argument whose main idea the reader will easily recognize.

Theorem 3. *For any $r \in \mathbb{N}$, the optimal integration errors for the classes $B(MH_r)$ and $B(MW_r^2)$ satisfy the lower estimates*

$$\delta_N(B(MH_r)) \gtrsim N^{-r}(\log N)^{d-1}, \tag{53}$$

$$\delta_N(B(MW_r^2)) \gtrsim N^{-r}(\log N)^{\frac{d-1}{2}}. \tag{54}$$

Proof. The main idea of the method is to construct a function which is zero at all nodes of the cubature formula, but whose integral is large. Similarly to Roth's original proof of (5), this is achieved by appropriately defining the function on the dyadic rectangles which contain no chosen points.

We start by proving (53). Fix any positive infinitely-differentiable function $b(x)$ of one variable supported on the interval $[0,1]$. For a dyadic box $R = R_1\times\ldots\times R_d \in \mathscr{D}^d$, where $R_j = \left[k_j 2^{-s_j},(k_j+1)2^{-s_j}\right)$, define the functions

$$b_R(x_1,\ldots,x_d) := \prod_{j=1}^d b(2^{s_j}x_j - k_j). \tag{55}$$

The function b_R is obviously supported on the rectangle R. As in (25) we choose n so that $2N < 2^n \le 4N$. For each choice of $\mathbf{r} \in \mathbb{H}_n^d$, out of 2^n dyadic boxes $R \in \mathscr{D}_{\mathbf{r}}^d$, at least a half, 2^{n-1}, do not contain any points of $\mathscr{P}_N$. Set

$$G(x_1,...,x_d) = c2^{-rn} \sum_{\mathbf{s}\in\mathbb{H}_n^d} \sum_{R\in\mathscr{D}_{\mathbf{s}}^d : R\cap\mathscr{P}_N=\emptyset} b_R(x_1,...,x_d) \tag{56}$$

for some small constant $c>0$. It is evident that $\Lambda(G,\mathscr{P}_N)=0$ because all the terms of G are supported on empty rectangles R, so that $G(p)=0$ for all $p\in\mathscr{P}_N$. At the same time, denoting $B=\int_0^1 b(x)dx$, we have

$$\int_{[0,1]^d} G(x)dx_1...dx_d \geq c2^{-rn}\cdot\#\mathbb{H}_n^d\cdot 2^{n-1}\cdot 2^{-n}B^d \gtrsim 2^{-rn}n^{d-1}. \tag{57}$$

Hence we obtain

$$\left|\Lambda(G,\mathscr{P}_N) - \int_{[0,1]^d} G(x)dx\right| \gtrsim 2^{-rn}n^{d-1} \approx N^{-r}(\log N)^{d-1}. \tag{58}$$

It only remains to check that $G\in B(MH_r)$. The Hölder norm of the r^{th} mixed derivative of G can be estimated in the following way

$$\begin{aligned}
\left\|\Delta_{\mathbf{t}}\left(\left(\frac{\partial^d}{\partial x_1...\partial x_d}\right)^r G\right)\right\|_\infty &\leq c\sum_{\mathbf{s}\in\mathbb{H}_n^d}\left\|\sum_{R\in\mathscr{D}_{\mathbf{s}}^d : R\cap\mathscr{P}_N=\emptyset} 2^{-rs_j}\Delta_{\mathbf{t}}\left(\left(\frac{\partial^d}{\partial x_1...\partial x_d}\right)^r b_R\right)\right\|_\infty \\
&\leq c\sum_{\mathbf{s}\in\mathbb{H}_n^d}\prod_{j=1}^d 2^{-rs_j}\|\Delta_{t_j}(2^{rs_j}b^{(r)}(2^{s_j}x_j))\|_\infty \\
&\lesssim \sum_{\mathbf{s}\in\mathbb{H}_n^d}\prod_{j=1}^d 2^{-rs_j}\min\{1,2^{rs_j}|t_j|\} \\
&\leq \prod_{j=1}^d\sum_{s_j=0}^\infty 2^{-rs_j}\min\{1,2^{rs_j}|t_j|\} \lesssim \prod_{j=1}^d |t_j|,
\end{aligned} \tag{59}$$

where we have used the fact that rectangles $R\in\mathscr{D}_{\mathbf{r}}^d$ are disjoint for fixed $\mathbf{r}$, the product structure of the functions b_R, and the estimate (52). Therefore $G\in B(MH_r)$ if the constant c is small enough and hence (53) is proved.

We turn to the proof of (54). As one can guess from the right-hand side of this inequality, it will resemble Roth's proof of the L^2 discrepancy estimate (5) even more. The argument will proceed along the same lines as the proof of (53), but the choice of the analog of the function b_R will be more delicate.The r^{th} mixed derivatives of these functions should form an orthogonal family. Unfortunately, we cannot start with the Haar function, because even in one dimension its r^{th} antiderivative $(\mathscr{T}_1)^r h_I$ is not compactly supported anymore if $r\geq 2$. In order to fix this problem, we can define auxiliary functions inductively depending on r. For a dyadic interval I, whose left and right halves are denoted by I_l and I_r, let us set $h_I^0=h_I$, $h_I^1=h_{I_l}-h_{I_r}$, and proceeding in a similar fashion $h_I^r=h_{I_l}^{r-1}-h_{I_r}^{r-1}$.

This construction creates the following effect: not only h_I^r itself, but also all of its antiderivatives $(\mathscr{T}_1)^k h_I^r$ of order $k\leq r-1$ are supported on I and have mean zero, therefore the r^{th} antiderivative $(\mathscr{T}_1)^r h_I^r$ is supported on the interval I. Set $\phi_{[0,1)}^r=$

$(\mathscr{T}_1)^r h^r_{[0,1)}$. For a dyadic interval $I \in \mathscr{D}^d$, $I = \big[k2^{-j},(k+1)2^{-j}\big)$, we define $\phi^r_I(x) = \phi_{[0,1)}(2^j x - k)$, assuming that $\phi^r_{[0,1)}$ is zero outside $[0,1)$. Then we have

$$\big(\phi^r_I\big)^{(r)}(x) = 2^{jr}\big(\phi^r_{[0,1)}\big)^{(r)}(2^j x - k) = 2^{jr} h^r_{[0,1)}(2^j x - k) = 2^{jr} h^r_I(x) = |I|^{-r} h^r_I(x), \tag{60}$$

i.e. $\phi^r_I = |I|^{-r}(\mathscr{T}_1)^r h^r_I$. As usually, in the multivariate case, for a dyadic box R we define

$$\phi^r_R(x_1,\dots,x_d) = \prod_{j=1}^d \phi^r_{R_j}(x_j), \qquad h^r_R(x_1,\dots,x_d) = \prod_{j=1}^d h^r_{R_j}(x_j). \tag{61}$$

The one-dimensional case then implies that $\left(\frac{\partial^d}{\partial x_1 \dots \partial x_d}\right)^r \phi^r_R(x) = |R|^{-r} h^r_R(x)$. Next, we choose n as before, $2N < 2^n \le 4N$, and define a function similar to (56) and (41)

$$W(x_1,\dots,x_d) = \gamma 2^{-rn} n^{-\frac{d-1}{2}} \sum_{\mathbf{s}\in\mathbb{H}^d_n} \sum_{R\in\mathscr{D}^d_{\mathbf{s}} : R\cap\mathscr{P}_N=\emptyset} \phi^r_R(x_1,\dots,x_d). \tag{62}$$

From the definition of ϕ^r_R, we have $\int_R \phi^r_R(x) = |R| \int_{[0,1]^d} \phi^r_{[0,1)^d}(x)dx$. Repeating the previous reasoning verbatim we find that $\Lambda(W,\mathscr{P}_N) = 0$ and

$$\begin{aligned} \left|\int_{[0,1]^d} W(x)dx\right| &\gtrsim 2^{-rn} n^{-\frac{d-1}{2}}\, \#\mathbb{H}^d_n\, 2^{n-1}|R| \left|\int_{[0,1]^d} \phi^r_{[0,1)^d}(x)dx\right| \\ &\approx 2^{-rn} n^{\frac{d-1}{2}} \approx N^{-r}(\log N)^{\frac{d-1}{2}}. \end{aligned} \tag{63}$$

To see that $W \in B(MW^2_r)$, we first observe that h^r_R form an orthogonal system. Obviously $W \in MW^2_r$ since each $\phi^r_R = (\mathscr{T}_k)^r h^r_R$. We use orthogonality to estimate the norm of the r^{th} mixed derivative.

$$\begin{aligned} \left\|\left(\frac{\partial^d}{\partial x_1 \dots \partial x_d}\right)^r W\right\|_2^2 &= \gamma^2 2^{-2rn} n^{-(d-1)} \sum_{\mathbf{s}\in\mathbb{H}^d_n} \sum_{R\in\mathscr{D}^d_{\mathbf{s}} : R\cap\mathscr{P}_N=\emptyset} 2^{2rn}\|h^r_R\|_2^2 \\ &\approx n^{-(d-1)} \cdot n^{d-1} \cdot 2^n \cdot 2^{-n} \approx 1. \end{aligned} \tag{64}$$

Hence $W \in B(MW^2)$ if γ is sufficiently small. This finishes the proof of (54). □

We would like to point out that in order for this proof to be extended to the classes $B(MW^p_r)$ for $p \in (1,\infty)$, one should estimate the L^p norm of the mixed derivative of W, which, by the way, has a very similar structure to the test function (41) used by Roth. This can be done in a straightforward way using the material of the next section – Littlewood–Paley theory. The computation leading to this estimate is almost identical to (88). A more detailed account of various lower bounds for the errors of cubature formulas in classes of functions with mixed smoothness can be found, for example, in [109, 114]. The recent books [43] and [79], as well as the chapters of this book written by the same authors, give very nice accounts of the connections between discrepancy and numerical integration.

3 Littlewood–Paley Theory and applications to discrepancy

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 [99], see the discussion after (46).

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 applications of this theory to the irregularities of distribution.

We would like to point out that in general Littlewood–Paley theory is a vast subject in harmonic analysis which arises in various fields and settings, has numerous applications, and is available in many different variations. For the purposes of our exposition we are restricting the discussion just to the dyadic Littlewood–Paley theory, i.e. its version related to the Haar function expansions and other similar dyadic orthogonal decompositions. Other versions of this theory (on Euclidean spaces $\mathbb{R}^n$, on domains, for trigonometric (Fourier) series, in the context of complex analysis) can be found in many modern books on harmonic analysis, e.g. [105, 50]. A more detailed treatment of the dyadic Littlewood–Paley theory can be enjoyed in [81].

3.1 One-dimensional dyadic Littlewood–Paley theory

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 defined as

$$Sf(x) = \left(\Big| \int_0^1 f(t)dt \Big|^2 + \sum_{I\in\mathscr{D}} \frac{|\langle f, h_I\rangle|^2}{|I|^2} \mathbf{1}_I(x) \right)^{\frac{1}{2}} \tag{65}$$
$$= \left(\left| \int_0^1 f(t)dt \right|^2 + \sum_{k=0}^{\infty} \left(\sum_{I\in\mathscr{D},\,|I|=2^{-k}} \frac{\langle f, h_I\rangle}{|I|} h_I(x) \right)^2 \right)^{\frac{1}{2}}$$

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 $I \in \mathscr{D}$. In particular, if the function has the Haar expansion $f = \sum_{I\in\mathscr{D}_*} a_I h_I$, then its square function is

$$Sf = \left(\sum_{I \in \mathscr{D}_*} a_I^2 \mathbf{1}_I \right)^{\frac{1}{2}} = \left(\sum_{k=-1}^{\infty} \left(\sum_{I \in \mathscr{D}: |I| = 2^{-k}} a_I h_I \right)^2 \right)^{\frac{1}{2}}. \tag{66}$$

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

$$\|Sf\|_2 = \|f\|_2. \tag{67}$$

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.

Theorem 4 (Littlewood–Paley inequalities, [118]). *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*

$$A_p \|Sf\|_p \le \|f\|_p \le B_p \|Sf\|_p. \tag{68}$$

The asymptotic behavior of the constants A_p and B_p is known [118] and is very useful in numerous arguments, especially when (68) 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: [118, 105, 23].

The dyadic square function arises naturally in probability theory. Denote by $\mathscr{D}_k$ the collection of dyadic intervals in $[0,1]$ of fixed length 2^{-k}. We shall slightly abuse notation and also denote the σ-algebra generated by this family by $\mathscr{D}_k$. Let f be an L^2 function on $[0,1]$. We construct the sequence of conditional expectations of f with respect to the families $\mathscr{D}_k$,

$$f_k = \mathbb{E}(f|\mathscr{D}_k) = \sum_{I \in \mathscr{D}_k} \frac{1}{|I|} \int_I f(x) dx \cdot \mathbf{1}_I. \tag{69}$$

The sequence $\{f_k\}_{k \ge 0}$ forms a martingale, meaning that $\mathbb{E}(f_{k+1}|\mathscr{D}_k) = f_k$. As usually for a dyadic interval I of length $2^{-(k-1)}$ denote by I_l and I_r its left and right dyadic "children" of length 2^{-k} and let $\langle f \rangle_I$ stand for the average of f over I. Keeping in mind that $2\langle f \rangle_I = \langle f \rangle_{I_l} + \langle f \rangle_{I_r}$, it is then easy to check that the martingale differences for $k \ge 1$ satisfy

$$\begin{aligned} d_k := f_k - f_{k-1} &= \sum_{I \in \mathscr{D}_{k-1}} \Big(\big(\langle f \rangle_{I_l} \mathbf{1}_{I_l} + \langle f \rangle_{I_r} \mathbf{1}_{I_r} \big) - \langle f \rangle_I \mathbf{1}_I \Big) \\ &= \sum_{I \in \mathscr{D}_{k-1}} \frac{1}{2} (-\langle f \rangle_{I_l} + \langle f \rangle_{I_r})(-\mathbf{1}_{I_l} + \mathbf{1}_{I_r}) = \sum_{I \in \mathscr{D}_{k-1}} \frac{\langle f, h_I \rangle}{|I|} h_I. \end{aligned} \tag{70}$$

Setting $d_0 = f_0$, we define the *martingale difference square function* :

$$Sf = \left(\sum_{k=0}^{\infty} |d_k|^2 \right)^{\frac{1}{2}}. \tag{71}$$

One can see from (66) that it is exactly the same object as the dyadic Littlewood–Paley square function defined in (65).

Littlewood–Paley square function estimates (68) can also be viewed as a generalization of the famous Khintchine inequality. Indeed, consider the Rademacher functions $r_k(x) = \sum_{I\in\mathscr{D}_k} h_I(x)$. Then at any point $x\in[0,1]^d$, since dyadic intervals in $\mathscr{D}_k$ are disjoint, the square function of a linear combination of Rademacher functions is constant:

$$S\big(\sum_k \alpha_k r_k\big)(x) = \bigg(\sum_{k=0}^{\infty}\sum_{I\in\mathscr{D}_k} |\alpha_k|^2 \mathbf{1}_I\bigg)^{\frac12} = \big(\sum_k |\alpha_k|^2\big)^{\frac12}. \tag{72}$$

Therefore, Littlewood–Paley inequalities imply

$$\big\|\sum_k \alpha_k r_k\big\|_p \approx \big\|S\big(\sum_k \alpha_k r_k\big)\big\|_p = \big(\sum_k |\alpha_k|^2\big)^{\frac12}, \tag{73}$$

which is precisely the Khintchine inequality for $p>1$.

3.1.1 The Chang–Wilson–Wolff inequality

The Littlewood–Paley inequalities are tightly related to the famous Chang–Wilson–Wolff inequality, which states that if the square function of f is bounded, then f is exponentially square integrable (subgaussian).

To formulate it rigorously we need to introduce exponential Orlicz function classes. For a convex function $\psi:\mathbb{R}_+\to\mathbb{R}_+$ with $\psi(0)=0$, the Orlicz norm of a function f on the domain D is defined as

$$\|f\|_\psi := \inf\bigg\{K>0 : \int_D \psi\bigg(\frac{|f(x)|}{K}\bigg)dx \le 1\bigg\} \tag{74}$$

The corresponding Orlicz space is the space of functions for which the above norm is finite. For example, if $\psi(t)=t^p$, one recovers the usual L^p spaces. In the case when $\psi(t)=e^{t^\alpha}$ for large values of t (if $\alpha\ge 1$, one may take $\psi(t)=e^{|t|^\alpha}-1$, however for $\alpha<1$ convexity near zero would be violated) the arising Orlicz spaces are denoted $\exp(L^\alpha)$. One of the most important members of this scale of function spaces is $\exp(L^2)$, often referred to as the space of *exponentially square integrable* or *subgaussian* functions. It is a standard fact that exponential Orlicz norms can be characterized in the following ways

$$\|F\|_{\exp(L^\alpha)} \approx \sup_{q>1} q^{-1/\alpha}\|F\|_q \approx \sup_{\lambda>0} -\lambda^{-\alpha}\log\big|\{x : |F(x)|>\lambda\}\big| \tag{75}$$

The first equivalence here can be easily established using Taylor series for e^x and Stirling's formula, while the second one is a simple computation involving distribution functions, see a similar calculation in (120). The last expression explains the term *subgaussian* in the context of functions $f \in \exp(L^2)$: in this space, $\mathbb{P}(|f| > \lambda) \lesssim e^{-c\lambda^2}$.

We can now state the Chang–Wilson–Wolff inequality:

Theorem 5 (Chang–Wilson–Wolff inequality, [26]). *The following estimate holds:*

$$\|f\|_{\exp(L^2)} \lesssim \|Sf\|_\infty. \tag{76}$$

This fact can be derived extremely easily as a consequence of the Littlewood–Paley inequality (68) with sharp constants and the characterization (75) of the exponential norm.

$$\|f\|_{\exp(L^2)} \approx \sup_{p\geq 1} p^{-\frac{1}{2}} \|f\|_p \lesssim \sup_{p\geq 1} p^{-\frac{1}{2}} \cdot \sqrt{p}\|Sf\|_p = \sup_{p\geq 1} \|Sf\|_p \leq \|Sf\|_\infty, \tag{77}$$

which proves (76). □

Observe that this bound strongly resembles the Khintchine inequality. Indeed, if we use the Littlewood–Paley inequality with sharp constants in (73), much in the same fashion as in (77), we obtain the exponential form of the Khintchine inequality

$$\big\| \sum_k \alpha_k r_k \big\|_{\exp(L^2)} \lesssim \big(\sum_k |\alpha_k|^2 \big)^{\frac{1}{2}}. \tag{78}$$

In other words, a linear combination of independent ± 1 random variables obeys a subgaussian estimate. For a precise quantitative distributional version of this statement see (119).

3.2 *From vector-valued inequalities to the multiparameter setting*

It is very important for our further discussion that the Littlewood–Paley 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, which was proved in [46], allows one to extend the Littlewood–Paley inequalities to the multiparameter setting in a fairly straightforward way by successively applying (68) in each dimension while treating the other dimensions as vector-valued coefficients [82, 105].

We note that in the general case one would apply the one dimensional Littlewood–Paley inequality d times – once in each coordinate, see §3.3. 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 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 [82, 17]), we postpone this to (87) and first illustrate the use of this approach by a simple example, important to the topic of our discussion.

Recall that the test function (41) in Roth's proof was constructed as $F = \sum_{\mathbf{r}\in\mathbb{H}_n^d} f_{\mathbf{r}} = \sum_{R:\,|R|=2^{-n}} \varepsilon_R h_R$, where $\varepsilon_R = \pm 1$. We want to estimate the L^q norm of F. Notice that we can rewrite it as $F = \sum_{I\in\mathcal{D}} \alpha_I h_I(x_1)$, where

$$\alpha_I = \sum_{\substack{R:\,|R|=2^{-n}\\ R_1=I}} \varepsilon_R \prod_{j=2}^{\infty} h_{R_j}(x_j), \tag{79}$$

which allows one to apply the one-dimensional Littlewood–Paley square function (66) in the first coordinate x_1 to obtain

$$\|F\|_q = \Big\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \Big\|_q \le B_q \Bigg\| \Bigg[\sum_{r_1=0}^{n} \Big| \sum_{\substack{|R|=2^{-n}\\ |R_1|=2^{-r_1}}} \varepsilon_R h_R \Big|^2 \Bigg]^{1/2} \Bigg\|_q. \tag{80}$$

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

$$\sum_{r_1=1}^{n} \Big| \sum_{\substack{|R|=2^{-n}\\ |R_1|=2^{-r_1}}} \varepsilon_R h_R(x) \Big|^2 = \sum_{r_1=1}^{n} \sum_{\substack{|R|=2^{-n}\\ |R_1|=2^{-r_1}}} |\varepsilon_R|^2 \mathbf{1}_R(x) = \sum_{\substack{R\in\mathcal{D}^2,\\ |R|=2^{-n}}} \mathbf{1}_R(x) = \#\mathbb{H}_n^2 \approx n, \tag{81}$$

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 (80) can be viewed as a Hilbert space-valued function. Indeed, fix all the coordinates except x_2 and define an ℓ^2-valued function

$$F_2(x_2) = \sum_{I\in\mathcal{D}} \Bigg\{ \sum_{\substack{|R|=2^{-n},\, R_2=I\\ |R_1|=2^{-r_1}}} \varepsilon_R \prod_{j\ne 2} h_{R_j}(x_j) \Bigg\}_{r_1=1}^{n} h_I(x_2). \tag{82}$$

Then the expression inside the L^q norm on the right hand side of (80) is exactly $\|F_2(x_2)\|_{\ell^2}$. Applying the Hilbert space-valued Littlewood–Paley inequality in the second coordinate, we get

$$\begin{aligned} \|F\|_q &= \Big\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \Big\|_q \le B_q \Big\| \|F_2\|_{\ell^2} \Big\|_q \\ &\le B_q^2 \Bigg\| \Bigg[\sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \Big| \sum_{\substack{|R|=2^{-n}\\ |R_j|=2^{-r_j},\ j=1,2}} \varepsilon_R h_R \Big|^2 \Bigg]^{1/2} \Bigg\|_q. \end{aligned} \tag{83}$$

And if $d = 3$, then an analog of (81) 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

$$\begin{aligned} \|F\|_q &= \Big\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \Big\|_q \leq \dots \\ &\leq B_q^{d-1} \Bigg\| \Bigg[\sum_{r_1=1}^{n} \cdots \sum_{r_{d-1}=1}^{n} \Bigg| \sum_{\substack{|R|=2^{-n} \\ |R_j|=2^{-r_j},\, j=1,\dots,d-1}} \varepsilon_R h_R \Bigg|^2 \Bigg]^{1/2} \Bigg\|_q . \end{aligned} \tag{84}$$

Just as explained in (81), in this case all the rectangles in the innermost summation are disjoint and thus

$$\|F\|_q \leq B_q^{d-1} \Bigg\| \Bigg[\sum_{R\in\mathscr{D}^d,\, |R|=2^{-n}} |\varepsilon_R|^2 \mathbf{1}_R \Bigg]^{\frac{1}{2}} \Bigg\|_q = B_q^{d-1} \left(\#\mathbb{H}_n^d \right)^{\frac{1}{2}} \approx n^{\frac{d-1}{2}} . \tag{85}$$

3.3 Multiparameter (product) Littlewood–Paley theory

For a function of the form $f = \sum_{R\in\mathscr{D}_*^d} a_R h_R$ on $[0,1]^d$, the expression

$$S_d f(x) = \Bigg[\sum_{R\in\mathscr{D}_*^d} |a_R|^2 \mathbf{1}_R(x) \Bigg]^{\frac{1}{2}} = \Bigg(\sum_{\mathbf{r}\in\{\{-1\}\cup\mathbb{Z}_+\}^d} \Bigg| \sum_{R\in\mathscr{D}_{\mathbf{r}}^d} a_R h_R(x) \Bigg|^2 \Bigg)^{\frac{1}{2}} \tag{86}$$

is called the *product* dyadic square function of f. We remind that $\mathscr{D}_{\mathbf{r}}^d$ is the collection of dyadic rectangles R whose shape is defined by $|R_j| = 2^{-r_j}$ for $j = 1,\dots,d$ and the rectangles in this family are disjoint.

The product Littlewood–Paley inequalities (whose proof is essentially identical to the argument presented above) state that

$$A_p^d \|S_d f\|_p \leq \|f\|_p \leq B_p^d \|S_d f\|_p . \tag{87}$$

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

$$\|F\|_q = \Big\| \sum_{|R|=2^{-n}} \varepsilon_R h_R \Big\|_q \approx \|S_d f\|_q = \Bigg\| \Bigg[\sum_{|R|=2^{-n}} |\varepsilon_R|^2 \mathbf{1}_R \Bigg]^{\frac{1}{2}} \Bigg\|_q = \left(\#\mathbb{H}_n^d \right)^{\frac{1}{2}} \approx n^{\frac{d-1}{2}} . \tag{88}$$

We have chosen to include a separate illustrative proof of this estimate earlier in order to demonstrate the essence of the product Littlewood–Paley theory. In addition, the argument leading to (85) 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 one plays an important role in some further estimates, see (99).

The proof of Schmidt's L^p lower bound (7) 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 (41). Then, replacing Cauchy–Schwarz with Hölder's inequality in (44) and using (88), we obtain:

$$\|D_N\|_p \geq \frac{\langle D_N, F\rangle}{\|F\|_q} \gtrsim n^{\frac{d-1}{2}} \approx \big(\log N\big)^{\frac{d-1}{2}}. \tag{89}$$

An analog of the second proof (45) of Roth's estimate (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 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 $M \in \mathbb{N}$*

$$\mu\left(\{x \in [0,1]^d : \sum_{k=1}^M \mathbf{1}_{A_k}(x) \geq \frac{1}{2}cM\}\right) > \frac{1}{2}c. \tag{90}$$

Proof. Assuming this is not true, we immediately arrive to a contradiction

$$\begin{aligned} cM \leq \int \sum_{k=1}^M \mathbf{1}_{A_k}(x)\, dx < \frac{1}{2}cM \cdot \mu\left(\sum_{k=1}^M \mathbf{1}_{A_k} < \frac{1}{2}cM\right) \\ + M \cdot \mu\left(\sum_{k=1}^M \mathbf{1}_{A_k} \geq \frac{1}{2}cM\right) \leq \frac{1}{2}cM + M \cdot \frac{1}{2}c = cM. \square \end{aligned} \tag{91}$$

We shall apply the lemma as follows: for each $\mathbf{r} \in \mathbb{H}_n^d$, let $A_{\mathbf{r}}$ be the union of rectangles $R \in \mathscr{D}_{\mathbf{r}}^d$ which do not contain points of $\mathscr{P}_N$. Then $\mu(A_{\mathbf{r}}) \geq c = \frac{1}{2}$ 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 (34):

$$S_d D_N(x) = \left[\sum_{R \in \mathscr{D}^d} \frac{\langle D_N, h_R\rangle^2}{|R|^2} \mathbf{1}_R(x)\right]^{\frac{1}{2}} \gtrsim (M \cdot N^2\, 2^{-2n})^{\frac{1}{2}} \approx n^{\frac{d-1}{2}}. \tag{92}$$

Integrating this estimate over E and applying the Littlewood–Paley inequality (87) finishes the proof of (7):

$$\|D_N\|_p \gtrsim \|S_d D_N\|_p \gtrsim n^{\frac{d-1}{2}} \approx (\log N)^{\frac{d-1}{2}}. \square \tag{93}$$

3.4 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. Discrepancy theory has recently witnessed a surge of activity in this direction. We shall give a very brief overview of estimates and conjectures related to various function spaces. All of the results described below are direct descendants of Theorem 1 and Roth's method as every single one of them makes use of the Haar coefficients of the discrepancy function.

3.4.1 Hardy spaces H^p

In particular, a direct extension of the above argument provides a lower bound of the discrepancy function in product Hardy spaces H^p, $0 < p \le 1$. These spaces are generalizations due to Chang and R. Fefferman of the classical classes introduced by Hardy, see [24, 25]. The discussion of these spaces in the multiparameter dyadic setting, which is relevant to our situation, can be found in [14]. The Hardy space H^p norm of a function $f = \sum_{R\in\mathscr{D}^d} \alpha_R h_R$ is equivalent to the norm of its square function in L^p, i.e.

$$\|f\|_{H^p} \approx \|S_d f\|_p. \tag{94}$$

The following result about the Hardy space norm of the discrepancy function was obtained by Lacey [65]. For $0 < p \le 1$,

$$\|\widetilde{D}_N\|_{H^p} \ge C_{d,p} (\log N)^{\frac{d-1}{2}}, \tag{95}$$

where $\widetilde{D}_N = \sum\limits_{R\in\mathscr{D}^d} \dfrac{\langle D_N, h_R\rangle}{|R|} h_R$, in other words, $\widetilde{D}_N$ is the discrepancy function D_N modified so as to have mean zero over every subset of coordinates. The proof of this result is a verbatim repetition of the previous proof (92) – one simply estimates the norm of the square function. Observe that a duality argument in the spirit of (44) would not have worked in this case, as H^p is only a *quasi*-Banach space for $p < 1$ and thus no duality argumens are available.

As this example clearly illustrates, in harmonic analysis Hardy spaces H^p serve as a natural substitute for L^p spaces when $p \le 1$. Indeed, numerous analytic tools, such as square functions, maximal functions, atomic decompositions [105], allow one to extend the L^p estimates to the H^p setting for $0 < p \le 1$. Similarly, the L^p asymptotics of the discrepancy is continued by the H^p estimates when $p \le 1$.

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 asymptotic bounds in N for *all* values of $p > 0$, which includes Conjecture 3 as a subcase.

Conjecture 4. For all $p \in (0,1]$ the discrepancy function satisfies the estimate

$$\|D_N\|_p \ge C_{d,p} (\log N)^{\frac{d-1}{2}}. \tag{96}$$

3.4.2 The behavior of discrepancy in and near L^1

The only currently available information regarding the conjecture above is the result of Halász [52] who proved that (96) indeed holds in dimension $d = 2$ for the L^1 norm:

$$\|D_N\|_1 \geq C\sqrt{\log N}. \tag{97}$$

We shall discuss his method in §4. Halász was also 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 [65] has proved that if one replaces L^1 with the Orlicz space $L(\log L)^{\frac{d-2}{2}}$, then the conjectured bound holds

$$\|D_N\|_{L(\log L)^{(d-2)/2}} \geq C_d (\log N)^{\frac{d-1}{2}}. \tag{98}$$

We remark that an adaptation of the proof of Schmidt's L^p bound given in the previous subsection, specifically estimate (85), can easily produce a slightly weaker inequality

$$\|D_N\|_{L(\log L)^{(d-1)/2}} \geq C_d (\log N)^{\frac{d-1}{2}} \tag{99}$$

Indeed, let F once again be as defined in (41). It is well known that (see e.g. [73]) 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.

We recall that the constant arising in the Littlewood–Paley inequalities (68) is $B_q \approx \sqrt{q}$ for large q and the implicit constant in (85) is B_q^{d-1}. Thus using the equivalence between the exponential Orlicz norm and the growth of L^p norms (75) we obtain

$$\begin{aligned} \|F\|_{\exp\left(L^{2/(d-1)}\right)} &\approx \sup_{q>1} q^{-\frac{d-1}{2}} \|F\|_q \lesssim \sup_{q>1} q^{-\frac{d-1}{2}} \cdot B_q^{d-1} n^{\frac{d-1}{2}} \\ &\approx \sup_{q>1} q^{-\frac{d-1}{2}} \cdot q^{\frac{d-1}{2}} n^{\frac{d-1}{2}} = n^{\frac{d-1}{2}}, \end{aligned} \tag{100}$$

and (99) immediately follows by duality. Notice that a more straightforward bound (88) would not suffice for this estimate, since in the general d-parameter inequality the constant is of the order $q^{d/2}$, not $q^{(d-1)/2}$. These estimates are similar in spirit to the Chang–Wilson–Wolff inequality discussed in §3.1.1.

3.4.3 Besov space estimates

In a different vein, Triebel has recently studied the behavior of the discrepancy in Besov spaces [115, 116]. He proves, among other things, that

$$\|D_N\|_{S^r_{p,q}B([0,1]^d)} \geq C_{d,p,q,r} N^r (\log N)^{\frac{d-1}{q}}, \tag{101}$$

$$1 < p, q < \infty, \quad \frac{1}{p} - 1 < r < \frac{1}{p}. \tag{102}$$

Here the space $S^r_{p,q}B([0,1]^d)$ is the Besov space with dominating mixed smoothness. The exact original definition of this class is technical and would take our discussion far afield. There exists, however, a characterization of the Besov norms in terms of the Haar expansion (which is reminiscent of the Littlewood–Paley square function Sf). For a function $f = \sum_{R \in \mathscr{D}^d_*} \frac{\alpha_R}{|R|} h_R$, we have

$$\|f\|_{S^r_{p,q}B([0,1]^d)} \approx \left(\sum_{\mathbf{s} \in (\{-1\} \cup \mathbb{Z}_+)^d} 2^{(s_1+\ldots+s_d)(r-1/p+1)q} \left(\sum_{\substack{R \in \mathscr{D}^d_*: \\ |R_j| = 2^{-s_j}}} |\alpha_R|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \tag{103}$$

whenever the right-hand side is finite.

To give a better idea about these spaces, we would 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 and are similar to the previously defined spaces $MW^p_r([0,1]^d)$, see §2.3 . Furthermore, when $r = 0$, $S^0_{p,2}B([0,1]^d)$ is nothing but $L^p([0,1]^d)$. In particular, in the case $p = q = 2$, $r = 0$, the characterization (103) simply states that $\{h_R\}_{R \in \mathscr{D}^d_*}$ is an orthogonal basis of L^2.

Thus, if $q = 2$ and $r = 0$, one recovers Roth's L^2 and Schmidt's L^p estimates from (101). Inequalities (101) are sharp in all dimensions ([57] $d = 2$, [76] $d \geq 3$), see §6. For more details, the reader is directed towards Triebel's recent book [116] 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.

3.4.4 Weighted L^p estimates

The recent work of Ou [80] 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 the following condition (initially introduced by Muckenhoupt [78]) holds

$$\sup_{R \in \mathscr{D}^d} \left(\int_R \omega(x) dx \right) \left(\int_R \omega^{-\frac{1}{p-1}}(x) dx \right)^{p-1} < \infty. \tag{104}$$

The space $L^p(\omega)$ is then defined as the L^p space with respect to the measure $\omega(x)\,dx$. The class of A_p weights plays a tremendously important role in harmonic analysis: they give the largest reasonable class of measures such that the standard boundedness properties of classical opertors (such as maximal functions, singular integrals, square functions) continue to hold in L^p spaces built on these measures. By an adaptation of the square function argument (92), Ou was able to show that

$$\|D_N\|_{L^p(\omega)} \geq C_{d,p,\omega} (\log N)^{\frac{d-1}{2}}, \tag{105}$$

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

3.4.5 Approaching L^∞: BMO and exponential Orlicz spaces

Moving toward 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 [20] 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*), i.e. the space of functions f for which the following norm is finite:

$$\|f\|_{\mathrm{BMO}} = \sup_{U\subset[0,1]^d} \left(\frac{1}{|U|} \sum_{R\in\mathscr{D}^d} \frac{|\langle f, h_R\rangle|^2}{|R|} \right)^{\frac{1}{2}}, \tag{106}$$

where the supremum is extended over all measurable subsets of $[0,1]^d$ with positive measure. Notice that in the case $d=1$, when U is a dyadic interval, the expression inside the parentheses is actually equal to $\frac{1}{|U|}\int_U \left| f(x) - f_U \right|^2 dx$, where f_U is the mean of f over U, which yields exactly the standard one-dimensional BMO. The definition above, introduced by Chang and Fefferman [24], is a proper generalization of the classical BMO space to the dyadic multiparameter setting. In particular, the classical $H^1 - \mathrm{BMO}$ 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 than L^∞:

$$\|D_N\|_{\mathrm{BMO}} \geq C_d (\log N)^{\frac{d-1}{2}}. \tag{107}$$

In fact, this estimate is not hard to obtain with the help of the same test function F (41) that we have used several times already – all we have to do is estimate its dual (H^1) norm. Just as in (88):

$$\|F\|_{H^1} \approx \|SF\|_1 = \left\| \left[\sum_{R\in\mathscr{D}^d,\, |R|=2^{-n}} |\varepsilon_R|^2 \mathbf{1}_R \right]^{\frac{1}{2}} \right\|_1 = \left(\#\mathbb{H}_n^d\right)^{\frac{1}{2}} \approx n^{\frac{d-1}{2}}, \tag{108}$$

which immediately yields the result.

In addition, the authors prove lower bounds in the aforementioned exponential Orlicz spaces, see (75). These spaces $\exp(L^\alpha)$ serve as an intermediate scale between the L^p spaces, $p < \infty$, and L^∞. In particular, for all $\alpha > 0$ and for all $1 < p < \infty$, we have $L^\infty \subset \exp(L^\alpha) \subset L^p$. The following estimate is contained in [20]: *in dimension $d = 2$ for all $2 \leq \alpha < \infty$ we have*

$$\|D_N\|_{\exp(L^\alpha)} \geq C(\log N)^{1-\frac{1}{\alpha}}. \tag{109}$$

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 (128) below), with the obvious modification that the test function has to be estimated in the dual space $\big(\exp(L^\alpha)\big)^* = L(\log L)^{1/\alpha}$. Hence the same problems and obstacles that arise when dealing with the star-discrepancy prevent straightforward extensions of this estimate to higher dimensions. We finish this discussion by mentioning that both of these estimates, (107) and (109), 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 $\mathscr{P}_N$ 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 [96] who proved

$$\|D_N\|_\infty \geq C \log N. \tag{110}$$

Other proofs of this inequality have been later given by Liardet [72, 1979], Béjian [13, 1982] (who produced the best currently known value of the constant $C = 0.06$), and Halász [52, 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.

4.1 The small ball conjecture: formulations and simple estimates

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 of the star-discrepancy (11). 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.

Unfortunately, this inequality does not (more precisely, has not been proved to) directly imply the lower bound for the L^∞ norm of the discrepancy function. It is, however, linked to discrepancy through Roth's orthogonal function method. Even though no formal connections are known, most arguments designed for this inequality can be transferred to the discrepancy setting. In a certain sense, it can be viewed as a linear version of the star-discrepancy estimate.

We now state the conjectured inequality:

*Conjecture 5 (***The small ball conjecture).** In dimensions $d \geq 2$, for any choice of the coefficients α_R one has the following inequality:

$$n^{\frac{d-2}{2}} \Big\| \sum_{R\in\mathscr{D}^d:\,|R|=2^{-n}} \alpha_R h_R \Big\|_\infty \gtrsim 2^{-n} \sum_{R:\,|R|=2^{-n}} |\alpha_R|. \tag{111}$$

The challenge and the point of interest of the conjecture is the precise value of the exponent of n on the left-hand side. If one replaces $n^{(d-2)/2}$ by $n^{(d-1)/2}$, the inequality becomes almost trivial, and, in fact, holds even for the L^2 norm:

$$n^{\frac{d-1}{2}} \Big\| \sum_{R\in\mathscr{D}^d:\,|R|=2^{-n}} \alpha_R h_R \Big\|_2 \gtrsim 2^{-n} \sum_{R:\,|R|=2^{-n}} |\alpha_R|. \tag{112}$$

Proof of (112). Indeed, using the orthogonality of Haar functions and keeping in mind that $\|h_R\|_2 = |R|^{1/2}$, we obtain

$$\begin{aligned} \Big\| \sum_{R\in\mathscr{D}^d:\,|R|=2^{-n}} \alpha_R h_R \Big\|_2 &= \bigg(\sum_{|R|=2^{-n}} |\alpha_R|^2 2^{-n} \bigg)^{\frac{1}{2}} \\ &\gtrsim \frac{\sum_{|R|=2^{-n}} |\alpha_R| 2^{-n/2}}{\left(n^{d-1}2^n\right)^{\frac{1}{2}}} = n^{-\frac{d-1}{2}} \cdot 2^{-n} \sum_{|R|=2^{-n}} |\alpha_R|, \end{aligned} \tag{113}$$

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 $n^{d-1}2^n$.

Alternatively, this inequality can be proved by duality. Consider the familiar function $F = \sum_{\mathbf{r}\in\mathbb{H}_n^d} f_{\mathbf{r}} = \sum_{|R|=2^{-n}} \varepsilon_R h_R$, where $\varepsilon_R = \operatorname{sgn}(\alpha_R)$. We know very well by now, see (42), that $\|F\|_2 \approx n^{\frac{d-1}{2}}$. On the other hand, by orthogonality,

$$\Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, F \Big\rangle = \sum_{|R|=2^{-n}} |\alpha_R| \|h_R\|_2^2 = 2^{-n} \sum_{|R|=2^{-n}} |\alpha_R|, \tag{114}$$

which immediately implies (112). □

As we have already witnessed on several occasions, 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.

The L^2 inequality (112) and the conjecture (111) should be compared to Roth's L^2 discrepancy estimate (5) and Conjecture 1. The computations just presented are very close to the proof (45) and (44) of (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 (112) becomes

$$\Big\| \sum_{R \in \mathscr{D}^d : |R|=2^{-n}} \alpha_R h_R \Big\|_2 \gtrsim n^{\frac{d-1}{2}}, \tag{115}$$

while the conjectured L^∞ inequality (111) for $\alpha_R = \pm 1$ turns into

*Conjecture 6 (***The signed small ball conjecture).** If all the coefficients $\alpha_R = \pm 1$, we have the inequality

$$\Big\| \sum_{R \in \mathscr{D}^d : |R|=2^{-n}} \alpha_R h_R \Big\|_\infty \gtrsim n^{\frac{d}{2}}. \tag{116}$$

Recalling that n in Roth's argument was chosen to be approximately $\log_2 N$, one immediately sees the similarity of these inequalities to (5) and (11).

We would like to add a few comments about the signed small ball conjecture. There are some indications that this restricted version may turn out to be significantly simpler to prove than the more general Conjecture 5, see §5.5. However, this variation of the conjecture, unlike its full form, does not appear to have any real applications. On the other hand, one can formulate a slightly more generic statement of the conjecture by allowing some coefficients to equal zero, but not allowing the left-hand side to degenerate:

*Conjecture 7 (***Generic signed small ball conjecture).** Assume that the coefficients α_R are either ± 1 or 0, and no more than half of all the coefficients are zero. Then we have the inequality

$$\Big\| \sum_{R \in \mathscr{D}^d : |R|=2^{-n}} \alpha_R h_R \Big\|_\infty \gtrsim n^{\frac{d}{2}}. \tag{117}$$

This form of the conjecture is strong enough to yield applications, see §4.8. Unfortunately, it seems to be just as hard as the general small ball conjecture (111).

4.2 Sharpness of the small ball conjecture

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. [17] or [112]. We include the proof of the sharpness of inequality (111) here for the sake of completeness.

Lemma 3 (Sharpness of the small ball conjecture). *Let $\{\alpha_R\}_{R\in\mathscr{D}^d:\,|R|=2^{-n}}$ be independent ± 1 random variables. Then, on the average, the converse of the small ball inequality holds, i.e.*

$$\mathbb{E}\Big\| \sum_{|R|=2^{-n}} \alpha_R h_R(x) \Big\|_\infty \lesssim n^{-\frac{d-2}{2}} 2^{-n} \sum_{|R|=2^{-n}} |\alpha_R| = n^{d/2}. \tag{118}$$

Proof. The function $\sum_{|R|=2^{-n}} \alpha_R h_R(x)$ is constant on dyadic cubes Q_k of sidelength $2^{-(n+1)}$. The total number of such cubes is $M = 2^{(n+1)d}$. Let us define M random variables $X_k = \sum_{|R|=2^{-n}} \alpha_R h_R|_{Q_k}$. Since X_k is a sum of $\#\mathbb{H}_n^d$ independent ± 1 random variables, by the Khintchine inequality we have $\mathbb{E}|X_k| \approx n^{(d-1)/2}$. Moreover, by a standard inequality (usually attributed to Bernstein, Hoeffding, Chernoff, or Azuma, see e.g. [59]), concerning sums of random variables, we have

$$\mathbb{P}(|X_k| > t) \le 2\exp\big(-t^2/(4\cdot \#\mathbb{H}_n^d)\big). \tag{119}$$

Recalling that $\#\mathbb{H}_n^d \approx n^{d-1}$, it is easy to deduce from this inequality that for some constant $C>0$, the random variables $Y_k = \frac{1}{Cn^{(d-1)/2}} X_k$ have bounded $\exp(L^2)$ norm, in other words $\|X_k\|_{\exp(L^2)} \lesssim n^{(d-1)/2}$ (this is essentially the exponential form of the Khintchine inequality, see (78)). Indeed, denoting $\psi(t) = \exp(t^2)$, we obtain

$$\begin{aligned} \mathbb{E}\,\psi(Y_k) &= \int_0^\infty \mathbb{P}\big(\psi(Y_k) > t\big) = \int_0^\infty \mathbb{P}\big(|X_k| > Cn^{(d-1)/2}\sqrt{\log t}\big)\,dt \\ &\le \int_0^\infty \min\{1, 2\exp(-C^2 n^{d-1} \log t/(4\cdot \#\mathbb{H}_n^d))\}\,dt \\ &\le \int_0^\infty \min\{1, t^{-K}\}\,dt \lesssim 1, \end{aligned} \tag{120}$$

where $K>1$, if C is large enough. Therefore, applying Jensen's inequality with the convex function ψ, we get

$$\begin{aligned} \psi\Big(\mathbb{E} \sup_{k=1,\dots,M} |Y_k|\Big) &\le \mathbb{E}\psi\Big(\sup_{k=1,\dots,M} |Y_k|\Big) \le \mathbb{E} \sup_{k=1,\dots,M} \psi(|Y_k|) \\ &\le \mathbb{E} \sum_{k=1}^M \psi(|Y_k|) \lesssim M = 2^{(n+1)d}. \end{aligned} \tag{121}$$

Since $\psi^{-1}(t) = \sqrt{\log t}$, we arrive to

$$\mathbb{E}\Big\| \sum_{|R|=2^{-n}} \alpha_R h_R(x) \Big\|_\infty = C n^{\frac{d-1}{2}}\, \mathbb{E} \sup |Y_k| \lesssim n^{\frac{d-1}{2}} \cdot \psi^{-1}(2^{(n+1)d}) \approx n^{d/2}, \quad (122)$$

which finishes the proof. □

The sharpness of the Small Ball Conjecture provides evidence that perhaps the correct estimate for the star-discrepancy should be Conjecture 1: $\|D_N\|_\infty \gtrsim (\log N)^{d/2}$. To validate the evidence we shall now illustrate the connection between this inequality and the discrepancy estimates. As mentioned earlier, the connection is not direct, but rather comes from the method of proof. We have already discussed the similarities between the proofs of the L^2 inequalities. Let us now turn to the case of L^∞.

The small ball conjecture (111) has been verified in $d = 2$ by M. Talagrand [106] in 1994. In 1995, V. Temlyakov [110] (see also [111, 112]) has given another, very elegant proof of this inequality in two dimensions, which closely resembled the argument of Halász [52] for (8). We shall present Temlyakov's proof first as it is somewhat "cleaner" and avoids some technicalities. Then we shall explain which adjustments need to be made in order to translate this argument into Halász's proof of Schmidt's estimate for $\|D_N\|_\infty$.

4.3 *Proof of the small ball conjecture in dimension* $d = 2$

The proof is based on Riesz products. An important feature of the two-dimensional case is the following *product rule*.

Lemma 4 (Product rule). *Assume that $R, R' \in \mathscr{D}^2$ are not disjoint, $R \neq R'$, and $|R| = |R'|$, then*

$$h_R \cdot h_{R'} = \pm h_{R\cap R'}, \quad (123)$$

i.e. the product of two Haar functions is again a Haar function.

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, $h_{R_1} \cdot h_{R_1'} = h_{R_1}^2 = \mathbf{1}_{R_1}$, so we lose orthogonality in the first coordinate. Since, as the reader will see below, we shall be considering very long products, the orthogonality may be lost completely. The fact that the product rule fails in higher dimensions is a major obstruction on the path to solving the conjecture.

For each $k = 0, \dots, n$ 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. Hence these are really r-functions, $f_k = f_{\mathbf{r}}$ with $\mathbf{r} = (k, n-k)$ and $\varepsilon_R = \operatorname{sgn}(\alpha_R)$. We are now ready to construct the test function as a Riesz product:

$$\Psi := \prod_{k=0}^{n} \Big(1 + f_k\Big). \quad (124)$$

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 (124). 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 (124) and using the product rule, one can see that

$$\Psi = 1 + \sum_{R \in \mathscr{D}^d :\, |R| = 2^{-n}} \operatorname{sgn}(\alpha_R) h_R \, + \, \Psi_{>n}, \tag{125}$$

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 the trivial case of Hölder's inequality, $p = \infty$, $q = 1$,

$$\Big\| \sum_{R \in \mathscr{D}^d :\, |R| = 2^{-n}} \alpha_R h_R \Big\|_\infty \geq \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \Psi \Big\rangle \tag{126}$$

$$= \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \sum_{|R|=2^{-n}} \operatorname{sgn}(\alpha_R) h_R \Big\rangle$$

$$= \sum_{|R|=2^{-n}} \alpha_R \cdot \operatorname{sgn}(\alpha_R) \cdot \|h_R\|_2^2 = \, 2^{-n} \cdot \sum_{|R|=2^{-n}} |\alpha_R|, \tag{127}$$

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

4.4 *Halász's proof of Schmidt's lower bound for the discrepancy*

We now explain how the same idea can be used to prove a discrepancy estimate. This argument has, in fact, been created by Halász [52, 1981] even earlier than Temlyakov's proof of the small ball inequality in $d = 2$. 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 (44) of the L^2 estimate (5) and whose existence is guaranteed by Lemma 1. The test function is then constructed in a fashion very similar to (124):

$$\Phi := \prod_{k=0}^{n} \Big(1 + \gamma f_k \Big) - 1 = \gamma \sum_{k=0}^{n} f_k \, + \, \Phi_{>n}, \tag{128}$$

where $\gamma > 0$ is a small constant, and $\Phi_{>n}$, by the product rule (123), is in the span of Haar functions with support of area less than 2^{-n}. In complete analogy with the previous proof, we find that $\|\Phi\|_1 \leq 2$. Also,

$$\left\langle D_N, \sum_{k=0}^{n} f_k \right\rangle \geq c(n+1) \geq C' \log N. \tag{129}$$

Up to this point the argument repeated the proof of the two-dimensional small ball conjecture word for word. 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 which is a natural extension of Lemma 1.

Lemma 5. *Let $f_{\mathbf{s}}$ be any* r*-function with parameter* $\mathbf{s}$*. Denote $s = \|\mathbf{s}\|_1$. Then, for some constant $\beta_d > 0$,*

$$\langle D_N, f_{\mathbf{s}} \rangle \leq \beta_d N 2^{-s}. \tag{130}$$

Proof. It follows from (33), that the area part of D_N satisfies $|\langle N x_1 \cdot \ldots \cdot x_d, f_{\mathbf{s}} \rangle| \lesssim 2^s \cdot N 2^{-2s} = N 2^{-s}$. As to the counting part, it follows from the proof of Lemma 1 that $\mathbf{1}_{[p,\mathbf{1}]}$ is orthogonal to the functions h_R for all $R \in \mathscr{D}^d_{\mathbf{s}}$ except for the rectangle R which contains the point p. It is then easy to check that

$$\langle \mathbf{1}_{[p,\mathbf{1})}, f_{\mathbf{s}} \rangle = \langle \mathbf{1}_{[p,\mathbf{1})}, h_R \rangle \lesssim |R| = 2^{-s}. \tag{131}$$

The estimate for the counting part of D_N then follows by summing over all the points of $\mathscr{P}_N$. □

We now estimate the higher order terms in $\langle D_N, \Phi \rangle$. Write $\Phi_{>n} = F_2 + F_3 + \ldots + F_n$, where

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

Notice that, due to the product rule, the product $f_{j_1} \cdot f_{j_2} \cdots f_{j_k}$ is an r-function with parameter $\mathbf{s} = (n - j_1, j_k)$, so $s = n - j_1 + j_k$. We reorganize the sum according to the parameter s, $n+1 \leq s \leq 2n$. To obtain a term which yields an r-function corresponding to a fixed value of s, we need to have $j_k = j_1 + s - n \leq n$. 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 $2 \leq k \leq s - n + 1$. We obtain

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

where we used that $N \leq 2^{n-1}$. Since $n \leq \log_2 N + 2$, by making γ very small we can assure that this quantity is less than $\frac{1}{2} C' \log N$, a half of (129). We finally obtain that

$$\|D_N\|_\infty \geq \frac{1}{2}\langle D_N, \Phi\rangle \geq \frac{1}{2}\left(C'\log N - \frac{1}{2}C'\log N\right) \gtrsim \log N, \tag{132}$$

which finishes the proof of Schmidt's bound. □

4.5 The proof of the L^1 discrepancy bound

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 (15) for the discrepancy function contained in the same fascinating paper by Halász [52] (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

$$\Gamma := \prod_{k=0}^{n}\left(1+\frac{i\gamma}{\sqrt{\log N}}f_k\right) - 1 = \frac{i\gamma}{\sqrt{\log N}}\sum_{k=0}^{n} f_k + \Gamma_{>n}, \tag{133}$$

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

$$\|\Gamma\|_\infty \leq \left(1+\frac{\gamma^2}{\log N}\right)^{\frac{n}{2}} + 1 \leq e^{\gamma^2/2} + 1 \lesssim 1. \tag{134}$$

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

$$\|D_N\|_1 \gtrsim |\langle D_N, \Gamma\rangle| \gtrsim \frac{\gamma}{\sqrt{\log N}}\langle D_N, \sum_{k=0}^{n} f_k\rangle \gtrsim \frac{n+1}{\sqrt{\log N}} \approx \sqrt{\log N}, \tag{135}$$

which finishes the proof of (15). □

4.6 Riesz products. Lacunary Fourier series

It is not surprising that the Riesz product approach is effective in these problems. As discussed earlier, the extremal values of the discrepancy function (as well as of hyperbolic Haar sums) are achieved on very thin sets. 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 §6.1.3.

But there is an even better explanation of the reason behind the successfull application of the Riesz products in these contexts. In order to understand its roots we turn to classical Fourier analysis. Riesz products have initially appeared in connection with lacunary Fourier series [84, 100, 119] and have proved to be an extremely important tool for these objects. It would be interesting to compare the estimates whose proofs we have just discussed with a classical theorem about lacunary Fourier series due to Sidon [100, 101]. Its proof can be found in almost every book on Fourier analysis, e.g. [119, 60, 50]. We shall reproduce it here in order to convince the reader that the proofs of the three previous inequalities (the small ball inequality (126) and lower bounds for $\|D_N\|_\infty$ (132) and $\|D_N\|_1$ (135) in dimension $d = 2$) are natural.

Recall that an increasing sequence $\{\lambda_j\}_{j=1}^\infty \subset \mathbb{N}$ is called *lacunary* if there exists $q > 1$ so that $\lambda_{j+1}/\lambda_j > q$. Let f be a 1-periodic function. We say that f has lacunary Fourier series if there exists a lacunary sequence Λ such that the Fourier coefficients of f,

$$\widehat{f}(k) = \int_0^1 f(x) e^{-2\pi i k x} dx, \tag{136}$$

are supported on the sequence Λ. In other words, $\widehat{f}(k) = 0$ whenever $k \notin \Lambda$. We have the following theorem.

Theorem 6 (Sidon, [100, 101]).

1. *Let f be a bounded* 1*-periodic function with lacunary Fourier series. Then we have*

$$\|f\|_\infty \gtrsim \sum_{k=1}^{\infty} |\widehat{f}(k)|. \tag{137}$$

2. *Assume that a function $f \in L^1[0,1]$ has lacunary Fourier series. Then*

$$\|f\|_1 \gtrsim \|f\|_2. \tag{138}$$

In both cases, the implicit constant depends only on the constant of lacunarity $q > 1$.

Proof. The reader will easily recognize the arguments that follow: the previous proofs in this section are their direct offspings. We shall initially operate under the assumption that $q \geq 3$. This condition guarantees that any integer n can be represented in the form $n = \sum_k \varepsilon_k \lambda_k$, $\varepsilon_k = -1, 0, 1$, in at most one way.

We begin by proving the first part of the theorem. Construct the following Riesz product

$$P_N(x) = \prod_{k=1}^{N} \big(1 + \cos(2\pi \lambda_k x + \delta_k)\big), \tag{139}$$

where δ_k is chosen so that $e^{i\delta_k} = \widehat{f}(k)/|\widehat{f}(k)|$. Obviously, $P_N(x)$ is non-negative for all x. It is also easy to see that $\widehat{P_N}(0) = \int_0^1 P(x)\,dx = 1$. Indeed, writing $\cos t = \frac{1}{2}(e^{it} + e^{-it})$ and multiplying the product out, we see that the leading term is 1 and all others have integral zero. Hence, $\|P_N\| = 1$.

Moreover, for $k \leq N$, we have $\widehat{P_N}(\lambda_k) = \frac{1}{2} e^{i\delta_k}$. This again follows from expanding the Riesz product. We obtain a trigonometric polynomial, in which, due to our assumption that $q \geq 3$, the term $e^{2\pi i \lambda_k x}$ can only arise from the product of the cosine in the k^{th} factor with the 1's coming from all the other factors. Besides, for $k > N$, evidently $\widehat{P_N}(\lambda_k) = 0$. Therefore we can apply the Parseval identity:

$$\|f\|_\infty \geq \left| \int_0^1 f(x)\overline{P_N(x)}dx \right| = \left| \sum_{k\in\mathbb{Z}} \widehat{f}(k)\overline{\widehat{P_N}(k)} \right| = \frac{1}{2}\sum_{k=1}^{N} |\widehat{f}(\lambda_k)|. \qquad (140)$$

Clearly, we can now take the limit as $N \to \infty$. The restriction $q \geq 3$ may be removed in the following fashion. Find the smallest n such that $q^n > 3$, $1 - \frac{1}{q^n-1} > \frac{1}{q}$, $1 + \frac{1}{q^n-1} < q$ and subdivide the sequence $\{\lambda_j\}_{j=1}^\infty$ into n subsequences of the form $\Lambda_m = \{\lambda_{m+jn}\}_{j=1}^\infty$, $m = 0, 1, \dots, n-1$. Then, repeating the argument above, we can prove an analog of (140) for Λ_m, i.e., $\|f\|_\infty \gtrsim \sum_{k\in\Lambda_m} |\widehat{f}(k)|$, see [60, Chapter V] for details. Summing these estimates over m finishes the proof.

We now turn to the proof of the second part of the theorem. It will also be achieved using a Riesz product. We first assume that $q \geq 3$. Let $a_N^2 = \sum_{k=1}^N |\widehat{f}(\lambda_k)|^2$ and $c_k = |\widehat{f}(\lambda_k)|/a_N$. Define the function

$$Q_N(x) = \prod_{k=1}^{N} \Big(1 + ic_k \cos(2\pi i \lambda_k x + \theta_k) \Big). \qquad (141)$$

It is then clear that $|Q_N(x)| \leq \prod_{k=1}^N (1+c_k^2)^{1/2} \leq e^{\frac{1}{2}\sum c_k^2} = \sqrt{e}$, i.e. $\|Q_N\|_\infty \leq \sqrt{e}$. If $q \geq 3$, we can easily show that $\widehat{Q_N}(\lambda_k) = \frac{1}{2} ic_k e^{i\theta_k} = \frac{1}{2a_N}\overline{\widehat{f}(\lambda_k)}$ for a proper choice of θ_k. Parseval's identity then yields

$$\|f\|_1 \gtrsim \int_0^1 f(x)\overline{Q_N(x)}dx = \sum_{k\in\mathbb{Z}} \widehat{f}(k)\overline{\widehat{Q_N}(k)} = \frac{1}{2a_N}\sum_{k=1}^{N} |\widehat{f}(\lambda_k)|^2 = \frac{1}{2}\left(\sum_{k=1}^{N} |\widehat{f}(\lambda_k)|^2 \right)^{\frac{1}{2}}.$$

We finish the proof of (138) by letting N approach infinity and recalling that $\|f\|_2^2 = \left(\sum_{k=1}^N |\widehat{f}(\lambda_k)|^2 \right)^{\frac{1}{2}}$. The assumption $q \geq 3$ is removed in exactly the same way as in the first case. □

One cannnot help but notice extremely close similarities between the constructions of Riesz products for the small ball inequality and discrepancy estimates in dimension $d = 2$ and the ones just used in the proof of Sidon's theorem. Indeed, the constructions (124) and (128) bear strong resemblance to the product (139) used to estimate $\|f\|_\infty$, while the idea of the product (133) is nearly identical to the Riesz product (141) which produces the bound for $\|f\|_1$.

The absolute efficiency of Riesz products in the two-dimensional cases of the small ball inequality and the L^∞ discrepancy bound 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 ($|R_1| = 2^{-k}$, $k = 0, 1, \dots, n$, in other words, the consecutive frequencies differ by a factor of 2). As we saw in this subsection, historically Riesz products were specifically designed to work in such settings (lacunary Fourier series, see e.g. [119], [84, 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 5 is taken in [21].

Before we proceed to the discussion of the recent progress in the multidimensional case, we would like to briefly explain the connections of Conjecture 5 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.

4.7 Probability: the small ball problem for the Brownian sheet

Having read thus far, the reader is perhaps slightly confused by the name *small ball inequality*. It would be worthwhile to explain this nomenclature at this point. It comes from probability theory, namely the small ball problem for the Brownian sheet, which is concerned with finding the exact asymptotic behavior of the small deviation probability $\mathbb{P}(\|\mathbb{B}\|_{L^\infty([0,1]^d)} < \varepsilon)$ as $\varepsilon \to 0$, where $\mathbb{B}$ is the Brownian sheet, i.e. a centered multiparameter Gaussian process characterized by the covariance relation

$$\mathbb{E}\mathbb{B}(s) \cdot \mathbb{B}(t) = \prod_{j=1}^{d} \min(s_j, t_j) \tag{142}$$

for $s, t \in [0,1]^d$. It is known that the paths of $\mathbb{B}$ are almost surely continuous, so we can safely write $L^\infty([0,1]^d)$ and $C([0,1]^d)$ norms interchangeably.

The circle of small deviation (or small ball) problems is an active and rapidly developing area of modern probability theory. The common goal of all of these problems is computing the probability that the values of a random variable or a random process deviate little from the mean in various senses (i.e. stay in a *small ball* for a certain norm). This field is far less understood than the classical area of *large deviation* estimates, and numerous fundamental questions about small deviations are still open. A detailed account of small ball probabilities for Gaussian processes can be found in a nice survey [71]. The Brownian sheet $\mathbb{B}$, being the basic example of a multiparameter process and a natural generalization of the Brownian motion, presents special interest.

For the sake of brevity, let us denote the logarithm of the probability of the small deviation of $\mathbb{B}$ in the sup-norm by $\varphi(\varepsilon) := -\log \mathbb{P}(\|\mathbb{B}\|_{L^\infty([0,1]^d)} < \varepsilon)$. It is well known that in the case when $d = 1$, i.e. $\mathbb{B}$ is the Brownian *motion*, $\varphi(\varepsilon) \approx \varepsilon^{-2}$ for small ε. Moreover, even the precise value of the implicit constant is known in this case: $\lim_{\varepsilon\to 0} \frac{\varphi(\varepsilon)}{\varepsilon^{-2}} = \frac{\pi^2}{8}$, see [47]. In higher dimensions, however, the situation becomes more complicated due to the appearance of logarithmic factors in this asymptotitcs. In dimension $d = 2$, it was shown by Bass [4, 1988] that $\varphi(\varepsilon) \lesssim \frac{1}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^3$. This estimate was later extended to all dimensions by Dunker, Kühn, Lifshits, and Linde [44]:

$$\varphi(\varepsilon) \lesssim \frac{1}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^{2d-1}. \tag{143}$$

On the other hand, it was established much earlier [40, 1982] that the probability of the small deviation in the L^2 norm in all dimensions $d \geq 2$ satisfies

$$-\log\mathbb{P}(\|\mathbb{B}\|_{L^2([0,1]^d)} < \varepsilon) \approx \frac{1}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^{2d-2}, \tag{144}$$

and since $\|\mathbb{B}\|_{L^2} \leq \|\mathbb{B}\|_{L^\infty}$, this readily implies $\varphi(\varepsilon) \gtrsim \frac{1}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^{2d-2}$. Thus, one finds a gap of the order of $\log\frac{1}{\varepsilon}$ between the upper and the lower estimates, and the lower estimate is, in fact, an L^2 bound. This is a situation, which closely mirrors what happens in the case of discrepancy and the small ball inequality. For a while the experts were not sure which of the two bounds, if any, is correct (notice that the upper bound (143) is too big when $d = 1$). However, it is now generally believed that the upper bound (143) is sharp for $d \geq 2$.

Conjecture 8. In dimensions $d \geq 2$, for the Brownian sheet B we have

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

The lower bound for $d = 2$ in this conjecture has been obtained by Talagrand [106] using (111). The work of Bilyk, Lacey, and Vagharshakyan [17, 18] yields a decrease in the gap between lower and upper bounds in dimensions $d \geq 3$. Namely, there exists $\theta = \theta(d) > 0$ such that for small ε

$$-\mathbb{P}(\|\mathbb{B}\|_{C([0,1]^d)} < \varepsilon) \gtrsim \varepsilon^{-2}(\log 1/\varepsilon)^{2d-2+\theta}. \tag{145}$$

This improvement was based on the progress in the higher-dimensional small ball inequality (196). We should now explain how the small ball inequality for Haar functions (111) enters the picture in this problem. The argument presented here follows Talagrand's ideas.

Small ball inequality implies a lower bound for the small deviation probability

Consider the integration operator $\mathscr{T}_d$ acting on functions on the unit cube $[0,1]^d$ and defined as

$$(\mathscr{T}_d f)(x_1,\dots,x_d) := \int_0^{x_1}\dots\int_0^{x_d} f(y_1,\dots,y_d)\,dy_1\dots dy_d. \tag{146}$$

Let $\{u_k\}_{k\in\mathbb{N}}$ be any orthonormal basis of $L^2([0,1]^d)$ and set $\eta_k = \mathscr{T}_d u_k$. Then the Brownian sheet can be represented as

$$\mathbb{B} = \sum_{k\in\mathbb{N}} \gamma_k \eta_k, \tag{147}$$

where γ_k are independent $\mathscr{N}(0,1)$ (standard Gaussian) random variables. This idea goes back to Levy's construction of the Brownian motion [67]. The Gaussian structure is not hard to check. As to the covariance, writing $\eta_k(s) = \langle \mathbf{1}_{[\mathbf{0},s)}, u_k\rangle$ and taking into account independence of γ_k's, one can easily compute

$$\begin{aligned}\mathbb{E}\Big(\sum_{k\in\mathbb{N}} \gamma_k \eta_k(s)\Big)\Big(\sum_{k\in\mathbb{N}} \gamma_k \eta_k(t)\Big) &= \sum_{k\in\mathbb{N}} \mathbb{E}\gamma_k^2 \cdot \eta_k(s)\eta_k(t) \\ &= \sum_{k\in\mathbb{N}} \langle \mathbf{1}_{[\mathbf{0},s)}, u_k\rangle\langle \mathbf{1}_{[\mathbf{0},t)}, u_k\rangle = \langle \mathbf{1}_{[\mathbf{0},s)}, \mathbf{1}_{[\mathbf{0},t)}\rangle \\ &= \big|[\mathbf{0},s)\cap[\mathbf{0},t)\big| = \prod_{j=1}^{d} \min\{s_j, t_j\},\end{aligned} \tag{148}$$

where in the second line we use the fact that u_k's form an orthonormal basis.

We shall use specific functions u_k and η_k. In dimension 1, for a dyadic interval I, consider the function

$$u_I(x) = \frac{1}{|I|^{\frac{1}{2}}}\big(-\mathbf{1}_{I_1}(x) + \mathbf{1}_{I_2}(x) + \mathbf{1}_{I_3}(x) - \mathbf{1}_{I_4}(x)\big), \tag{149}$$

where I_j, $j = 1,\dots,4$ are four quarters of I: successives dyadic subintervals of I of length $\frac{1}{4}|I|$. The point of this choice of u is that both u and its antiderivative $\mathscr{T}_1 u$ behave similarly to the Haar function. In particular, the system $\{u_I\}_{I\in\mathscr{D}}$ is also an orthonormal basis of $L^2([0,1])$. Observe that, up to the normalization, these functions are identical to the functions h_I^r with $r = 1$, defined in §2.3 in the proof of the lower bounds for the errors of cubature formulas in the class $B(MW_r^2)$, Theorem 3. In dimensions $d \geq 2$, one defines the basis functions indexed by dyadic rectangles $R = R_1 \times \cdots \times R_d \in \mathscr{D}^d$ as a tensor product

$$u_R(x_1,\dots,x_d) = u_{R_1}(x_1)\cdot\ldots\cdot u_{R_d}(x_d). \tag{150}$$

The functions $\eta_R = \mathscr{T}_d u_R$ are then continuous; moreover, their mixed derivative $\frac{\partial^d}{\partial x_1 \ldots \partial x_d}\eta_R = u_R$ has L^2 norm equal to 1. We shall now formulate a version of the small ball conjecture for these continuous wavelets.

Conjecture 9. In all dimensions $d \geq 2$, for any choice of coefficients α_R, we have the inequality

$$n^{\frac{d-2}{2}} \Big\| \sum_{|R|=2^{-n}} \alpha_R \eta_R \Big\|_\infty \gtrsim 2^{-\frac{3n}{2}} \sum_{|R|=2^{-n}} |\alpha_R|. \tag{151}$$

Notice that the factor $2^{-\frac{3n}{2}}$ is different from the one in the inequality (111). This is a result of normalization: while we have used L^∞-normalized Haar functions, $\|h_R\|_\infty = 1$, the sup-norm of the functions η_R is smaller, $\|\eta_R\|_\infty \approx 2^{-|R|/2} = 2^{-n/2}$.

Even though this conjecture is at the first glance somewhat harder than the small ball conjecture for the Haar functions, the proofs are usually similar. In fact, Talagrand in his paper [106] proves this conjecture for $d = 2$, but first he presents the proof of Conjecture 5 for the Haar functions, (111), despite the fact that strictly speaking it was not necessary – it is simply more transparent, less obstructed by the technicalities, and clearly explains the main ideas. The Riesz product arguments can also be adapted to this case. One can even still use Riesz products built with Haar functions, which brings the amount of technical complications to an absolute minimum (see the discussion on the last page of [18]).

For now let us assume that the conjectured inequality (151) holds. We shall now show how it implies a lower bound for the small deviation problem. First, we shall need a well-known fact from probability theory, which we state here in a very simple form.

Lemma 6 (Anderson's lemma, [3]). *Let X_t, Y_t, $t \in T$ be independent centered Gaussian random processes. Then for any bounded measurable function $\theta : \mathbb{R} \to \mathbb{R}$*

$$\mathbb{P}(\sup_{t\in T} |X_t + \theta(t)| < c) \leq \mathbb{P}(\sup_{t\in T} |X_t| < c) \qquad \text{and} \tag{152}$$

$$\mathbb{P}(\sup_{t\in T} |X_t + Y_t| < c) \leq \mathbb{P}(\sup_{t\in T} |X_t| < c). \tag{153}$$

The first inequality of this lemma reflects a general intuition that Gaussian measures are concentrated near zero. The second inequality can be deduced by simply applying the first one conditionally.

We now employ Anderson's lemma to extract just one layer of η_R's from the decomposition (147) of $\mathbb{B}$ – namely, we shall leave only those functions η_R which are supported on dyadic boxes of volume $|R| = 2^{-n}$ for a carefully chosen value of n. This idea strongly resonates with Roth's principle (24): just as in the case of the discrepancy function D_N, the behavior of the small ball probabilities of $\mathbb{B}$ is essentially defined by its projection onto the part of the basis which corresponds to rectangles with fixed volume. We apply (153) with $X_t = \sum_{|R|=2^{-n}} \gamma_R \eta_R$ and $Y_t = \sum_{|R|\neq 2^{-n}} \gamma_R \eta_R$. This would enable us to use the small ball inequality (151) as our next step.

$$\begin{aligned}\mathbb{P}(\|\mathbb{B}\|_{L^\infty([0,1]^d)} < \varepsilon) &\leq \mathbb{P}\left(\Big\| \sum_{|R|=2^{-n}} \gamma_R \eta_R \Big\|_\infty < \varepsilon \right) \\ &\leq \mathbb{P}\left(C n^{-\frac{d-2}{2}} 2^{-\frac{3n}{2}} \sum_{|R|=2^{-n}} |\gamma_R| < \varepsilon \right),\end{aligned} \tag{154}$$

where C is the implied constant in (151). We are left with a standard object in probability theory: the sum of absolute values of independent $\mathcal{N}(0,1)$ random variables.

Using the exponential form of Chebyshev's inequality we can write for a sequence of independent standard Gaussians γ_k:

$$\mathbb{P}\left(\sum_{k=1}^{M}|\gamma_k| \le A\right) \le e^A\,\mathbb{E}e^{-\sum_{k=1}^{M}|\gamma_k|} = e^A\left(\mathbb{E}e^{-|\gamma|}\right)^M. \tag{155}$$

We now apply this inequality with $M = \#\{R \in \mathcal{D}^d : |R| = 2^{-n}\} = 2^n \cdot \#\mathbb{H}_n^d \approx 2^n n^{d-1}$ and $A = \frac{\varepsilon}{C} n^{\frac{d-2}{2}} 2^{\frac{3n}{2}}$ in order to be able to finish (154). We see that the right-hand side of (155) is then bounded by $\exp(\frac{1}{C}\varepsilon n^{\frac{d-2}{2}} 2^{\frac{3n}{2}} - C_1 2^n n^{d-1})$. Choosing n to be the maximal integer such that

$$\frac{1}{C}\varepsilon n^{\frac{d-2}{2}} 2^{\frac{3n}{2}} \le \frac{1}{2}C_1 2^n n^{d-1}, \quad \text{i.e.} \quad \varepsilon \le CC_1 2^{-\frac{n}{2}} n^{\frac{d}{2}}, \tag{156}$$

we find that, since in this case $\varepsilon \approx 2^{-\frac{n}{2}} n^{\frac{d}{2}}$,

$$\mathbb{P}\left(\sum_{|R|=2^{-n}}|\gamma_R| < \frac{\varepsilon}{C} n^{\frac{d-2}{2}} 2^{\frac{3n}{2}}\right) \le e^{-\frac{1}{2}C_1 2^n n^{d-1}} \le e^{-\frac{C''}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^{2d-1}}. \tag{157}$$

Therefore,

$$\varphi(\varepsilon) = -\log\mathbb{P}\left(\|\mathbb{B}\|_\infty < \varepsilon\right) \gtrsim \frac{1}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^{2d-1}. \tag{158}$$

This finishes the proof of the lower bound in Conjecture 8 assuming that the smooth (or, rather, continuous) version of the small ball conjecture, Conjecture 9, holds. □

Notice that, in another close parallel to Roth's method in discrepancy theory, we chose $n \approx \log\frac{1}{\varepsilon}$, although the exact choice of its value here was more delicate.

4.8 Approximation theory: entropy of classes with mixed smoothness

Consider the integration operator $\mathcal{T}_d$ as described in (146). Let us define the function space $MW^p([0,1]^d) = \mathcal{T}_d(L^p([0,1]^d))$ and set $B(MW^p) = \mathcal{T}_d(B(L^p))$ to be the image of the unit ball of $L^p([0,1]^d)$ under the action of $\mathcal{T}_d$. In other words, $MW^p([0,1]^d)$ can be viewed as 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 $B(MW^p)$ is its unit ball. These function classes have already been defined in §2.3. It is not hard to see that $B(MW^p)$ is compact in the L^∞ metric. Its compactness may be quantified using the notion of *covering numbers*. Let B_∞ denote the unit ball of $L^\infty([0,1]^d)$ and define

$$N(\varepsilon,p,d) := \min\left\{N : \exists\{x_k\}_{k=1}^N \subset B(MW^p), B(MW^p) \subset \bigcup_{k=1}^N (x_k + \varepsilon B_\infty)\right\} \tag{159}$$

to be the least number N of L^∞ balls of radius ε needed to cover the unit ball $B(MW^p([0,1]^d))$, or, equivalently, the size of the smallest ε-net of $B(WM^p)$ in the uniform norm. The task at hand is to determine the correct order of growth of these numbers as $\varepsilon \downarrow 0$. The quantity

$$\psi(\varepsilon) = \log N(\varepsilon, p, d) \tag{160}$$

is referred to as the *metric entropy* of $B(MW^p)$ with respect to the L^∞ norm. The inverse of this quantity is known in the literature as *entropy numbers*:

$$\varepsilon_m := \inf\left\{\varepsilon : \exists \{x_k\}_{k=1}^{2^m} \subset B(MW^p),\, B(MW^p) \subset \bigcup_{k=1}^{2^m} (x_k + \varepsilon B_\infty)\right\}, \tag{161}$$

in other words, the smallest value of ε for which $\psi(\varepsilon) \le m$. It is clear that estimates of metric entropy or covering numbers may be reformulated in terms of the entropy numbers, however we shall mostly resort to the former.

Kuelbs and Li [63] have discovered a tight connection between the small ball probabilities and the properties of the corresponding reproducing kernel Hilbert space, which in the case of the Brownian sheet is $WM^2([0,1]^d)$, see §4.9. We state a partial form of their result tailored to the topic of our presentation.

Theorem 7 (Kuelbs, Li, [63]). *The rates of asymptotic growth of the metric entropy $\psi(\varepsilon)$ of the space $MW^2([0,1]^d)$ and the logarithm of the small ball probability of the d-dimensional Brownian sheet $\varphi(\varepsilon) = -\log \mathbb{P}(\|\mathbb{B}\|_\infty < \varepsilon)$ are related in the following way. For $\alpha > 0$,*

$$\varphi(\varepsilon) \approx \varepsilon^{-2}\big(\log \frac{1}{\varepsilon}\big)^{\alpha} \quad \text{if and only if} \quad \psi(\varepsilon) \approx \varepsilon^{-1}\big(\log \frac{1}{\varepsilon}\big)^{\alpha/2}. \tag{162}$$

We shall explore this connection in a more general setting in §4.9 and, in particular, prove this theorem. For more information and a wider spectrum of inequalities relating the small deviation probabilities and metric entropy, the reader is referred to [63, 71, 74].

Theorem 7 together with Conjecture 8 yields an equivalent conjecture:

Conjecture 10. For $d \ge 2$, we have

$$\log N(\varepsilon, 2, d) \simeq \varepsilon^{-1}(\log 1/\varepsilon)^{d-1/2}, \tag{163}$$

as $\varepsilon \downarrow 0$.

Just as in the case of the small ball probabilities for the Brownian sheet, the conjecture is resolved in dimension $d = 2$, which follows from the work of Talagrand [106]. The upper bound is known in all dimensions [44]. The lower bound of the order $\frac{1}{\varepsilon}\big(\log\frac{1}{\varepsilon}\big)^{d-1+\theta/2}$ in dimensions $d \ge 3$ can be 'translated' from the corresponding inequality (145) for the Brownian sheet.

We would now like to discuss the relation between this conjecture and the small ball inequality, Conjecture 5. Of course, one can combine the arguments of the previous subsection for the Brownian sheet with the Kuelbs–Li equivalence to demonstrate that the lower bound in Conjecture 10 follows from the small ball conjecture (111) or, more precisely, its continuous counterpart (151). However, we would like to illustrate how one can use the small ball inequality to directly deduce the lower bound for the metric entropy.

Small ball conjecture implies a lower bound for metric entropy

Estimates akin to the small ball inequality (111) or (151) have been known for a long time to be useful for obtaining bounds of various approximation theory characteristics, such as metric entropy, entropy numbers, Kolmogorov widths etc, see [111, 112]. We present one possible approach to this connection.

We shall use the basis functions u_R (see (150)) and their antiderivatives $\eta_R = \mathscr{T}_d u_R$ defined in the previous subsection. Let $\sigma : \{R \in \mathscr{D}^d, |R| = 2^{-n}\} \to \{\pm 1\}$ be a choice of signs on the set of dyadic rectangles with fixed volume 2^{-n}. Define the functions

$$F_\sigma = \frac{c}{2^{\frac{n}{2}} n^{\frac{d-1}{2}}} \sum_{|R|=2^{-n}} \sigma_R \eta_R, \tag{164}$$

where $c > 0$ is a small constant. Then by the orthonormality of the functions u_R we have

$$\left\| \frac{\partial^d F_\sigma}{\partial x_1 \dots \partial x_d} \right\|_2^2 = \left\| \frac{c}{2^{\frac{n}{2}} n^{\frac{d-1}{2}}} \sum_{|R|=2^{-n}} \sigma_R u_R \right\|_2^2 = \frac{c^2}{2^n n^{d-1}} 2^n \cdot \# \mathbb{H}_n^d \le 1, \tag{165}$$

if c is sufficiently small. Since $\eta_R = \mathscr{T}_d u_R$, this estimate implies that $F_\sigma \in B(MW^2)$. Now assume that the continuous version of the small ball conjecture, Conjecture 9 holds for the functions η_R. Take two different choices of signs σ and σ'. Then (151) would imply:

$$\begin{aligned} \left\| F_\sigma - F_{\sigma'} \right\|_\infty &= \left\| \frac{c}{2^{\frac{n}{2}} n^{\frac{d-1}{2}}} \sum_{|R|=2^{-n}} (\sigma_R - \sigma'_R) \eta_R \right\|_\infty \\ &\gtrsim 2^{-\frac{n}{2}} n^{-\frac{d-1}{2}} \cdot n^{-\frac{d-2}{2}} 2^{-\frac{3n}{2}} \sum_{|R|=2^{-n}} |\sigma_R - \sigma'_R| \\ &\gtrsim n^{-\frac{2d-3}{2}} 2^{-2n} \cdot 2^n \# \mathbb{H}_n^d \approx n^{1/2} 2^{-n}, \end{aligned} \tag{166}$$

where we have additionally assumed that σ and σ' differ on a large portion (e.g., one quarter) of all dyadic rectangles with volume 2^{-n}. We see that, in this case, F_σ and $F_{\sigma'}$ are ε-separated in L^∞ with $\varepsilon = 2^{-n} n^{1/2}$.

In order to construct a large ε net for the set $B(MW^2([0,1]^d))$, it would be therefore sufficient to produce a large collection $\mathscr{A}$ of choices of sign σ such that any two elements of $\mathscr{A}$ are sufficiently different, i.e. coincide at most on a fixed portion of the rectangles.

Coding theory comes in handy in this setup. In fact, a reader familiar with its basic notions perhaps already recognized the concept of Hamming distance in the previous sentence. Consider a binary code X of length m, i.e. $X \subset \{0,1\}^m$ is just a collection of strings of m zeros and ones. For any two elements x, $y \in X$, their Hamming distance is defined as

$$d_H(x,y) = \#\{j = 1,...,m : x_j \neq y_j\}, \tag{167}$$

in other words, the number of components in which x and y do not coincide. The minimum Hamming distance (weight) of the code X is then defined as the smallest Hamming distance between its elements, $\min_{x,y\in X, x\neq y} d_H(x,y)$. The following classical result in coding theory, which we state in the simplest form adapted to our exposition, provides a lower bound on the size of the maximal code with large minimum Hamming weight.

Lemma 7 (Gilbert–Varshamov bound [49, 117]). *Let $A(m,k)$ denote the maximal size of a binary code of length m with the minimum Hamming distance at least k. Then*

$$A(m,k) \geq \frac{2^m}{\sum_{j=0}^{k-1}\binom{m}{j}}. \tag{168}$$

The proof of this estimate is so beautifully simple that we decided to include it here.

Proof. We first observe that given an m-bit string $x \in \{0,1\}^m$, there are precisely $\binom{m}{j}$ strings $y \in \{0,1\}^m$ such that $d_H(x,y) = j$. Indeed, we need to choose j bits out of m that are to be changed. Hence the size of $B_H(x,k)$, the neighborhood of x of radius k in the Hamming metric (all elements y with $d_H(x,y) < k$), is equal to $\sum_{j=0}^{k-1}\binom{m}{j}$.

Let now X be the maximal code of length m with minimum Hamming weight k. Then $\cup_{x\in X} B_H(x,k) = \{0,1\}^m$, for otherwise there would exist another element whose distance to all points of X is at least k, which would violate the maximality of X. Thus,

$$2^m = \#\bigcup_{x\in X} B_H(x,k) \leq \sum_{x\in X} \#B_H(x,k) = \#X \cdot \sum_{j=0}^{k-1}\binom{m}{j}, \tag{169}$$

which proves the lemma. □

We shall apply this lemma to codes X indexed by the family of dyadic rectangles $\{R \in \mathscr{D}^d, |R| = 2^{-n}\}$. Hence, the length of the code is $m = 2^n \#\mathbb{H}_n^d \approx 2^n n^{d-1}$. For any element of such a code $x \in X$, we can define a choice of sign σ^x by setting $\sigma_R^x = (-1)^{x_R}$. We would like the code to have the minimal Hamming weight of the same order of magnitude as the length of the code, i.e. $k \approx m \approx 2^n n^{d-1}$. Take, for example, $k = \frac{m}{4}$. One can easily check using Stirling's formula $m! \approx \frac{1}{\sqrt{2\pi m}}\left(\frac{m}{e}\right)^m$ that

$\binom{m}{m/4} \approx \frac{1}{\sqrt{m}}\left(\frac{1}{4}\right)^{-m/4}\left(\frac{3}{4}\right)^{-3m/4}$. The Gilbert-Varshamov bound then guarantees that there exists such a code X with size at least

$$\begin{aligned} \#X \geq \frac{2^m}{\sum_{j=0}^{k-1}\binom{m}{j}} \geq \frac{2^m}{k\cdot\binom{m}{k}} = \frac{2^m}{m/4\cdot\binom{m}{m/4}} \\ \approx \frac{1}{\sqrt{m}}\cdot 2^m \left(\frac{1}{4}\right)^{m/4}\left(\frac{3}{4}\right)^{3m/4} \gtrsim C^m \end{aligned} \tag{170}$$

for some constant $C > 1$ when m is large, since $2\cdot(1/4)^{1/4}\cdot(3/4)^{3/4} > 1$. To summarize, we can find a code such that its Hamming weight is roughly the same as its length m and its size is roughly the same as the size of the largest possible code, $\{0,1\}^m$ (both are exponential in m).

Having chosen such a code X, we define the collection $\mathscr{A} = \{\sigma^x : x \in X\}$ and consider the set of functions $\mathscr{F} = \{F_\sigma\}_{\sigma\in\mathscr{A}}$. According to (166) this family is an ε-net of $B(MW^2)$ in the L^∞ norm with $\varepsilon = 2^{-n}n^{1/2}$. The cardinality of this family satisfies

$$\log\#\mathscr{F} = \log\#X \gtrsim m \approx 2^n n^{d-1} = 2^n n^{-1/2}\cdot n^{d-\frac{1}{2}} \approx \frac{1}{\varepsilon}\left(\log\frac{1}{\varepsilon}\right)^{d-\frac{1}{2}}, \tag{171}$$

which yields precisely the lower bound in Conjecture 10. □

In the end we would like to observe that in the proof of this implication we have employed only a restricted form of the small ball inequality. In the computation (166), the coefficients $\alpha_R = \sigma_R - \sigma_R'$ take only three values: ± 2 and 0. Besides, zeros are not allowed to occur too often (at most a fixed proportion of all coefficients). This (up to a factor of 2) is exactly the setting of the *generic* signed small ball conjecture, Conjecture 7. Therefore, this version of the conjecture (but with smooth wavelets η_R in place of the Haar functions) is already sufficient for applications. However, unlike Conjecture 6 (the purely signed variant of the inequality, see §5.5), the generic setting does not seem to produce any real simplifications.

4.9 The equivalence of small ball probabilities and metric entropy

The equivalence of Conjecture 8 in probability and Conjecture 10 in approximation theory proved by Kuelbs and Li [63] is a fascinating connection between two problems, which at first glance have little in common. We strongly agree with Michel Talagrand who stated [106]:

> It certainly would be immoral to deprive the reader of a discussion of this beautifully simple fact (that once again demonstrates the power of abstract methods).

Therefore we would like devote a portion of this chapter to the discussion of the proof of this equivalence.

Before we are able to explain the argument however, we need to recall some classical results from the theory of Gaussian measures, which we shall state here without proof. Complete details and background information may be found in such excellent references as [22], [69], or [74]. We, rather than giving the most general definitions and statements, will mostly specialize to the particular problem at hand.

Let $\mathbf{P}$ be a Gaussian measure on the Banach space X. The *small ball problem* for the measure $\mathbf{P}$ is concerned with the asymptotic behavior of the quantity

$$\varphi(\varepsilon) = -\log \mathbf{P}(\varepsilon B_X), \tag{172}$$

where B_X is the unit ball of the space X.

In the case we are interested in, the Brownian sheet, the space X is $C([0,1]^d)$ and the measure $\mathbf{P}$ is the law of the Brownian sheet $\mathbb{B}$, i.e. for a set $A \in C([0,1]^d)$, $\mathbf{P}(A) = \mathbb{P}(\mathbb{B} \in A)$. In this notation, the definition of $\varphi(\varepsilon)$ above coincides with the one given in §4.7

$$\varphi(\varepsilon) = -\log \mathbb{P}(\|\mathbb{B}\|_{L^\infty([0,1]^d)} < \varepsilon) = -\log \mathbf{P}(B_\infty(0,\varepsilon)), \tag{173}$$

where $B_\infty(a,r)$ is the L^∞ ball of radius $r > 0$ centered at $a \in C([0,1]^d)$. Recall that $\mathbb{B}$ has mean zero, so the measure $\mathbf{P}$ is centered.

Assume that X, as in our case, is a space of real valued functions on a domain $D \subset \mathbb{R}^d$ with the property that point evaluations $L_x(f) = f(x)$, $x \in D$, are continuous linear functionals on X. We can then introduce the covariance kernel of $\mathbf{P}$, the function $K_{\mathbf{P}} : D \times D \to \mathbb{R}$ defined by

$$K_{\mathbf{P}}(s,t) = \int_X f(s) f(t) \mathbf{P}(df). \tag{174}$$

By definition, see (142), the covariance kernel of the Brownian sheet $\mathbb{B}$ is given by $K_{\mathbf{P}}(s,t) = \mathbb{E}\mathbb{B}(s)\mathbb{B}(t) = \prod_{j=1}^d \min\{s_j,t_j\}$.

The *reproducing kernel Hilbert space* $H_{\mathbf{P}}$ is then defined as the Hilbert space of functions $f \in X$ with the property that the reproducing kernel of $H_{\mathbf{P}}$ is precisely the covariance kernel of $\mathbf{P}$, i.e. for $t \in D$ and any $f \in H_{\mathbf{P}}$, the function evaluation of f at t can be represented as the inner product of f and $K_{\mathbf{P}}(\cdot,t)$,

$$f(t) = \langle f, K_{\mathbf{P}}(\cdot,t) \rangle. \tag{175}$$

In the case when $X = C([0,1]^d)$ and $\mathbf{P}$ is the law of the Brownian sheet, this space happens to be precisely the Sobolev space of functions with mixed derivative in L^2 as defined in the previous subsection, $H_{\mathbf{P}} = MW^2([0,1]^d)$. Indeed, MW^2 is a Hilbert space with the inner product given by

$$\langle f, g \rangle_{MW^2} = \langle \phi_f, \phi_g \rangle_{L^2} = \int_{[0,1]^d} \frac{\partial^d f}{\partial x_1 \ldots \partial x_d}(x) \cdot \frac{\partial^d g}{\partial x_1 \ldots \partial x_d}(x) dx, \tag{176}$$

where $\phi_f \in L^2([0,1]^d)$ is such that $f = \mathscr{T}_d \phi_f$, in other words ϕ_f is the mixed derivative of f. It is easy to see that in $d = 1$, $\min\{s,t\} = \int_0^s \mathbf{1}_{[0,t)}(\tau)\, d\tau = \big(\mathscr{T}_1 \mathbf{1}_{[0,t)}\big)(s)$

for $s,t \in [0,1]$. Therefore, in d dimensions $K_{\mathbf{P}}(s,t) = \big(\mathscr{T}_d\big(\prod_{j=1}^d \mathbf{1}_{[0,t_j)}\big)\big)(s_j)$ and for any $f \in MW^2([0,1]^d)$ we have

$$\langle f, K_{\mathbf{P}}(\cdot,t)\rangle_{MW^2} = \int_0^{t_1} \dots \int_0^{t_d} \phi_f(s)ds = f(t), \tag{177}$$

hence $K_{\mathbf{P}}(s,t)$ is in fact the reproducing kernel of MW^2.

In a certain sense, $H_{\mathbf{P}}$ is a subspace of X which carries most of the information about the measure $\mathbf{P}$. We shall need two standard facts which relate the Gaussian measure and its reproducing kernel Hilbert space.

Lemma 8. *Let $\mathbf{P}$ be a centered Gaussian measure on a Banach space X, let $H_{\mathbf{P}}$ be its reproducing kernel Hilbert space and $h \in H_{\mathbb{P}}$. Then, for any symmetric set $A \in X$ we have*

$$\exp\big(-\|h\|_{H_{\mathbf{P}}}^2/2\big) \cdot \mathbf{P}(A) \le \mathbf{P}(A+h) \le \mathbf{P}(A). \tag{178}$$

The right inequality here is simply a restatement of Anderson's lemma, (153), which is intuitively natural since a Gaussian measure is concentrated around the mean. The left bound, known as Borell's inequality, shows that the measure of a shifted set decays not too fast, in a fashion suggested by the Gaussian structure of the measure. The assumption $h \in H_{\mathbf{P}}$ is crucial for Borell's inequality as the shifted measure $\mathbf{P}(\cdot + h)$ is not even absolutely continuous with respect to $\mathbf{P}$ unless h lies in the reproducing kernel Hilbert space.

The second fact that we shall rely upon is the isoperimetric inequality.

Theorem 8 (Gaussian isoperimetric inequality). *Let $\mathbf{P}$ be a centered Gaussian measure on the Banach space X and K be the unit ball of $H_{\mathbf{P}}$. For a measurable set $A \subset X$ and $\lambda > 0$, we have*

$$\Phi^{-1}(\mathbf{P}(A+\lambda K)) \ge \Phi^{-1}(\mathbf{P}(A)) + \lambda, \tag{179}$$

where Φ is the distribution function of a $\mathcal{N}(0,1)$ (standard Gaussian) random variable, i.e. $\Phi(x) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^x e^{-t^2/2}dt$. The equality in (179) *holds whenever A is a half-space.*

This inequality is a proper extension of the classical Euclidean isoperimetric inequality to the infinite dimensional setting, where $\mathbb{R}^d$ is replaced by a Banach space X, the volume by the Gaussian measure $\mathbf{P}$, and the surface measure of A by $\lim_{\lambda\downarrow 0} \frac{1}{\lambda}(\mathbf{P}(A+\lambda K) - \mathbf{P}(A))$. Observe that in the Gaussian case the role of Euclidean balls is played by half-spaces.

Such a correspondence allows one to transfer geometric volume arguments to Banach spaces, where volume is not available. Indeed, if one wants to establish the connection between the covering numbers and the size of the small balls, the first impulse is to attempt to compare volumes. We have already given an argument along these lines in the proof of the Gilbert–Varshamov bound (168). In the general case, Gaussian measures provide an appropriate substitution for the notion of volume, while the above estimates (178) and (179) provide the necessary tools.

We are now ready to give the proof of the equivalence between the metric entropy and small ball probability estimates.

Let $N(\varepsilon,K)$ be the covering number of K, the unit ball of $H_{\mathbf{P}}$, with respect to the norm of X, that is the smallest number N such that for some $\{x_k\}_{k=1}^N \subset K$ we have $K \subset \cup_{k=1}^N B_X(x_k,\varepsilon)$, where $B_X(a,r) = \{x \in X : \|x-a\|_X < r\}$. Consider the quantities $\psi(\varepsilon) = \log N(\varepsilon,K)$ (the *metric entropy*) and $\varphi(\varepsilon) = -\log \mathbf{P}(\varepsilon B_X)$.

Lemma 9. *We have the following two estimates relating the metric entropy and the small ball probability:*

$$\psi(\sqrt{2}\varepsilon/\sqrt{\varphi(\varepsilon)}) \le 2\varphi(\varepsilon), \tag{180}$$

$$\psi(\varepsilon/\sqrt{2\varphi(\varepsilon)}) \ge \varphi(2\varepsilon) - \log 2. \tag{181}$$

Proof. Fix a parameter $\lambda > 0$ to be chosen later. Let $M = M(\varepsilon)$ be the largest number of *disjoint* balls of X of radius ε with centers in λK: $B_X(x_k,\varepsilon)$, $x_k \in \lambda K$, $k = 1,...,M$. Then $N(2\varepsilon,\lambda K) = N(2\varepsilon/\lambda,K) \le M(\varepsilon)$. Indeed, doubling the radii of all $M(\varepsilon)$ disjoint balls we obtain a covering of λK by balls of radius 2ε (if some point x of λK is not covered, then $B_X(x,\varepsilon)$ does not intersect any of the original balls, which contradicts the maximality assumption: we have chosen the *largest* disjoint family). By Borell's inequality, we have $\mathbf{P}(B_X(x_k,\varepsilon)) \ge e^{-\lambda^2/2}\mathbf{P}(\varepsilon B_X)$. Therefore, by disjointness of the balls $B(x_k,\varepsilon)$,

$$1 = \mathbf{P}(X) \ge \sum_{k=1}^M \mathbf{P}(x_k,\varepsilon) \ge N(2\varepsilon/\lambda,K)\cdot e^{-\lambda^2/2} P(\varepsilon B_X). \tag{182}$$

Hence, taking logarithms, one obtains

$$\psi(2\varepsilon/\lambda) \le \frac{\lambda^2}{2} + \varphi(\varepsilon). \tag{183}$$

Choosing $\lambda = \sqrt{2\varphi(\varepsilon)}$ results in $\psi(\sqrt{2}\varepsilon/\sqrt{\varphi(\varepsilon)}) \le 2\varphi(\varepsilon)$, which proves (180).

In the opposite direction, let the family of balls $\{B(x_k,\varepsilon)\}_{k=1}^N$, $x_k \in \lambda K$, be a covering of λK. Then $N \ge N(\varepsilon,\lambda K) = N(\varepsilon/\lambda,K)$. Besides, the doubled balls $\{B(x_k,2\varepsilon)\}_{k=1}^N$ obviously form a covering of a "thickened" set $\lambda K + \varepsilon B_X$. Therefore, using Anderson's lemma (the second inequality in (178)), we arrive to

$$\mathbf{P}(\lambda K + \varepsilon B_X) \le \sum_{k=1}^N \mathbf{P}(B_X(x_k,2\varepsilon)) \le N(\varepsilon/\lambda,K)\cdot \mathbf{P}(2\varepsilon B_X). \tag{184}$$

We now only need to show that the left-hand side is bounded below by some constant. Notice that the thickening was necessary, since $\mathbf{P}(\lambda K) = 0$. We shall apply the isoperimetric inequality (179) with $A = \varepsilon B_X$ and $\lambda = \sqrt{2\varphi(\varepsilon)}$. We have

$$\begin{aligned}\mathbf{P}(\lambda K + \varepsilon B_X) &\ge \Phi(\Phi^{-1}(\mathbf{P}(\varepsilon B_X)) + \lambda) = \Phi\big(\Phi^{-1}\big(e^{-\varphi(\varepsilon)}\big) + \sqrt{2\varphi(\varepsilon)}\big) \\ &\ge \Phi\big(-\sqrt{2\varphi(\varepsilon)} + \sqrt{2\varphi(\varepsilon)}\big) = \Phi(0) = \frac{1}{2},\end{aligned} \tag{185}$$

where we have used the fact that $\Phi(-x) \leq e^{-x^2/2}$. Therefore it follows from (184) that $\psi(\varepsilon/\sqrt{2\varphi(\varepsilon)}) \geq \varphi(2\varepsilon) - \log 2$, which is precisely (181). □

Proof of Theorem 7. We now specialize these estimates to the Brownian sheet $\mathbb{B}$ and its reproducing kernel Hilbert space MW^2. In this situation $\mathbf{P}(\varepsilon B_X) = \mathbb{P}(\|\mathbb{B}\|_{C([0,1]^d)} < \varepsilon)$ and $N(\varepsilon, K) = N(\varepsilon, MW^2([0,1]^d)) = N(\varepsilon, 2, d)$.

Assume that, as suggested by the discussion in §4.7, $\varphi(\varepsilon) \approx \varepsilon^{-2}\big(\log\frac{1}{\varepsilon}\big)^{\alpha}$. Setting $\delta = \frac{\sqrt{2}\varepsilon}{\sqrt{\varphi(\varepsilon)}} \approx \frac{\varepsilon^2}{(\log\frac{1}{\varepsilon})^{\alpha/2}}$ and using (180), we obtain

$$\psi(\delta) \lesssim \varepsilon^{-2}\bigg(\log 1/\varepsilon\bigg)^{\alpha} \approx \delta^{-1}\bigg(\log\frac{1}{\delta}\bigg)^{\alpha/2}. \tag{186}$$

The other parts of the equivalence (162) are proved analogously. □

4.10 Trigonometric polynomials with frequencies in the hyperbolic cross

Finally we would like to give a short overview of a different, but closely related analog of the small ball inequality, namely its version for trigonometric polynomials. Consider periodic functions defined on $\mathbb{T}^d$. For an integrable function on $\mathbb{T}^d$, its Fourier coefficients are defined as $\widehat{f}_{\mathbf{k}} = \int_{\mathbb{T}} f(x)e^{-2\pi i\mathbf{k}\cdot x}\,dx$ where $\mathbf{k} = (k_1, ..., k_d) \in \mathbb{Z}^d$. In the case of trigonometric polynomials, unlike the case of Haar functions, frequencies are not readily dyadic. Hence it will be useful to split the frequencies into dyadic blocks. For a vector $\mathbf{s} = (s_1, ..., s_d) \in \mathbb{Z}^d_+$ we denote

$$\rho(\mathbf{s}) := \{\mathbf{k} \in \mathbb{Z}^d : \big[2^{s_j-1}\big] \leq k_j < 2^{s_j},\ j = 1, ..., d\}, \tag{187}$$

where $[x]$ stands for the integer part of x. We then define the dyadic blocks of a function $f \in L^1(\mathbb{T}^d)$ as parts of the Fourier expansion of f which correspond to $\rho(\mathbf{s})$:

$$\delta_{\mathbf{s}}(f)(x) := \sum_{\mathbf{k}:\,|\mathbf{k}|\in\rho(\mathbf{s})} \widehat{f}_{\mathbf{k}} e^{2\pi i\mathbf{k}\cdot x}, \tag{188}$$

where we put $|\mathbf{k}| = (|k_1|, ..., |k_d|)$. These blocks play a similar role to the expressions $\sum_{R\in\mathscr{D}^d_{\mathbf{s}}} a_R h_R$, where the summation runs over the family of disjoint dyadic rectangles R with $|R_j| = 2^{-s_j}$ for $j = 1, ..., d$. Such linear combinations appeared naturally in the definitions of the r-functions (28) and the dyadic Littlewood–Paley square function (86).

The Littlewood–Paley inequalities adapted to this trigonometric setting read

$$\|f\|_p \approx \bigg\|\bigg(\sum_{\mathbf{s}\in\mathbb{Z}^d_+} |\delta_{\mathbf{s}}(f)|^2\bigg)^{\frac{1}{2}}\bigg\|_p, \tag{189}$$

which bears a strong resemblance to (86)-(87). In particular, when $d = 1$ one recovers the classical Littlewood–Paley inequalities for Fourier series.

For an even number n, denote by $\mathbf{Y}_n = \{\mathbf{s} \in (2\mathbb{Z}_+)^d : s_1 + ... + s_d = n\}$ the set of vectors with even coordinates and ℓ_1 norm equal to n. This is essentially the familiar set $\mathbb{H}_n^d$ slightly modified for technical reasons. We shall also define the dyadic *hyperbolic cross* as

$$\mathbb{Q}_n = \bigcup_{\mathbf{s} \in \mathbb{Z}_+^d : s_1 + ... + s_d \leq n} \rho(\mathbf{s}). \tag{190}$$

In dimension $d = 2$, roughly speaking, it consists of the integer points that lie under the parabola $xy = 2^n$ and satisfy $x, y < 2^n$. Considering integer vectors $\mathbf{k}$ with $|k| \in \mathbb{Q}_n$ produces a symmetrization which visualizes the meaning of the name dyadic *hyperbolic cross*. The pure hyperbolic cross is defined as $\Gamma(N) = \{\mathbf{k} \in \mathbb{Z}^d : \prod_{j=1}^d \max\{1, |k_j|\} \leq N$, which makes the term even more obvious.

The trigonometric analog of the small ball inequality (111) in dimension $d = 2$

$$\Big\| \sum_{\mathbf{s} \in \mathbb{Y}_n} \delta_{\mathbf{s}}(f) \Big\|_\infty \gtrsim \sum_{\mathbf{s} \in \mathbb{Y}_n} \big\| \delta_{\mathbf{s}}(f) \big\|_1 \tag{191}$$

was obtained by Temlyakov [111] via a Riesz product argument very similar to §4.3. One can notice easily that the small ball inequality for the Haar functions can be rewritten in a very similar form

$$n^{\frac{d-2}{2}} \Big\| \sum_{R:\,|R| = 2^{-n}} \alpha_R h_R \Big\|_\infty \gtrsim \sum_{\mathbf{r} \in \mathbb{H}_n^d} \Big\| \sum_{R \in \mathscr{D}_{\mathbf{r}}^d} \alpha_R h_R \Big\|_1 . \tag{192}$$

In fact, (191) can be improved to a somewhat stronger version stronger version. Define the *best hyperbolic cross approximation* of f as

$$E_{\mathbb{Q}_n}(f)_p = \inf_{t \in T(\mathbb{Q}_n)} \|f - t\|_p, \tag{193}$$

where $T(\mathbb{Q}_n) = \{t : t(x) = \sum_{\mathbf{k}:\,|\mathbf{k}| \in \mathbb{Q}_n} c_k e^{2\pi i \mathbf{k} \cdot x}\}$ is the family of trigonometric polynomials with frequencies in the hyperbolic cross $\mathbb{Q}_n$. Then almost the same argument that proves (191) also yields

$$E_{\mathbb{Q}_{n-3}}(f)_\infty \gtrsim \sum_{\mathbf{s} \in \mathbb{Y}_n} \big\| \delta_{\mathbf{s}}(f) \big\|_1 . \tag{194}$$

To draw a parallel with the Haar function version, the reader can check that (192) holds if the summation on the left-hand side is extended to include rectangles of size $|R| \geq 2^{-n}$ – the proof given in §4.3 need not even be changed.

Inequalities (191), (4.10) have been applied in [111, 112] to obtain estimates of entropy numbers and Kolmogorov widths of certain function classes with mixed smoothness. It was also shown in [112] that inequality (191) cannot hold unless $d = 2$. Moreover, it cannot even hold if we replace the L^∞ norm on the left by L^p,

$p < \infty$ or the L^1 norm on the right by L^q, $q > 1$. Analogously to Conjecture 5, we can formulate

Conjecture 11 (The trigonometric small ball conjecture). In dimensions $d \geq 2$, the following inequality holds

$$n^{\frac{d-2}{2}} \Big\| \sum_{\mathbf{s}\in\mathbb{Y}_n} \delta_{\mathbf{s}}(f) \Big\|_\infty \gtrsim \sum_{\mathbf{s}\in\mathbb{Y}_n} \big\| \delta_{\mathbf{s}}(f) \big\|_1 \tag{195}$$

The sharpness of (195) has been established in [112] by a probabilistic argument of the same flavor as the one presented in §4.2. For more information about these inequalities, their applications, and hyperbolic cross approximations the reader is invited to consult [111, 112] as well as the monographs [108, 109].

5 Higher dimensions

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. After all, this could be just an artifact of the method.

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 (116). 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 almost 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, at some points all functions have the same sign and $\Big\| \sum_{|R|=2^{-n}} \alpha_R h_R \Big\|_\infty \approx n^{d-1}$)

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 [8], 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, although he did not state this result and apparently was not aware of the connections. In turn, the fact that Beck's work implicitly contains progress on Conjectures 8 and 10 in dimension $d = 3$ eluded most of the experts in small deviation probabilities and metric entropy.

In 2008, largely building upon Beck's work and enhancing it with new ideas and methods, the author, M. Lacey, and A. Vagharshakyan [17], [18], obtained the first significant improvement over the 'trivial' estimate in all dimensions greater than two:

Theorem 9. *In all dimensions $d \geq 3$ there exists $\eta(d) > 0$ such that for all choices of coefficients we have the inequality:*

$$n^{\frac{d-1}{2}-\eta(d)} \Big\| \sum_{R:\,|R|=2^{-n}} \alpha_R h_R \Big\|_\infty \gtrsim 2^{-n} \sum_{R:\,|R|=2^{-n}} |\alpha_R|. \tag{196}$$

A modification of the argument to the discrepancy framework (in a way analogous to the one described in the previous section) was also used to obtain an improvement (10) over Roth's estimate (5) in all dimensions $d \geq 3$. (This theorem has already been stated in the introduction, see Theorem 2; we simply restate it here in order to show the whole spectrum of theorems obtained by the method.)

Theorem 10. *There exists a constant $\eta = \eta(d)$, such that in all dimensions $d \geq 3$, for any set $\mathscr{P}_N \subset [0,1]^d$ of N points, the discrepancy function satisfies*

$$\|D_N\|_\infty \gtrsim (\log N)^{\frac{d-1}{2}+\eta}. \tag{197}$$

The inequality (196) also directly translates into improved lower bounds of the small deviation probabilities for the Brownian sheet (Conjecture 10) and the metric entropy of the mixed derivative spaces (Conjecure 8).

Theorem 11. *There exists a constant $\theta = \theta(d)$, such that in all dimensions $d \geq 3$, the small ball probability for the Brownian sheet satisfies*

$$-\log \mathbb{P}(\|\mathbb{B}\|_{C([0,1]^d)} < \varepsilon) \gtrsim \frac{1}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^{2d-2+\theta}. \tag{198}$$

Theorem 12. *In dimensions $d \geq 3$, the metric entropy of the unit ball of $MW^2([0,1]^d)$ with respect to the L^∞ norm satisfies*

$$\log N(\varepsilon, 2, d) \gtrsim \frac{1}{\varepsilon}\left(\log\frac{1}{\varepsilon}\right)^{d-1+\theta/2}. \tag{199}$$

Due to the equivalence between the two problems, the value of $\theta = \theta(d)$ is the same in both theorems above.

Since complete technical details of the proof of (196), which can be found in [17, 18, 19] as well as Lacey's notes on the subject [66], would take up more space than the rest of this chapter, we shall simply present the main ideas of the argument and the heuristics behind them. An interested reader can then follow the complete proof in the listed references.

5.1 A short Riesz product

The Riesz product constructed in (124) for the proof of the two-dimensional small ball conjecture turns out to be just too long for a higher dimensional problem.

Consider a very simple example when all $\alpha_R > 0$ and the dimension d is even (or $\alpha_R < 0$ for odd d). If we take the same product as in (124), $\prod_{\mathbf{r}\in\mathbb{H}_n^d}(1+f_{\mathbf{r}})$ with $f_{\mathbf{r}} = \sum_{|R|=2^{-n}} \operatorname{sgn}(\alpha_R) h_R$, we can easily see that on the dyadic cube of sidelength $2^{-(n+1)}$ adjacent to the origin all the functions $f_{\mathbf{r}}$ are positive, hence all the factors of the Riesz product are equal to 2. Therefore,

$$\Big\| \prod_{\mathbf{r}\in\mathbb{H}_n^d} (1+f_{\mathbf{r}}) \Big\|_1 \gtrsim 2^{\#\mathbb{H}_n^d} \cdot 2^{-d(n+1)}. \tag{200}$$

This number becomes huge for large n as $\#\mathbb{H}_n^d \approx n^{d-1}$. Therefore, this construction does not stand a chance in dimensions $d \geq 3$.

Following the idea of Beck, the test function is constructed as a "short" Riesz product. For $\mathbf{r} \in \mathbb{H}_n^d$, we consider the r-functions $f_{\mathbf{r}} = \sum_{R\in\mathcal{D}_{\mathbf{r}}^d} \operatorname{sgn}(\alpha_R) h_R$. 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,..., I_q numbered in increasing order. Let $\mathbb{A}_j := \{\mathbf{r} \in \mathbb{H}_n^d \mid r_1 \in I_j\}$. Each group $\mathbb{A}_j$ then contains $\#\mathbb{A}_j \approx n^{d-1}/q$ vectors. 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 $\|\mathbf{r}\|_1 = n$. We construct the functions

$$F_j = \sum_{\mathbf{r}\in\mathbb{A}_j} f_{\mathbf{r}}. \tag{201}$$

Due to orthogonality, $\|F_j\|_2 \approx \sqrt{\#\mathbb{A}_j} \approx n^{(d-1)/2}/\sqrt{q}$. We now introduce the "false" L^2 normalization: $\widetilde{\rho} = aq^{1/4} n^{-(d-1)/2}$ ($a > 0$ is a small constant), whereas the "true" normalization would be somewhat larger, $\rho = \sqrt{q} n^{-\frac{d-1}{2}}$. We are now ready to define the Riesz product

$$\Psi := \prod_{j=1}^{q} (1+\widetilde{\rho} F_j). \tag{202}$$

Let us explain the effects that this construction creates and compare it to the two-dimensional Temlyakov's test function (124).

First of all, the grouping of r-functions by the values of the first coordinate mildly mirrors the construction of (124). 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. In particular, if $i < j$, then, in the first coordinate, the Haar functions involved in F_j are supported on intervals strictly smaller than those that support the Haar functions in F_i. It follows that for any $k \leq q$ and $1 \leq j_1 < j_2 < \ldots < j_k \leq q$

$$\int_{[0,1]^d} F_{j_1}(x) \cdot \ldots \cdot F_{j_k}(x) = 0, \tag{203}$$

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

$$\int_{[0,1]^d} \Psi(x)dx = 1, \tag{204}$$

as (203) implies that all the higher order terms have mean zero. By comparison, Beck's [8] 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 (124) 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 (202). This heuristic is quantified in (212).

However, we cannot take Ψ to be the 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 (123), problems arise when the rectangles supporting the Haar functions coincide in one of the coordinates, in other words, when for two vectors $\mathbf{r}$, $\mathbf{s} \in \mathbb{H}_n^d$ and for some $k = 1,\dots,d$, we have $r_k = s_k$. We say that a *coincidence* occurs in this situation. We say that vectors $\{\mathbf{r}_j\}_{j=1}^m \subset \mathbb{H}_n^d$ are *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 $r_{i,k} \neq r_{j,k}$. We can then write

$$\Psi = 1 + \Psi^{sd} + \Psi^{\neg sd}, \text{ where} \tag{205}$$

$$\Psi^{sd} = \sum_{k=1}^{q} \widetilde{\rho}^k \sum_{1\le j_1<j_2<\dots<j_k\le q} \left(\widetilde{\sum} f_{\mathbf{r}_{j_1}} \cdot \dots \cdot f_{\mathbf{r}_{j_k}} \right), \tag{206}$$

and the tilde above the innermost sum indicates that the sum is extended over all collections of vectors $\{\mathbf{r}_{j_t} \in \mathbb{A}_{j_t} : t = 1,\dots,k\}$ 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 don't.

5.2 *The Beck gain*

The function Ψ^{sd} is then taken to be the test function. Since all the coincidences are eliminated, the product rule (123) is applicable and an argument similar to (126)-(127) can be carried out, provided we can show that $\|\Psi^{sd}\|_1 \lesssim 1$.

5.2.1 Simple coincidences

An enormous part of of the proof of Theorem 9 in [17, 18] is devoted to the study of analytic and combinatorial aspects of coincidences, i.e. the behavior of $\Psi^{\neg sd}$. An important starting point is the following non-trivial lemma, which as a tribute to József Beck's ideas [8] we call the *Beck gain*:

Lemma 10 (Beck gain). *For every $p \geq 2$ we have the following inequality*

$$\Big\| \sum_{\substack{\mathbf{r}\neq\mathbf{s}\in\mathbb{H}_n^d \\ r_1=s_1}} f_{\mathbf{r}} \cdot f_{\mathbf{s}} \Big\|_p \lesssim p^{\frac{2d-1}{2}}\, n^{\frac{2d-3}{2}}. \tag{207}$$

The main aspect of this lemma is the precise power of n in the estimate. The exponent $\frac{2d-3}{2}$ is in fact very natural. Indeed, d-dimensional vectors $\mathbf{r}$ and $\mathbf{s}$ have d parameters each. The condition $\|\mathbf{r}\|_1 = \|\mathbf{s}\|_1 = n$ 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 in the sum is of the order of n^{2d-3} and (207) essentialy 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, see e.g. computation (227) and the discussion thereafter.

Another intuitive explanation may be given from the following point of view. It is not hard to see that

$$\begin{aligned} \Big\| \sum_{\mathbf{r}\neq\mathbf{s}\in\mathbb{H}_n^d} f_{\mathbf{r}} \cdot f_{\mathbf{s}} \Big\|_p &= \Big\| \Big(\sum_{\mathbf{r}\in\mathbb{H}_n^d} f_{\mathbf{r}} \Big)^2 - \sum_{\mathbf{r}\in\mathbb{H}_n^d} f_{\mathbf{r}}^2 \Big\|_p \\ &\leq \Big\| \sum_{\mathbf{r}\in\mathbb{H}_n^d} f_{\mathbf{r}} \Big\|_{2p}^2 + \#\mathbb{H}_n^d \approx n^{d-1}, \end{aligned} \tag{208}$$

since $\#\mathbb{H}_n^d \approx n^{d-1}$ and the L^{2p} norm of $F = \sum_{\mathbf{r}\in\mathbb{H}_n^d} f_{\mathbf{r}}$ is of the order $n^{\frac{d-1}{2}}$ as was shown in (88) using the Littlewood–Paley inequalities. Therefore, by imposing the condition $r_1 = s_1$ one *gains* $\sqrt{n}$ in the estimate, which explains the name that the authors have given to this 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 [8]. In his argument, in order to compute the L^2 norm, Beck expands the square of the sum:

$$\Big\| \sum_{\substack{\mathbf{r}\neq\mathbf{s}\in\mathbb{H}_n^d \\ r_1=s_1}} f_{\mathbf{r}} \cdot f_{\mathbf{s}} \Big\|_2^2 = \sum_{\substack{\mathbf{r}\neq\mathbf{s},\,\mathbf{u}\neq\mathbf{v} \\ r_1=s_1,\,u_1=v_1}} \int_{[0,1]^d} f_{\mathbf{r}} \cdot f_{\mathbf{s}} \cdot f_{\mathbf{u}} \cdot f_{\mathbf{v}}\, dx. \tag{209}$$

Notice that each integral above is equal to zero unless the four-tuple of vectors $(\mathbf{r},\mathbf{s},\mathbf{u},\mathbf{v}) \in (\mathbb{H}_n^d)^4$ has a coincidence in each coordinate. Careful and lengthy combinatorial analysis of the arising patterns of coincidences then leads to the desired inequality.

The extension and generalization obtained in [17, 18] is achieved by replacing the process of expanding the square by the applications of the Littlewood–Paley square function (66), 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 (68) yields a constant $B_p \approx \sqrt{p}$. The lemma was initally proved in $d = 3$ [17] and then extended to $d \geq 3$ [18] by a tricky induction argument. The reader is invited to see see [17, Lemma 8.2], [18, Lemma 5.2] for complete details.

5.2.2 Long coincidences

As we shall see, Lemma 10 is very powerful and yields important consequences, e.g. (212)-(213). Yet it is only a starting point in the analysis. 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

$$\Big\|\sum f_{\mathbf{r_1}} \cdot \ldots \cdot f_{\mathbf{r_k}}\Big\|_p \lesssim p^{\alpha M} n^{\frac{M}{2}}, \tag{210}$$

where the sum is extended over all k-tuples $\mathbf{r_1}$, ..., $\mathbf{r_k}$ 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.

These patterns of coincidences may be described by d-colored graphs $G = (V,E)$, where the set of vertices $V = \{1,...,k\}$ corresponds to vectors $\mathbf{r_1},...,\mathbf{r_k}$, and two vertices i and j are connected by an edge of color m, $m = 1,...,d$ if the vectors $\mathbf{r_i}$ and $\mathbf{r_j}$ have a coincidence in the m^{th} coordinate: $r_{i,m} = r_{j,m}$.

In the case of a single coincidence, when $k = 2$ and the graph describing the coincidence consists of two vertices and one edge, estimate (210) turns precisely into inequality (207) of Lemma 207. At present, inequality (210) in full generality is only a conjecture. In [17, 18] a partial result with a larger power of n is obtained for $k > 2$. Namely, it is proved that, if the summation is taken over a fixed pattern of coincidences of length k, the following estimate holds for some $\gamma > 0$

$$\Big\|\sum f_{\mathbf{r_1}} \cdot \ldots \cdot f_{\mathbf{r_k}}\Big\|_p \lesssim p^{Ck} n^{\left(\frac{d-1}{2}-\gamma\right)\cdot k}. \tag{211}$$

In other words, we have a gain proportional to the total length of the coincidence. This would later allow one to sum the estimates over all possible patterns of coincidences.

Roughly speaking, this inequality is proved by choosing a large *matching* (disjoint collection of edges) in the associated graph. Each edge in the matching corresponds to a simple coincidence to which an analog of the Beck gain lemma (207) may be applied, see [18, Theorem 8.3] for details. This approach, in particular, puts a restriction on the size of the gain. Consider, for example, a star-like graph with d edges of d distinct colors, which connect a single vertex (center) to d other vertices. The largest matching in such a graph consists of one edge. Therefore, in general, one cannot expect a matching of size more than k/d, which immediately yields $\gamma \lesssim 1/d$.

5.3 *The proof of Theorem 9*

In this subsection we shall outline the main steps and ideas of the proof of Theorem 9 based on the construction of the short Riesz product and the Beck gain.

The ultimate goal of constructing the short Riesz product Ψ (202) was to produce an L^1 test function Ψ^{sd}. The fact that Ψ^{sd} has bounded L^1 norm is proved through a series of estimates which are gathered in the following technical lemma (see Lemma 4.8 in [18]):

Lemma 11. *We have the following estimates:*

$$\mu(\{\Psi < 0\}) \lesssim \exp(-A\sqrt{q})\,; \tag{212}$$

$$\|\Psi\|_2 \lesssim \exp(a'\sqrt{q})\,; \tag{213}$$

$$\int \Psi(x)dx = 1\,; \tag{214}$$

$$\|\Psi\|_1 \lesssim 1\,; \tag{215}$$

$$\|\Psi^{\neg sd}\|_1 \lesssim 1\,; \tag{216}$$

$$\|\Psi^{sd}\|_1 \lesssim 1\,, \tag{217}$$

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

While we shall not give complete proofs of most of these inequalities, some remarks, explaining their nature and the main ideas, are in order.

We start with the first two inequalities (212)-(213) which are consequences of the Beck gain (207) for simple coincidences.

5.3.1 The distributional estimate (212)

Inequality (212) 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. Indeed, since $\Psi = \prod_{j=1}^{q}(1 + \widetilde{\rho}F_j)$, we have

$$\mu(\{\Psi < 0\}) \le \sum_{j=1}^{q} \mu(\{\widetilde{\rho}F_j < -1\}) = \sum_{j=1}^{q} \mu\big(\{\rho F_j < -\frac{1}{a}\sqrt[4]{q}\}\big), \tag{218}$$

where we have replaced the 'false' L^2 normalization $\widetilde{\rho} = aq^{1/4}/n^{(d-1)/2}$ by the 'true' one $\rho = \sqrt{q}/n^{\frac{d-1}{2}}$. Let us view the functions F_j as a sum of ± 1 random variables. If all of them were independent, we would be able to deduce estimate (212) immediately using the large deviation bounds of Chernoff-Hoeffding type, see e.g. (119), much in the same way as in (120). However, the presence of coincidences destroys independence. The Beck gain estimate (207) allows one to surpass this obstacle.

In fact, a weaker version of (212) can be proved without referring to the Beck gain. We have, for all $p > 1$,

$$\|\rho F_j\|_p = \frac{\sqrt{q}}{n^{\frac{d-1}{2}}} \Big\| \sum_{\mathbf{r}\in\mathbb{A}_j} f_{\mathbf{r}} \Big\|_p \lesssim \frac{\sqrt{q}}{n^{\frac{d-1}{2}}} p^{\frac{d-1}{2}} (\#\mathbb{A}_j)^{\frac{1}{2}} \lesssim p^{\frac{d-1}{2}}. \tag{219}$$

This estimate follows from successive applications of the Littlewood–Paley inequality (68) in the first $d-1$ coordinates (the last one is not needed due to the restriction $|R| = 2^{-n}$) and is identical to the calculation leading to (85). A constant of the order $\sqrt{p}$ arises each time we apply the square function. This shows, using the equivalent definitions of the exponential Orlicz norms (75), that $\|\rho F_j\|_{\exp(L^{2/(d-1)})} \lesssim 1$ and hence $\mu\big(\{\rho F_j < -\frac{1}{a}\sqrt[4]{q}\}\big) \lesssim \exp(-Cq^{1/2(d-1)})$.

To get the desired $\exp(L^2)$ bound, one would have to use Littlewood–Paley just once in order to get the constant of $p^{1/2}$ on the right-hand side. Therefore, the strategy to obtain the sharper inequality (212) is the following: we apply the Littlewood–Paley inequality to ρF_j just in the first coordinate. The "diagonal" terms yield a constant, while the rest of the terms are precisely the ones that have a coincidence in the first coordinate and are governed by the Beck gain. To be more precise, recall that $F_j = \sum_{\mathbf{r}:\, r_1\in I_j} f_{\mathbf{r}}$ and apply the Littlewood–Paley square function in the first coordinate

$$\begin{aligned} \|\rho F_j\|_p &\lesssim \sqrt{p}\|S_1(F_j)\|_p = \sqrt{p}\Bigg\| \Big(\sum_{t\in I_j} \rho^2 \big(\sum_{\mathbf{r}:\, r_1=t} f_{\mathbf{r}} \big)^2 \Big)^{1/2} \Bigg\|_p \\ &= \sqrt{p}\Bigg\| \rho^2 \sum_{\mathbf{r}\in\mathbb{A}_j} f_{\mathbf{r}}^2 + \rho^2 \sum_{\substack{\mathbf{r}\neq\mathbf{s}\in\mathbb{A}_j \\ r_1=s_1}} f_{\mathbf{r}}\cdot f_{\mathbf{s}} \Bigg\|_{p/2}^{1/2} \\ &\lesssim \sqrt{p}\left(1 + \rho^2 \Bigg\| \sum_{\substack{\mathbf{r}\neq\mathbf{s}\in\mathbb{A}_j \\ r_1=s_1}} f_{\mathbf{r}}\cdot f_{\mathbf{s}} \Bigg\|_{p/2}^{1/2} \right). \end{aligned} \tag{220}$$

The diagonal term above is bounded by a constant since $f_{\mathbf{r}}^2 = 1$ and $\rho^{-2} = n^{d-1}/q$ is roughly equal to the number of elements of $\mathbb{A}_j$. The Beck gain estimate (207) can be applied to the second term to obtain

$$\rho^2 \Big\| \sum_{\substack{\mathbf{r}\neq\mathbf{s}\in\mathbb{A}_j \\ r_1=s_1}} f_{\mathbf{r}} \cdot f_{\mathbf{s}} \Big\|_{p/2} \lesssim \frac{q}{n^{d-1}} p^{\frac{2d-1}{2}} n^{\frac{2d-3}{2}} = q p^{d-\frac{1}{2}} n^{-\frac{1}{2}} \lesssim 1 \qquad (221)$$

when p is not too big. Hence for relatively small values of p, the Beck gain term will not dominate over the diagonal term. For this range of exponents p we obtain $\|\rho F_j\|_p \lesssim \sqrt{p}$. This inequality for the full range of p by (75) would have implied $\|\rho F_j\|_{\exp(L^2)} \lesssim 1$. Even though this estimate cannot be deduced in full generality, repeating the proof of (75) we can find that $\mu\big(\{\rho F_j < -t\}\big) \lesssim \exp(-Ct^2)$ for moderate values of t, and (212) will follow. □

5.3.2 The L^2 bound (213)

An explanation for the L^2 bound (213) may again be given using the heuristics of probability theory. If F_j's were independent random variables, we would immediately have (213):

$$\begin{aligned} \int \prod_{j=1}^{q} (1+\widetilde{\rho} F_j)^2 \, dx &= \prod_{j=1}^{q} \int (1+\widetilde{\rho} F_j)^2 \, dx \qquad (222) \\ &\leq \prod_{j=1}^{q} (1+\widetilde{\rho}^2 \|F_j\|_2^2) \leq \left(1+\frac{a^2}{\sqrt{q}}\right)^q \leq e^{a^2\sqrt{q}}. \end{aligned}$$

While they are not independent, one can apply a conditional expectation argument and Beck gain (207), since the lack of independence is the result of coincidences.

We can see from the discussion of the first two conclusions of Lemma 11 that, from the probabilistic point of view, the Beck gain estimate (207) compensates for the lack of independence.

5.3.3 The integral and the L^1 norm of the Riesz product Ψ (214)-(215)

Equality (214) has already been explained, see (204). It follows from the fact that the functions F_j, $j = 1, \dots, q$ are orthogonal already in the first coordinate, since they consist of Haar functions of different frequencies.

Even though Ψ is not positive unlike in the two-dimensional case, the L^1estimate (215) easily follows from the previous three inequalities (212)-(214) using Cauchy–Schwarz inequality:

$$\begin{aligned} \|\Psi\|_1 &= \int \Psi(x)dx - 2\int_{\{\Psi<0\}} \Psi(x)dx \leq 1 + 2\mu(\{\Psi<0\})^{1/2} \cdot \|\Psi\|_2 \qquad (223) \\ &\lesssim 1 + \exp(-A\sqrt{q}/2 + a'\sqrt{q}) \lesssim 1. \quad \square \end{aligned}$$

5.3.4 The L^1 norm of coincidences (216)

Estimate (216) is the deepest part of this result and follows from the scrupulous analysis of coincidences which was outlined in §5.2, especially the bounds for long coincidences.

Recall that, as explained in §5.2.2, we describe long coincidences by d-colored graphs. Let the set of vertices be $V = V(G) \subset \{1,...,q\}$ and impose an additional condition that $s \in V(G)$ implies $\mathbf{r_s} \in \mathbb{A}_s$. This assumption reflects the way the vectors in the Riesz product are grouped. Denote by

$$SumProduct(G) := \sum \prod_{s \in V(G)} f_{\mathbf{r_s}}, \tag{224}$$

where the sum is extended over all tuples of vectors $\{r_s\}_{s\in V(G)}$ with $\mathbf{r_s} \in \mathbb{A}_s$ whose pattern of coincidences is described by the graph G. This is precisely the object whose norm is estimated in the Beck gain inequality for longer coincidences (211).

We can then represent the non-distinct part of the Riesz product Ψ as a sum over all possible configurations of coincidences as follows

$$\Psi^{\neg sd} = \sum_G \widetilde{\rho}^{|V(G)|}(-1)^{\mathrm{ind}(G)+1} \cdot SumProd(G) \cdot \prod_{s \notin V(G)} (1+\widetilde{\rho} F_j). \tag{225}$$

Here the sum is taken over all ‘admissible’ graphs – graphs that describe a realizable pattern of coincidences. The parameter $\mathrm{ind}(G)$ is simply a proper parameter needed in order to take care of the overlaps of different patterns of coincidences and to produce a correct version of the inclusion-exclusion formula. It is defined as the total number of equalities which describe the given arrangement of coincidences.

For a given pattern G, the factor $SumProd(G)$ gives all possible products arising from this pattern, while $\prod_{s\notin V(G)}(1+\widetilde{\rho}F_j)$ is the part of the Riesz product which is not involved in the given configuration of coincidences. Observe that in general the function $\prod_{s\notin V(G)}(1+\widetilde{\rho}F_j)$ satisfies more or less the same estimates as the full Riesz product Ψ itself, since it is of nearly identical form.

We shall interpolate between L^1 (215) and L^2 (213) estimates of the Riesz product $\prod_{s\notin V(G)}(1+\widetilde{\rho}F_j)$ to bound its L^p norm and find that, when p gets sufficiently close to 1, it is bounded by a constant. This is quite natural since its L^1 norm is bounded by a constant and it is a limit of L^p norms as p approaches 1. To be more precise, we take $p = (\sqrt{q})' = \frac{\sqrt{q}}{\sqrt{q}-1}$. In this case, $\frac{1}{p} = \frac{\sqrt{q}-2}{\sqrt{q}} \cdot 1 + \frac{2}{\sqrt{q}} \cdot \frac{1}{2}$. For the sake of brevity we denote $\Psi_{V(G)^c} := \prod_{s\notin V(G)}(1+\widetilde{\rho}F_j)$. We then obtain

$$\|\Psi_{V(G)^c}\|_{(\sqrt{q})'} \le \|\Psi_{V(G)^c}\|_1^{(\sqrt{q}-2)/\sqrt{q}} \cdot \|\Psi_{V(G)^c}\|_2^{2/\sqrt{q}} \lesssim 1 \cdot e^{a'\sqrt{q}\cdot\frac{2}{\sqrt{q}}} \lesssim 1. \tag{226}$$

We are now ready to estimate the L^1 norm of $\Psi^{\neg sd}$, the non-distinct part of Ψ. From (225) we have

$$\|\Psi^{\neg sd}\|_1 \le \sum_G \widetilde{\rho}^{|V(G)|} \|SumProd(G)\cdot \Psi_{V(G)^c}\|_1 \tag{227}$$

$$
\begin{aligned}
&\leq \sum_{G} \widetilde{\rho}^{|V(G)|} \|SumProd(G)\|_{\sqrt{q}} \cdot \|\Psi_{V(G)^c}\|_{(\sqrt{q})'} \\
&\lesssim \sum_{v=2}^{q} \sum_{G:|V(G)|=v} q^{\frac{v}{4}} n^{-\frac{d-1}{2}\cdot v} \cdot q^{Cv} n^{\left(\frac{d-1}{2}-\gamma\right)\cdot v} \cdot 1 \\
&\lesssim \sum_{v=2}^{q} \binom{q}{v} q^{vd} n^{(C'\varepsilon-\gamma)\cdot v} \leq \left(n^{-\gamma'}+1\right)^{q} - 1 \\
&\lesssim q n^{-\gamma'} \lesssim 1
\end{aligned}
$$

provided that ε is small enough. Here we have applied the Beck gain for long coincidences (211) to $SumProd(G)$ and the interpolation estimate (226) to $\Psi_{V(G)^c}$.

The number of admissible graphs with the given vertex set of v vertices can be controlled in the following way. Let us initially look at coincidences in a single coordinate. The number of ways to have a single coincidence is at most 2^v (every vertex either participates in a coincidence or not), for two coincidences the number of possibilities is at most 3^v (every vertex is in the first, second, or none of the coincidences), etc. Hence the total number of possibilities is no more than $2^v + 3^v + \dots + (v/2)^v \lesssim v^v$. If we now consider coincidences in all coordinates, the number of patterns is bounded by $(v^v)^d \leq q^{vd}$.

This computation reveals the motivation for some of the previously discussed estimates as well as the arising limitations.

- First of all, we see from the last two lines of the inequality that the amount of gain in (196) is forced to be bounded by the amount of the Beck gain (211). More precisely, we need $\varepsilon \lesssim \gamma$ in order to have $q^C \ll n^\gamma$. Moreover, the estimate on the total number of graphs $q^{vd} \approx n^{\varepsilon vd}$ suggests that $\varepsilon \lesssim \frac{1}{d}\gamma$. Since $\gamma \lesssim \frac{1}{d}$ as explained in §5.2.2, this tells us that the gain $\eta(d)$ in Theorem 9 coming from thus argument is at most $\varepsilon \lesssim 1/d^2$.
- Besides, we see that in order to bound the norm of $\Psi_{V(G)^c}$, the index $(\sqrt{q})'$ needs to be sufficiently close to 1, hence $\sqrt{q}$ is rather large. Therefore, it is really important to be able not only to estimate the norm of the terms involving coincidences $SumProd(G)$ in L^p spaces for $p \geq 2$, but also to be able to track how the implicit constants depend on the integrability index p.

The computation(227) finishes the proof of (216) and leaves us just one little step away from the proof of Theorem 9.

5.4 *The* L^1 *norm of* Ψ^{sd} (217) *and the conclusion of the proof*

Since $\Psi^{sd} = \Psi - 1 - \Psi^{\neg sd}$, the sought bound (217) is trivially implied by the previous two. We can now conclude the proof of Theorem 9 following the lines of (126)-(127):

$$\Big\| \sum_{R\in\mathscr{D}^d:\,|R|=2^{-n}} \alpha_R h_R \Big\|_\infty \gtrsim \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \Psi^{sd} \Big\rangle \tag{228}$$
$$= \Big\langle \sum_{|R|=2^{-n}} \alpha_R h_R, \widetilde{\rho} \sum_{|R|=2^{-n}} \operatorname{sgn}(\alpha_R) h_R \Big\rangle$$
$$= \widetilde{\rho} \sum_{R\in\mathscr{D}^d:\,|R|=2^{-n}} \alpha_R \cdot \operatorname{sgn}(\alpha_R) \cdot \|h_R\|_2^2$$
$$\approx n^{-\frac{d-1}{2}+\frac{\varepsilon}{4}}\, 2^{-n} \cdot \sum_{|R|=2^{-n}} |\alpha_R|,$$

so, (196) holds with $\eta = \varepsilon/4$. □

5.5 The signed small ball inequality

The *signed* small ball inequality, i.e. a version with $\alpha_R = \pm 1$ for each R, see (116) may be viewed as a toy model of Conjecture 5. It avoids numerous technicalities, while preserving most of the complications arising from the combinatorial complexity of the higher dimensional dyadic boxes. In [19], the same authors came up with a significant simplification of the arguments in [17, 18] for the signed case – it only required the simplest estimate for coincidences (207), and not the more complicated (210). It yielded the bound

$$\Big\| \sum_{|R|=2^{-n}} \alpha_R h_R \Big\|_\infty \gtrsim n^{\frac{d-1}{2}+\eta} \tag{229}$$

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 fact, given Lemma 11, it is quite easy to produce a proof of the "improvement of the L^2 estimate" (229), which is just (196) restricted to the signed case $\alpha_R = \pm 1$. We provide this proof below.

5.5.1 The proof of the signed version of Theorem 9

We shall use the same short Riesz product $\Psi = \prod_{j=1}^{q}(1+\widetilde{\rho} F_j)$ defined in (202). Recall that $F_j = \sum_{\mathbf{r}\in\mathbb{A}_j} f_{\mathbf{r}}$, where $\mathbb{A}_j := \{\mathbf{r} \in \mathbb{H}_n^d : \frac{n(j-1)}{q} \le r_1 < \frac{nj}{q}\}$, i.e. the first component of $\mathbf{r}$ lies in the j^{th} subinterval of $\{1,2,\dots,n\}$. Notice that in the signed case the expression inside the L^∞ norm in the small ball inequality simply equals the sum of all F_j's:

$$H_n = \sum_{|R|=2^{-n}} \alpha_R h_R = \sum_{j=1}^{q} F_j. \tag{230}$$

Unlike the general case, we can now take the product Ψ itself to be the dual test function, rather than extracting its "coincidence-free" part. The coincidences will be taken care of inside the argument. We first look at the inner product of a single F_j with Ψ. Denote by $\Psi_{\neq j} = \prod_{i=1, i\neq j}^{q} (1+\widetilde{\rho} F_i)$ the part of the Riesz product which consists of all factors except the j^{th} one. Note that its structure is virtually indistinguishable from that of the full product, hence $\Psi_{\neq j}$ satisfies essentially the same estimates as Ψ itself, see Lemma 11. Another important observation is that F_j is orthogonal to $\Psi_{\neq j}$: because of the structure of the product there are no coincidences in the first coordinate, thus, in the first component, F_j and Ψ_j consist of Haar functions of different frequencies. We then obtain

$$\begin{aligned} \langle F_j, \Psi \rangle &= \langle \sum_{\mathbf{r}\in \mathbb{A}_j} f_{\mathbf{r}}, \Psi \rangle = \sum_{\mathbf{r}\in \mathbb{A}_j} \langle f_{\mathbf{r}}, (1+\widetilde{\rho} F_j)\cdot \Psi_{\neq j} \rangle \\ &= \widetilde{\rho} \sum_{\mathbf{r}\in \mathbb{A}_j} \langle f_{\mathbf{r}}^2, \Psi_{\neq j} \rangle + \widetilde{\rho} \langle \Phi_j, \Psi_{\neq j} \rangle \\ &= \widetilde{\rho} (\# \mathbb{A}_j) + \widetilde{\rho} \langle \Phi_j, \Psi_{\neq j} \rangle, \end{aligned} \tag{231}$$

where Φ_j is exactly the expression arising in the Beck gain estimate (207)

$$\Phi_j = \sum_{\mathbf{r},\mathbf{s}\in \mathbb{A}_j, \mathbf{r}\neq \mathbf{s}, r_1 = s_1} f_{\mathbf{r}} \cdot f_{\mathbf{s}}. \tag{232}$$

The second line of the above computation (231) reflects the fact that the integral of $f_{\mathbf{r}}\Psi$ is equal to zero unless we get a coincidence in the first coordinate; and this coincidence may arise in two ways – when $f_{\mathbf{r}}$ hits itself (in which case, since $f_{\mathbf{r}}^2 = 1$ and $\int \Psi_{\neq j} = 1$, we simply pick up the cardinality of $\mathbb{A}_j$) or when it is paired with a different vector from $\mathbb{A}_j$ with the same first coordinate (so that the Beck gain (207) may be applied). We shall see that the former will be the main term in the estimate, while the latter may be treated as the error term.

Just as in (226), interpolation between L^1 (215) and L^2 (213) estimates of the Riesz product $\Psi_{\neq j}$ shows that the L^p norm $\|\Psi_{\neq j}\|_p$ is at most a constant. Copying (226), we need to take $p = (\sqrt{q})' = \frac{\sqrt{q}}{\sqrt{q}-1}$ to obtain

$$\|\Psi_{\neq j}\|_{(\sqrt{q})'} \le \|\Psi_{\neq j}\|_1^{(\sqrt{q}-2)/\sqrt{q}} \|\Psi_{\neq j}\|_2^{2/\sqrt{q}} \lesssim 1 \cdot e^{a'\sqrt{q}\cdot \frac{2}{\sqrt{q}}} \lesssim 1. \tag{233}$$

We can now apply Hölder's inequality, the Beck gain (207), and the previous inequality to estimate

$$|\langle \Phi_j, \Psi_{\neq j} \rangle| \le \|\Phi\|_{\sqrt{q}} \|\Psi_{\neq j}\|_{(\sqrt{q})'} \lesssim (\sqrt{q})^{d-1/2} n^{d-3/2} \cdot q^{-1/2}, \tag{234}$$

where the extra factor of $q^{-1/2}$ comes from the restriction $\mathbf{r}, \mathbf{s} \in \mathbb{A}_j$, which means that the parameter $r_1 = s_1 \in \left[\frac{n(j-1)}{q}, \frac{nj}{q}\right)$ can actually be chosen in n/q ways rather than n. Recalling that $\widetilde{\rho} \approx q^{1/4} n^{-(d-1)/2}$, $q \approx n^{\varepsilon}$, and $(\#\mathbb{A}_j) \approx n^{d-1}/q$, together with the fact that $\|\Psi\|_1 \lesssim 1$, we obtain

$$\|H_n\|_\infty \gtrsim |\langle H_n, \Psi \rangle| = \left| \sum_{j=1}^{q} \langle F_j, \Psi \rangle \right| \gtrsim q\widetilde{\rho} \left(\#\mathbb{A}_j - q^{\frac{d}{2}-\frac{3}{4}} n^{d-\frac{3}{2}} \right)$$
$$\approx q^{\frac{1}{4}} n^{\frac{d-1}{2}} - q^{\frac{d+1}{2}} n^{\frac{d-2}{2}} \gtrsim n^{\frac{d-1}{2}+\frac{\varepsilon}{4}}, \tag{235}$$

provided that ε is small enough, so that the second term is of smaller order of magnitude than the first one. This happens precisely when $\varepsilon < \frac{2}{d}$ which already yields the aforementioned restriction $\eta(d) \approx \frac{1}{d}$. An even more stringent condition on ε (yet still yielding the same rate of decay in terms of the dimension) arises from the proof of the L^2 estimate of Ψ (213).

The reader is reminded that estimate (215), $\|\Psi\|_1 \lesssim 1$, which is used in this proof, only relied on the L^2 bound (213), which in turn exploited only the simplest case of the Beck gain (207) for a single coincidence. In other words, one does not need to consider long coincidences – dealing with $\Psi^{\neg sd}$ can be avoided altogether. Thus the proof of estimate (196) in the signed case circumvents the heavy analytic and combinatorial investigation of coincidences and indeed allows for tremendous simplifications of the argument.

We close the discussion of the signed version of the small ball inequality by outlining two other potential points of view and approaches to the problem.

5.5.2 A new approach: independence and conditional expectation

In dimension $d = 2$, the signed small ball inequality (111) can be easily proved as a consequence of the independence of the random variables $f_{\mathbf{r}}$, which is easy to check. Independence implies

$$\mathbb{P}(f_{\mathbf{r}} = 1 : \mathbf{r} = (k, n-k),\, k = 0, ..., n) = \prod_{k=0}^{n} \mathbb{P}(f_{(k,n-k)} = 1) = \frac{1}{2^{n+1}} > 0, \tag{236}$$

i.e. on a set of positive measure the all the functions $f_{\mathbf{r}}$ are positive. On this set therefore

$$\sum_{|R|=2^{-n}} \varepsilon_R h_R(x) = \sum_{k=0}^{n} f_{(k,n-k)}(x) = n+1, \tag{237}$$

which proves Conjecture 6 in dimension $d = 2$.

In higher dimensions, due to possible coincidences in vectors $\mathbf{r} \in \mathbb{H}_n^d$, independence of the functions $f_{\mathbf{r}}$ no longer holds. This shortcoming can be partially compensated for by delicate conditional expectation arguments. The proof of the three-dimensional version of inequality (229) in [21] 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.

5.5.3 L^1 approximation

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 $\|H_n\|_2 \le \|H_n\|_1^{1/3} \cdot \|H_n\|_4^{2/3}$. It is easy to see that $\|H_n\|_2 \approx \|H_n\|_4 \approx n^{(d-1)/2} = n^{1/2}$ (the computation of the L^4 norm is identical to (88)). The estimate for the L^1 norm of H_n then follows.

Equality (125), on the orther hand, implies that the L^1 norm of $H_n - (-\Psi_{>n})$ is at most $1 + \|\Psi\|_1 = 2$, i.e. H_n, the hyperbolic sum of Haar functions 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 5 would follow if we can prove that for any choice of $\alpha_R = \pm 1$ we have

$$\mathrm{dist}_{L^1}\left(\sum_{R:\,|R|=2^{-n}} \alpha_R h_R,\, H_{>n}\right) \lesssim n^{\frac{d-2}{2}}, \tag{238}$$

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

These ideas are not new. In fact, in [107, 1980] (see also [108]), more than ten years prior to the proof of the small ball inequality (111) in dimension $d = 2$, Temlyakov has used a very similar Riesz product construction in order to prove an analog of the statement described above, namely, that trigonometric polynomials with frequencies in a hyperbolic cross (see §4.10) can be well approximated in the L^1 norm by a linear combination of harmonics of higher order.

6 Low discrepancy distributions and dyadic analysis

Most of the content of this chapter so far has been concerned with proofs of various lower bounds for the 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.

6.1 The van der Corput set

We recall a very standard construction, the so-called "digit-reversing" van der Corput set [39], also known as the Hammersley point set. This distribution of points is constructed in the following simple, yet very clever fashion. For $N = 2^n$ define a set $\mathcal{V}_n$ consisting of 2^n points

$$\mathcal{V}_n = \{(0.x_1x_2\ldots x_n, 0.x_nx_{n-1}\ldots x_1) : x_k = 0, 1; k = 1, \ldots, n\}, \tag{239}$$

where the coordinates are written as binary fractions. That means that the binary digits of the y-coordinate are exactly the digits of the x-coordinate written in the reverse order. Very roughly speaking, the effect of this construction is the following: if the x-coordinate changes a little, the y-coordinate changes significantly (although this is not exactly true), hence this set is well spreaded over the unit square.

Indeed, its star-discrepancy is optimal in the order of magnitude,

$$\|D_{\mathscr{V}_n}\|_\infty \leq n+1 \approx \log N. \tag{240}$$

This fact has been shown by van der Corput for the corresponding one-dimensional infinite sequence. Halton [53] and Hammersley [55] later transferred the idea to the multidimensional setting to construct the sets with the best currently known order of magnitude of the star-discrepancy, $(\log N)^{d-1}$.

A crucial property of the van der Corput set, which allows one to deduce such a favorable discrepancy bound is the fact that it forms a dyadic (or binary) net of order n: any dyadic rectangle R of area $|R| = 2^{-n}$ contains precisely one point of $\mathscr{V}_n$, and hence the discrepancy of $\mathscr{V}_n$ with respect to such rectangles is zero. For more information on nets, their constructions, and properties, the reader is referred to §3 of the chapter by J. Dick and F. Pillichshammer in the current book as well as [43].

6.1.1 The L^2 discrepancy of the van der Corput set

Different norms of the discrepancy function of variations of this set have been studied by many authors: [15],[20], [36], [39], [45], [54], [57], [61], [68], [86] to name just a few. We do not claim to present a complete survey of these results here – a comprehensive survey of numerous interesting properties of this elementary and at the same time wonderful set is yet to be written. Instead, we concentrate only on some estimates which we find most relevant to the theme of this chapter. Naturally, we shall start with the L^2 discrepancy.

It is well known that, while $\mathscr{V}_n$ 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

$$\int D_{\mathscr{V}_n}(x)dx = \frac{n}{8} + \mathscr{O}(1) \approx \log N \tag{241}$$

as observed in [54, 36, 15, 57]. Therefore, of course, $\|D_{\mathscr{V}_n}\|_2 \gtrsim \log N$.

One can look at this from a different point of view: (241) means that in any reasonable orthogonal (Haar, Fourier, wavelet, Walsh etc) decomposition of $D_{\mathscr{V}_n}$ the zero-order coefficient is already too big, so, by Plancherel's theorem, the L^2 norm is big. However, it turns out that the input of all the other coefficients is exactly of the right order, see e.g. [54, 20, 15], hence (241) is the *only* obstacle. Halton and Zaremba [54] showed that

$$\left\|D_{\mathscr{V}_n}\right\|_2^2 = \frac{n^2}{2^6} + \mathscr{O}(n), \tag{242}$$

which in conjunction with (241) proves this point.

There are several standard remedies which allow one to alter the van der Corput set so as to achieve the optimal order of the L^2 discrepancy. All of them, explicitly or implicitly, deal with reducing the quantity $\int D_{\mathscr{V}_n}(x)dx$. Here is a brief list of these methods.

(i) **Random shifts:**

Roth [86, 87] has demonstrated that there exists a shift of $\mathscr{V}_n$ modulo 1 which achieves optimal L^2 discrepancy. The proof was probabilistic: it was shown that the expectation over random shifts has the right order of magnitude

$$\mathbb{E}_\alpha \|D_{\mathscr{V}_{n,\alpha}}\|_2 \lesssim \sqrt{\log N}, \tag{243}$$

where $\mathscr{V}_{n,\alpha} = \big\{\big((x+\alpha) \bmod 1, y\big) : (x,y) \in \mathscr{V}_n\big\}$. A straightforward calculation shows that $\mathbb{E}_\alpha \int D_{\mathscr{V}_n}(x)dx = \mathscr{O}(1)$. A deterministic example of such a shift was constructed recently in [15].

(ii) **Symmetrization.**

This idea was introduced by Davenport [42] to construct the first example of a set with optimal order of L^2 discrepancy in dimension $d = 2$. His example was a symmetrized irrational lattice, i.e. $\left\{\left(\pm\frac{k}{N}, \{k\alpha\}\right)\right\}_{k=1}^{N}$, where α is an irrational number with bounded partial quotients of the continued fraction expansion and $\{x\}$ is the fractional part of x. Roughly speaking, the symmetrization 'cancels out' the zero-order term of the Fourier expansion of D_N. A similar idea was applied to the van der Corput set in [36].

(iii) **Digit shifts (digit scrambling).**

The method goes back to [54] in dimension $d = 2$ and [30] in higher dimensions. In the case of the van der Corput set it works extremely well and may be easily described. Fix an n-bit sequence of zeros and ones $\sigma = (\sigma_k)_{k=1}^n \in \{0,1\}^n$. We alter $\mathscr{V}_n$ as follows:

$$\begin{aligned}\mathscr{V}_n^\sigma = \big\{&(0.x_1x_2\dots x_n,\ 0.(x_n\oplus\sigma_n)(x_{n-1}\oplus\sigma_{n-1})\dots(x_1\oplus\sigma_1)) : \\ &x_k = 0,\ 1;\ k = 1,\dots,n\big\},\end{aligned} \tag{244}$$

where $\oplus$ denotes addition modulo 2. To put this definition into simple words, we can say that after flipping the digits, we also change some of them to the opposite (we 'scramble' or 'shift' precisely those digits for which $\sigma_k = 1$).

This procedure has been thoroughly studied for the van der Corput set. It is well known that it improves its distributional qualities in many different senses [61, 45]. In particular, when approximately half of the digits are shifted, i.e. $\sum \sigma_k \approx \frac{n}{2}$, this set has optimal order of magnitude of the L^2 discrepancy [54, 62, 20].

There is a natural explanation for this phenomenon which continues the line of reasoning started by (241). If one views the digits x_i as independent $0-1$ random variables and tries to compute the quantity $\int D_{\mathscr{V}_n}(x)dx$, one inevitably encounters expressions of the type $\mathbb{E}x_i \cdot x_j$. And while for $i \neq j$ we obtain $\mathbb{E}x_i \cdot x_j = \frac{1}{4}$, in the 'diagonal' case this quantity is twice as big, $\mathbb{E}x_i^2 = \frac{1}{2}$. And this occurs $n = \log_2 N$ times which leads to the estimate (241). However, if the digit x_i is scrambled, we have $\mathbb{E}x_i \cdot (1-x_i) = 0$. Therefore, one should scramble approximately one half of all digits in order to compensate for the 'diagonal' effect. The details are left to the reader and can be also found in the aforementioned references.

The nice dyadic structure of this set makes it perfectly amenable to the methods of harmonic analysis. For example, in [15] it is analyzed using Fourier series, in [68, 35, 36, 43] the authors exploit Walsh functions (the Walsh analysis of the van der Corput sets is nicely described in the chapter by W. Chen and M. Skriganov in the current volume), while the estimates in [20, 57] are based on the Haar coefficients of $D_{\mathscr{V}_n}$. We shall focus on the latter results as they directly relate to Roth's method, the main topic of our chapter, and complement previously discussed lower bounds.

6.1.2 Discrepancy of the van der Corput set in other function spaces

It has been shown in [20] that the BMO (107) and $\exp(L^\alpha)$ (109) lower estimates in dimension $d=2$, which we presented in §3.4 are sharp. In particular, for the digit-shifted van der Corput set $\mathscr{V}_n^\sigma$ with $\sum \sigma_k \approx \frac{n}{2}$ and for $\alpha \geq 2$ we have

$$\|D_{\mathscr{V}_n^\sigma}\|_{\exp(L^\alpha)} \lesssim (\log N)^{1-\frac{1}{\alpha}}. \tag{245}$$

In the case of the BMO norm, the standard van der Corput set satisfies

$$\|D_{\mathscr{V}_n}\|_{\mathrm{BMO}} \lesssim \sqrt{\log N}. \tag{246}$$

These inequalities were based on estimates of the Haar coefficients of the discrepancy function, namely

$$|\langle D_{\mathscr{V}_n^\sigma}, h_R\rangle| \lesssim \min\left\{1/N,\, |R|\right\}. \tag{247}$$

This estimate for small rectangles is straightforward. The counting and linear part can be bounded separately. The estimate for the counting part relies on the fact that $\mathscr{V}_n^\sigma$ is a dyadic net and thus there cannot be too many points in a small dyadic box, while the contribution of the linear part as computed in (33) is of the order

$N|R|^2 \lesssim |R|$. In turn, coefficients corresponding to large rectangles involve subtle cancellations suggested by the structure and self-similarities of $\mathcal{V}_n^{\sigma}$. We point out that, in accordance with Roth's principle (24), the cutoff between 'small' and 'large' rectangles occurs at the scale $|R| \approx \frac{1}{N}$. The BMO and $\exp(L^{\alpha})$ can then be obtained by applying arguments of Littlewood–Paley type.

Almost simultaneously to these results, the Besov norm of the same digit-shifted van der Corput set has been estimated using a very similar method in [57], see also [76]. In fact, this work went much further: all the Haar coefficients of $D_{\mathcal{V}_n^{\sigma}}$ have been computed exactly. This led to showing that the lower Besov space estimate (101) of Triebel [115] is sharp in $d = 2$, more precisely

$$\|D_{\mathcal{V}_n^{\sigma}}\|_{S_{pq}^r B([0,1]^d)} \lesssim N^r \left(\log N\right)^{\frac{1}{q}} \tag{248}$$

for $1 \leq p, q \leq \infty$, $0 \leq r < \frac{1}{p}$.

6.1.3 The structure of the Riesz product and the van der Corput set

We close our discussion of the van der Corput set with an amusing observation which pinpoints yet another connection between the small ball inequality (111) and discrepancy.

Consider the two-dimensional case of the small ball inequality and assume that all the coefficients α_R are non-negative. Recall Temlyakov's test function (124): $\Psi = \prod_{k=1}^{n} \left(1 + f_k\right)$. In this case, since $\mathrm{sgn}(\alpha_R) = +1$, the r-functions $f_k = \sum_{|R|=2^{-n}, |R_1|=2^{-k}} h_R$ are actually Rademacher functions. As explained in the very beginning of §4.6, the Riesz product Ψ captures the set where all the functions f_k are positive. To be more precise, $\Psi = 2^{n+1} \mathbf{1}_E$, where $E = \{x \in [0,1]^2 : f_k(x) = +1, k = 0, 1, \ldots, n\}$.

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, either both 0, or both 1. Therefore, $(x_1, x_2) \in E$ when this holds for all $k = 0, 1, \ldots, n$, 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 $\mathcal{V}_{n+1}$! Therefore

$$E = \mathcal{V}_{n+1} + [0, 2^{-(n+1)}) \times [0, 2^{-(n+1)}), \tag{249}$$

i.e. the Riesz product, which produces the proof of the small ball conjecture (5), is essentialy supported on the standard van der Corput set. Notice also that replacing f_k by $-f_k$ results in 'scrambling' the k^{th} digit in the van der Corput set.

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**, 266–271 (1945)
2. van Aardenne-Ehrenfest, T.: On the impossibility of a just distribution. Nederl. Akad. Wetensch. Proc. **52**, 734–739 (1949); Indagationes Math. **11**, 264–269 (1949)
3. Anderson, T. W.: The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. **6**, 170–176 (1955)
4. Bass, R.F.: Probability estimates for multiparameter Brownian processes. Ann. Probab. **16** 1, 251–264 (1988)
5. Beck, J.: Balanced two-colorings of finite sets in the square. I. Combinatorica **1** 4, 327–335 (1981)
6. Beck, J.: Irregularities of distribution. I. Acta Math. **159** 1-2, 1–49 (1987)
7. Beck, J.: Irregularities of distribution. II. Proc. London Math. Soc. (3) **56** 1, 1–50 (1988)
8. Beck, J.: A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution. Compositio Math. **72** 3, 269–339 (1989)
9. Beck, J.: The modulus of polynomials with zeros on the unit circle: a problem of Erd ös. Ann. of Math. (2) **134**, 609–651 (1991)
10. Beck, J., Chen, W. W. L.: Note on irregularities of distribution. Mathematika **33**, 148–163 (1986)
11. Beck, J., Chen, W.W.L.: Irregularities of distribution. Cambridge University Press, Cambridge (1987)
12. Beck, J., Chen, W. W. L.: Note on irregularities of distribution. II. Proc. London Math. Soc. (3) **61** 2, 251–272 (1990)
13. Béjian, R.: Minoration de la discrépance d'une suite quelconque sur T. Acta Arith. **41** 2, 185–202 (1982)
14. Bernard, A.: Espaces H1 de martingales à deux indices. Dualité avec les martingales de type "BMO". Bull. Sci. Math. (2) **103** 3, 297–303
15. Bilyk, D.: Cyclic shifts of the van der Corput set. Proc. Amer. Math. Soc. **137** 8, 2591–2600 (2009)
16. Bilyk, D.: Uniform Distribution Theory **6** 1, 143–184 (2011)
17. Bilyk, D., Lacey, M.: On the small ball Inequality in Three Dimensions. Duke Math. J. **143** 1, 81–115 (2008)
18. Bilyk, D., Lacey, M., Vagharshakyan, A.: On the small ball inequality in all dimensions. J. Funct. Anal. **254** 9, 2470-2502 (2008)
19. Bilyk, D., Lacey, M., Vagharshakyan, A.: On the signed small ball inequality. Online J. of Analytic Combinatorics **3**, (2008)
20. Bilyk, D., Lacey, M., Parissis, I., Vagharshakyan, A..: Exponential squared integrability of the discrepancy function in two dimensions. Mathematika **55** 1-2, 1–27 (2009)
21. Bilyk, D., Lacey, M., Parissis, I., Vagharshakyan, A.: A three-dimensional signed small ball inequality. In: Dependence in probability, analysis and number theory, pp. 73-87. Kendrick Press, Heber City, UT, 2010 (2010)
22. Bogachev, V.I.: Gaussian Measures. American Mathematical Society. (1998)
23. Burkholder, D.L.: Sharp inequalities for martingales and stochastic integrals. Astérisque **157-158**, 75–94 (1988)
24. Chang, S.-Y. A., Fefferman, R.: A continuous version of duality of H^1 with BMO on the bidisc. Ann. of Math. (2) **112** 1, 179–201 (1980)
25. 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** 1, 1– 43 (1985)
26. Chang, S.-Y. A., Wilson, J. M., Wolff, T. H.: Some weighted norm inequalities concerning the Schrödinger operators. Comment. Math. Helv. **60** 2, 217– 246 (1985)
27. Chazelle, B.: The Discrepancy Method. Randomness and complexity. Cambridge University Press., Cambridge (2000)

28. Chazelle, B.: Complexity bounds via Roth's method of orthogonal functions.In: Analytic number theory, pp. 144–149. Cambridge Univ. Press, Cambridge (2009)
29. Chen, W. W. L.: On irregularities of distribution. Mathematika. **27** 2, 153–170 (1981)
30. Chen, W. W. L.: On irregularities of distribution. II. Quart. J. Math. Oxford Ser. (2) **34** 135, 257–279 (1983)
31. Chen, W. W. L.: On irregularities of distribution. III. J. Austral. Math. Soc. Ser. A **60** 2, 228–244 (1996)
32. Chen, W. W. L.: On irregularities of distribution. IV J. Number Theory **80** 1, 44–59 (2000)
33. Chen, W. W. L.: Fourier techniques in the theory of irregularities of point distribution. In: Fourier analysis and convexity, Appl. Numer. Harmon. Anal., pp. 59–82. Birkhäuser, Boston (2004)
34. Chen, W. W. L.: private communication. Palo Alto, CA (2008)
35. 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**, 67–95 (2002)
36. Chen, W. W. L., Skriganov, M.M.: Davenport's theorem in the theory of irregularities of point distribution. J. Math. Sci. (N. Y.) **115** 1, 2076–2084 (2003)
37. Chen, W.W.L., Travaglini, G.: Some of Roth's ideas in discrepancy theory. In: Analytic number theory, pp. 150–163. Cambridge Univ. Press , Cambridge (2009)
38. van der Corput, J. G.: Verteilungsfunktionen I. Akad. Wetensch. Amdterdam, Proc. **38**, 813–821 (1935)
39. van der Corput, J. G.: Verteilungsfunktionen II. Akad. Wetensch. Amdterdam, Proc. **38**, 1058– 1068 (1935)
40. Csáki, E.: On small values of the square integral of a multiparameter Wiener process. In: Statistics and probability (Visegrád, 1982), pp. 1926. Reidel, Dordrecht (1984)
41. Daubechies, I.: Ten lectures on wavelets. SIAM, Philadelphia. (1992)
42. Davenport, H.: Note on irregularities of distribution. Mathematika **3**, 131–135 (1956)
43. Dick, J., Pillichshammer, F.: Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration. Cambridge University Press, Cambridge (2010)
44. Dunker, T., Kühn, T., Lifshits, M., Linde, W.: Metric entropy of the integration operator and small ball probabilities for the Brownian sheet. C. R. Acad. Sci. Paris Sér. I Math. **326** 3, 347–352 (1998)
45. Faure, H., Pillichshammer, F.: L_p discrepancy of generalized two-dimensional Hammersley point sets. Monatsh. Math. **158** 1, 31–61 (2009)
46. Fefferman, R., Pipher, J.: Multiparameter operators and sharp weighted inequalities. Amer. J. Math. **119** 2, 337–369 (1997)
47. Feller, W.: An Introduction to Probability Theory and Its Applications. Wiley, New York (1966)
48. Frolov, K.K.: Upper bounds on the discrepancy in L_p, $2 \le p < \infty$. Dokl. Akad. Nauk USSR **252**, 805–807 (1980)
49. Gilbert, E. N.: A comparison of signalling alphabets. Bell System Technical Journal **3**, 504–522 (1952)
50. Grafakos, L.: Classical and Modern Fourier Analysis. Prentice Hall, Upper Saddle River, NJ (2003)
51. Haar, A.: Zur Theorie der orthogonalen Funktionensysteme. Math. Annal. **69**, 331–371 (1910)
52. Halász, G.: On Roth's method in the theory of irregularities of point distributions. In: Recent progress in analytic number theory, Vol. 2, pp. 79–94. Academic Press, London (1981)
53. Halton, J.H.: On the efficiency of certain quasirandom sequences of points in evaluating multidimensional integrals. Num. Math. **2**, 84– 90 (1960)
54. Halton, J.H., Zaremba, S.C.: The extreme and L^2 discrepancies of some plane sets. Monatsh. Math. **73**, 316– 328 (1969)
55. Hammersley, J.M.: Monte Carlo methods for solving multivariable problems. Ann. New York Acad. Sci. **86** 844– 874 (1960)
56. Heinrich, S.: Some open problems concerning the star-discrepancy. J. Complexity **19** 1, 416–419 (2003)

57. Hinrichs, A.: Discrepancy of Hammersley points in Besov spaces of dominating mixed smoothness. Math. Nachr. **283** 3, 478-488 (2010)
58. Hinrichs, A., Markhasin, L.: On lower bounds for the L_2 discrepancy. Journal of Complexity, **27** 2, 127–132 (2011)
59. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. **58**, 13–30 (1963)
60. Katznelson, Y.: An introduction to harmonic analysis. Cambridge University Press, Cambridge (2004)
61. Kritzer, P.: On some remarkable properties of the two-dimensional Hammersley point set in base 2. J. Théor. Nombres Bordeaux **18** 1, 203–221 (2006)
62. Kritzer, P., Pillichshammer, F.: An exact formula for the L_2 discrepancy of the shifted Hammersley point set. Unif. Distrib. Theory **1** 1, 1–13 (2006)
63. Kuelbs, J., Li, W. V.: Metric entropy and the small ball problem for Gaussian measures, J. Funct. Anal. **116** 1, 133– 157 (1993)
64. Kuipers, L., Niederreiter, H.: Uniform distribution of sequences John Wiley & Sons, New York-London-Sydney (1974)
65. Lacey, M.: On the discrepancy function in arbitrary dimension, close to L^1. Analysis Math. **34** 2, 119–136 (2008)
66. Lacey, M.: Small Ball and Discrepancy Inequalities (2008). Available on www.arxiv.org.
67. Lévy, P.: Problèmes concrets d'analyse fonctionelle. Gautheir-Villars (1951)
68. Larcher, G., Pillichshammer, F.: Walsh series analysis of the L_2-discrepancy of symmetrisized point sets. Monatsh. Math. **132** 1, 1–18, (2001)
69. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: Isoperimetry and Processes. Springer, New York (1991)
70. Lerch, M.: Question 1547. L'Intermediaire Math. **11**, 144–145 (1904)
71. Li, W. V., Shao, Q.-M.: Gaussian processes: inequalities, small ball probabilities and applications. In: Handbook of Statist., **19**, pp. 533–597. North-Holland, Amsterdam. (2001)
72. Liardet, P.: Discrépance sur le cercle. Primaths. I, Univ. Marseille (1979)
73. Lindenstrauss, J., Tzafriri, L.: Classical Banach spaces. I. Springer-Verlag, Berlin (1977)
74. Lifshits, M.A.: Gaussian Random Functions. Kluwer (1995)
75. Matoušek, J.: Geometric Discrepancy: An Illustrated Guide. Springer-Verlag, Berlin (1999)
76. Markhasin, L.: Quasi-Monte Carlo methods for integration of functions with dominating mixed smoothness in arbitrary dimension, preprint (2012). Available on www.arxiv.org.
77. Montgomery, H. L.: Ten lectures on the interface between analytic number theory and harmonic analysis (CBMS regional conference series in mathematics No. 84) Amer. Math. Soc., Providence (1994)
78. Muckenhoupt, B.: Weighted norm inequalities for the Hardy maximal function. Trans. Amer. Math. Soc. **165**, 207–226 (1972)
79. Novak, E., Woźniakowski, H.: Tractability of Multivariate Problems: Volume II: Standard Information for Functionals. EMS Tracts in Mathematics (2010)
80. Ou, W.: Irregularity of distributions and multiparameter A_p weights. Uniform Distribution Theory **5** 2, 131–139 (2010)
81. Pereyra, C.: Lecture notes on dyadic harmonic analysis. In: Pérez-Esteva, S., Villegas-Blas, C.: Second Summer school in analysis and mathematical physics. Topics in analysis: harmonic, complex, nonlinear and quantization, pp.1–61. Contemporary Mathematics 289 AMS (2001)
82. Pipher, J.: Bounded double square functions. Ann. Inst. Fourier (Grenoble) **36** 2, 69–82 (1986)
83. Polya, G.: How to solve it. 2nd ed. Princeton University Press, Princeton, NJ (1957)
84. Riesz, F.: Über die Fourierkoeffizienten einer stetigen Funktion von beschränkter Schwankung. Math. Z. **2** 3-4, 312–315 (1918)
85. Roth, K. F.: On irregularities of distribution. Mathematika **1**, 73–79 (1954)
86. Roth, K. F.: On irregularities of distribution. II. Comm. Pure Appl. Math. **29** 6, 739–744 (1976)

87. Roth, K. F.: On irregularities of distribution. III. Acta Arith. **35**, 373–384 (1979)
88. Roth, K. F.: On irregularities of distribution. IV. Acta Arith. **37**, 67–75 (1980)
89. Roth, K. F.: On a theorem of Beck. Glasgow Math. J. **27**, 195–201 (1985)
90. Schmidt, W.M.: Irregularities of distribution. Quart. J. Math. Oxford Ser. (2) **19**, 181–191 (1968)
91. Schmidt, W.M.: Irregularities of distribution. II. Trans. Amer. Math. Soc. **136**, 347–360 (1969)
92. Schmidt, W.M.: Irregularities of distribution. III. Pacific J. Math. **29**, 225–234 (1969)
93. Schmidt, W.M.: Irregularities of distribution. IV. Invent. Math. **7**, 55–82 (1969)
94. Schmidt, W.M.: Irregularities of distribution. V. Proc. Amer. Math. Soc. **25**, 608–614 (1970)
95. Schmidt, W.M.: Irregularities of distribution. VI. Compositio Math. **24**, 63–74 (1972)
96. Schmidt, W.M.: Irregularities of distribution. VII. Acta Arith. **21**, 45–50 (1972)
97. Schmidt, W.M.: Irregularities of distribution. VIII. Trans. Amer. Math. Soc. **198**, 1–22 (1974)
98. Schmidt, W.M.: Irregularities of distribution. IX. Acta Arith. **27**, 385–396 (1975)
99. Schmidt, W.M.: Irregularities of distribution. X. In: Number theory and algebra, pp. 311–329. Academic Press, New York (1977)
100. Sidon, S.: Verallgemeinerung eines Satzes über die absolute Konvergenz von Fourierreihen mit Lücken. Math. Ann. **97**, 675–676 (1927)
101. Sidon, S.: Ein Satz über trigonometrische Polynome mit Lücken und seine Anwendung in der Theorie der Fourier-Reihen. J. Reine Angew. Math. **163**, 251–252 (1930)
102. Skriganov, M.M.: Harmonic analysis on totally disconnected groups and irregularities of point distributions. J. Reine Angew. **600**, 25–49 (2006)
103. Skriganov, M.M.: private communication. Palo Alto (2008)
104. Sobol, M.I.: Multidimensional Quadrature Formulas and Haar Functions. Nauka, Moscow (1969) (in Russian).
105. Stein, E.M.: Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals. Princeton University Press, Princeton, NJ (1993)
106. Talagrand, M.: The small ball problem for the Brownian sheet. Ann. Probab. **22** 3, 1331–1354 (1994)
107. Temlyakov, V.N.: Approximation of periodic functions of several variables with a bounded mixed difference. Math. USSR Sb., **41**, 53–66 (1982)
108. Temlyakov, V.N.: Approximation of functions with bounded mixed derivative. Proc. Steklov Inst. Math. **1 (178)**, (1989)
109. Temlyakov, V.N.: Approximation of periodic functions. Nova Science, New York (1994)
110. Temlyakov, V.N.: Some Inequalities for Multivariate Haar Polynomials. East Journal on Approximations **1** 1, 61 –72 (1995)
111. Temlyakov, V.N.: An inequality for trigonometric polynomials and its application for estimating the entropy numbers. J. Complexity **11** 2, 293– 307 (1995)
112. Temlyakov, V.N.: An Inequality for Trigonometric Polynomials and its Application for Estimating the Kolmogorov Widths. East Journal on Approximations **2**, 253–262 (1996)
113. Temlyakov, V. N.: On two problems in the multivariate approximation. East J. Approx. **4** 4, 505–514 (1998).
114. Temlyakov, V.N.: Cubature formulas, discrepancy, and nonlinear approximation. J. Complexity **19** 3, 352 – 391 (2003)
115. Triebel, H.: Numerical integration and discrepancy, a new approach. Math. Nachr. **283** 1, 139–159 (2010)
116. Triebel, H.: Bases in Function Spaces, Sampling, Discrepancy, Numerical Integration. European Math. Soc. Publishing House, Zürich (2010)
117. Varshamov, R.R.: Estimate of the number of signals in error correcting codes. Dokl. Acad. Nauk SSSR **117**, 739–741 (1957)
118. Wang, G.: Sharp square-function inequalities for conditionally symmetric martingales. Trans. Amer. Math. Soc. **328** 1, 393–419 (1991)
119. Zygmund, A.: Trigonometric series. Vol. I, II. Cambridge University Press, Cambridge (2002)