Harmonicity and Invariance on Slices of the Boolean Cube

Yuval Filmus[1] and Elchanan Mossel[2]

1   Department of Computer Science, Technion – Israel Institute of Technology,
    Haifa, Isreal
    yuvalfi@cs.technion.ac.il
2   The Wharton School, University of Pennsylvania, Philadelphia, USA; and
    University of California, Berkeley, USA
    mossel@wharton.upenn.edu

──── **Abstract** ────

In a recent work with Kindler and Wimmer we proved an invariance principle for the slice for low-influence, low-degree functions. Here we provide an alternative proof for *general* low-degree functions, with no constraints on the influences. We show that any real-valued function on the slice, whose degree when written as a harmonic multi-linear polynomial is $o(\sqrt{n})$, has approximately the same distribution under the slice and cube measure.

Our proof is based on a novel decomposition of random increasing paths in the cube in terms of martingales and reverse martingales. While such decompositions have been used in the past for stationary reversible Markov chains, ours decomposition is applied in a non-reversible non-stationary setup. We also provide simple proofs for some known and some new properties of harmonic functions which are crucial for the proof.

Finally, we provide independent simple proofs for the known facts that 1) one cannot distinguish between the slice and the cube based on functions of $o(n)$ coordinates and 2) Boolean symmetric functions on the cube cannot be approximated under the uniform measure by functions whose sum of influences is $o(\sqrt{n})$.

1   Introduction

The basic motivating question for our work is the following:

▶ **Question 1.1.** *Assume n is even. How distinguishable are the uniform measure μ on $\{0,1\}^n$ and the measure ν given by the uniform measure on $\{0,1\}^n$ conditioned on $\sum_i x_i = n/2$?*

More generally: how distinguishable are the product measure μ_p on $\{0,1\}^n$ where each coordinate takes the value 1 independently with probability p and ν_{pn} given by the uniform measure on $\{0,1\}^n$ conditioned on $\sum_i x_i = pn$ (assuming pn is an integer)?

Note that the two measures are easily distinguished using the simple sum of coordinates test. However, our interest is in understanding if the two measures are distinguishable using restricted families of tests, such as *low-depth circuits* or *low-degree polynomials*.

We call $\{0,1\}^n$ the *cube*, the support of the distribution ν_{pn} the *slice*, and the support of ν the *middle slice*. For exposition purposes, the introduction will only discuss the middle

slice, though all results (previous and ours) extend for the case of μ_p and ν_{pn} for every fixed p.

1.1 Low-degree polynomials

In a recent joint work with Kindler and Wimmer [6] we provided a partial answer to Question 1.1 by extending the non-linear invariance principle of [14]. Suppose that f is a low-degree low-influence multilinear polynomial which satisfies $\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} = 0$ (such polynomials are called *harmonic*[1]). The invariance principle of [6] establishes that the distribution of f under the measure ν is close to its distribution under the product measure μ on $\{0,1\}^n$, as well as to its distribution under the product space $\mathbb{R}^n$ equipped with the product Gaussian measure $\mathcal{G} = \mathrm{N}(1/2, 1/4)^{\otimes n}$.

The restriction to multilinear harmonic functions is quite natural in the slice — as every function on the slice has a unique representation as a harmonic multilinear function (this fact, due to Dunkl [3], is proved in Section 3). It is the analog of the implicit restriction to multilinear polynomial in the original non-linear invariance principle.

Both the invariance principle proven in [14] and the one proven in [6] require that the functions have low influences. Indeed, a function like x_1 has a rather different distribution under μ compared to $\mathcal{G}$. Similarly the function $x_1 - x_2$ has a rather different distribution under ν compared to $\mathcal{G}$.

However, note that the distribution of $x_1 - x_2$ under ν is quite similar to its distribution under μ. It is natural to speculate that low-degree harmonic functions have similar distributions under ν and μ. Unfortunately, the proof of the invariance principle in [6] goes through Gaussian space, rendering the low-influence condition necessary even when comparing ν and μ.

Our main result in this paper is a direct proof of the invariance principle on the slice showing that the distribution of a low-degree harmonic function on the slice is close to its distribution on the corresponding cube. Our results do not require the condition of low influences.

▶ **Theorem 1.2.** *Let $f \colon \{-1,1\}^n \to \mathbb{R}$ be a harmonic multilinear polynomial of degree $o(\sqrt{n})$ and variance 1. Then for any 1-Lipschitz function φ (i.e., one satisfying $|\varphi(x) - \varphi(y)| \le |x - y|$),*

$$|\mathbb{E}_{\nu}[\varphi(f)] - \mathbb{E}_{\mu}[\varphi(f)]| = o(1),$$

and the Lévy distance[2] between the distribution of f under μ and the distribution of f under ν is $o(1)$.

See Section 2 for the definition of harmonic functions and Theorem 4.1 as well as Corollary 4.2 for a more quantitative bounds and more general statements (which apply in particular for any i.i.d. measure on the cube and the corresponding slice). In Subsection 4.1 we show that the results cannot be extended to polynomials whose degree is much bigger than $\sqrt{n}$.

---

[1]   This somewhat unfortunate terminology is borrowed from Bergeron [1, Section 8.4], in which an S_n-harmonic polynomial is one which is annihilated by $\sum_{i=1}^{n} \frac{\partial^k}{\partial x_n^k}$ for all k. For multilinear polynomials, both definitions coincide.

[2]   The Lévy distance between two real random variables X, Y is the infimal τ such that for all $x \in \mathbb{R}$ it holds that $\Pr[X \le x - \tau] - \tau \le \Pr[Y \le x] \le \Pr[X \le x + \tau] + \tau$.

Novel proof ingredients

The basic idea of the proof is to use the coupling method by showing that the distribution of f on different slices of the cube is at most indentical, as long as the slices are of distance at most roughly $\sqrt{n}$ (here the distance between the kth slice and the ℓth slice is $|k - \ell|$). In fact, for our proof to work we crucially need to allow distances which are somewhat larger than $\sqrt{n}$.

To construct the coupling, we use a uniform random increasing path to couple level ℓ to level k above it. The main novel technique is representing the difference between the two levels as a difference of two martingales. Such representations have been used before in the analysis of *stationary reversible* Markov chains in Banach spaces [15], and even earlier in the analysis of stochastic integrals [13]. However, all previous decompositions were for stationary reversible chains while ours is neither. Our novel representation of the differences might be of interest in other applications.

Properties of harmonic functions

Complementing the analytic techniques, we also provide an elegant self-contained treatment of harmonic functions on the slice which is independent, simpler, and contains more results than the first author's earlier paper [5]. In particular we make crucial use of a two-sided Poincaré inequality, see e.g. Lemma 3.11.

Applications

Except for the natural interpretation of Theorem 1.2 in terms of distinguishing between distributions, it can be used to prove results in extremal combinatorics in the same way the main result of Theorem [6] is used. For example, in Proposition 4.5 we give a proof of the Kindler–Safra theorem on the slice, first proved in [6].

1.2 Influences, symmetric functions and circuits

We prove a few other results that give partial answers to Question 1.1.

- Using direct computation of the total variation distance we prove the following theorem:

▶ **Theorem 1.3.** *Let f be a function on $\{0,1\}^n$ depending on $o(n)$ coordinates and satisfying $\|f\|_\infty \leq 1$. Then*

$$|\mathbb{E}_\nu[f] - \mathbb{E}_\mu[f]| = o(1).$$

- We prove that symmetric functions cannot be approximated by functions whose total influence is $o(\sqrt{n})$.

▶ **Theorem 1.4.** *There exists a constant $\delta > 0$ such that if f is a symmetric Boolean function such that $\frac{1}{3} \leq \mathbb{E}_{\mu_p}[f] \leq \frac{2}{3}$ then $\Pr_{\mu_p}[f \neq g] > \delta$ for every Boolean function g satisfying $\mathrm{Inf}[g] = o(\sqrt{n})$.*

Since it is well-known [2] based on arguments from [12, 7] that a depth d size m circuit has total influence at most $O((\log m)^{d-1})$, our result immediately implies circuit lower bounds for such function. However, much better bounds are known, see e.g. [18] for general symmetric functions and [16] for the case of the majority function. Nevertheless, Theorem 1.4 is more general as it holds for functions f that are not necessarily the majority function and for functions g that are not necessarily in AC^0. Moreover, the proof of Theorem 1.4 is based on a new and simple probabilistic argument.

Exact formulations and proofs of these results appear in the full version of the paper.

1.3   Other results comparing the cube to the slice

Question 1.1 is not a new question. So we conclude the introduction with a few classical results related to this question:

- The limiting behavior of the partial sum $W(s) = \frac{1}{\sqrt{s}} \sum_{i=1}^{s} (x_i - \frac{1}{2})$ as $s \to \infty$ under the cube and the slice measures are well-studied. It is well-known that under the cube measure $W(s)$ converges to brownian motion, while under the slice measure it converges to a brownian bridge.
- It is well-known that the partial sums $W(s)$ are at least as concentrated in the slice as they are in the cube [8].
- It is well-known that Lipchitz function of the random variables $x_1, \ldots, x_n$ are concentrated both in the cube and in the slice. The results for the slice follow from the hypercontractive estimates by Lee and Yau [11]. These are also needed in our proofs.

Paper organization

A few useful definitions appear in Section 2. Harmonic functions are defined and analyzed in Section 3. We outline the proof of our invariance principle for Lipschitz functions in Section 4, in which we also give an invariance principle for the function $\varphi(x) = (|x| - 1)^2$ and illustrate it by a proof of a Kindler–Safra theorem for the slice, first proved in [6].

Many proofs have been ommitted from this extended abstract. Complete proofs appear in the full version of the paper, available online at `http://arxiv.org/abs/1507.02713`.

2   Definitions

Notation

We employ the falling power notation $n^{\underline{k}} = n(n-1)\cdots(n-k+1)$. The notation $\mathbf{1}_E$ equals 1 if the condition E holds, and 0 otherwise. The sign function is denoted sgn. The *L2 triangle inequality* is $(a+b)^2 \leq 2(a+b)^2$ or its generalization $(\sum_{i=1}^{n} a_i)^2 \leq n \sum_{i=1}^{n} a_i^2$.

A monomial is *squarefree* if it is not divisible by a square of a variable. (Thus there are 2^n squarefree monomials on n variables.) A polynomial is *multilinear* if all monomials are squarefree. A polynomial is *homogeneous* if all monomials have the same total degree. The dth homogeneous part of a polynomial $f = \sum c_m m$, denote $f^{=d}$, is the sum of $c_m m$ over all monomial m of total degree d. A polynomial f over $x_1, \ldots, x_n$ is *harmonic* if $\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} = 0$.

A univariate function f is *C-Lipschitz* if $|f(x) - f(y)| \leq C|x - y|$. A function is *Lipschitz* if it is 1-Lipschitz.

The expectation and variance of a random variable are denoted $\mathbb{E}, \mathbb{V}$, and $\|\cdot\|$ denotes its L2 norm $\|X\| = \sqrt{\mathbb{E}[X^2]}$. To signify that expectation is taken with respect to a distribution α, we write $\mathbb{E}_\alpha[X]$, $\mathbb{V}_\alpha[x]$, and $\|\cdot\|_\alpha$. A normal distribution with mean μ and variance σ^2 is denoted $\mathrm{N}(\mu, \sigma^2)$. A binomial distribution with n trials and success probability p is denoted $\mathrm{B}(n, p)$.

The symmetric group on $[n] = \{1, \ldots, n\}$ is denoted S_n. A distribution on $\mathbb{R}^n$ is *exchangeable* if it is invariant under the action of S_n (that is, under permutation of the coordinates); a discrete distribution is exchangeable if the probability of $(x_1, \ldots, x_n)$ depends only on $x_1 + \cdots + x_n$. For a function f on $\mathbb{R}^n$ and a permutation π, we define $f^\pi(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$.

The slice

The n-dimensional Boolean cube is the set $\{0,1\}^n$. For an integer $0 \le k \le n$, the kth slice of the n-dimensional Boolean cube is the set

$$\binom{[n]}{k} = \left\{ (x_1, \ldots, x_n) \in \{0,1\}^n : \sum_{i=1}^n x_i = k \right\}.$$

Probability measures

Our work involves two main probability measures, where n is always understood:
- The uniform measure on the slice $\binom{[n]}{k}$ is ν_k.
- The product measure μ_p on the Boolean cube is given by $\mu_p(x) = p^{\sum_i x_i}(1-p)^{\sum_i (1-x_i)}$. Note that $\nu_k, \mu_{k/n}$ have the same marginal distributions.

3 Harmonic functions

A basic and easy result states that every function on $\{-1,1\}^n$ has a unique representation as a multilinear polynomial, known as the *Fourier expansion*. It is easy to see that a multilinear polynomial has the same mean and variance with respect to the uniform measure on $\{-1,1\}^n$ and with respect to the standard n-dimensional Gaussian measure. In this section we describe the corresponding canonical representation on the slice, due to Dunkl [3, 4] and elaborated by Srinivasan [17] and Filmus [5].

Every function on the slice $\binom{[n]}{k}$ can be represented as a multilinear polynomial, but this representation is not unique. However, as found by Dunkl [3, 4], we can make it unique by demanding that it be *harmonic* in the sense of the following definition.

▶ **Definition 3.1.** A polynomial P over $x_1, \ldots, x_n$ is *harmonic* if

$$\sum_{i=1}^n \frac{\partial P}{\partial x_i} = 0.$$

In other words, P is harmonic if $\Delta P = 0$, where Δ is the differential operator $\sum_{i=1}^n \frac{\partial}{\partial x_i}$.

▶ **Definition 3.2.** A *basic function* is a (possibly empty) product of factors $x_i - x_j$ on disjoint indices. A function is *elementary* if it is a linear combination of basic functions.

Most, but not all, of the harmonic polynomials we consider will be multilinear. Here are some basic properties of harmonic polynomials.

▶ **Lemma 3.3.** *The set of harmonic polynomials is an algebra of polynomials, and is closed under partial derivatives, under permutations of the coordinates, and under taking homogeneous parts. In particular, all elementary functions are harmonic.*

Proof. Suppose f, g are harmonic. Then $\Delta(\alpha f + \beta g) = \alpha \Delta f + \beta \Delta g = 0$; $\Delta(fg) = f\Delta g + g\Delta f = 0$; $\Delta \frac{\partial f}{\partial x_i} = \frac{\partial \Delta f}{\partial x_i} = 0$; and $\Delta(f^\pi) = (\Delta f)^\pi = 0$. Finally, since $\Delta(\sum_{d=0}^n f^{=d}) = \sum_{d=0}^n \Delta f^{=d}$ and $\Delta f^{=d}$ is homogeneous of degree $d-1$, we see that $\Delta f^{=d} = 0^{=d-1} = 0$. ◀

▶ **Lemma 3.4.** *A polynomial f is harmonic if and only if for all $x_1, \ldots, x_n, c$ we have*

$$f(x_1 + c, \ldots, x_n + c) = f(x_1, \ldots, x_n).$$

Proof. Given $x_1, \ldots, x_n$, define a function

$$\phi(x_1, \ldots, x_n, c) = f(x_1 + c, \ldots, x_n + c).$$

The chain rule implies that $\frac{\partial \phi}{\partial c} = \Delta f$. Hence $\Delta f = 0$ iff ϕ is independent of c.   ◀

Our first theorem states that every function on the slice has a unique representation as a harmonic multilinear polynomial of degree at most $\min(k, n - k)$.

▶ **Theorem 3.5.** *Let $0 \le k \le n$. Every function on the slice $\binom{[n]}{k}$ has a unique representation as a harmonic multilinear polynomial of degree at most $\min(k, n - k)$.*

Similarly, we can prove that every harmonic multilinear polynomial is elementary.

▶ **Lemma 3.6.** *A multilinear polynomial is harmonic iff it is elementary. In particular, a harmonic multilinear polynomial over $x_1, \ldots, x_n$ has degree at most $n/2$.*

As we stated in the introduction to this section, multilinear polynomials enjoy the useful property of having the same mean and variance with respect to all product measures with fixed marginal mean and variance. The corresponding property for harmonic multilinear polynomials is stated in the following theorem, which also follows from the work of the first author [5].

▶ **Theorem 3.7.** *Let f, g be homogeneous harmonic multilinear polynomials of degree d_f, d_g, respectively, and let α be an exchangeable measure. If $d_f \ne d_g$ then $\mathbb{E}_\alpha[fg] = 0$. If $d_f = d_g = d$ then there exists a constant $C_{f,g}$ independent of α such that*

$$\mathbb{E}_\alpha[fg] = C_{f,g} \, \mathbb{E}_\alpha[(x_1 - x_2)^2 \cdots (x_{2d-1} - x_{2d})^2].$$

▶ **Corollary 3.8.** *Let f be a harmonic multilinear polynomial of degree at most d with constant coefficient $f^{=0}$. Suppose that α, β are exchangeable measures and $C > 0$ is a constant that for $t \le d$ satisfies*

$$\mathbb{E}_\alpha[(x_1 - x_2)^2 \cdots (x_{2t-1} - x_{2t})^2] \le C \, \mathbb{E}_\beta[(x_1 - x_2)^2 \cdots (x_{2t-1} - x_{2t})^2].$$

Then $\mathbb{E}_\alpha[f] = f^{=0}$, $\|f\|_\alpha^2 \le C \|f\|_\beta^2$, and $\mathbb{V}_\alpha[f] \le C \, \mathbb{V}_\beta[f]$.

The following lemma computes $\mathbb{E}[(x_1 - x_2)^2 \cdots (x_{2d-1} - x_{2d})^2]$ for the measures ν_k, μ_p.

▶ **Lemma 3.9.** *Let $p = k/n$. We have*

$$\mathbb{E}_{\nu_k}[(x_1 - x_2)^2 \cdots (x_{2d-1} - x_{2d})^2] = 2^d \frac{k^{\underline{d}}(n-k)^{\underline{d}}}{n^{\underline{2d}}} = (2p(1-p))^d \left(1 \pm O\left(\frac{d^2}{p(1-p)n}\right)\right),$$

$$\mathbb{E}_{\mu_p}[(x_1 - x_2)^2 \cdots (x_{2d-1} - x_{2d})^2] = (2p(1-p))^d.$$

This straightforward computation appears in [5, Theorem 4.1] and [6, Lemma 2.9]. Qualitatively, the lemma states that the norm of a low degree basic function is similar in both ν_k and μ_p. This is not surprising: the coordinates in the slice are almost independent, and a low degree basic function depends only on a small number of them.

We proceed by stating the so-called two-sided Poincaré inequality, starting with the following fact.

▶ **Lemma 3.10.** *Let f be a harmonic multilinear polynomial. Then*

$$\sum_{i<j} f^{(i\,j)} = \sum_{d=0}^{n/2} \left[\binom{n}{2} - d(n-d+1) \right] f^{=d},$$

where $f^{=d}$ is the dth homogeneous part of f.

▶ **Lemma 3.11.** *Let f be a harmonic multilinear polynomial of degree at most d. Then with respect to any exchangeable measure,*

$$n\,\mathbb{V}[f] \le \frac{1}{2} \sum_{i<j} \|f - f^{(i\,j)}\|^2 \le d(n-d+1)\,\mathbb{V}[f].$$

Finally, we state another two-sided Poincaré inequality, this time for derivatives. We start with the following surprising corollary of Theorem 3.7.

▶ **Lemma 3.12.** *Let f, g be homogeneous harmonic multilinear polynomials of degree d. Then for any exchangeable measure α,*

$$\frac{\sum_{i=1}^{n} \mathbb{E}_\alpha\left[\frac{\partial f}{\partial x_i} \frac{\partial g}{\partial x_i} \right]}{\mathbb{E}_\alpha[fg]} = 2d \frac{\mathbb{E}_\alpha[(x_1-x_2)^2 \cdots (x_{2d-3}-x_{2d-2})^2]}{\mathbb{E}_\alpha[(x_1-x_2)^2 \cdots (x_{2d-1}-x_{2d})^2]}.$$

We deduce the following two-sided Poincaré inequality.

▶ **Lemma 3.13.** *Let f be a multilinear polynomial of degree d, and let α be an exchangeable measure. Suppose that for $1 \le t \le d$ we have*

$$m \le 2t \frac{\mathbb{E}_\alpha[(x_1-x_2)^2 \cdots (x_{2t-3}-x_{2t-2})^2]}{\mathbb{E}_\alpha[(x_1-x_2)^2 \cdots (x_{2t-1}-x_{2t})^2]} \le M.$$

Then also

$$m\,\mathbb{V}[f] \le \sum_{i=1}^{n} \left\| \frac{\partial f}{\partial x_i} \right\|^2 \le M\,\mathbb{V}[f].$$

The following lemma computes m, M for the measures ν_k, μ_p.

▶ **Lemma 3.14.** *Let $p = k/n$. We have*

$$2d \frac{\mathbb{E}_{\nu_k}[(x_1-x_2)^2 \cdots (x_{2d-3}-x_{2d-2})^2]}{\mathbb{E}_{\nu_k}[(x_1-x_2)^2 \cdots (x_{2d-1}-x_{2d})^2]} = d \frac{(n-2d+2)(n-2d+1)}{(k-d+1)(n-k-d+1)}$$

$$= \frac{d}{p(1-p)} \left(1 \pm O\left(\frac{d}{p(1-p)n} \right) \right),$$

$$2d \frac{\mathbb{E}_{\mu_p}[(x_1-x_2)^2 \cdots (x_{2d-3}-x_{2d-2})^2]}{\mathbb{E}_{\mu_p}[(x_1-x_2)^2 \cdots (x_{2d-1}-x_{2d})^2]} = \frac{d}{p(1-p)}.$$

4 Invariance principle

In this section we state and outline the proof of an invariance principle showing that the distribution of a low-degree function on a slice $\binom{[n]}{k}$ is similar to its distribution on the Boolean cube with respect to the measure $\mu_{k/n}$. For convenience, we analyze the similarity in distribution via Lipschitz test functions, and derive similarity in more conventional terms as a corollary. The basic idea is to show that the distribution of a low degree function on a

given slice $\binom{[n]}{k}$ is similar to its distribution on nearby slices $\binom{[n]}{\ell}$. If we can show this for all slices satisfying $|k - \ell| \leq B$ for some $B = \omega(\sqrt{n})$, then the invariance follows by decomposing the Boolean cube as a union of slices.

Here is the formal statement of our invariance principle.

▶ **Theorem 4.1.** *There exists a constant $K > 0$ such that the following holds.*
Let f be a harmonic multilinear polynomial of degree d satisfying $d^2 \leq K \frac{p(1-p)n}{\log[p(1-p)n]}$ such that $\mathbb{V}[f]_{\nu_{pn}} = 1$. For any Lipschitz function φ,

$$\left| \underset{\nu_{pn}}{\mathbb{E}} [\varphi(f)] - \underset{\mu_p}{\mathbb{E}} [\varphi(f)] \right| = O\left(\sqrt{\frac{d}{\sqrt{p(1-p)n}} \log^{1/2} \frac{\sqrt{p(1-p)n}}{d}} \right).$$

As a corollary, we can estimate the Lévy distance between $f(\nu_{pn})$ and $f(\mu_p)$, along the lines of [14, Theorem 3.19(28)].

▶ **Corollary 4.2.** *Let f be a harmonic multilinear polynomial of degree d satisfying $d^2 \leq K \frac{p(1-p)n}{\log[p(1-p)n]}$ such that $\mathbb{V}[f]_{\nu_{pn}} = 1$, where $K > 0$ is the constant from Theorem 4.1. The Lévy distance between $f(\nu_{pn})$ and $f(\mu_p)$ is at most*

$$\epsilon = O\left(\sqrt[4]{\frac{d}{\sqrt{p(1-p)n}} \log^{1/2} \frac{\sqrt{p(1-p)n}}{d}} \right).$$

That is, for all y it holds that

$$\underset{\nu_{pn}}{\Pr}[f \leq y - \epsilon] - \epsilon \leq \underset{\mu_p}{\Pr}[f \leq y] \leq \underset{\nu_{pn}}{\Pr}[f \leq y + \epsilon] + \epsilon.$$

We conjecture that $f(\nu_{pn})$ and $f(\mu_p)$ are also close in CDF distance, but unfortunately the method of proof of [14, Theorem 3.19(30)] fails in this case.

The complete proof of Theorem 4.1 appears in the full version of the paper. In the remainder of this section, we explain the intuition behind the proof.

Our argument concerns the following objects:
- A harmonic multilinear polynomial f of degree d and unit norm. We think of d as "small".
- A Lipschitz functional φ.
- A slice $\binom{[n]}{pn}$. We think of p as constant, though the argument even works for subconstant p.

Our goal is to show that $\mathbb{E}_{\mu_p}[\varphi(f)] \approx \mathbb{E}_{\nu_{pn}}[\varphi(f)]$. The first step is to express μ_p as a mixture of ν_ℓ for various ℓ:

$$\underset{\mu_p}{\mathbb{E}}[\varphi(f)] = \sum_{\ell=0}^{n} \binom{n}{\ell} p^\ell (1-p)^{n-\ell} \underset{\nu_\ell}{\mathbb{E}}[\varphi(f)].$$

Applying the triangle inequality, this shows that

$$\left| \underset{\mu_p}{\mathbb{E}}[\varphi(f)] - \underset{\nu_{pn}}{\mathbb{E}}[\varphi(f)] \right| \leq \sum_{\ell=0}^{n} \binom{n}{\ell} p^\ell (1-p)^{n-\ell} \left| \underset{\nu_\ell}{\mathbb{E}}[\varphi(f)] - \underset{\nu_{pn}}{\mathbb{E}}[\varphi(f)] \right|.$$

In general we expect $|\mathbb{E}_{\nu_\ell}[\varphi(f)] - \mathbb{E}_{\nu_{pn}}[\varphi(f)]|$ to grow with $|\ell - pn|$, and our strategy would be to consider separately slices close to pn, say $|pn - \ell| \leq \delta$, and slices far away from pn, say $|pn - \ell| > \delta$. We will bound the contribution of slices close to pn directly. If δ is large enough then we expect the contribution of slices far away from pn to be small, essentially

since μ_p is concentrated on slices close to pn. For this argument to work, we need to choose δ so that $\delta = \omega(\sqrt{n})$.

It remains to bound $|\mathbb{E}_{\nu_\ell}[\varphi(f)] - \mathbb{E}_{\nu_{pn}}[\varphi(f)]|$ for ℓ close to pn. One strategy to obtain such a bound is to bound instead $|\mathbb{E}_{\nu_s}[\varphi(f)] - \mathbb{E}_{\nu_{s+1}}[\varphi(s)]|$ for various s, and use the triangle inequality. To this end, it is natural to consider the following coupling: let $(\mathbf{X}(s), \mathbf{X}(s+1)) \in \binom{[n]}{s} \times \binom{[n]}{s+1}$ be chosen uniformly at random under the constraint $\mathbf{X}(s) \subset \mathbf{X}(s+1)$. We can then bound

$$|\underset{\nu_s}{\mathbb{E}}[\varphi(f)] - \underset{\nu_{s+1}}{\mathbb{E}}[\varphi(s)]| = |\mathbb{E}[\varphi(f(\mathbf{X}(s))) - \varphi(f(\mathbf{X}(s+1)))]| \leq$$

$$\mathbb{E}[|\varphi(f(\mathbf{X}(s))) - \varphi(f(\mathbf{X}(s+1)))|] \leq \mathbb{E}[|f(\mathbf{X}(s)) - f(\mathbf{X}(s+1))|].$$

Denoting $\boldsymbol{\pi}(s+1) = \mathbf{X}(s+1) \setminus \mathbf{X}(s)$ and using the mulilinearity of f, this shows that

$$|\underset{\nu_s}{\mathbb{E}}[\varphi(f)] - \underset{\nu_{s+1}}{\mathbb{E}}[\varphi(s)]| \leq \mathbb{E}\left[\left|\frac{\partial f}{\partial x_{\boldsymbol{\pi}(s+1)}}(\mathbf{X}(s))\right|\right] = \mathbb{E}\left[\frac{1}{n-s}\sum_{i \notin \mathbf{X}(s)}\left|\frac{\partial f}{\partial x_i}(\mathbf{X}(s))\right|\right].$$

While we cannot bound $\sum_i |\frac{\partial f}{\partial x_i}|$ directly, Lemma 3.13 implies that $\sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 = O(d)$. Applying Cauchy–Schwartz, we get that for s close to pn,

$$|\underset{\nu_s}{\mathbb{E}}[\varphi(f)] - \underset{\nu_{s+1}}{\mathbb{E}}[\varphi(s)]| \leq \frac{1}{\Theta(n)}\mathbb{E}\left[\sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}(\mathbf{X}(s))\right|\right]$$

$$\leq \frac{1}{\Theta(n)}\mathbb{E}\left[\sqrt{n}\sqrt{\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(\mathbf{X}(s))^2}\right] = O\left(\sqrt{\frac{d}{n}}\right).$$

Recall now that our original goal was to bound $|\mathbb{E}_{\nu_\ell}[\varphi(f)] - \mathbb{E}_{\nu_{pn}}[\varphi(f)]|$ for $|\ell - pn| \leq \delta$, and our intended δ satisfied $\delta = \omega(\sqrt{n})$. Unfortunately, the idea just described only gives a bound of the form $|\mathbb{E}_{\nu_\ell}[\varphi(f)] - \mathbb{E}_{\nu_{pn}}[\varphi(f)]| = O(\delta\sqrt{d/n})$, which is useless for our intended δ.

One way out is to take $\delta = C\sqrt{n}$. This allows us to obtain meaningful bounds both on the contribution of slices close to pn and on the contribution of slices far away from pn. Although this only gives a constant upper bound on $|\mathbb{E}_{\mu_p}[\varphi(f)] - \mathbb{E}_{\nu_{pn}}[\varphi(f)]|$ if applied directly, this idea can be used in conjuction with the invariance principle for the Boolean cube [14] to give an invariance principle for the slice, and this is the route chosen in the prequel [6]. One drawback of this approach is that the invariance principle for the Boolean cube requires all influences to be small.

Our approach, in contrast, considers a coupling $(\mathbf{X}(0), \dots, \mathbf{X}(n))$ of *all* slices. Analogous to $f(\mathbf{X}(s+1)) - f(\mathbf{X}(s))$, we consider the quantity

$$\mathbf{C}(s) = (n-s)(f(\mathbf{X}(s+1)) - f(\mathbf{X}(s))) - s(f(\mathbf{X}(s-1)) - f(\mathbf{X}(s))).$$

As before, we can bound $\mathbb{E}[|\mathbf{C}(s)|] = O(\sqrt{dn})$. Moreover,

$$\sum_{u=s}^{t}\mathbf{C}(u) = (n-t)f(\mathbf{X}(t+1)) + (t-1)f(\mathbf{X}(t)) - (n-s-1)f(\mathbf{X}(s)) - sf(\mathbf{X}(s-1)),$$

and so we can bound $|\mathbb{E}_{\nu_\ell}[\varphi(f)] - \mathbb{E}_{\nu_{pn}}[\varphi(f)]|$ by bounding the expectation of $\sum_{u=pn}^{\ell}\mathbf{C}(u)$ or of $\sum_{u=\ell}^{pn}\mathbf{C}(u)$. The triangle inequality gives $|\sum_{u=s}^{t}\mathbf{C}(u)| = O(|s-t|\sqrt{dn})$, which suffers

from the same problem that we encountered above. However, by expressing $\mathbf{C}(s)$ as a difference of two martingales, we are able to improve on the triangle inequality, showing that

$$\left| \sum_{u=s}^{t} \mathbf{C}(u) \right| = O(\sqrt{|s-t|dn}),$$

a bound which is useful for $|s-t| = o(n/d)$ rather than for $|s-t| = o(\sqrt{n/d})$ as before.

In more detail, we define

$$\mathbf{U}(u) = f(\mathbf{X}(u+1)) - f(\mathbf{X}(u)) - \mathbb{E}[f(\mathbf{X}(u+1)) - f(\mathbf{X}(u))|\mathbf{X}(u)],$$
$$\mathbf{D}(u) = f(\mathbf{X}(u-1)) - f(\mathbf{X}(u)) - \mathbb{E}[f(\mathbf{X}(u-1)) - f(\mathbf{X}(u))|\mathbf{X}(u)],$$

both martingales by construction, $\mathbf{U}(u)$ for increasing u, and $\mathbf{D}(u)$ for decreasing u. We claim that $\mathbf{C}(u) = (n-u)\mathbf{U}(u) - u\mathbf{D}(u)$. If this holds, then using the fact that $\mathbb{E}[\mathbf{U}(u)\mathbf{U}(v)] = \mathbb{E}[\mathbf{D}(u)\mathbf{D}(v)] = 0$ for $u \neq v$ and the L2 triangle inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we get

$$\mathbb{E}\left[\left(\sum_{u=s}^{t} \mathbf{C}(u)\right)^2\right] \leq 2\,\mathbb{E}\left[\left(\sum_{u=s}^{t}(n-u)\mathbf{U}(u)\right)^2\right] + 2\,\mathbb{E}\left[\left(\sum_{u=s}^{t} u\mathbf{D}(u)\right)^2\right]$$

$$= 2\sum_{u=s}^{t}(n-u)^2\,\mathbb{E}[\mathbf{U}(u)^2] + 2\sum_{u=s}^{t} u^2\,\mathbb{E}[\mathbf{D}(u)^2].$$

This shows that $\mathbb{E}[(\sum_{u=s}^{t}\mathbf{C}(u))^2]$ scales linearly in $t-s$ rather than quadratically in $t-s$, which is what we would get if we just applied the triangle inequality. Since the L1 norm is bounded by the L2 norm, we conclude that $\mathbb{E}[|\sum_{u=s}^{t}\mathbf{C}(u)|] = O(\sqrt{|s-t|dn})$.

Finally, let us explain why $\mathbf{C}(u) = (n-u)\mathbf{U}(u) - u\mathbf{D}(u)$. In view of our previous expression for $\mathbf{C}(u)$, this boils down to proving that

$$(n-u)\,\mathbb{E}[f(\mathbf{X}(u+1)) - f(\mathbf{X}(u))|\mathbf{X}(u)] - u\,\mathbb{E}[f(\mathbf{X}(u-1)) - f(\mathbf{X}(u))|\mathbf{X}(u)] = 0.$$

We can rewrite the left-hand side as

$$\mathbb{E}\left[\sum_{i \notin \mathbf{X}(u)}[f(\mathbf{X}(u) \cup \{i\}) - f(\mathbf{X}(u))] - \sum_{i \in \mathbf{X}(u)}[f(\mathbf{X}(u) \setminus \{i\}) - f(\mathbf{X}(u))]\right].$$

Since f is multilinear, we can replace the differences with derivatives:

$$\mathbb{E}\left[\sum_{i \notin \mathbf{X}(u)} \frac{\partial f}{\partial x_i}(\mathbf{X}(u)) - \sum_{i \in \mathbf{X}(u)} -\frac{\partial f}{\partial x_i}(\mathbf{X}(u))\right] = \mathbb{E}\left[\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{X}(u))\right].$$

However, the last expression clearly vanishes, since f is harmonic. This completes the outline of the proof.

4.1 High-degree functions

Theorem 4.1 requires that $d = o(\sqrt{p(1-p)n})$. Indeed, Lemma 3.9, which implies that the norm of a low degree function is approximately the same under both μ_p and ν_{pn}, already requires the degree to be $o(\sqrt{p(1-p)n})$. For $d = \omega(\sqrt{p(1-p)n})$ and constant $p \neq 1/2$ we exhibit below a $0/\pm 1$-valued function f which satisfies $\|f\|_{\mu_p} = 1$ while $\|f\|_{\nu_{pn}} = o(1)$. This shows that for constant $p \neq 1/2$ the dependence on the degree is essential in Theorem 4.1,

since $|\mathbb{E}_{\nu_{pn}}[|f|] - \mathbb{E}_{\mu_p}[|f|]| = \|f\|^2_{\mu_p} - \|f\|^2_{\nu_{pn}} = 1 - o(1)$. We do not know whether this dependence is necessary for $p = 1/2$. Indeed, Lemma 3.9 can be extended above $\sqrt{n}$ in this case, as the calculation below shows.

Let $d = \omega(\sqrt{p(1-p)n})$, and assume further that $d = o((p(1-p)n)^{2/3})$. We consider the function $f = (2p(1-p))^{-d/2}(x_1 - x_2) \cdots (x_{2d-1} - x_{2d})$, whose μ_p-norm is 1 according to Lemma 3.9. The lemma also gives its ν_k-norm (where $k = pn$) as

$$\|f\|^2_{\nu_k} = (p(1-p))^{-d} \frac{k^{\underline{d}}(n-k)^{\underline{d}}}{n^{\underline{2d}}}.$$

We estimate this expression using Stirling's approximation, starting with $k^{\underline{d}}$:

$$k^{\underline{d}} = \frac{k!}{(k-d)!} = \left(\frac{k}{k-d}\right)^{k-d+1/2} \frac{k^d}{e^d} e^{O(1/k)-O(1/(k-d))} = \left(1 + \frac{d}{k-d}\right)^{k-d} \frac{k^d}{e^d}(1 \pm o(1)).$$

The Taylor series $\log(1+x) = x - x^2/2 + O(x^3)$ shows that

$$\left(1 + \frac{d}{k-d}\right)^{k-d} = \exp\left[d - \frac{d^2}{2(k-d)} + o(1)\right] = \exp\left[d - \frac{d^2}{2k} + o(1)\right],$$

and so $k^{\underline{d}} = k^d e^{-d^2/2k}(1 \pm o(1))$. We can similarly estimate $(n-k)^{\underline{d}} = (n-k)^d e^{-d^2/2(n-k)}(1 \pm o(1))$ and $n^{\underline{2d}} = n^{2d} e^{-2d^2/n}(1 \pm o(1))$, concluding that

$$\|f\|^2_{\nu_k} = (p(1-p))^{-d} \frac{k^d(n-k)^d}{n^{2d}} e^{-d^2/2k - d^2/2(n-k) + 2d^2/n}(1 \pm o(1))$$

$$= \exp\left[\frac{d^2}{2p(1-p)n}\left(-1 + 4p(1-p)\right) \pm o(1)\right].$$

If $p \neq 1/2$ is fixed, we immediately conclude that $\|f\|_{\nu_k} = o(1)$.

4.2 Approximately Boolean functions

Theorem 4.1 only applies to Lipschitz test functions, but in many applications we are interested in functions which grow faster, for example the distance-from-$\{-1,1\}$ function $\varphi(x) = (|x| - 1)^2$. Using hypercontractivity, we can obtain an invariance principle for $\varphi(x) = (|x| - 1)^2$.

▶ **Theorem 4.3.** *Let f be a harmonic multilinear polynomial of degree d satisfying $d^2 \leq K \frac{p(1-p)n}{\log[p(1-p)n]}$ such that $\|f\|_{\nu_{pn}} = 1$, where K is the constant in Theorem 4.1. We have*

$$\left|\mathbb{E}_{\nu_{pn}}[(|f| - 1)^2] - \mathbb{E}_{\mu_p}[(|f| - 1)^2]\right| = O\left(\frac{d^{1/4}\log^{1/8} n}{n^{1/8}}(p(1-p))^{-O(d)}\right).$$

As an illustration of this theorem, we give an alternative proof of [6, Theorem 7.5], a Kindler–Safra theorem for the slice.

▶ **Definition 4.4.** *A function f on a given domain is Boolean if on the domain it satisfies $f \in \{\pm 1\}$. If the domain is a cube, we use the term cube-Boolean. If it is a slice, we use the term slice-Boolean.*

▶ **Proposition 4.5.** *Let f be a multilinear polynomial of degree d such that $\mathbb{E}_{\mu_p}[(|f| - 1)^2] = \epsilon$. There exists a cube-Boolean function g on $(p(1-p))^{-O(d)}$ coordinates such that $\|f - g\|^2_{\mu_p} = O((p(1-p))^{-O(d)}\epsilon)$.*

Proof. This is essentially proved in [10, 9]. Explicitly, they prove the same result without the guarantee that g is cube-Boolean. In order to get our version, let $F = \operatorname{sgn} f$ and $G = \operatorname{sgn} g$. By definition $\|F - f\|^2 = \epsilon$, and so $\|F - g\|^2 = O((p(1-p))^{-O(d)}\epsilon)$. Since F is cube-Boolean, this implies that $\|F - G\|^2 = O((p(1-p))^{-O(d)}\epsilon)$. We conclude that $\|f - G\|^2 = O((p(1-p))^{-O(d)}\epsilon)$. ◄

▶ **Theorem 4.6.** *For every $d > 0$ there is a parameter $N_{d,p}$ depending continuously on p such that if $n > N_{d,p}$ then the following holds.*

Let f be a slice-Boolean harmonic multilinear polynomial such that $\|f^{>d}\|_{\nu_{pn}}^2 = \epsilon$, where $f^{>d} = \sum_{i>d} f^{=i}$. There exists a slice-Boolean harmonic multilinear polynomial g that depends on $(p(1-p))^{-O(d)}$ coordinates (invariant to permutations of the other coordinates) satisfying

$$\|f - g\|_{\nu_{pn}}^2 \leq O((p(1-p))^{-O(d)}\epsilon) + \tilde{O}\left(\frac{1}{n^{1/8}}\right),$$

where $\tilde{O}$ hides polylogarithmic factors.

Proof. Let $\tilde{f} = f^{\leq d}$ (that is, $\tilde{f} = \sum_{i \leq d} f^{=i}$), so that $\mathbb{E}_{\nu_{pn}}[(|\tilde{f}| - 1)^2] \leq \mathbb{E}_{\nu_{pn}}[(\tilde{f} - f)^2] = \epsilon$. Notice that $\tilde{f}$ is a harmonic multilinear polynomial of degree at most d. Theorem 4.3 implies that

$$\mathbb{E}_{\mu_p}[(|\tilde{f}| - 1)^2] \leq \underbrace{\epsilon + O\left(\frac{d^{1/4}\log^{1/8} n}{n^{1/8}}(p(1-p))^{-O(d)}\right)}_{\epsilon_1}.$$

Proposition 4.5 implies that there exists a cube-Boolean function g on a set J of $M = O((p(1-p))^{-O(d)})$ coordinates such that $\epsilon_2 \triangleq \mathbb{E}_{\mu_p}[(\tilde{f} - g)^2] = O((p(1-p))^{-O(d)}\epsilon_1)$. The function g is also slice-Boolean, but it is not necessarily harmonic. Let h be the unique harmonic multilinear function of degree at most $\min(k, n-k)$ that agrees with g on $\binom{[n]}{pn}$. Note that h also depends only on the coordinates in J, and in particular it has degree at most M (in fact, [6, Lemma 3.1] implies that $\deg h \leq \deg g$). Invoking [6, Theorem 3.3], we see that $\|g - h\|_{\mu_p}^2 = O(\frac{M^2 2^M}{p(1-p)n})$, and so

$$\epsilon_3 = \|\tilde{f} - h\|_{\mu_p}^2 = O\left(\frac{M^2 2^M}{p(1-p)n} + \epsilon_2\right).$$

Corollary 3.8 and Lemma 3.9 imply that $\|\tilde{f} - h\|_{\nu_{pn}}^2 = O(\epsilon_3)$, using the fact that $\deg(\tilde{f} - h) \leq M$. The proof is completed by noticing that $\|f - h\|_{\nu_{pn}}^2 = O(\epsilon + \epsilon_3)$. ◄

The proof of [6, Theorem 7.5] contains an additional argument guaranteeing that $\deg g \leq d$. The same argument can be applied here.

References

1 Fránćois Bergeron. *Algebraic Combinatorics and Coinvariant Spaces*. CMS Treatises in Mathematics. A K Peters, 2009.

2 Ravi B Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997.

3 Charles F. Dunkl. A Krawtchouk polynomial addition theorem and wreath products of symmetric groups. *Indiana Univ. Math. J.*, 25:335–358, 1976.

4 Charles F. Dunkl. Orthogonal functions on some permutation groups. In *Relations between combinatorics and other parts of mathematics*, volume 34 of *Proc. Symp. Pure Math.*, pages 129–147, Providence, RI, 1979. Amer. Math. Soc.

5 Yuval Filmus. An orthogonal basis for functions over a slice of the boolean hypercube. *Elec. J. Comb.*, 23(1):P1.23, 2016.

6 Yuval Filmus, Guy Kindler, Elchanan Mossel, and Karl Wimmer. Invariance principle on the slice. In *31st Conf. Comp. Comp.*, 2016.

7 Johan Håstad. Almost optimal lower bounds for small depth circuits. In Silvio Micali, editor, *Randomness and Computation*, volume 5 of *Advances in Computing Research*, pages 143–170. JAI Press, 1989.

8 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

9 Guy Kindler. *Property testing, PCP and Juntas*. PhD thesis, Tel-Aviv University, 2002.

10 Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas, 2004. Unpublished manuscript.

11 Tzong-Yau Lee and Horng-Tzer Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Prob.*, 26(4):1855–1873, 1998.

12 N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

13 Terry J. Lyons and T. S. Zhang. Decomposition of Dirichlet processes and its application. *Ann. Probab.*, 22(1):494–524, 1994.

14 Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Ann. Math.*, 171:295–341, 2010.

15 Assaf Naor, Yuval Peres, Oded Schramm, and Scott Sheffield. Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces. *Duke Math J.*, 134(1):165–197, 2006.

16 Ryan O'Donnell and Karl Wimmer. Approximation by DNF: Examples and counterexamples. In *Automata, Languages and Programming*, volume 4596 of *Lecture Notes in Computer Science*, pages 195–206. Springer Berlin Heidelberg, 2007.

17 Murali K. Srinivasan. Symmetric chains, Gelfand–Tsetlin chains, and the Terwilliger algebra of the binary Hamming scheme. *J. Algebr. Comb.*, 34(2):301–322, 2011.

18 Avishay Tal. Tight bounds on the Fourier spectrum of AC^0. Manuscript, 2014.