Recent Work in Discrepancy Theory

Lilian Matthiesen

Abstract

We describe recent work of D. Bilyk and M. T. Lacey that provides an improvement to the previously best known lower bound on the sup-norm of the box discrepancy function in three dimensions. While Bilyk and Lacey make use of the ℓ_2-valued Littlewood-Paley inequality, we show that the real-valued Littlewood-Paley is sufficient to prove the result. We also give a different proof to the last argument that is needed in the 'conclusion of the proof'.

The first section is concerned with the proof of the L_1-norm bound of the short Riesz product that forms the heart of the proof.

The second section describes how to replace the ℓ_2-valued by the real-valued Littlewood-Paley inequality in the proof.

The third section contains an outline of the proof of the actual discrepancy theory result.

The fourth provides a complete proof of the real-valued Littlewood-Paley inequality.

Contents

1 Introduction

We are mainly interested in the discrepancy theory result that can be deduced from the Riesz Product Bilyk and Lacey consider in [2].

Building up on ideas of Roth and Halász, Haar functions are used to construct the Riesz product. One-dimensional Haar functions are defined with respect to dyadic intervals as follows: Let $I \subset [0,1]$ be a dyadic interval, then

$$h_I : [0,1] \to \mathbb{R}; \quad h_I = -\chi_{I_{left}} + \chi_{I_{right}} \ .$$

For any axis-parallel box $R = I_1 \times \cdots \times I_d$ in the d-dimensional unit cube all sides of which are dyadic (let's call it a dyadic box) we define

$$h_R : [0,1]^d \to \mathbb{R}; \quad h_R(x) = \prod_{j=1}^{d} h_{I_j}(x_j).$$

What makes these Haar functions so special? Together with $\chi_{[0,1]}$, the one-dimensional Haar functions define an orthogonal basis for the L_2-functions on $[0,1]$. More importantly we can view them as independent identically distributed random variables in the following sense:

Let $f_r = \sum_{I \in \mathcal{D}_r} \varepsilon_I h_I$, where $\mathcal{D}_r$ denotes the dyadic intervals of length 2^{-r}, and $\varepsilon_I = \pm 1$. Then the f_r form a family of Rademacher random variables: they are independent, identically distributed and $\mathbb{P}(f_r = 1) = \mathbb{P}(f_r = -1) = \frac{1}{2}$. Viewing the f_r as random variables provides a way to intuitively understand how this Riesz product proof works. Haar functions proved to be useful in obtaining bounds on discrepancy functions. They, for example, let us split the unit cube into small rectangles to consider local discrepancies. Halasz found a tight lower bound on the L_1-norm of the 2-dimensional discrepancy function, using Riesz product.

His Riesz product has a factor for each r-function (generalizations of the above to higher dimensions). Since the number of isomorphism types of boxes grows exponentially in the dimension (it is about n^{d-1}), the number of terms in the expansion of the product gets very large. Proofs of the discrepancy results via Hölder's inequality and a Riesz product as test function require an upper bound on an appropriate norm of the Riesz product. Since this is achieved by bounding each term in the expansion individually, we run into problems in higher dimensions.

To hold against this accumulation of terms, Beck introduced a short Riesz product, considering sums of r-functions at once instead of individual r-functions in each factor.

2 Small Ball Inequality - Main Line of Argument

The main theorem in Bilyk and Lacey [2] is the small ball inequality:

Theorem 1 (1.5 in [2])**.** *In dimension 3 there is $\eta > 0$ such that*

$$2^{-n} \sum_{|R|=2^{-n}} |\alpha(R)| \ll n^{1-\eta} \left\| \sum_{|R|=2^{-n}} \alpha(R) h_R \right\|_\infty .$$

The relation of this inequality to the discrepancy result lies in the proofs: both can be proven using Hölder's inequality with similar Riesz products as test functions. The main part—which both proofs share—is to establish an L_1-norm bound of the Riesz product.

Example and conjectural tight bound. Let us consider an example to get an idea of what this theorem says: Let the coefficients $\alpha(R)$ be random choices of ± 1. Then the left hand side is at most $n^{d-1} = n^2$, since every point belongs to about n^{d-1} (half-open) dyadic rectangles of volume 2^{-n}, and there are 2^n rectangles of each isomorphism type. By the central limit theorem we have for fixed x

$$\mathbb{E} \left| \sum_{|R|=2^{-n}} \alpha(R) h_R(x) \right| \simeq \sqrt{n^{d-1}} .$$

So there has to be a small rectangle (of volume $2^{-d(n-1)}$; this is the smallest volume that occurs as intersection of dyadic rectangles of volume 2^{-n}) for which the sum $\sum_{|R|=2^{-n}} \alpha(R) h_R$ is rather large. The theorem says that there necessarily appear large deviations. One may ask how large they have to be. This is the question for a tight bound in the theorem.

It is possible to give an upper bound on the RHS in the special case considered above[1]:

We need the following two theorems (3.1.3 and 3.2.1 in [3]). The first will be used to find a bound on the sup-norm.

Theorem 2. *Let $X_1, \dots, X_n \in \exp(L^2)$ be functions of $\exp(L^2)$-norm at most one, then*

$$\mathbb{E} \sup_{n \le N} |X_n| \ll \sqrt{\log N + 1} .$$

[1] cf. [3, p.10].

The second, Khintchine's inequality, will be used to ensure that the conditions of the first theorem hold in our application.

Theorem 3 (Khintchine's inequality)**.** *Let $\{r_k\}_{k\in\mathbb{N}}$ be a family of Rademacher variables. Then for all finite sequences of coefficients $\{a_k\}$*

$$\left\|\sum_k a_k r_k\right\|_{\exp(L^2)} \ll \left(\sum_k a_k^2\right)^{1/2} .$$

As noted above, the sup-norm $\left\|\sum_{|R|=2^{-n}} \alpha(R) h_R\right\|_\infty$ is attained on one of the $2^{d(n-1)}$ rectangles of volume $2^{-d(n-1)}$. That is, we may take a grid $\mathcal{X}$ of $2^{d(n-1)}$ points, one in each of the rectangles, and write the sup-norm as the maximum absolute value which $\sum_{|R|=2^{-n}} \alpha(R) h_R$ attains on the grid. To apply Theorem 2 we need to ensure that the $\exp(L^2)$-norms of the functions the supremum of which we take are $\ll 1$. By Khintchine's inequality

$$\left\|\frac{\sum_{|R|=2^{-n}} \alpha(R) h_R}{\sqrt{n^{d-1}}}\right\|_{\exp(L^2)} \ll 1,$$

and hence by Theorem 2

$$\mathbb{E}\left\|\sum_{|R|=2^{-n}} \alpha(R) h_R\right\|_\infty \ll \sqrt{\log 2^{d(n-1)}}\sqrt{n^{d-1}} \ll n^{d/2} .$$

It is conjectured that in the d-dimensional case the following holds:

$$2^{-n} \sum_{|R|=2^{-n}} |\alpha(R)| \ll n^{\frac{1}{2}(d-2)} \left\|\sum_{|R|=2^{-n}} \alpha(R) h_R\right\|_\infty .$$

Comparing this conjecture with the previous considerations, where

$$2^{-n} \sum_{|R|=2^{-n}} |\alpha(R)| \simeq n^{d-1} \quad \text{and} \quad \left\|\sum_{|R|=2^{-n}} \alpha(R) h_R\right\|_\infty \ll n^{d/2} ,$$

shows that the conjecture is sharp.

2.1 The Riesz product

Having looked at an example, let us now turn towards the construction of the Riesz product. The construction is close to the ideas Beck used, but adapted in a way that allows to use the Littlewood-Paley inequalities which fit into the frame of Haar functions very naturally.

To make the Riesz product short, one considers sums of r-functions in each factor instead of individual r-functions. Since the first order terms in the expansion of the product are most important this essentially does not effect the way it is applied.

To be able to apply the Littlewood-Paley inequality later, we have to ensure that one can easily retrieve a Haar expansion for the expansion of the Riesz product (cf. section 2.7). This is achieved in choosing the sums as 'ordered' sums in the following sense:

To make the Riesz product short we consider all r-functions $f_{\vec{r}}$ with first coordinate in a short interval at once: Let $[n] = I_1 \cup \cdots \cup I_q$, where the I_j are adjacent disjoint intervals of length n/q. $\mathbb{H}_n$ denotes the set of all vectors in $\mathbb{N}^d$ with norm $n = |\vec{r}| = r_1 + \cdots + r_d$. Let $\mathbb{A}_t$ denote the subset with $r_1 \in I_t$:

$$\mathbb{A}_t = \{\vec{r} \in \mathbb{H}_n : r_1 \in I_t\} \ .$$

Then we consider the following 'short' Riesz product

$$\Psi = \prod_{t=1}^{q} (1 + \tilde{\rho} F_t)$$

where

$$\begin{aligned} F_t &= \sum_{\vec{r} \in \mathbb{A}_t} f_{\vec{r}} \\ f_{\vec{r}} &= \sum_{R \in \mathcal{R}_{\vec{r}}} \operatorname{sgn}(\alpha(R)) h_R \ . \end{aligned}$$

The factor $\tilde{\rho}$ has the purpose of normalizing terms in the expanded product. It is chosen small enough so that one expects that $\|\Psi\|_1 \approx \mathbb{E}\Psi$. $\tilde{\rho}$ is defined by

$$q = an^{\varepsilon}, \quad b = \frac{1}{6}, \quad \tilde{\rho} = aq^b n^{-1} \ .$$

Here a is a small positive constant and $0 < \varepsilon < \frac{1}{2}$. Further, we put

$$\rho = \sqrt{q} n^{-1} \ .$$

Expanding Ψ, one is mainly interested in those terms of the sum that are products of *strongly distinct* r-functions. A family of r-functions is called *strongly distinct* if the corresponding family of vectors in $\mathbb{H}_n$ satisfies: Restricted to each coordinate we obtain a set of distinct values, i.e. no coincidences appear.

Why only these terms? Firstly, these products are r-functions again themselves (however, for every product of length ≥ 2 the associated vectors have norm $> n$). Though, it should be noted that these are not the only terms that turn out to be r-functions. A second reason lies in the approach to the norm estimates of the not wanted terms. This is done by means of the inclusion-exclusion principle which requires all not strongly distinct products to be involved.

With this in mind, let Ψ_u^{sd} denote the sum over all u-fold strongly distinct products appearing in the expansion of Ψ. Then the terms we are interested in are

$$\Psi^{sd} = \sum_{u=1}^{q} \Psi_u^{sd} \tilde{\rho}^u \ .$$

Proof of the small ball inequality. At this point, let us see how the Riesz product can be used to deduce the theorem.

The only terms in Ψ^{sd} that are associated to vectors in $\mathbb{H}_n$ are those in Ψ_1^{sd}. Let

$$H_n = \sum_{|R|=2^{-n}} \alpha(R) h_R$$

be the function whose sup-norm appears in Theorem 1 and recall that

$$f_{\vec{r}} = \sum_{R \in \mathcal{R}_{\vec{r}}} \operatorname{sgn}(\alpha(R)) h_R \ .$$

Then

$$\begin{aligned} \langle H_n, \Psi^{sd} \rangle &= \langle H_n, \Psi_1^{sd} \rangle \\ &= \tilde{\rho} 2^{-n} \sum_{|R|=2^{-n}} |\alpha(R)| \\ &\gg q^b n^{-1} 2^{-n} \sum_{|R|=2^{-n}} |\alpha(R)| \ . \end{aligned}$$

By Hölder's inequality

$$\langle H_n, \Psi^{sd} \rangle \leq \|H_n\|_\infty \left\|\Psi^{sd}\right\|_1 \ ,$$

in order to prove Theorem 1, it then remains to show that

$$\left\|\Psi^{sd}\right\|_1 \ll 1 \ .$$

This is the aim of the remaining part of this chapter. This result also forms the main part of the discrepancy result.

2.2 Dividing into two lemmata

Instead of attacking $\left\|\Psi^{sd}\right\|_1 \ll 1$ directly, it is split into two lemmata:

Lemma 1. $\|\Psi\|_1 \ll 1$.

And with the notation $\Psi = 1 + \Psi^{sd} + \Psi^{\neg}$ the second lemma is

Lemma 2. $\|\Psi^{\neg}\|_1 \ll 1$.

Therefore, what we are aiming for will follow:

$$\left\|\Psi^{sd}\right\|_1 = \|\Psi - 1 - \Psi^{\neg}\|_1 \leq 1 + \|\Psi\|_1 + \|\Psi^{\neg}\|_1 \ll 1 \ .$$

It is the proof of Lemma 2 that makes use of the principle of inclusion and exclusion and requires all not strongly distinct sums.

2.3 Proof of Lemma 1, $\|\Psi\|_1 \ll 1$.

In this section the main steps of the proof of Lemma 1 will be considered. We seek a bound on $\|\Psi\|_1 = \mathbb{E}\,|\Psi|$. It seems to be easier to establish a bound on $\mathbb{E}\Psi$, as one can make use of the orthogonality of Haar functions. So one may rewrite the L_1-norm as follows and apply Cauchy-Schwarz:

$$\begin{aligned} \|\Psi\|_1 &= \mathbb{E}\Psi - 2\mathbb{E}(\Psi\chi_{\Psi<0}) \\ &\leq \mathbb{E}\Psi + 2\mathbb{P}(\Psi < 0)^{1/2}\,\|\Psi\|_2 \ . \end{aligned}$$

To continue the proof, three basic steps are required:

Proposition 1. *(i)* $\mathbb{E}\Psi = 1$.

(ii) $\mathbb{P}(\Psi < 0) \ll \exp(-Aq^{1/2-b})$ *where* $A > 1$ *and* $A \to \infty$ *as* $a \to 0$ *(recall* $\tilde{\rho} = aq^b n^{-1}$*).*

(iii) $\|\Psi\|_2 \ll \exp(a'q^{2b})$ *where* $0 < a' < 1$ *and* $a' \to 0$ *as* $a \to 0$.

This proposition then yields

$$\begin{aligned}\|\Psi\|_1 &\ll 1+\exp(Aq^{1/2-b}/2+a'q^{2b}) \\ &\ll 1 \quad \text{for sufficiently small } \alpha,\end{aligned}$$

where in the last step the properties of A and a, and the fact that $1/2-b=2b$ were used[2].

2.4 Proof of Proposition 1 – some technical details

2.4.1 $\mathbb{E}\Psi = 1$

Expanding the product Ψ gives the leading term one and a sum of products over Haar functions. For each product the first components of associated vectors come from distinct intervals I_j, and we obtain a one-dimensional r-function in the first coordinate. Hence, expectations of the products all vanish.

2.4.2 $\mathbb{P}(\Psi < 0) \ll \exp(-Aq^{1/2-b})$

To obtain a proof of this, an $\exp L^1$-norm bound is needed:

Theorem 4 (Bilyk and Lacey [2], Theorem 1.4.1)**.** *In dimension $d \geq 2$ we have the estimate*

$$\left\| \sum_{|R|=2^{-n}} \alpha(R) h_R \right\|_{\exp(L^{2/(d-1)})} \ll \left\| \left(\sum_{|R|=2^{-n}} \alpha(R)^2 \chi_R \right)^{1/2} \right\|_{\infty} .$$

Theorem 4 follows from a series of applications of the ℓ_2-valued case of the Littlewood-Paley inequality (see [2] for a proof). Note that we are going to apply this theorem to a sum of r-functions. In this special case, the theorem is immediate from Khintchine's inequality, which does not require the Littlewood-Paley inequality and in particular not its ℓ_2-valued version (see Section 3 for more details).

To find a bound on the probability $\mathbb{P}(\Psi < 0)$ we will use the exponential version of Markov's inequality. The above theorem comes in to handle the expectation in the Markov bound. Recall that

$$\Psi = \prod_{t=1}^{q} (1+\tilde{\rho} F_t) .$$

[2]This is the reason for the exact choice of $b = \frac{1}{6}$; cf. [2, p.12].

Thus

$$\mathbb{P}(\Psi < 0) \leq \sum_{t=1}^{q} \mathbb{P}(\tilde{\rho} F_t < -1) \ .$$

As F_t has $\ll \frac{n}{q} n$ terms the right-hand side of the exp L^1-norm bound applied to

$$-\rho F_t = \sqrt{q} n^{-1} F_t$$

is $\ll 1$. Let

$$c := \max_t \|-\rho F_t\|_{\exp L} \ll 1 \ ,$$

where we used Theorem 4. Continuing the estimate,

$$\begin{aligned}
\mathbb{P}(\Psi < 0) &\leq \sum_{t=1}^{q} \mathbb{P}(\rho F_t < -a^{-1} q^{1/2-b}) \\
&\leq \sum_{t=1}^{q} \mathbb{P}\left(\exp \frac{-\rho F_t}{c} > \exp \frac{a^{-1} q^{1/2-b}}{c} \right) \\
&\leq \sum_{t=1}^{q} \mathbb{E}\left(\exp \frac{-\rho F_t}{c} \right) \exp(-\frac{a^{-1} q^{1/2-b}}{c}) \\
&\leq q \exp \frac{-a^{-1} q^{1/2-b}}{c} \\
&\leq q \exp \frac{-a^{-1} q^{1/2-b}}{\tilde{c}},
\end{aligned}$$

where $\tilde{c}$ is the implicit constant of $c \ll 1$.

2.4.3 $\|\Psi\|_2 \ll \exp(a' q^{2b})$

By orthogonality of Haar functions we expect the 2-norm to be small. The proof uses conditional expectation arguments, an application of norm interpolation and an L_p estimate of sums of products of r-functions that have a coincidence in their second coordinate, the latter being derived by an application of the Littlewood-Paley inequality.

2.5 Lemma 2, $\|\Psi^{\neg}\| \ll 1$.

The proof builds up on inclusion-exclusion and a sequence of alternate applications of Littlewood-Paley and triangle inequalities.

Each term in the expansion of Ψ involves expressions of the form

$$\mathrm{Prod}(\mathcal{C}) := \sum_{(\vec{r}_1, \ldots, \vec{r}_k) \in \mathcal{C}} \prod_{j=1}^{k} f_{\vec{r}_j}$$

for subsets $\mathcal{C} \subset \mathbb{H}_n^k$.

Given $V \subset \{1, \ldots, q\}$ let $\mathrm{NSD}(V)$ denote the set of not strongly distinct $|V|$-tuples in $\prod_{j \in V} \mathbb{A}_j$ that furthermore satisfy the stronger condition that every element of each tuple is involved in a coincidence:

$$\begin{aligned}\mathrm{NSD(V)} =\{(\vec{r}_1, \ldots, \vec{r}_{|V|}) \in \prod_{j \in V} \mathbb{A}_j \mid & \text{ for each } i \in \{1, \ldots, |V|\} \\ & \text{there are } k \neq i \text{ and } l \in \{2,3\} \text{ such that } r_{il} = r_{kl})\} \ .\end{aligned}$$

With this notation we can express the not strongly distinct terms $\Psi^{\neg}$ of Ψ, using PIE, as follows:

$$\Psi^{\neg} = \sum_{V \subset [q], |V| \geq 2} (-1)^{|V|+1} \tilde{\rho}^{|V|} \mathrm{Prod}(\mathrm{NSD}(V)) \prod_{t \in [q] \smallsetminus V} (1 + \tilde{\rho} F_t) \ .$$

From this, our task to bound $\|\Psi^{\neg}\|$ can be simplified: Applications of triangle inequality, Hölder's inequality and the estimate[3]

$$\sup_{V \subset \{1,\ldots,q\}} \left\| \prod_{t \in V} (1 + \tilde{\rho} F_t) \right\|_{1+q^{-2b}} \ll 1$$

reduce it to finding L_p estimates of $\mathrm{Prod}(\mathrm{NSD}(V))$.

The L_p estimate found in [2] is the following

Theorem 5. *There are positive constants C_0, C_1, C_2, C_3 and κ such that*

$$\rho^{|V|} \left\| \mathrm{Prod}(\mathrm{NSD}(V)) \right\|_p \ll \left(C_0 \left| V \right|^{C_1} p^{C_2} q^{C_3} n^{-\kappa} \right)^{|V|} \ .$$

In the above estimate, this theorem is applied with $|V| \leq q$, $p \ll q^{2b} = q^{1/3} \simeq n^{\varepsilon''}$. Therefore, the polynomial growth in p—which might not be best possible—has no effect to the bound obtained.

2.6 Reduction from $\mathrm{NSD}(V)$ to less complex sets

The bound of Theorem 5 is crude enough to allow quite generous use of the triangle inequality and further applications of PIE. The set $\mathrm{NSD}(V)$ admits further decomposition.

For this, let us analyse the structure of the set $\mathrm{NSD}(V)$. It is essential to know about the 'pattern' of coincidences. Since these use very little information, they can be more easily described in the language of graphs:

[3] This is derived from part (*iii*) of Proposition 1 which was obtained on the way to prove Lemma 1.

For $V \subset [q]$ and an element of $\mathrm{NSD}(V)$ let us consider the *coincidence graph* (V, E_2, E_3) on V. E_2 is a '2'-coloured edge set that consists of exactly those edges $\vec{r}_i\vec{r}_j$ for which $r_{i2} = r_{j2}$, and, similarly, E_3 is the '3'-coloured edge set E_3 consisting of $\vec{r}_i\vec{r}_j$ for which $r_{i3} = r_{j3}$.

Note that

(*i*) E_2 and E_3 are disjoint since coincidences in two coordinates imply $\vec{r}_i = \vec{r}_j$.

(*ii*) Both, E_2 and E_3, are vertex-disjoint unions of cliques by transitivity of coincidence.

(*iii*) Each vertex belongs to a clique by definition of $\mathrm{NSD}(V)$ (each vertex is involved in a coincidence).

For an application of PIE, this construction is not yet entirely suitable. We will also consider graphs with vertex sets $E'_2 \subset E_2$ and $E'_3 \subset E_3$ that satisfy properties (*i*)-(*iii*) for the smaller subsets. Let's call them extended coincidence graphs.

Any graph G on $V \in [q]$ satisfying (*i*)-(*iii*) defines a subset $X(G)$ of $\mathrm{NSD}(V)$ of those tuples for which it is an extended coincidence graph, i.e.

$$X(G) = \{\{\vec{r}_v : v \in V\} \in \prod_{v \in V} \mathbb{A}_v \mid kl \in E_j \implies r_{kj} = r_{lj}\} \ .$$

A graph G on $V \in [q]$ that satisfies (*i*)-(*iii*) is called *admissible*.

2.6.1 Reduction to admissible graphs

The aim is to pass down from the set $\mathrm{NSD}(V)$ to sets $X(G)$ for admissible graphs and then, in the next paragraph, further to connected admissible graphs, for which the bound from Theorem 5 can be obtained.

For two admissible graphs G_1 and G_2 on the same vertex set V, let $G_1 \wedge G_2$ be, in case it exists, the smallest (w.r.t. $|E_2| + |E_3|$) admissible graph that contains both G_1 and G_2 as subgraphs. k-fold products are defined inductively.

Let $\mathcal{G}_1$ (the 'primes') be the set of all admissible graphs on V which cannot be written as a product of distinct admissible graphs. Let $\mathcal{G}_k$ be the set admissible graphs on V for which k is the minimum value such that these graphs can be written as k-fold product of primes.

It suffices to consider $k \leq q$ since each new factor increases the order of some clique.

Now we can apply PIE again to obtain

$$\mathrm{Prod}(\mathrm{NSD}(V)) = \sum_{k=1}^{q} (-1)^{k-1} \sum_{G \in \mathcal{G}_k} \mathrm{Prod}(X(G)) \ .$$

The number of admissible graphs is easily taken care of by the generous bound in Theorem 5: there are at most $2^{2|V|} |V|^{2|V|}$ of them. To see this, enumerate V—there are $|V|! < |V|^{|V|}$ ways to do so—and insert separators for the E_2-cliques, viewing the last segment as the set of vertices that do not belong to an E_2-clique. This gives another factor of $2^{|V|}$. Squaring takes care for E_3 and gives the bound.

It remains to consider admissible graphs.

2.6.2 Reduction to connected graphs

Proposition 2 (10.7 in [2])**.** *Let $G_1, \ldots, G_k$ be admissible graphs on pairwise disjoint vertex sets $V_1, \ldots, V_k$, and let $G = \bigcup G_i$ be their disjoint union. Then*

$$\mathrm{Prod}(X(G)) = \prod_{j=1}^{k} \mathrm{Prod}(X(G_j)) \ .$$

Therefore, if we know that connected admissible graphs $\tilde{G}$ satisfy

$$\rho^{|V|} \left\| \mathrm{Prod}(X(\tilde{G}) \right\|_p \ll \left(C_0 \left|V\right|^{C_1} p^{C_2} q^{C_3} n^{-\kappa} \right)^{|V|} \ ,$$

then, writing G as a disjoint union of connected admissible graphs $G_1, \ldots, G_k$ leads to

$$\begin{aligned} \rho^{|V|} \left\| \mathrm{Prod}(X(G)) \right\|_p &\leq \prod_{j=1}^{k} \rho^{|V_i|} \left\| \mathrm{Prod}(X(G_j)) \right\|_{kp} \\ &\leq \prod_{j=1}^{k} \left(C_0 \left|V_j\right|^{C_1} (kp)^{C_2} q^{C_3} n^{-\kappa} \right)^{|V_j|} \\ &\overset{(k \leq q)}{\leq} \left(C_0 \left|V\right|^{C_1} p^{C_2} q^{C_2 + C_3} n^{-\kappa} \right)^{|V|} \ . \end{aligned}$$

2.7 Littlewood-Paley

Next we will see how one can apply the real-valued Littlewood-Paley inequality to obtain an estimate on $\left\| \mathrm{Prod}(X(G)) \right\|_p$ for connected admissible graphs.

Let $v_1 < \cdots < v_l$ be the vertices of G and consider any of the products we are summing over in $\mathrm{Prod}(X(G))$:

$$f_{\vec{r}_1} \dots f_{\vec{r}_l} \ .$$

This product further splits into a sum of products of Haar functions. Can we retrieve—just from the information we have—the Fourier expansion with respect to the Haar basis in some coordinate, or, if not that, then at least to the extend that is needed to compute the square function? It turns out we can; the way the 'short' Riesz product is constructed using 'ordered' sums makes this possible for the first coordinate.

The first coordinates of the vectors $\vec{r}_j$ satisfy $r_{j,1} \in I_{v_j}$, that is $r_{1,1} < \cdots < r_{l,1}$. The Haar functions in the expansion for the first coordinate are therefore determined by those that appear in the last factor, i.e. $f_{\vec{r}_l}$. For a fixed value $b_l \in I_{v_l}$ in the last interval, the coefficients of the sum of Haar functions on intervals of length 2^{-b_l} are given, up to the sign, by $\mathrm{Prod}(X(G, b_l))$, where

$$X(G, b_l) = \{(\vec{r}_1 \dots \vec{r}_l) \in X(G) \mid r_{l,1} = b_l\}.$$

However, to compute the square function we only need the squares of coefficients:

$$S(f) = \left(\mathbb{E}^2 f + \sum_{I \in \mathcal{D}} \frac{|\langle f, h_I \rangle|^2}{|I|^2} \chi_I \right)^{1/2} \ .$$

Thus, we can apply the Littlewood-Paley inequality $\|f\|_p \ll \sqrt{p}\, \|S(f)\|_p$ in the first component:

$$\|\mathrm{Prod}(X(G))\|_p \ll \sqrt{p} \left\| \left(\sum_{b_l \in I_{v_l}} |\mathrm{Prod}(X(G, b_l))|^2 \right)^{1/2} \right\|_p \ .$$

What we are left with now is still built up by many Haar functions. Can we apply Littlewood-Paley another time? That is, can we apply it to those Haar functions in the first coordinate that are determined by the second-largest vertex? The right-hand side of the last inequality almost looks like an expression to which one can apply the ℓ_2-valued Littlewood-Paley inequality: We only need to completely eliminate factors that come from $f_{\vec{r}_l}$. A way to do so, is to apply the triangle inequality and specify second (or third) coordinates for all $\vec{r}_l$ with fixed first coordinate b_l simultaneously, so that we can pull that factor out of the sum $\mathrm{Prod}(X(G, b_l))$.

Define sets with specified second coordinates as follows:

$$X(G; b_l, c_Q) = \{(\vec{r}_1 \dots \vec{r}_{l-1}) \mid (\vec{r}_1 \dots \vec{r}_{l-1}, (b_l, c_Q, n - b_l - c_Q) \in X(G, b_l))\} \ .$$

Now, we can proceed as described above:

$$\begin{aligned}
\left\| \left(\sum_{b_l \in I_{v_l}} |\mathrm{Prod}(X(G, b_l))|^2 \right)^{1/2} \right\|_p &= \left\| \left(\sum_{b_l \in I_{v_l}} \left(\sum_{c_Q} \mathrm{Prod}(X(G; b_l, c_Q)) \right)^2 \right)^{1/2} \right\|_p \\
&\overset{\text{(C-S)}}{\leq} \left\| \left(\sum_{b_l \in I_{v_l}} n \sum_{c_Q} \mathrm{Prod}(X(G; b_l, c_Q))^2 \right)^{1/2} \right\|_p \\
&\overset{\triangle-\text{inequ.}}{\leq} \sqrt{n} \left(\sum_{c_Q} \sum_{b_l \in I_{v_l}} \left\| \mathrm{Prod}(X(G; b_l, c_Q))^2 \right\|_{p/2} \right)^{1/2} \\
&\leq n^{3/2} q^{-1/2} \sup_{c_Q \in [n], b_l \in I_{v_l}} \left\| \mathrm{Prod}(X(G; b_l, c_Q)) \right\|_p \ .
\end{aligned}$$

Note: Bilyk and Lacey do not apply the triangle inequality to b_l. Instead, they continue by applying the vector-valued Littlewood-Paley inequality. However, they pick up the additional factors $(n/q)^{1/2}$ later when the algorithm we are about to describe terminates, and therefore obtain the same bound.

We are in the position of applying Littlewood-Paley again in the first coordinate. This time the Haar functions are determined by the second largest vertex v_{l-1}. However, the second or third coordinate may already be determined. This happens whenever v_{l-1} and v_l belong to the same clique. In this case Cauchy-Schwarz and the application of the triangle inequality with respect to the second or third coordinate is not needed. The factor after application of Littlewood-Paley and triangle inequality therefore is

$$p^{1/2} n^{1/2} q^{-1/2} \ .$$

In both of the above situations we fix for the vertex v to which we apply Littlewood-Paley to a vector. Since the second (resp. third) coordinates are constant on E_2 (resp. E_3) cliques, this fixes the *value* (of the second/third coordinate) of all cliques that contain v. Continuing in this manner of applying Littlewood-Paley to the highest available vertex and applications of C-S and triangle inequality as needed we will eventually reach a situation where the whole set of vectors corresponding to the remaining vertices is determined by the previous choices of b_j's and c_Q's. That is, each remaining vertex belongs to a clique in E_2 and one in E_3 the values of which have already been chosen. Therefore $\mathrm{Prod}(X(G, b_l, \dots, c_{Q_l}, \dots))$ consists of just one term, and consequently has L_p-norm 1.

Our bound then is given by the product of factors we collected in each step. Let $V_{3/2}$ be the set of vertices where all inequalities were applied, i.e. those vertices that do not belong to any cliques with higher vertices, and let $V_{1/2}$ be the set of vertices at which only Littlewood-Paley and the triangle inequality were applied, i.e. those that are contained in exactly one clique that also contains a higher vertex. The bound obtained is then

$$\|\mathrm{Prod}(X(G))\|_p \ll p^{(\left|V_{3/2}\right|+\left|V_{1/2}\right|)/2} n^{3/2\left|V_{3/2}\right|} n^{1/2\left|V_{1/2}\right|} q^{-1/2(\left|V_{3/2}\right|+\left|V_{1/2}\right|)} .$$

Multiplying this estimate by $\rho^{|V|} = q^{1/2|V|} n^{-|V|}$ we see that all that remains to show that Theorem 5 holds, is to show that there is $\kappa > 0$ such that

$$|V|^{-1} \left(3/2 \left|V_{3/2}\right| + 1/2 \left|V_{1/2}\right| - |V|\right) < -\kappa .$$

2.8 Conclusion of the proof

We seek a fixed negative upper bound on

$$\begin{aligned} &|V|^{-1} \left(3/2 \left|V_{3/2}\right| + 1/2 \left|V_{1/2}\right| - |V|\right) \\ &= |V|^{-1} \left(1/2 \left|V_{3/2}\right| - 1/2 \left|V_{1/2}\right| - \left|V - V_{1/2} - V_{3/2}\right|\right) . \end{aligned}$$

Our main tool is the fact that the graph G we are looking at is connected. Let T be a normal spanning tree. Every path in T that intersects $V_{3/2}$ exactly in its endvertices v_1 and v_2 contains a vertex from $V - V_{1/2} - V_{3/2}$: Consider the set S of all vertices on the $v_1 - v_2$-path that belong to two cliques. Since v_1 and v_2 belong to $V_{3/2}$ they are introduced earlier (i.e. have large values) than all their neighbours. In particular, v_1 and v_2 cannot be neighbours and the set S contains some vertex that is smaller than both v_1 and v_2. Let v be the minimal element of S, then v belongs to two cliques the values of each of which were chosen before. Thus v is determined and thereby belongs to $V - V_{1/2} - V_{3/2}$.

We would like to use this information to compare the sizes of $\left|V_{3/2}\right|$ and $\left|V - V_{1/2} - V_{3/2}\right|$. The only problem that could occur are intersecting paths that may hinder us to employ a counting argument by injection. By normality of the tree and the fact that each vertex belongs to at most two cliques, all vertices have degree ≤ 3 in T and every vertex of degree 3 belongs to a clique containing at least two of its neighbours. Suppose $s \notin V_{3/2}$ is a vertex of degree 3, and let v_1, v_2, v_3 be $V_{3/2}$-vertices so that the $v_j - s$-paths in T only intersect in v and contain no further $V_{3/2}$-vertices. Let a and b be neighbours of s belonging to the same clique. Replacing the segment asb by ab in the $v_i - v_j$-path containing a and b, there no longer is a vertex that belongs to all three $v_i - v_j$-paths. By the previous argument, we find at least 2 distinct $(V - V_{1/2} - V_{3/2})$-vertices on $v_i - v_j$-paths in T.

Now, we can injectively assign to each but maybe one $V_{3/2}$-vertex v a vertex in $V - V_{1/2} - V_{3/2}$ that lies on the path in T starting in v and ending in a chosen root of T. That is,

$$\left|V_{3/2}\right| \leq \left|V - V_{1/2} - V_{3/2}\right| + 1 \ . \tag{1}$$

This shall suffice us. We obtain

$$\begin{aligned} |V| &= \left|V_{3/2}\right| + \left|V_{1/2}\right| + \left|V - V_{1/2} - V_{3/2}\right| \\ &\leq 2\left|V - V_{1/2} - V_{3/2}\right| + 1 + \left|V_{1/2}\right| \ . \end{aligned}$$

Thus,

$$-\frac{1}{2}\left|V - V_{1/2} - V_{3/2}\right| \leq \frac{1 + \left|V_{1/2}\right| - |V|}{4} \ . \tag{2}$$

For the expression we aim to bound, this yields

$$\begin{aligned} & |V|^{-1}\left(\frac{1}{2}\left|V_{3/2}\right| - \frac{1}{2}\left|V_{1/2}\right| - \left|V - V_{1/2} - V_{3/2}\right|\right) \\ & \overset{(by (1))}{\leq} |V|^{-1}\left(-\frac{1}{2}\left|V - V_{1/2} - V_{3/2}\right| + \frac{1}{2} - \frac{1}{2}\left|V_{1/2}\right|\right) \\ & \overset{(by (2))}{\leq} |V|^{-1}\left(\frac{3}{4} - \frac{\left|V_{1/2}\right| + |V|}{4}\right) \\ & \leq -\frac{1}{4} + \frac{3}{4|V|} \ . \end{aligned}$$

For $|V| \geq 4$ this last expression is $\leq -\frac{1}{16}$. The cases $|V| = 2$ or 3 can be treated separately.

3 We Only Need the Real-Valued Littlewood-Paley Inequality

The points where the ℓ_2-valued Littlewood-Paley inequality is applied in [2] are in the algorithm (cf. section 2.7, and [2, p.26/27, 'A General Estimate']), and to obtain $\exp(L^1)$-norm bounds in the proof of Lemma 7.8, part (7.9) in [2] – see Proposition 1, part (*ii*) of these notes.

We have already seen that one can get along just with the real-valued version in the first case. For the second, we can deduce the bound we need from Khintchine's inequality:

Theorem 6 (Khintchine's inequality)**.** *Let $\{r_k\}_{k\in N}$ be a family of Rademacher variables. Then for all finite sequences of coefficients $\{a_k\}$*

$$\left\| \sum_k a_k r_k \right\|_{\exp(L^2)} \ll \left(\sum_k a_k^2 \right)^{1/2} .$$

The result used in section 2.4 is the following:

Theorem 7 (Bilyk and Lacey [2], Theorem 1.4.1)**.** *In dimension $d \geq 2$ we have the estimate*

$$\left\| \sum_{|R|=2^{-n}} \alpha(R) h_R \right\|_{\exp(L^{2/(d-1)})} \ll \left\| \left(\sum_{|R|=2^{-n}} \alpha(R)^2 \chi_R \right)^{1/2} \right\|_\infty .$$

The case where Theorem 7 is applied to is that of $d = 3$ and f being a sum of r-functions, i.e. a sum of Rademacher random variables.

First note, that

$$\|f\|_{\exp(L^\alpha)} \simeq \sup_{p\geq 1} p^{-1/\alpha} \|f\|_p ,$$

where

$$\|f\|_{\exp(L^\alpha)} = \inf\{c > 0 \mid \mathbb{E}(e^{f^\alpha/c^\alpha} - 1) \leq 1\} .$$

Proof. Let $c = \|f\|_{\exp(L^\alpha)}$. Then

$$
\begin{aligned}
2 \geq \int_0^1 e^{f^\alpha/c^\alpha} dx &= \int_0^1 \sum_{k=0}^{\infty} \frac{(fc^{-1})^{\alpha k}}{k!} dx \\
&= \sum_{k=0}^{\infty} \frac{\|f\|_{\alpha k}^{\alpha k}}{k!\; c^{\alpha k}} \\
&\sim \sum_{k=0}^{\infty} \frac{\|f\|_{\alpha k}^{\alpha k}}{\sqrt{2\pi k}(k/e)^k c^{\alpha k}} \\
&= \sum_{k=0}^{\infty} \left(\frac{\|f\|_{\alpha k}^{\alpha}}{(2\pi k)^{1/2k}(k/e) c^\alpha} \right)^k .
\end{aligned}
$$

For the last series to be convergent and ≤ 2, we have to have

$$
\begin{aligned}
& \sup_{k \geq 1} \frac{\|f\|_{\alpha k}^{\alpha}}{(2\pi k)^{1/2k}(k/e) c^\alpha} < 1 \\
\iff & \sup_{k \geq 1} \frac{\|f\|_{\alpha k} (\alpha k)^{-1/\alpha}}{(2\pi k)^{1/2k\alpha}} \ll c \\
\iff & \sup_{k \geq 1} \|f\|_{\alpha k} (\alpha k)^{-1/\alpha} \ll c \\
\iff & \sup_{p \geq 1} p^{-1/\alpha} \|f\|_p \ll c .
\end{aligned}
$$

Since c is defined to be the infimum over all $c > 0$ that satisfy the above condition $\mathbb{E}(e^{f^\alpha/c^\alpha} - 1) \leq 1$, we obtain $c \ll \sup_p p^{-1/\alpha} \|f\|_p$ as well. $\square$

This equivalence admits to compare $\exp(L^\alpha)$-norms for different values of α. In particular

$$
\|f\|_{\exp(L)} \simeq \sup_{p \geq 1} \frac{1}{p} \|f\|_p \leq \sup_{p \geq 1} \frac{1}{\sqrt{p}} \|f\|_p \simeq \|f\|_{\exp(L^2)} .
$$

Thus, bounds on the $\exp(L^2)$-norm obtained by Khintchine's inequality are bounds for the $\exp(L^1)$-norm as well. In the special case of linear combinations of r-functions (this is the only case of interest to us) Khintchine's inequality implies Theorem 7.

4 The Discrepancy Theory Result

In this section we will be concerned with Bylik and Lacey's discrepancy result.

Let $\mathcal{A}_N \subset [0,1)^3$ be a set of N points in the 3-dimensional unit cube. Define a discrepancy function of this point distribution as follows:

$$D_N : [0,1)^3 \to \mathbb{R}; \quad D_N(x) = \big|A_N \cap [0,x)\big| - N \cdot \mathrm{vol}\big([0,x)\big),$$

where $[0,x)$ denotes the box in $[0,1)^3$ with diagonally opposite corners 0 and x. By means of the Riesz product defined to prove the small ball result, Bilyk and Lacey obtain the following lower bound on the sup-norm of the box discrepancy for corners:

Theorem 8 (Theorem 2.1.9 in [3])**.** *There is $0 < \eta < \frac{1}{2}$ such that for all collections $\mathcal{A}_N$ of N points in $[0,1)^3$*

$$\|D_N\|_\infty \gg (\log N)^{1+\eta} \ .$$

This is proven by Hölder's inequality with Ψ^{sd} as test function. Note that the proof of $\left\|\Psi^{sd}\right\|_1 \ll 1$ did not depend on the particular choice of r-functions for each $\vec{r} \in \mathbb{H}_n$ used to define Ψ. For this application, we will choose the r-functions in a (fairly common) way that is appropriate to handle the discrepancy function. The main fact we need is that $\left\|\Psi^{sd}\right\|_1 \ll 1$. Then

$$\begin{aligned}
\|D_N\|_\infty &\gg \|D_N\|_\infty \left\|\Psi^{sd}\right\|_1 \\
&\geq \left|\langle D_N, \Psi^{sd}\rangle\right| \\
&\geq \sum_{k=1}^{q} \left|\langle D_N, \Psi^{sd}_k\rangle\right| \ .
\end{aligned}$$

We expect the main contribution to come from the terms Ψ^{sd}_1 of order 1, since for all further terms the supports of the involved Haar functions are smaller, while for the discrete part of the discrepancy function only those Haar functions with support containing one of the N points make a contribution to $\langle D_N, \Psi^{sd}_k\rangle$, and while for the linear part the measure of the support matters as well.

Therefore, we will choose the $f_{\vec{r}}$ in a way that helps to bound $\left|\langle D_N, \Psi^{sd}_1\rangle\right|$ below. The choice of n the below proof suggests is n such that $2N \leq 2^n < 4N$, i.e. $n \simeq \log N$.

First, let us consider the linear part: This is easily handled on observing that

$$\int_{[0,1)^d} h_R(x)\mathrm{vol}\big([0,x)\big)dx = 4^{-d}\,|R|^2 \ .$$

Concerning the discrete part, the terms that appear are of the form

$$\langle \big| A_N \cap [0,x) \big|, \pm h_R \rangle \ .$$

When R contains no point of $\mathcal{A}_n$, then this inner product vanishes. Call such rectangles R 'good'. For the 'bad' rectangles, that do contain points of $\mathcal{A}_n$, we cannot say much about the inner product, and therefore would like to ignore them. To do so, since we are trying to establish a lower bound the absolute value of inner product $\langle D_N, \Psi_1^{sd} \rangle$, we ensure that both $\langle D_N, \Psi_1^{sd} \rangle$ and $\langle \big| A_N \cap [0,x) \big|, \pm h_R \rangle$ are positive, so that we can drop the latter. This is achieved by

$$f_{\vec{r}} := \sum_{\substack{R \in \mathcal{R}_{\vec{r}} \\ R \text{ good}}} -h_R + \sum_{\substack{R \in \mathcal{R}_{\vec{r}} \\ R \text{ bad}}} \operatorname{sgn}(\langle D_N, h_R \rangle) h_R \ .$$

The minus sign ensures that the linear contributions are positive. What remains, is to count: There are at most N bad rectangles since each of them contains at least one of the N points. The number of rectangles in total is $2^n \geq 2N$, so we have $\simeq 2^n$ good ones.

$$\begin{aligned}
|\langle D_N, f_{\vec{r}} \rangle| &= \langle D_N, f_{\vec{r}} \rangle \\
&\geq \sum_{\substack{R \in \mathcal{R}_{\vec{r}} \\ R \text{ good}}} \langle D_N, -h_R \rangle \\
&= \sum_{\substack{R \in \mathcal{R}_{\vec{r}} \\ R \text{ good}}} N \int_{[0,1)^d} h_R(x) \operatorname{vol}\big([0,x)\big) dx \\
&\gg_d N 2^{-2n} 2^n \gg 1 \ .
\end{aligned}$$

This proves the following proposition:

Proposition 3 (2.3.1 in [3]). *There is a constant $c > 0$, only depending on the dimension, such that, given a fixed set $\mathcal{A}_N$ of N points in $[0,1)^d$, we can find for each $\vec{r} \in \mathbb{H}_n$ some r-function $f_{\vec{r}}$ satisfying*

$$\langle D_N, f_{\vec{r}} \rangle \geq c \ .$$

With the choices of r-functions provided by Proposition 3, put

$$\Psi = \prod_{t=1}^{q} (1 + \tilde{\rho} F_t)$$

and let Ψ^{sd} be the sum of strongly regular products in the expansion.
Then, since there are about n^2 elements in $\mathbb{H}_n$, we obtain by the above

Proposition 3

$$\langle D_N, \Psi_1^{sd} \rangle = \langle D_N, \tilde{\rho} \sum_{\vec{r} \in \mathbb{H}_n} f_{\vec{r}} \rangle \gg \tilde{\rho} n^2 = a q^b n \simeq a n^{1+\varepsilon/6} .$$

Each of the terms in the sum Ψ_k^{sd} for $k \geq 2$ is a product of k strongly distinct r-functions (times a factor $\tilde{\rho}^k$). Therefore, each such term is itself an r-function corresponding to some vector $\vec{s}$ of norm strictly greater than n. Here the following proposition applies:

Proposition 4 (2.3.4 in [3]). *Let $f_{\vec{s}}$ be any r-function with $|\vec{s}| > n$. Then*

$$|\langle D_N, f_{\vec{s}} \rangle| \ll N 2^{-|\vec{s}|} .$$

When aiming to apply this proposition, we also need to know how many terms each sum Ψ_k^{sd} has, that is, given our collection $\{f_{\vec{r}} : \vec{r} \in \mathbb{H}_n\}$, how many choices of strongly distinct k-tuples are there? And, with view to the proposition just stated: how many choices of strongly distinct k-tuples are there whose product is an r-function corresponding to a specific vector $\vec{s}$. Let this number of choices be denoted by $\mathrm{Count}(\vec{s}, k)$.

Using straightforward counting arguments, it turns out that for $k \geq 3$

$$\begin{aligned} \mathrm{Count}(\vec{s}, k) &\ll (|\vec{s}| - n) \binom{(|\vec{s}| - n)^2}{k-2} + (|\vec{s}| - n)^3 \binom{(|\vec{s}| - n)^2}{k-3} \\ &\ll (|\vec{s}| - n)^3 \binom{(|\vec{s}| - n)^2}{k-3} \end{aligned}$$

and for $k = 2$

$$\mathrm{Count}(\vec{s}, 2) \ll |\vec{s}| - n .$$

We disregard the condition of being strongly distinct. How many vectors $\vec{r} \in \mathbb{H}_n$ are there such that $r_j \leq s_j$ for $j = 1, 2, 3$? Note that $\vec{s} - \vec{r} \in \mathbb{H}_{|\vec{s}|-n}$. Thus, there are at most $(|\vec{s}| - n)^{d-1} = (|\vec{s}| - n)^2$ such vectors $\vec{r}$. For 2 or 3 vectors in each k-tuple one or two coordinates agree with those of $\vec{s}$. These vectors can be chosen in at most $(|\vec{s}| - n)^3$ ways. It remains to choose $k - 2$ or $k - 3$ vectors all satisfying $r_j < s_j$ for $j = 1, 2, 3$. Thus the bound follows.

Using these facts and the previous proposition one shows that $\left|\langle D_N, \Psi_2^{sd} \rangle\right|$ and $\sum_{k=3}^{q} \left|\langle D_N, \Psi_k^{sd} \rangle\right|$ are small compared to the bound on $\left|\langle D_N, \Psi_1^{sd} \rangle\right|$.

For $k \geq 3$ one obtains for instance

$$\begin{aligned}
&\sum_{k=3}^{q} \left|\langle D_N, \Psi_k^{sd}\rangle\right| \\
&\leq \sum_{k=3}^{q}\sum_{h=k}^{2n}\sum_{\vec{s}:|\vec{s}|=n+h} \tilde{\rho}^k N 2^{-n-h}\mathrm{Count}(\vec{s},k) \\
&\ll \sum_{k=3}^{q}\sum_{h=k}^{2n}(n+h)^2\tilde{\rho}^k 2^{-h}h^3\binom{h^2}{k-3} && \text{estimate for Count} \\
&= \sum_{k=3}^{q}\sum_{h=k}^{2n} n^2(aq^b/n)^k 2^{-h}h^3\binom{h^2}{k-3} && \text{Def. } \tilde{\rho},\quad (n+h)\ll n \\
&\ll (aq^b/n)^3\sum_{j=0}^{q}\sum_{h=j+2}^{2n} n^2(aq^b/n)^j 2^{-h}h^3\binom{h^2}{j} && \text{change of var.} \\
&\leq (aq^b)^3\sum_{h=3}^{2n}2^{-h}h^3\sum_{j=0}^{q}(aq^b/n)^j\binom{h^2}{j} && \begin{array}{c}\text{reverse order of summation}\\ \text{add a few terms in doing so}\end{array} \\
&\leq (aq^b)^3\sum_{h=3}^{2n}2^{-h}h^3\sum_{j=0}^{h^2}(aq^b/n)^j[2-aq^b/n]^{h^2-j}\binom{h^2}{j} && \begin{array}{c}\text{the new factors are } >1\\ \text{since } (2-aq^b/n)>1\end{array} \\
&\leq (aq^b)^3\sum_{h=3}^{2n}2^{-h}h^3\cdot 2 && \text{binomial thm} \\
&\ll (aq^b)^3 \simeq n^{\varepsilon/2}\ .
\end{aligned}$$

5 Littlewood-Paley Inequality

The Littlewood-Paley inequalities relate the p-norm of a function to the p-norm of its square function. The square function is the pointwise ℓ_2-norm of the Haar expansion:

$$S(f)(x) = \left(\mathbb{E}(f)^2 + \sum_{I \in \mathcal{D}} \frac{|\langle f, h_I \rangle|^2}{|I|^2} \chi_I \right)^{1/2} .$$

The Haar functions are not L_2 normalised. The normalised system $\{\frac{h_I}{\sqrt{|I|}} \mid I \in \mathcal{D}\} \cup \chi_I$ forms an ONB for L_2 and we obtain for $p = 2$,

$$\|S(f)\|_2^2 = |\mathbb{E}(f)|^2 + \sum_{I \in \mathcal{D}} \frac{|\langle f, h_I \rangle|^2}{|I|} = \|f\|_2^2 ,$$

which is just Parseval's identity.

The Littlewood-Paley inequalities state:

Theorem 9 (Littlewood-Paley inequalities). *For $1 < p < \infty$ there are constants A_p and B_p, where B_p can be chosen to satisfy $B_p \ll \sqrt{p}$, such that*

$$A_p \|S(f)\|_p \leq \|f\|_p \leq B_p \|S(f)\|_p .$$

We will only be interested in the second inequality and only prove that. There are a number of different ways to prove these inequalities[4]. The one chosen here follows in part[5] Lacey [3] and in part Bañuelos and Moore [1].

The main tools of the proof are conditional expectation arguments and martingales. The idea is, instead of proving the result directly, to replace f by a function that satisfies $|f| \leq f^*$ (at least almost everywhere) and about which we know a little more. We are done in case we can show that $\|f^*\|_p \ll \sqrt{p} \|S(f)\|_p$.

We will work with the partial Fourier series:

$$f_n = \mathbb{E}f + \sum_{|I| > 2^{-n}} \frac{\langle f, h_I \rangle}{\sqrt{|I|}} \frac{h_I}{\sqrt{|I|}} .$$

[4]See E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Univ. Press, Princeton, N.J., 1970 for other approaches.

[5]The problem we have with the version in [3] is that it is not clear how the good-λ inequality follows from the Chang-Wilson-Wolff theorem. In fact, Chang, Wilson and Wolff need another lemma for this step (see Corollary 3.1 in S. Y. A. Chang, J. M. Wilson, and T. H. Wolff, *Some weighted norm inequalities concerning the Schrödinger operators*, Comment. Math. Helv. **60** (1985), 217-286) and they use the martingale maximal function instead of the dyadic maximal function.

These can be rewritten as[6]

$$f_n = \mathbb{E}(f \mid \mathcal{F}_n) \ ,$$

where $\mathcal{F}_n$ is the σ-algebra generated by dyadic intervals of length 2^{-n}. At this point martingales come in already: the f_n form a martingale with respect to the filtration given by $\mathcal{F}_n$.

Another fact that makes the partial Fourier series suitable to work with in our particular case is that we can express the square function easily in terms of the f_n:

$$S(f) = \left(\sum_n (f_{n+1} - f_n)^2 \right)^{1/2} \ .$$

The function which f is going to be replaced by is the *martingale maximal function* which is defined as

$$f^* = \sup_n |f_n| \ .$$

Since the Haar expansion of any function $s \in L_1$ converges to s almost everywhere, we have $\lim_{n\to\infty} |f_n| = |f|$ a.e. and hence for the martingale maximal function

$$|f| \leq f^* \quad \text{a.e.}$$

Our aim will be to prove the *good-λ inequality*. From that inequality the Littlewood-Paley inequality follows easily. The proof of the good-λ inequality involves two further propositions, which are treated in the next two sections. The first of them splits into two parts, one of which we defer to the last subsection of this section in order to not distract too much from the main proof.

Let us state the good-λ inequality first and consider how this inequality enables us to prove the Littlewood-Paley inequality:

Theorem 10. *For all* $0 < \varepsilon < 1/2$

$$\mathbb{P}(f^* > 2\lambda; S(f) < \varepsilon\lambda) \leq Ce^{-c\varepsilon^{-2}} \mathbb{P}(f^* > \lambda) \ ,$$

where c *can be taken to be* $1/4$.

---

[6]Recall: $\mathbb{E}(f \mid \mathcal{G}) = H$ means that H is $\mathcal{G}$-measurable, $\|H\|_2 < \infty$ and among all H' with these properties H minimizes $\|f - H\|_2$—that is exactly the property partial Fourier series satisfy.

We can express the L_p-norm, by means of Fubini's theorem, as

$$\|g\|_p^p = \int_0^\infty p\lambda^{p-1}\mathbb{P}(g > \lambda)d\lambda \ .$$

Therefore

$$\begin{aligned}
\|f^*\|_p^p &= \int_0^\infty p\lambda^{p-1}\mathbb{P}(f^* > \lambda)d\lambda \\
&= 2^p \int_0^\infty p\lambda^{p-1}\mathbb{P}(f^* > 2\lambda)d\lambda \\
&\le 2^p \int_0^\infty p\lambda^{p-1}\Big(\mathbb{P}(S(f) > \varepsilon\lambda) + \mathbb{P}(f^* > 2\lambda; S(f) < \varepsilon\lambda)\Big)d\lambda \\
&\le (2/\varepsilon)^p \int_0^\infty p\lambda^{p-1}\mathbb{P}(S(f) > \lambda)d\lambda + Ce^{-c\varepsilon^{-2}}2^p \int_0^\infty p\lambda^{p-1}\mathbb{P}(f^* > \lambda)d\lambda \ .
\end{aligned}$$

Letting

$$\varepsilon \simeq p^{-1/2} \ ,$$

the above yields

$$\|f^*\|_p \ll \sqrt{p}\,\|S(f)\|_p \ ,$$

and hence proves the Littlewood-Paley inequality.

5.1 Maximal function and Doob's inequality for submartingales

The good-λ inequality involves a probability of the form $\mathbb{P}(f^* > \lambda)$. Recall that f^* is defined as $f^* = \sup_n |f_n|$. A tool suitable in this situation is Doob's inequality for submartingales (the proofs of martingale results are omitted here; they may be found in [4]):

Theorem 11 (Doob)**.** *Let Z be a non-negative submartingale, i.e.*

$$\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) \ge Z_n \quad a.s.,$$

then for $c > 0$

$$\mathbb{P}(\sup_{k\le n} Z_k \ge c) \le \frac{1}{c}\mathbb{E}(Z_n) \ .$$

As we are interested in an exponential bound for the good-λ inequality, we would like to apply Doob's inequality in the situation

$$\mathbb{P}(\sup_{k\le n} Z_k \ge c) = \mathbb{P}(e^{t\cdot\sup_{k\le n} Z_k} \ge e^{tc}) \ ,$$

where Z is a martingale.

This, indeed, is justified by the following elementary lemma (see [4] for a proof):

Lemma 3. *If M is a martingale and g a convex function such that $\mathbb{E}\,|g(M_n)| < \infty$ for all n, then $g(M)$ is a submartingale.*

We will look at both

$$\mathbb{P}(\sup_{k\leq n} f_k > \lambda) \quad \text{and} \quad \mathbb{P}(\sup_{k\leq n}(-f_k) > \lambda)$$

for $n \to \infty$ and thereby obtain a bound for

$$\mathbb{P}(f^* = \sup_{k\leq n} |f_k| > \lambda) \ .$$

Proposition 5. *Suppose that $f_0 = \mathbb{E}f = 0$. Then*

$$\mathbb{E}\left(e^{tf_n - t^2 S(f)^2}\right) \leq 1 \quad \textit{for all } n \ , \tag{3}$$

and

$$\mathbb{P}(f^* > \lambda) \leq Ce^{-c\frac{\lambda^2}{\|S(f)\|_\infty^2}} \ . \tag{4}$$

Note: The proof of (3) is deferred to section 5.3.

By Doob's inequality:

$$\begin{aligned}\mathbb{P}(\sup_{k\leq n} f_k > \lambda) &\leq \mathbb{P}(e^{t\cdot \sup_{k\leq n} Z_k} \geq e^{tc}) \\ &\overset{\text{Doob}}{\leq} e^{-t\lambda}\mathbb{E}(e^{tf_n}) \\ &\leq e^{-t\lambda}e^{t^2\|S(f)\|_\infty} \cdot \mathbb{E}\left(e^{tf_n - t^2 S(f)^2}\right) \\ &\leq e^{-t\lambda}e^{t^2\|S(f)\|_\infty} \ .\end{aligned}$$

If we set $t = \frac{1}{2}\frac{\lambda}{\|S(f)\|_\infty}$, let $n \to \infty$ and repeat with the martingale $\{-f_k\}$, we obtain

$$\mathbb{P}(f^* > \lambda) \leq 2e^{-\frac{1}{4}\frac{\lambda^2}{\|S(f)\|_\infty^2}} \ .$$

5.2 Another proposition and the proof of the good-λ inequality

Proposition 6.

$$\mathbb{P}(f^* > 2\lambda) \leq C\exp\left(-\frac{c\lambda^2}{\|S(f)\|_\infty^2}\right)\mathbb{P}(f^* > \lambda)$$

Proof. The event $\{f^* > \lambda\}$ is supported on a disjoint union $\bigcup_{I\in\mathcal{J}} I$ of dyadic intervals, since $f^* = \sup_n |f_n|$, and $|f_n|$ is constant on dyadic intervals of length 2^{-n}.

Define $\tau = \min\{t : |f_t| > \lambda\}$. Then $\{\tau = \infty\} = \{f^* > \lambda\}$, and note that $|f_{\tau-1}| \leq \lambda$ and that $f_{\tau-1}$ is constant on each interval I in $\mathcal{J}$.

With these definitions, we have

$$\begin{aligned}
\mathbb{P}(f^* > 2\lambda) &= \sum_{I\in\mathcal{J}} \mathbb{P}(f^* > 2\lambda \mid I)\mathbb{P}(I) \\
&\leq \sum_{I\in\mathcal{J}} \mathbb{P}(f^* - |f_{\tau-1}| > \lambda \mid I)\mathbb{P}(I) \\
&\leq \sum_{I\in\mathcal{J}} \mathbb{P}(\sup_{t\geq\tau} |f_t - f_{\tau-1}| > \lambda \mid I)\mathbb{P}(I)\ .
\end{aligned}$$

Now we can apply Proposition 5 to the martingale $f_t - f_{\tau-1}$ on I (this is a martingale since τ is a stopping time). Observing that the sup-norm of the square function of this martingale is at most $\|S(f)\|_\infty$, we obtain

$$\begin{aligned}
\mathbb{P}(f^* > 2\lambda) &\leq \sum_{I\in\mathcal{J}} \mathbb{P}(\sup_{t\geq\tau} |f_t - f_{\tau-1}| > \lambda \mid I)\mathbb{P}(I) \\
&\leq C \exp\left(-\frac{c\lambda^2}{\|S(f)\|_\infty^2}\right) \sum_{I\in\mathcal{J}} \mathbb{P}(I) \\
&= C \exp\left(-\frac{c\lambda^2}{\|S(f)\|_\infty^2}\right) \mathbb{P}(f^* > \lambda)\ .
\end{aligned}$$

□

As a corollary we can deduce the good-λ inequality:

Proof of the good-λ inequality. We define another stopping time to introduce the condition $S(f) < \varepsilon\lambda$: Let

$$\tau' = \min\{t : S(f_{t+1}) = \sqrt{\sum_{j=0}^{t} (f_{j+1} - f_j)^2} > \varepsilon\lambda\}$$

and let

$$\tilde{f}_t = f_{t\wedge\tau'}$$

be the corresponding stopped martingale. Its square function satisfies

$$\left\|S(\tilde{f})\right\|_\infty \leq \varepsilon\lambda\ .$$

Further, observe the equality of the following events

$$\{\tau' = \infty\} = \{S(f) \le \varepsilon\lambda\} = \{\forall t(\tilde{f}_t = f_t)\} \ .$$

With this in mind, we conclude

$$\begin{aligned}\mathbb{P}(f^* > 2\lambda; S(f) \le \varepsilon\lambda) &\le \mathbb{P}(\tilde{f}^* > 2\lambda) \\ &\le Ce^{-c\frac{\lambda^2}{\lambda^2\varepsilon^2}}\mathbb{P}(\tilde{f}^* > \lambda) \\ &\le Ce^{-c\varepsilon^{-2}}\mathbb{P}(f^* > \lambda) \ .\end{aligned}$$

□

5.3 Conclusion of the proof of Littlewood-Paley

All that remains is to prove the first part of Proposition 5, that for a martingale $\{f_k\}$ that starts at $f_0 = 0$, we have

$$\mathbb{E}\left(e^{tf_n - t^2 S(f)^2}\right) \le 1 \quad \text{for all } n \ .$$

Recall that $\mathcal{F}_n$ is the σ-field generated by dyadic intervals of length 2^{-n} and that the dyadic martingale is given by

$$f_n = \mathbb{E}(f \mid \mathcal{F}_n) = \mathbb{E}f + \sum_{|I| \ge 2^{-n}} \frac{\langle f, h_I\rangle}{|I|} h_I \ .$$

We define, for $t > 0$, a new martingale $\{q_n\}$ by

$$q_n = e^{tf_n}\left[\prod_{j=1}^{n-1} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j)\right]^{-1} \ .$$

This is a martingale: It follows inductively that q_n is $\mathcal{F}_n$-measurable and we check that $\mathbb{E}(q_{n+1} \mid \mathcal{F}_n) = q_n$:

$$\begin{aligned}\mathbb{E}(q_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}\left(e^{tf_{n+1}}\left[\prod_{j=1}^{n} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j)\right]^{-1} \Bigg|\mathcal{F}_n\right) \\ &= \left[\prod_{j=1}^{n} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j)\right]^{-1} \mathbb{E}\left(e^{tf_{n+1}}|\mathcal{F}_n\right) \\ &= \left[\prod_{j=1}^{n} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j)\right]^{-1} \mathbb{E}\left(e^{t(f_{n+1}-f_n)}|\mathcal{F}_n\right) e^{tf_n} \\ &= e^{tf_n}\left[\prod_{j=1}^{n-1} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j)\right]^{-1} = q_n \ .\end{aligned}$$

What do we gain from the knowledge that this is a martingale? We can make use of the tower property and deduce

$$\mathbb{E}(q_n) = \mathbb{E}[\mathbb{E}(q_{n+1} \mid \mathcal{F}_n)] = \mathbb{E}(q_{n+1}) \ .$$

Since $f_0 = \mathbb{E}f = 0$ we therfore have

$$\mathbb{E}(q_n) = \mathbb{E}(q_0) = 1 \quad \text{for all } n \ . \tag{5}$$

Finally, let us estimate the product that occurs in q_n:

$$\begin{aligned}&\prod_{j=1}^{n-1} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j) \\ &= \prod_{j=1}^{n-1} \sum_{|I| \geq 2^{-j}} \frac{\int \exp\left(t \sum_{|J|=2^{-j-1}} \frac{\langle f, h_J \rangle}{|J|} h_J\right) h_I dx}{|I|} h_I\end{aligned}$$

The inner sum

$$\sum_{|J|=2^{-j-1}} \frac{\langle f, h_J \rangle}{|J|} h_J$$

is constant on each dyadic interval of length 2^{-j-2} and changes the sign once on each dyadic interval of length 2^{-j-1}. The factor h_I within the integral is constant on each dyadic interval of length 2^{j-1}.

Thus, splitting the integral into pieces of length 2^{-j-1} we obtain

$$\begin{aligned}&\prod_{j=1}^{n-1} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j) \\ &= \prod_{j=1}^{n-1} \sum_{|I| \geq 2^{-j}} \sum_{|J|=2^{-j-1}} \chi_J \chi_I \frac{1}{2} \left(e^{-t\frac{\langle f, h_J \rangle}{|J|}} + e^{t\frac{\langle f, h_J \rangle}{|J|}}\right) \frac{|J|}{|I|} \ .\end{aligned}$$

Since $\sum_{|I| \geq 2^{-j}} \chi_I \frac{|J|}{|I|} \leq 1$, we are left with

$$\leq \prod_{j=1}^{n-1} \sum_{|J|=2^{-j-1}} \chi_J \frac{1}{2} \left(e^{-t\frac{\langle f, h_J \rangle}{|J|}} + e^{t\frac{\langle f, h_J \rangle}{|J|}}\right) \ .$$

Here we can apply the following elementary inequality, which follows on comparing the Taylor expansions of the involved terms:

$$\frac{1}{2}(e^{-\mu} + e^{\mu}) \leq e^{\mu^2} \ .$$

It yields

$$\begin{aligned}\prod_{j=1}^{n-1} \mathbb{E}(e^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j) &\le \prod_{j=1}^{n-1} \sum_{|J|=2^{-j-1}} \chi_J \exp\left(t^2 \frac{\langle f, h_J\rangle^2}{|J|^2}\right) \\ &= \prod_{j=1}^{n-1} \exp\left(\sum_{|J|=2^{-j-1}} t^2 \frac{\langle f, h_J\rangle^2}{|J|^2} \chi_J\right) \\ &\le e^{t^2 S(f)^2} ,\end{aligned}$$

and hence by (5),

$$\mathbb{E}e^{tf_n - t^2 S(f)^2} \le \mathbb{E}q_n = 1 \ .$$

The proof of the real-valued Littlewood-Paley inequality is complete.

References

[1] R. Bañuelos, C. N. Moore, *Probabilistic Behavior of Harmonic Functions*, Birkhäuser Verlag, Basel, Boston, Berlin, 1999.

[2] M. T. Lacey, D. Bilyk, On the Small Ball Inequality in Three Dimensions, preprint, arXiv:math/0609815. To appear in Duke Math J.

[3] M. T. Lacey, Small Ball and Discrepancy Inequalities, preprint, arXiv:math/0609816.

[4] D. Williams, *Probability with Martingales*, Cambridge Univ. Press, Cambridge, 1991.