

# Invariance principle on the slice

Yuval Filmus\*    Guy Kindler†    Elchanan Mossel‡    Karl Wimmer§

April 24, 2015

## Abstract

We prove an invariance principle for functions on a slice of the Boolean cube, which is the set of all vectors  $\{0, 1\}^n$  with Hamming weight  $k$ . Our invariance principle shows that a low-degree, low-influence function has similar distributions on the slice, on the entire Boolean cube, and on Gaussian space.

Our proof relies on a combination of ideas from analysis and probability, algebra and combinatorics.

Our result implies a version of *Majority is Stablest* for functions on the slice, a version of Bourgain’s tail bound, and a version of the Kindler–Safra theorem. As a corollary of the Kindler–Safra theorem, we prove a stability result of Wilson’s theorem for  $t$ -intersecting families of sets, improving on a result of Friedgut.

## 1 Introduction

Analysis of Boolean functions is an area at the intersection of theoretical computer science, functional analysis and probability theory, which traditionally studies Boolean functions on the Boolean cube  $\{0, 1\}^n$ . A recent development in the area is the development of a non-linear *invariance principle* of Mossel, O’Donnell and Oleszkiewicz [MOO10], a vast generalization of the fundamental Berry–Esseen theorem. The Berry–Esseen theorem is a quantitative version of the Central Limit Theorem, giving bounds on the speed of convergence of a sum  $\sum_i X_i$  to the corresponding Gaussian distribution. Convergence occurs as long as none of the summands  $X_i$  is too “prominent”. The invariance principle is an analog of the Berry–Esseen theorem for low-degree polynomials. Given a low-degree polynomial  $f$  on  $n$  variables in which none of the variables is too prominent (technically,  $f$  has low *influences*), the invariance principle states that the distribution of  $f(X_1, \dots, X_n)$  and  $f(Y_1, \dots, Y_n)$  is similar as long as each of the vectors  $(X_1, \dots, X_n)$  and  $(Y_1, \dots, Y_n)$  consists of independent coordinates, the distributions of  $X_i, Y_i$  have matching first and second moments, and the variables  $X_i, Y_i$  are hypercontractive.

The invariance principle came up in the context of proving a conjecture, *Majority is Stablest*, claiming that the majority function is the most noise stable among functions which have low influences. It is usually applied in the following setting: the  $X_i$  are skewed Bernoulli variables, and the  $Y_i$  are the matching normal distributions. The invariance principle allows us to analyze a function on the Boolean cube (corresponding to the  $X_i$ ) by analyzing its counterpart in Gaussian space (corresponding to the  $Y_i$ ), in which setting it can be analyzed using geometric methods. This approach has been used to prove many results in analysis of Boolean functions (see for example [Kho10]).

The proof of the invariance principle relies on the product structure of the underlying probability spaces. The challenge of proving an invariance principle for non-product spaces seems far from trivial.

---

\*Institute for Advanced Study, Princeton, NJ

†The Hebrew University of Jerusalem, Israel

‡The Wharton School, University of Pennsylvania and University of California, Berkeley

§Duquesne University, Pittsburgh, PA

Here we prove such an invariance principle for the distribution over  $X_1, \dots, X_n$  which is uniform over the *slice*  $\binom{[n]}{k}$ , defined as:

$$\binom{[n]}{k} = \{(x_1, \dots, x_n) \in \{0, 1\}^n : x_1 + \dots + x_n = k\}.$$

This setting arises naturally in hardness of approximation, see e.g. [DDG<sup>+</sup>15], and in extremal combinatorics (the Erdős–Ko–Rado theorem and its many extensions).

Our invariance principle states that if  $f$  is a low-degree function on  $\binom{[n]}{k}$  having low influences, then the distributions of  $f(X_1, \dots, X_n)$  and  $f(Y_1, \dots, Y_n)$  are close, where  $X_1, \dots, X_n$  is the uniform distribution on  $\binom{[n]}{k}$ , and  $Y_1, \dots, Y_n$  are either independent Bernoulli variables with expectation  $k/n$ , or independent Gaussians with the same mean and variance.

The classical invariance principle is stated only for low-influence functions. Indeed, high-influence functions like  $f(x_1, \dots, x_n) = x_1$  behave very differently on the Boolean cube and on Gaussian space. For the same reason, the condition of low-influence is necessary when comparing functions on the slice and on Gaussian space.

The invariance principle allows us to generalize two fundamental results to this setting: Majority is Stablest and Bourgain’s tail bound. Using Bourgain’s tail bound, we prove an analog of the Kindler–Safra theorem, which states that if a Boolean function is close to a function of constant degree, then it is close to a junta.

As a corollary of our Kindler–Safra theorem, we prove a stability version of the  $t$ -intersecting Erdős–Ko–Rado theorem, combining the method of Friedgut [Fri08] with calculations of Wilson [Wil84]. Friedgut showed that a  $t$ -intersecting family in  $\binom{[n]}{k}$  of almost maximal size  $(1 - \epsilon) \binom{n-t}{k-t}$  is close to an optimal family (a  $t$ -star) as long as  $\lambda < k/n < 1/(t+1) - \zeta$  (when  $k/n > 1/(t+1)$ ,  $t$ -stars are no longer optimal). We extend his result to the regime  $k/n \approx 1/(t+1)$ .

The classical invariance principle is stated for *multilinear* polynomials, implicitly relying on the fact that every function on  $\{0, 1\}^n$  can be represented (uniquely) as a multilinear polynomial, and that multilinear polynomials have the same mean and variance under any product distribution in which the individual factors have the same mean and variance. In particular, the classical invariance principle shows that the correct way to lift a low-degree, low-influence function from  $\{0, 1\}^n$  to Gaussian space is via its multilinear representation.

Our invariance principle crucially relies on tools developed in a recent paper [Fil14]. Dunkl [Dun76, Dun79] showed that every function on the slice has a unique representation as a multilinear polynomial annihilated by the operator  $\sum_{i=1}^n \frac{\partial}{\partial x_i}$ ; we call a polynomial satisfying this condition a *harmonic function*. The results of [Fil14] show that low-degree harmonic functions have *similar* mean and variance under both the uniform distribution on the slice and the corresponding Bernoulli and Gaussian product distributions. For this reason, our invariance principle is stated for harmonic functions.

How bad is the restriction to harmonic functions? Starting with an arbitrary multilinear polynomial  $f$ , there is a unique harmonic function  $\tilde{f}$  agreeing with  $f$  on a given slice. We show that as long as  $f$  depends on few variables, the two functions  $f$  and  $\tilde{f}$  are close as functions over the Boolean cube. This implies that  $f$  behaves similarly on the slice, on the Boolean cube, and on Gaussian space.

Our proof combines algebraic, geometric and analytic ideas. A coupling argument, which crucially relies on properties of harmonic functions, shows that the distribution of a low-degree, low-influence harmonic function  $f$  is approximately invariant when we move from the original slice to nearby slices. Taken together, these slices form a thin layer around the original slice, on which  $f$  has roughly the same distribution as on the original slice. The classical invariance principle implies that the distribution of  $f$  on the layer is close to its distribution on the Gaussian counterpart of the layer, which turns out to be *identical* to its distribution on all of Gaussian space, completing the proof.

A special case of our main result can be stated as follows.

**Theorem 1.1.** *For every  $\epsilon > 0$  and integer  $d \geq 0$  there exists  $\tau = \tau(\epsilon, d) > 0$  such that the following holds. Let  $n \geq 1/\tau$ , and let  $f$  be a harmonic multilinear polynomial of degree  $d$  such that with respect to the uniform measure  $\nu_{pn}$  on the slice  $\binom{[n]}{pn}$ , the variance of  $f$  is at most 1 and all influences of  $f$  are bounded by  $\tau$ .*

The CDF distance between the distribution of  $f$  on the slice  $\nu_{pn}$  and the distribution of  $f$  under the product measure  $\mu_p$  with marginals  $\text{Ber}(p)$  is at most  $\epsilon$ : for all  $\sigma \in \mathbb{R}$ ,

$$|\Pr_{\nu_{pn}}[f < \sigma] - \Pr_{\mu_p}[f < \sigma]| < \epsilon.$$

This result is proved in Section 4.2.

**Paper organization** Some preliminaries are described in Section 2. We examine harmonic multilinear polynomials in Section 3. We prove the invariance principle in Section 4. Section 5 proves *Majority is Stablest*, and Section 6 proves Bourgain’s tail bound, two applications of the main invariance principle. Section 7 deduces a version of the Kindler–Safra theorem from Bourgain’s tail bound. Our stability result for  $t$ -intersecting families appears in Section 8. Some open problems are described in Section 9.

## 2 Preliminaries

**Notation** The notation  $\mathbf{1}_E$  is the characteristic function of the event  $E$ . Expectation, variance and covariance are denoted by  $\mathbb{E}$ ,  $\mathbb{V}$  and  $\text{Cov}$ , respectively. The sign function is denoted  $\text{sgn}$ . The notation  $[n]$  denotes the set  $\{1, \dots, n\}$ . The *slice*  $\binom{[n]}{k}$  consists of all subsets of  $[n]$  of cardinality  $k$ . We often identify subsets of  $[n]$  with their characteristic vectors in  $\{0, 1\}^n$ .

The notation  $\text{Bin}(n, p)$  denotes a binomial distribution with  $n$  trials and success probability  $p$ . The notation  $\text{Po}(\lambda)$  denotes a Poisson distribution with expectation  $\lambda$ . The notation  $\text{N}(\mu, \Sigma^2)$  denotes a normal distribution with mean  $\mu$  and covariance matrix  $\Sigma^2$ . For a scalar  $p$ , we use  $\mathbf{p}$  to denote a constant  $p$  vector (of appropriate dimension which is clear from context) and  $I_n$  to denote the  $n \times n$  identity matrix.

For a probability distribution  $\pi$ ,  $\|f\| = \|f\|_\pi = \sqrt{\mathbb{E}_\pi[f^2]}$  is the L2 norm of  $f$  with respect to  $\pi$ . Note that  $\|f\|_1 = \mathbb{E}[|f|]$ .

The notation  $a^{\underline{b}}$  denotes the falling factorial function:  $a^{\underline{b}} = a(a-1) \cdots (a-b+1)$ .

Asymptotic notation ( $O(\cdot)$  and the like) will always denote non-negative expressions. When the expression can be positive or negative, we use the notation  $\pm O(\cdot)$ . The underlying limit is always  $n \rightarrow \infty$ . If the hidden constant depends on variables  $V$ , we use the notation  $O_V(\cdot)$ .

A  $C$ -Lipschitz functional is a function  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $|\psi(x) - \psi(y)| \leq C|x - y|$  which implies that for functions  $f, g$  on the same domain:

**Lemma 2.1.** *For every  $C$ -Lipschitz functional  $\psi$  and functions  $f, g$  on the same domain,*

$$|\mathbb{E}[\psi(f)] - \mathbb{E}[\psi(g)]| \leq C\|f - g\|.$$

**Probability distributions** Our argument will involve several different probability distributions on  $\mathbb{R}^n$  (where  $n$  will always be clear from context):

- $\mu_p$  is the product distribution supported on  $\{0, 1\}^n$  given by  $\mu_p(S) = p^{|S|}(1-p)^{n-|S|}$ .
- $\nu_k$  is the uniform distribution on the slice  $\binom{[n]}{k}$ .
- $\mathcal{G}_p$  is the Gaussian product distribution  $\text{N}((p, \dots, p), p(1-p)I_n)$ .
- $\gamma_{p,q} = \text{N}((q, \dots, q), \Sigma)$ , where  $\Sigma_{i,j} = \frac{n-1}{n}p(1-p)\delta(i=j) - \frac{n-1}{n}\frac{p(1-p)}{n-1}\delta(i \neq j)$  for  $1 \leq i, j \leq n$ .

As is well-known, the distribution  $\gamma_{p,q}$  results from conditioning  $\mathcal{G}_p$  on the sum being  $qn$ .

**Lemma 2.2.** *Let  $(X_1, \dots, X_n) \sim \mathcal{G}_p$ . The distribution of  $(X_1, \dots, X_n)$  conditioned on  $X_1 + \dots + X_n = qn$  is  $\gamma_{p,q}$ .*

*Proof.* Let  $S = X_1 + \dots + X_n$ , and consider the multivariate Gaussian distribution  $(X_1, \dots, X_n, S)$ , whose distribution is easily calculated to be  $N(\begin{pmatrix} \mathbf{p} & pn \\ \mathbf{1}' & n \end{pmatrix}, p(1-p))$ . Let  $(Y_1, \dots, Y_n)$  be the distribution of  $(X_1, \dots, X_n, S)$  conditioned on  $S = qn$ , which is well-known to be multivariate Gaussian. Using well-known formulas, the mean of this distribution is  $\mathbf{p} + \mathbf{1}n^{-1}(qn - pn) = \mathbf{q}$  (as can be derived directly), and its covariance matrix is  $p(1-p)(I_n - \mathbf{1}n^{-1}\mathbf{1}')$ . The diagonal elements are  $\mathbb{V}[Y_i] = p(1-p)(1 - \frac{1}{n})$  and the off-diagonal ones are  $\text{Cov}(Y_i, Y_j) = p(1-p)(-\frac{1}{n})$ .  $\square$

We can also go in the other direction.

**Lemma 2.3.** *Let  $(X_1, \dots, X_n) \sim \gamma_{p,q}$ , let  $Y \sim N(p-q, \frac{p(1-p)}{n})$ , and let  $Y_i = X_i + Y$ . Then  $(Y_1, \dots, Y_n) \sim \mathcal{G}_p$ .*

*Proof.* As is well-known,  $Y_1, \dots, Y_n$  is a multivariate Gaussian, and it is easy to see that its mean is  $\mathbf{p}$ . We have  $\mathbb{V}[Y_i] = \mathbb{V}[X_i] + \mathbb{V}[Y] = p(1-p)$  and  $\text{Cov}(Y_i, Y_j) = \text{Cov}(X_i, X_j) + \mathbb{V}[Y] = 0$ . The lemma follows.  $\square$

The distributions  $\mu_p$  and  $\nu_k$  are very close for events depending on  $o(\sqrt{n})$  coordinates.

**Lemma 2.4.** *Let  $A$  be an event depending on  $J$  coordinates, where  $J^2 \leq n$ . Then*

$$|\nu_{pn}(A) - \mu_p(A)| \leq \frac{J^2}{4p(1-p)n} \mu_p(A).$$

*Proof.* The triangle inequality shows that we can assume that  $A$  is the event  $x_1 = \dots = x_\ell = 0, x_{\ell+1}, \dots, x_J = 1$  for some  $\ell$ . Let  $k = pn$ . Clearly  $\mu_p(A) = (1-p)^\ell p^{J-\ell}$ , whereas

$$\nu_{pn}(A) = \frac{(n-k)^\ell k^{J-\ell}}{n^J}.$$

We have

$$\frac{n^J}{n^J} = \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{J-1}{n}\right) \geq 1 - \frac{1 + \dots + (J-1)}{n} \geq 1 - \frac{J^2}{2n}.$$

Therefore

$$\nu_{pn}(A) \leq \frac{(n-k)^\ell k^{J-\ell}}{n^J(1 - J^2/(2n))} \leq \mu_p(A) \left(1 + \frac{J^2}{n}\right),$$

using  $\frac{1}{1-x} \leq 1 + 2x$ , which is valid for  $x \leq 1/2$ .

Similarly,

$$\frac{(n-k)^\ell k^{J-\ell}}{(n-k)^\ell k^\ell} \geq 1 - \frac{\ell^2}{2(1-p)n} - \frac{(J-\ell)^2}{2pn} \geq 1 - \max\left(\frac{J^2}{2(1-p)n}, \frac{J^2}{2pn}\right) \geq 1 - \frac{J^2}{4p(1-p)n}.$$

Therefore

$$\nu_{pn}(A) \geq \frac{(n-k)^\ell k^\ell (1 - J^2/(4p(1-p)n))}{n^\ell} = \mu_p(A) \left(1 - \frac{J^2}{4p(1-p)n}\right).$$

This completes the proof.  $\square$

## 2.1 Harmonic multilinear polynomials

Our argument involves extending a function over a slice  $\binom{[n]}{k}$  to a function on  $\mathbb{R}^n$ , just as in the classical invariance principle, a function on  $\{0,1\}^n$  is extended to  $\mathbb{R}^n$  by writing it as a multilinear polynomial. In our case, the correct way of extending a function over a slice to  $\mathbb{R}^n$  is by interpreting it as a *harmonic multilinear polynomial*. Our presentation follows [Fil14], where the proofs of various results claimed in this section can be found. The basis in Definition 2.2 below also appears in earlier work of Srinivasan [Sri11], who constructed it and showed that it is orthogonal with respect to all exchangeable measures.

**Definition 2.1.** Let  $f \in \mathbb{R}[x_1, \dots, x_n]$  be a formal polynomial. We say that  $f$  is *multilinear* if  $\frac{\partial^2 f}{\partial x_i^2} = 0$  for all  $i \in [n]$ . We say that  $f$  is *harmonic* if

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

The somewhat mysterious condition of harmonicity arises naturally from the representation theory of the Johnson association scheme. Just as any function on the Boolean cube  $\{0, 1\}^n$  can be represented uniquely as a multilinear polynomial (up to an affine transformation, this is just the Fourier–Walsh expansion), every function on the slice  $\binom{[n]}{k}$  can be represented uniquely as a harmonic multilinear polynomial, using the identification

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

**Lemma 2.5** ([Fil14, Theorem 4.1]). *Every real-valued function  $f$  on the slice  $\binom{[n]}{k}$  can be represented uniquely as a harmonic multilinear polynomial of degree at most  $\min(k, n - k)$ .*

There is a non-canonical Fourier expansion defined for harmonic multilinear polynomials.

**Definition 2.2.** Let  $A = (a_1, \dots, a_d)$  and  $B = (b_1, \dots, b_d)$  be two sequences of some common length  $d$  of distinct elements of  $[n]$ . We say that  $A < B$  if:

- (a)  $A$  and  $B$  are disjoint.
- (b)  $B$  is monotone increasing:  $b_1 < \dots < b_d$ .
- (c)  $a_i < b_i$  for all  $i \in [d]$ .

A sequence  $B = (b_1, \dots, b_d)$  is a *top set* if  $A < B$  for some sequence  $A$ . The collection of all top sets of length  $d$  is denoted  $\mathcal{B}_{n,d}$ , and the collection of all top sets is denoted  $\mathcal{B}_n$ .

If  $A = (a_1, \dots, a_d)$  and  $B = (b_1, \dots, b_d)$  satisfy  $A < B$ , define

$$\chi_{A,B} = \prod_{i=1}^d (x_{a_i} - x_{b_i}).$$

For a top set  $B$ , define

$$\chi_B = \sum_{A < B} \chi_{A,B}.$$

Finally, define

$$\chi_d = \chi_{\{2,4,\dots,2d\}}.$$

**Lemma 2.6** ([Fil14, Theorem 3.1, Theorem 3.2]). *Let  $\pi$  be any exchangeable distribution on  $\mathbb{R}^n$  (that is,  $\pi$  is invariant under permutation of the coordinates). The collection  $\mathcal{B}_n$  forms an orthogonal basis for all harmonic multilinear polynomials in  $\mathbb{R}[x_1, \dots, x_n]$  (with respect to  $\pi$ ), and*

$$\|\chi_B\|_\pi^2 = c_B \|\chi_{|B|}\|_\pi^2, \text{ where } c_B = \prod_{i=1}^n \binom{b_i - 2(i-1)}{2},$$

and  $\|\chi_B\|_\pi$  denotes the norm of  $\chi_B$  with respect to  $\pi$ .

In particular, if  $f$  is a harmonic multilinear polynomial then  $\mathbb{E}[f]$  is the same under all exchangeable measures.

Lemma 2.3 and Lemma 2.6 put together have the surprising consequence that harmonic multilinear functions have exactly the same distribution under  $\mathcal{G}_p$  and  $\gamma_{p,q}$ .

**Lemma 2.7.** *Let  $f$  be a harmonic multilinear polynomial. The random variables  $f(\mathcal{G}_p)$ ,  $f(\gamma_{p,q})$  are identically distributed.*

*Proof.* According to Lemma 2.3, if  $(x_1, \dots, x_n) \sim \gamma_{p,q}$ ,  $y \sim N(q-p, \frac{p(1-p)}{n})$  and  $y_i = x_i + y$ , then  $(y_1, \dots, y_n) \sim \mathcal{G}_p$ . The lemma follows since  $y_i - y_j = x_i - x_j$  and a harmonic multilinear polynomial can be expressed as a function of the differences  $x_i - x_j$  for all  $i, j$ .  $\square$

Lemma 2.6 allows us to compare the norms of a harmonic multilinear function under various distributions.

**Corollary 2.8.** *Let  $\pi_1, \pi_2$  be two exchangeable distributions on  $\mathbb{R}^n$ , and let  $f$  be a harmonic multilinear polynomial of degree  $d$ . If for some  $\epsilon \geq 0$  and all  $0 \leq e \leq d$  it holds that  $(1-\epsilon)\|\chi_e\|_{\pi_1}^2 \leq \|\chi_e\|_{\pi_2}^2 \leq (1+\epsilon)\|\chi_e\|_{\pi_1}^2$ , then also  $(1-\epsilon)\|f\|_{\pi_1}^2 \leq \|f\|_{\pi_2}^2 \leq (1+\epsilon)\|f\|_{\pi_1}^2$ .*

The following lemma records the norms of basis elements for the distributions considered in this paper.

**Lemma 2.9.** *For all  $d$  we have*

$$\begin{aligned} \|\chi_d\|_{\mu_p}^2 &= \|\chi_d\|_{\mathcal{G}_p}^2 = (2p(1-p))^d, \\ \|\chi_d\|_{\nu_{pn}}^2 &= 2^d \frac{(pn)^d ((1-p)n)^d}{n^{2d}} = (2p(1-p))^d \left(1 \pm O\left(\frac{d^2}{p(1-p)n}\right)\right). \end{aligned}$$

*Proof.* The exact formulas for  $\|\chi_d\|_{\mu_p}^2$  and  $\|\chi_d\|_{\nu_{pn}}^2$  are taken from [Fil14, Theorem 4.1]. Since  $x_1, \dots, x_n$  are independent under  $\mathcal{G}_p$ , we have  $\|\chi_d\|_{\mathcal{G}_p}^2 = \mathbb{E}[(x_1 - x_2)^2]^d = (2p(1-p))^d$ .

It remains to prove the estimate for  $\|\chi_d\|_{\nu_{pn}}^2$ . The proof of [Fil14, Theorem 4.1] shows that

$$\|\chi_d\|_{\nu_{pn}}^2 = 2^d \frac{(pn)^d ((1-p)n)^d}{n^{2d}} = (2p(1-p))^d \frac{\left(1 - \frac{O(d^2)}{pn}\right) \left(1 - \frac{O(d^2)}{(1-p)n}\right)}{\left(1 - \frac{O(d^2)}{n}\right)}.$$

It follows that

$$\frac{\|\chi_d\|_{\nu_{pn}}^2}{(2p(1-p))^d} = 1 \pm O\left(\frac{d^2}{pn} + \frac{d^2}{(1-p)n} + \frac{d^2}{n}\right) = 1 \pm O\left(\frac{d^2}{p(1-p)n}\right). \quad \square$$

Lemma 2.9 and Corollary 2.8 imply an L2 invariance principle for low degree harmonic multilinear polynomials.

**Corollary 2.10.** *Suppose  $f$  is a harmonic multilinear polynomial of degree  $d$  on  $n$  variables. For any  $p \leq 1/2$  such that  $d \leq pn$  and any  $\pi \in \{\mu_p, \mathcal{G}_p\}$  we have*

$$\|f\|_{\nu_{pn}} = \|f\|_{\pi} \left(1 \pm O\left(\frac{d^2}{p(1-p)n}\right)\right).$$

## 2.2 Analysis of functions

We consider functions on three different kinds of domains: the Boolean cube  $\{0, 1\}^n$ , the slice  $\binom{[n]}{k}$ , and Gaussian space  $\mathbb{R}^n$ . We can view a multilinear polynomial in  $\mathbb{R}[x_1, \dots, x_n]$  as a function over each of these domains in the natural way.

For each of these domains, we proceed to define certain notions and state some basic results. The material for the Boolean cube and Gaussian space is standard, and can be found for example in [OD14].

**Functions on the Boolean cube** The Boolean cube is analyzed using the measure  $\mu_p$  for an appropriate  $p$ . The Fourier characters  $\omega_S$  and Fourier expansion of a function  $f: \{0, 1\}^n \rightarrow \mathbb{R}$  are given by

$$\omega_S(x_1, \dots, x_n) = \prod_{i \in S} \frac{x_i - p}{\sqrt{p(1-p)}}, \quad f = \sum_{S \subseteq [n]} \hat{f}(S) \omega_S.$$

We define  $f^{=k} = \sum_{|S|=k} \hat{f}(S) \omega_S$ , and so a multilinear polynomial  $f$  of degree  $d$  can be decomposed as  $f = f^{=0} + \dots + f^{=d}$ . Since the Fourier characters are orthogonal, the parts  $f^{=0}, \dots, f^{=d}$  are orthogonal. In the future it will be convenient to separate  $f$  into  $f = f^{\leq k} + f^{> k}$  for an appropriate  $k$ , where  $f^{\leq k} = f^{=0} + \dots + f^{=k}$  and  $f^{> k} = f^{=k+1} + \dots + f^{=d}$ .

Define  $f^{[i]}(x) = f(x^{[i]})$ , where  $x^{[i]}$  results from flipping the  $i$ th coordinate of  $x$ . The  $i$ th cube-influence is given by

$$\text{Inf}_i^c[f] = \|f - f^{[i]}\|^2 = \left\| \frac{\partial f}{\partial x_i} \right\|^2 = \frac{1}{p(1-p)} \sum_{S \ni i} \hat{f}(S)^2.$$

The total influence of  $f$  is  $\text{Inf}^c[f] = \sum_{i=1}^n \text{Inf}_i^c[f]$ , and it satisfies the Poincaré inequality

$$\mathbb{V}[f] \leq p(1-p) \text{Inf}^c[f] \leq (\deg f) \mathbb{V}[f].$$

The noise operator  $T_\rho$  is defined by

$$T_\rho f = \sum_{i=0}^{\deg f} \rho^i f^{=i}.$$

The noise stability of  $f$  at  $\rho$  is

$$\mathbb{S}_\rho^c[f] = \langle f, T_\rho f \rangle = \sum_{i=0}^{\deg f} \rho^i \|f^{=i}\|^2.$$

The noise operator (and so noise stability) can also be defined non-spectrally. We have  $(T_\rho f)(x) = \mathbb{E}[f(y)]$ , where  $y$  is obtained from  $x$  by letting  $y_i = x_i$  with probability  $\rho$ , and  $y_i \sim \mu_p$  otherwise.

**Functions on the slice** The slice  $\binom{[n]}{k}$  is analyzed using the measure  $\nu_k$ . The corresponding notion of Fourier expansion was described in Section 2.1. A harmonic multilinear polynomial  $f$  of degree  $d$  can be decomposed as  $f = f^{=0} + \dots + f^{=d}$ , where  $f^{=k}$  contains the homogeneous degree  $k$  part. The parts  $f^{=0}, \dots, f^{=d}$  are orthogonal.

The  $(i, j)$ th influence of a function  $f$  is  $\text{Inf}_{ij}^s[f] = \mathbb{E}[(f - f^{(ij)})^2]$ , where  $f^{(ij)}(x) = f(x^{(ij)})$ , and  $x^{(ij)}$  is obtained from  $x$  by swapping the  $i$ th and  $j$ th coordinates. We define the  $i$ th influence by  $\text{Inf}_i^s[f] = \frac{1}{n} \sum_{j=1}^n \text{Inf}_{ij}^s[f]$ , and the total influence by  $\text{Inf}^s[f] = \sum_{i=1}^n \text{Inf}_i^s[f]$ . The total influence satisfies the Poincaré inequality

$$\mathbb{V}[f] \leq \text{Inf}^s[f] \leq (\deg f) \mathbb{V}[f].$$

For a proof, see for example [Fil14, Lemma 5.6].

The noise operator  $H_\rho$  is defined by

$$H_\rho f = \sum_{d=0}^{\deg f} \rho^{d(1-(d-1)/n)} f^{=d}.$$

The noise stability of  $f$  at  $\rho$  is

$$\mathbb{S}_\rho^s[f] = \langle f, H_\rho f \rangle = \sum_{d=0}^{\deg f} \rho^{d(1-(d-1)/n)} \|f^{=d}\|^2.$$

The noise operator (and so noise stability) can also be defined non-spectrally. We have  $(H_\rho f)(x) = \mathbb{E}[f(y)]$ , where  $y$  is obtained from  $x$  by taking  $\text{Po}(\frac{n-1}{2} \log \frac{1}{\rho})$  random transpositions.

**Functions on Gaussian space** Gaussian space is  $\mathbb{R}^n$  under a measure  $\mathcal{G}_p$  for an appropriate  $p$ . In this paper, we mostly consider functions on  $\mathbb{R}^n$  given by multilinear polynomials, and these can be expanded in terms of the  $\omega_S$ . General functions can be expanded in terms of Hermite functions. Every square-integrable function can be written as  $f = \sum_{k \geq 0} f^{=k}$ , where  $f^{=k}$  satisfies  $f^{=k}(\alpha x_1 + p, \dots, \alpha x_n + p) = \alpha^k f^{=k}(x_1 + p, \dots, x_n + p)$ .

The distributions  $\mu_p$  and  $\mathcal{G}_p$  have the same first two moments, and this implies that  $\mathbb{E}_{\mu_p}[f] = \mathbb{E}_{\mathcal{G}_p}[f]$  and  $\|f\|_{\mu_p} = \|f\|_{\mathcal{G}_p}$  for every multilinear polynomial  $f$ . The Ornstein–Uhlenbeck operator  $U_\rho$  is defined just like  $T_\rho$  is defined for the cube. Noise stability is defined just like in the case of the cube, and we use the same notation  $\mathbb{S}^c$  for it.

The noise operator (and so noise stability) can also be defined non-spectrally. We have  $(U_\rho f)(x) = \mathbb{E}[f(y)]$ , where  $y = (1 - \rho)p + \rho x + \sqrt{1 - \rho^2} N(0, p(1 - p))$ . We can also define noise stability as  $\mathbb{S}_\rho^c[f] = \mathbb{E}[f(x)f(y)]$ , where  $(x, y) \sim N((p, p), \begin{pmatrix} p(1 - p) & \rho p(1 - p) \\ \rho p(1 - p) & p(1 - p) \end{pmatrix})$ .

**Homogeneous parts** For a function  $f$ , we have defined  $f^{=k}$  in three different ways, depending on the domain. When  $f$  is a harmonic multilinear polynomial, all three definitions coincide. Indeed, any harmonic multilinear polynomial is a linear combination of functions of the form  $\chi_{A,B}$ . We show that  $\chi_{A,B} = \chi_{A,B}^{=|B|}$  under all three definitions. Let  $A = a_1, \dots, a_k$  and  $B = b_1, \dots, b_k$ . Since  $\chi_{A,B}$  is homogeneous of degree  $k$  as a polynomial, we see that  $\chi_{A,B} = \chi_{A,B}^{=k}$  over the slice. Also,

$$\chi_{A,B} = (p(1 - p))^{k/2} \prod_{i=1}^k \left( \frac{x_{a_i} - p}{\sqrt{p(1 - p)}} - \frac{x_{b_i} - p}{\sqrt{p(1 - p)}} \right).$$

Opening the product into a sum of terms, we can identify each term with a basis function  $\omega_S$  for some  $S$  of size  $k$ . This shows that  $\chi_{A,B} = \chi_{A,B}^{=k}$  over the cube. Finally, since  $\chi_{A,B}$  is harmonic, in order to show that  $\chi_{A,B} = \chi_{A,B}^{=k}$  in Gaussian space, it suffices to show that  $\chi_{A,B}(\alpha x) = \alpha^k \chi_{A,B}(x)$ , which is true since  $\chi_{A,B}$  is homogeneous of degree  $k$  as a polynomial.

**Degrees** The following results state several ways in which degree for functions on the slice behaves as expected.

First, we show that degree is subadditive.

**Lemma 2.11.** *Let  $f, g$  be harmonic multilinear polynomials, and let  $h$  be the unique harmonic multilinear polynomial agreeing with  $fg$  on the slice  $\binom{[n]}{k}$ . Then  $\deg h \leq \deg f + \deg g$ .*

*Proof.* We can assume that  $\deg f + \deg g \leq k$ , since otherwise the result is trivial.

Let  $E_i$  be the operator mapping a function  $\phi$  on the slice to the function  $\phi^{=i}$  on the slice. That is, we take the harmonic multilinear representation of  $\phi$ , extract the  $i$ 'th homogeneous part, and interpret the result as a function on the slice. Also, let  $E_{\leq d} = \sum_{i=0}^d E_i$ . A function  $\phi$  on the slice has degree at most  $d$  if and only if it is in the range of  $E_{\leq d}$ .

Qiu and Zhan [QZ07] (see also Tanaka [Tan09]) show that  $fg$  is in the range of  $E_{\leq \deg f} \circ E_{\leq \deg g}$ , where  $\circ$  is the Hadamard product. The operators  $E_i$  are the primitive idempotents of the Johnson association scheme (see, for example, [BIS4, §3.2]). Since the Johnson association scheme is Q-polynomial (cometric), the range of  $E_{\leq \deg f} \circ E_{\leq \deg g}$  equals the range of  $E_{\leq \deg f + \deg g}$ , and so  $\deg fg \leq \deg f + \deg g$ .  $\square$

As a corollary, we show that ‘‘harmonic projection’’ doesn’t increase the degree.

**Corollary 2.12.** *Let  $f$  be a multilinear polynomial, and let  $g$  be the unique harmonic multilinear polynomial agreeing with  $f$  on the slice  $\binom{[n]}{k}$ . Then  $\deg g \leq \deg f$ .*



*Proof.* When  $f = x_1$ , one checks that  $g$  is given by the linear polynomial

$$g = \frac{1}{n} \sum_{i=1}^n (x_1 - x_i) + \frac{k}{n}.$$

The corollary now follows from Lemma 2.11 and from the easy observation  $\deg(\alpha F + \beta G) \leq \max(\deg F, \deg G)$ .  $\square$

An immediate corollary is that degree is substitution-monotone.

**Corollary 2.13.** *Let  $f$  be a harmonic multilinear polynomial, let  $g(x_1, \dots, x_n) = f(x_1, \dots, x_{n-1}, b)$  for  $b \in \{0, 1\}$ , and let  $h$  be the unique harmonic multilinear polynomial agreeing with  $g$  on the slice  $\binom{[n]}{k}$ . Then  $\deg h \leq \deg f$ .*

**Noise operators** We have considered two noise operators,  $H_\rho$  and  $T_\rho = U_\rho$ . Both can be applied syntactically on all multilinear polynomials. The following result shows that both operators behave the same from the point of view of Lipschitz functions.

**Lemma 2.14.** *Let  $f$  be a multilinear polynomial of degree at most  $n/2$ . For  $\delta < 1/2$  and any  $C$ -Lipschitz functional  $\psi$ , and with respect to any exchangeable measure,*

$$|\mathbb{E}[\psi(H_{1-\delta}f)] - \mathbb{E}[\psi(U_{1-\delta}f)]| = O\left(\frac{C\delta^{-2}}{n}\|f\|\right).$$

*Proof.* Let  $\rho = 1 - \delta$ . Lemma 2.1 shows that

$$|\mathbb{E}[\psi(H_\rho f)] - \mathbb{E}[\psi(U_\rho f)]|^2 \leq C^2 \|H_\rho f - U_\rho f\|^2 = C^2 \sum_{d=0}^{n/2} (\rho^{d(1-(d-1)/n)} - \rho^d)^2 \|f^{=d}\|^2.$$

Let  $R(x) = \rho^x$ . Then  $\rho^{d(1-(d-1)/n)} - \rho^d = \frac{d(d-1)}{n}(-R'(x))$  for some  $x \in [d(1 - (d-1)/n), d]$ . For such  $x$ ,  $R'(x) = \rho^x(-\log \rho) \leq \rho^{d(1-(d-1)/n)}(2\delta) \leq \rho^{d/2}(2\delta)$ , using  $\delta < 1/2$  and  $d \leq n/2$ . Therefore

$$\rho^{d(1-(d-1)/n)} - \rho^d \leq 2\delta \frac{d(d-1)}{n} \rho^{d/2}.$$

The expansion  $x^2/(1-x)^3 = \sum_{d=0}^{\infty} \binom{d}{2} x^d$  implies that  $d(d-1)\rho^{d/2} \leq 2\rho/(1-\sqrt{\rho})^3$ . Since  $1 - \sqrt{\rho} = 1 - \sqrt{1-\delta} \geq \delta/2$ , we conclude that

$$\rho^{d(1-(d-1)/n)} - \rho^d \leq \frac{32\delta^{-2}}{n}.$$

The lemma follows.  $\square$

### 3 On harmonicity

Let  $f$  be a function on the Boolean cube  $\{0, 1\}^n$ , and let  $\tilde{f}$  be the unique harmonic function agreeing with  $f$  on the slice  $\binom{[n]}{pn}$ . We call  $\tilde{f}$  the *harmonic projection* of  $f$  with respect to the slice  $\binom{[n]}{pn}$ . In this section we prove Theorem 3.3, which shows that when  $f$  depends on  $(1 - \epsilon)\log n$  variables, it is close to its harmonic projection under the measure  $\mu_p$ . Together with Corollary 2.10, this allows us to deduce properties of  $f$  on the slice given properties of  $f$  on the Boolean cube, an idea formalized in Corollary 3.4.

We start by examining single monomials.

**Lemma 3.1.** *Let  $m$  be a monomial of degree  $d$ , and let  $f$  be the unique harmonic multilinear polynomial agreeing with  $m$  on  $\binom{[n]}{k}$  (where  $d \leq k \leq n/2$ ). Then  $\deg f = d$  and the coefficient  $c_m$  of  $m$  in  $f$  is*

$$c_m = \frac{n - 2d + 1}{n - d + 1} = 1 - O\left(\frac{d}{n}\right).$$

*Proof.* Without loss of generality we can assume that  $m = x_{n-d+1} \cdots x_n$ . Let  $B = \{n-d+1, \dots, n\}$ . Recall that the basis element  $\chi_B$  is equal to

$$\chi_B = \sum_{a_1 \neq \dots \neq a_d \in [n-d]} (x_{a_1} - x_{n-d+1}) \cdots (x_{a_d} - x_n).$$

Let  $f$  be the unique harmonic multilinear polynomial agreeing with  $m$  on  $\binom{[n]}{k}$ . Corollary 2.12 shows that  $\deg f \leq d$ . The coefficient  $\hat{f}(B)$  of  $\chi_B$  in the Fourier expansion of  $f$  is given by the formula  $\hat{f}(B) = \langle f, \chi_B \rangle / \|\chi_B\|^2$ . Since  $\deg f \leq d$ , it is not hard to check that in the Fourier expansion of  $f$ , the monomial  $m$  only appears in  $\chi_B$ . Therefore the coefficient  $c_m$  of  $m$  in  $f$  is

$$c_m = (-1)^d (n-d)^d \frac{\langle f, \chi_B \rangle}{\|\chi_B\|^2},$$

since there are  $(n-d)^d$  summands in the definition of  $\chi_B$ . The value of  $\|\chi_B\|^2$  is given by Lemma 2.6 and Lemma 2.9:

$$\|\chi_B\|^2 = \binom{n-d+1}{2} \binom{n-d}{2} \cdots \binom{n-2d+2}{2} 2^d \frac{k^d (n-k)^d}{n^{2d}}.$$

We proceed to compute  $\langle f, \chi_B \rangle$ . Let  $S \in \binom{[n]}{k}$ . If  $f(S)\chi_B(S) \neq 0$  then  $B \subseteq S$ , which happens with probability  $k^d/n^d$ . The number of non-zero terms (each equal to  $(-1)^d$ ) is the number of choices of  $a_1, \dots, a_d \notin S$ , namely  $(n-k)^d$ . Therefore  $\langle f, \chi_B \rangle = (-1)^d k^d (n-k)^d / n^d$ , and so

$$\begin{aligned} c_m &= (n-d)^d \cdot \frac{k^d (n-k)^d}{n^d} \cdot \frac{n^{2d}}{(n-d+1)(n-d)^2 \cdots (n-2d+2)^2 (n-2d+1) k^d (n-k)^d} \\ &= \frac{(n-d)^d (n-d)^d}{(n-d+1)(n-d)^2 \cdots (n-2d+2)^2 (n-2d+1)} \\ &= \frac{n-2d+1}{n-d+1}. \end{aligned}$$

Finally, since  $c_m \neq 0$  and  $\deg f \leq d$ , we can conclude that  $\deg f = d$ . □

As a consequence, we obtain a result on Fourier characters on the cube.

**Lemma 3.2.** *Let  $\omega = \omega_S$  be a Fourier character with respect to the measure  $\mu_p$  of degree  $d$ , and let  $\tilde{\omega}$  be the unique harmonic multilinear polynomial agreeing with  $\omega$  on  $\binom{[n]}{np}$  (where  $d \leq np \leq n/2$ ). For  $\pi \in \{\mu_p, \mathcal{G}_p\}$  we have*

$$\|\omega - \tilde{\omega}\|_\pi^2 = O\left(\frac{d^2}{p(1-p)n}\right).$$

*Proof.* Recall that

$$\omega = \frac{1}{(p(1-p))^{d/2}} \prod_{i \in S} (x_i - p).$$

Lemma 3.1 shows that

$$\tilde{\omega} = \frac{c}{(p(1-p))^{d/2}} \prod_{i \in S} x_i + \eta, \quad c = 1 - O\left(\frac{d}{n}\right),$$

where  $\eta$  involves other monomials. In fact, since  $\tilde{\omega}$  is harmonic, it is invariant under shifting all the variables by  $p$ , and so

$$\tilde{\omega} = c\omega + \eta',$$

where  $\eta'$  involves other characters. Due to orthogonality of characters we have

$$\|\tilde{\omega} - \omega\|_{\pi}^2 = \|\tilde{\omega}\|_{\pi}^2 - (2c - 1)\|\omega\|_{\pi}^2 = \|\tilde{\omega}\|_{\pi}^2 - (2c - 1) = \|\tilde{\omega}\|_{\pi}^2 - 1 + O\left(\frac{d}{n}\right).$$

Since  $\tilde{\omega}$  is harmonic, Corollary 2.10 allows us to estimate  $\|\tilde{\omega}\|_{\pi}^2$  given  $\|\tilde{\omega}\|_{\nu_{pn}}^2$ , which we proceed to estimate:

$$\begin{aligned} \|\tilde{\omega}\|_{\nu_{pn}}^2 &= \frac{1}{(p(1-p))^d} \sum_{t=0}^d \binom{d}{t} \frac{k^t(n-k)^{d-t}}{n^d} (1-p)^{2t} p^{2(d-t)} \\ &= \frac{1}{(p(1-p))^d} \sum_{t=0}^d \binom{d}{t} p^t (1-p)^{d-t} (1-p)^{2t} p^{2(d-t)} \left(1 \pm O\left(\frac{d^2}{p(1-p)n}\right)\right) \\ &= 1 \pm O\left(\frac{d^2}{p(1-p)n}\right). \end{aligned}$$

Corollary 2.10 shows that the same estimate holds even with respect to  $\pi$ , and so

$$\|\tilde{\omega} - \omega\|_{\pi}^2 = \|\tilde{\omega}\|_{\pi}^2 - 1 + O\left(\frac{d^2}{p(1-p)n}\right) = O\left(\frac{d^2}{p(1-p)n}\right). \quad \square$$

We can now conclude that a multilinear polynomial depending on a small number of variables is close to its harmonic projection.

**Theorem 3.3.** *Let  $f$  be a multilinear polynomial depending on  $d$  variables, and let  $\tilde{f}$  be the unique harmonic multilinear polynomial agreeing with  $f$  on  $\binom{[n]}{pn}$ , where  $d \leq pn \leq n/2$ . For  $\pi \in \{\mu_p, \mathcal{G}_p\}$  we have*

$$\|f - \tilde{f}\|_{\pi}^2 = O\left(\frac{d^2 2^d}{p(1-p)n}\right) \|f\|_{\pi}^2.$$

*Proof.* We can assume without loss of generality that  $f$  depends on the first  $d$  coordinates. Express  $f$  as a linear combination of characters:  $f = \sum_{S \subseteq [d]} \hat{f}(S) \omega_S$ . Clearly  $\tilde{f} = \sum_{S \subseteq [d]} \hat{f}(S) \tilde{\omega}_S$ , where  $\tilde{\omega}_S$  is the unique function agreeing with  $\omega_S$  on  $\binom{[n]}{pn}$ . Lemma 3.2 together with the Cauchy–Schwartz inequality shows that

$$\|f - \tilde{f}\|_{\pi}^2 \leq 2^d \sum_{S \subseteq [d]} \hat{f}(S)^2 O\left(\frac{d^2}{p(1-p)n}\right) = O\left(\frac{d^2 2^d}{p(1-p)n}\right) \|f\|_{\pi}^2.$$

This completes the proof. □

Combining Theorem 3.3 with Corollary 2.10, we show how to deduce properties of  $f$  on the slice given its properties on the cube.

**Corollary 3.4.** *Let  $f$  be a multilinear polynomial depending on  $d$  variables, and let  $\tilde{f}$  be the unique harmonic multilinear polynomial agreeing with  $f$  on  $\binom{[n]}{pn}$ , where  $d \leq pn \leq n/2$ . Suppose that  $\|f\|_{\mu_p}^2 = \|f\|_{\mathcal{G}_p}^2 \leq 1$ . For  $\pi \in \{\mu_p, \mathcal{G}_p\}$  we have:*

1.  $|\mathbb{E}_{\pi}[f] - \mathbb{E}_{\nu_{pn}}[\tilde{f}]| = O_p\left(\frac{d^{d/2}}{\sqrt{n}}\right).$
2.  $\|\tilde{f}\|_{\nu_{pn}} = 1 \pm O_p\left(\frac{d^{d/2}}{\sqrt{n}}\right).$
3.  $\mathbb{V}[\tilde{f}]_{\nu_{pn}} = \mathbb{V}[f]_{\pi} \pm O_p\left(\frac{d^{d/2}}{\sqrt{n}}\right).$

4. For all  $\rho \in [0, 1]$ ,  $\mathbb{S}_\rho^c[\tilde{f}]_{\nu_{pn}} = \mathbb{S}_\rho^c[f]_\pi \pm O_p\left(\frac{d2^{d/2}}{\sqrt{n}}\right)$ .

5. For all  $\ell \leq d$ ,  $\|\tilde{f}^{=\ell}\|_{\nu_{pn}} = \|f^{=\ell}\|_{\mu_p} \pm O_p\left(\frac{d2^{d/2}}{\sqrt{n}}\right)$ .

*Proof.* Throughout the proof, we are using Corollary 2.10 to convert information on  $\tilde{f}$  with respect to  $\pi$  to information on  $\tilde{f}$  with respect to  $\nu_{pn}$ . All calculations below are with respect to  $\pi$ .

For the first item, note that

$$|\mathbb{E}[f] - \mathbb{E}[\tilde{f}]| \leq \|f - \tilde{f}\|_1 \leq \|f - \tilde{f}\|_2 = O_p\left(\frac{d2^{d/2}}{\sqrt{n}}\right).$$

The second item follows from the triangle inequality

$$\|f\| - \|f - \tilde{f}\| \leq \|\tilde{f}\| \leq \|f\| + \|f - \tilde{f}\|.$$

For the third item, notice first that  $|\mathbb{E}[f]| \leq \|f\|_1 \leq \|f\|_2 = 1$ . The item now follows from the previous two.

The fourth item follows from the fact that  $\mathbb{S}_\rho^c$  is 1-Lipschitz, which in turn follows from the fact that  $\mathbb{S}_\rho^c[f] = \|T_{\sqrt{\rho}}f\|^2$  and that  $T_{\sqrt{\rho}}$  is a contraction.

For the fifth item, assume that  $f$  depends on the first  $d$  variables, and write  $f = \sum_{S \subseteq [d]} c_S \omega_S$ . We have

$$\tilde{f}^{=\ell} = \sum_{S \subseteq [d]} c_S (\tilde{\omega}_S)^{=\ell} = \widetilde{f}^{=\ell} + \sum_{|S| > \ell} c_S (\tilde{\omega}_S)^{=\ell}.$$

Lemma 3.2 shows that for  $|S| > \ell$ ,  $\|(\tilde{\omega}_S)^{=\ell}\|^2 \leq \|\omega_S - \tilde{\omega}_S\|^2 = O_p\left(\frac{d^2}{n}\right)$ . Therefore

$$\|\tilde{f}^{=\ell} - \widetilde{f}^{=\ell}\|^2 \leq 2^d \sum_{|S| > \ell} c_S^2 O_p\left(\frac{d}{n}\right) = O_p\left(\frac{d^2 2^d}{n}\right).$$

The fifth item now follows from the triangle inequality and the second item.  $\square$

## 4 Invariance principle

In the sequel, we assume that parameters  $p \in (0, 1/2]$  and  $n$  such that  $pn$  is an integer are given. The assumption  $p \leq 1/2$  is without loss of generality.

We will use big O notation in the following way:  $f = O_p(g)$  if for all  $n \geq N(p)$ , it holds that  $f \leq C(p)g$ , where  $N(p), C(p)$  are continuous in  $p$ . In particular, for any choice of  $p_L, p_H$  satisfying  $0 < p_L \leq p_H < 1$ , if  $p \in [p_L, p_H]$  then  $f = O(g)$ . Stated differently, as long as  $\lambda \leq p \leq 1 - \lambda$ , we have a uniform estimate  $f = O_\lambda(g)$ . Similarly, all constants depending on  $p$  (they will be of the form  $A_p$  for various letters  $A$ ) depend continuously on  $p$ .

**Proof sketch** Let  $\psi$  be a Lipschitz functional and  $f$  a harmonic multilinear polynomial of unit variance, low slice-influences, and low degree  $d$ . A simple argument shows that  $f$  also has low cube-influences, and this implies that

$$\mathbb{E}_{\nu_k}[\psi(f)] \approx \mathbb{E}_{\nu_{pn}}[\psi(f)] \pm O_p\left(\frac{|k - np|}{\sqrt{n}} \cdot \sqrt{d}\right).$$

The idea is now to apply the multidimensional invariance principle jointly to  $f$  and to  $S = \frac{x_1 + \dots + x_n - np}{\sqrt{p(1-p)n}}$ , deducing

$$\mathbb{E}_{\mu_p}[\psi(f)\mathbf{1}_{|S| \leq \sigma}] = \mathbb{E}_{\mathcal{G}_p}[\psi(f)\mathbf{1}_{|S| \leq \sigma}] \pm \epsilon.$$

An application of Lemma 2.7 shows that

$$\mathbb{E}_{\mathcal{G}_p}[\psi(f)\mathbf{1}_{|S|\leq\sigma}] = \Pr_{\mathcal{G}_p}[|S|\leq\sigma] \mathbb{E}_{\mathcal{G}_p}[\psi(f)].$$

Similarly,

$$\mathbb{E}_{\mu_p}[\psi(f)\mathbf{1}_{|S|\leq\sigma}] = \Pr_{\mu_p}[|S|\leq\sigma] (\mathbb{E}_{\mu_p}[\psi(f)] \pm O_p(\sigma\sqrt{d})).$$

Since  $\Pr_{\mathcal{G}_p}[|S|\leq\sigma] \approx \Pr_{\mu_p}[|S|\leq\sigma] = \Theta_p(\sigma)$ , we can conclude that

$$\mathbb{E}_{\nu_{pn}}[\psi(f)] \approx \mathbb{E}_{\mathcal{G}_p}[\psi(f)] \pm O_p(\sigma\sqrt{d} + \frac{\epsilon}{\sigma}).$$

By choosing  $\sigma$  appropriately, we balance the two errors and obtain our invariance principle.

For minor technical reasons, instead of using  $\mathbf{1}_{|S|\leq\sigma}$  we actually use a Lipschitz function supported on  $|S|\leq\sigma$ .

**Main theorems** Our main theorem is Theorem 4.6, proved in Section 4.1 on page 16. This is an invariance principle for low-degree, low-influence functions and Lipschitz functionals, comparing the uniform measure on the slice  $\nu_{pn}$  to the measure  $\mu_p$  on the Boolean cube and to the Gaussian measure  $\mathcal{G}_p$ .

Some corollaries appear in Section 4.2 on page 17. Corollary 4.7 gives a bound on the Lévy distance between the distributions  $f(\nu_{pn})$  and  $f(\mathcal{G}_p)$  for low-degree, low-influences functions. Corollary 4.9 gives a bound on the CDF distance between the distributions  $f(\nu_{pn})$  and  $f(\mathcal{G}_p)$  for low-degree, low-influences functions. Corollary 4.10 extends the invariance principle to functions of arbitrary degree to which a small amount of noise has been applied.

## 4.1 Main argument

We start by showing that from the point of view of L2 quantities, distributions similar to  $\mu_p$  behave similarly.

**Definition 4.1.** Let  $p \in (0, 1)$ . A parameter  $q \in (0, 1)$  is *p-like* if  $|p - q| \leq \sqrt{p(1-p)/n}$ . A distribution is *p-like* if it is one of the following:  $\mu_q, \nu_{qn}, \mathcal{G}_q$ , where  $q$  is *p-like*.

**Lemma 4.1.** Let  $f$  be a harmonic multilinear polynomial of degree  $d \leq \sqrt{n}$ , and let  $\pi_1, \pi_2$  be two *p-like* distributions. Then

$$\|f\|_{\pi_1}^2 = \|f\|_{\pi_2}^2 (1 \pm O_p(d/\sqrt{n})).$$

The same holds if we replace  $\|f\|^2$  with  $\text{Inf}_{ij}^s[f] = \|f - f^{(ij)}\|^2$  or  $\text{Inf}_i^c[f] = \|\frac{\partial f}{\partial x_i}\|^2$ .

Furthermore, there is a constant  $S_p$  such that if  $d \leq S_p\sqrt{n}$  then for all *p-like* distributions  $\pi_1, \pi_2$ ,

$$\frac{1}{2} \leq \frac{\|f\|_{\pi_1}^2}{\|f\|_{\pi_2}^2} \leq 2.$$

*Proof.* Let  $\alpha_D(q) = (2q(1-q))^D$ , where  $D \leq d$ . An easy calculation shows that  $\alpha'_D(q) = 2(1-2q)D(2q(1-q))^{D-1}$ , and in particular  $|\alpha'_D(q)| = O(D\alpha_D(q)/q(1-q))$ . It follows that for *p-like*  $q$ ,  $\alpha_D(q) = \alpha_D(p)(1 \pm O_p(D/\sqrt{n}))$ . Lemma 2.9 thus shows that for  $\pi \in \{\mu_q, \nu_{qn}, \mathcal{G}_q\}$  and all  $D \leq d$ ,

$$\|\chi_D\|_{\pi}^2 = \alpha_D(q) \left(1 \pm O_p\left(\frac{D^2}{n}\right)\right) = (2p(1-p))^D \left(1 \pm O_p\left(\frac{D}{\sqrt{n}} + \frac{D^2}{n}\right)\right).$$

Since  $D \leq d$  and  $d \leq \sqrt{n}$  implies  $d^2/n \leq d/\sqrt{n}$ , we conclude that

$$\|\chi_D\|_{\pi}^2 = (2p(1-p))^D \left(1 \pm O_p\left(\frac{d}{\sqrt{n}}\right)\right).$$

The lemma now follows from Corollary 2.8. □

We single out polynomials whose degree satisfies  $d \leq S_p \sqrt{n}$ .

**Definition 4.2.** A polynomial has *low degree* if its degree is at most  $S_p \sqrt{n}$ , where  $S_p$  is the constant in Lemma 4.1.

We can bound the cube-influence of a harmonic multilinear polynomial in terms of its slice-influence.

**Lemma 4.2.** *Let  $f$  be a harmonic multilinear polynomial of low degree  $d$ , and let  $\pi$  be a  $p$ -like distribution. For all  $i \in [n]$ , with respect to  $\pi$ :*

$$\text{Inf}_i^c[f] \leq O_p \left( \frac{d}{n} \mathbb{V}[f] + \text{Inf}_i^s[f] \right).$$

*Proof.* We will show that for the product measure  $\pi = \mu_p$  it holds that

$$\text{Inf}_i^c[f] \leq \frac{2d}{p(1-p)(n-d)} \mathbb{V}[f] + \frac{2n}{p(1-p)(n-d)} \text{Inf}_i^s[f]$$

which will imply the statement of the lemma by Lemma 4.1.

The idea is to come up with an explicit expression for  $\text{Inf}_i^s[f]$ . Let  $j \neq i$ . For  $S$  not containing  $i, j$  we have

$$\omega_S^{(ij)} = \omega_S, \quad \omega_{S \cup \{i\}}^{(ij)} = \omega_{S \cup \{j\}}, \quad \omega_{S \cup \{j\}}^{(ij)} = \omega_{S \cup \{i\}}, \quad \omega_{S \cup \{i, j\}}^{(ij)} = \omega_{S \cup \{i, j\}}.$$

Therefore

$$\text{Inf}_{ij}^s[f] = \|f - f^{(ij)}\|^2 = \sum_{i, j \notin S} (\hat{f}(S \cup \{i\}) - \hat{f}(S \cup \{j\}))^2.$$

On the other hand, we have

$$p(1-p) \text{Inf}_i^c[f] = \sum_{S \ni i} \hat{f}(S)^2 \leq \frac{1}{n-d} \sum_{S \ni i} \hat{f}(S)^2 (n - |S|) = \frac{1}{n-d} \sum_{j \neq i} \sum_{i, j \notin S} \hat{f}(S \cup \{i\})^2.$$

The L2 triangle inequality shows that  $\hat{f}(S \cup \{i\})^2 \leq 2\hat{f}(S \cup \{j\})^2 + 2(\hat{f}(S \cup \{i\}) - \hat{f}(S \cup \{j\}))^2$ , and so

$$\begin{aligned} p(1-p) \text{Inf}_i^c[f] &\leq \frac{2}{n-d} \sum_{j \neq i} \sum_{i, j \notin S} \hat{f}(S \cup \{j\})^2 + \frac{2}{n-d} \sum_{j \neq i} \text{Inf}_{ij}^s[f] \\ &\leq \frac{2p(1-p)}{n-d} \sum_{j \neq i} \text{Inf}_j^c[f] + \frac{2n}{n-d} \text{Inf}_i^s[f] \\ &\leq \frac{2d}{n-d} \mathbb{V}[f] + \frac{2n}{n-d} \text{Inf}_i^s[f], \end{aligned}$$

using the Poincaré inequality. Rearranging, we obtain the statement of the lemma.  $\square$

Using Lemma 4.2, we can show that the behavior of a low degree function isn't too sensitive to the value of  $q$  in  $\nu_{qn}$ .

**Lemma 4.3.** *Let  $f$  be a harmonic multilinear polynomial of low degree  $d$ , and let  $\ell$  be an integer such that  $\nu_\ell$  is  $p$ -like. For every  $C$ -Lipschitz functional  $\psi$ ,*

$$\left| \mathbb{E}_{x \sim \nu_\ell} [\psi(f(x))] - \mathbb{E}_{x \sim \nu_{\ell+1}} [\psi(f(x))] \right| = O_p \left( C \sqrt{\frac{d}{n} \mathbb{V}[f]_{\nu_{pn}}} \right).$$

*Proof.* Let  $q = \ell/n$ , which is  $p$ -like. For  $i \in [n]$ , let  $(X^i, Y^i)$  be the distribution obtained by choosing a random  $X^i \in \binom{[n]}{\ell} \setminus \{i\}$  and setting  $Y^i = X^i \cup \{i\}$ . Note that  $f(X^i) - f(Y^i) = (f - f^{[i]})(X^i)$ . Since  $\Pr_{\nu_\ell}[x_i = 0] = 1 - q$ , we have

$$\mathbb{E}[(f(X^i) - f(Y^i))^2] \leq (1 - q)^{-1} \text{Inf}_i^c[f]_{\nu_\ell} = O_p \left( \frac{d}{n} \mathbb{V}[f]_{\nu_{pn}} + \text{Inf}_i^s[f]_{\nu_{pn}} \right),$$

using Lemma 4.2 and Lemma 4.1.

Consider now the distribution  $(X, Y)$  supported on  $\binom{[n]}{\ell} \times \binom{[n]}{\ell+1}$  obtained by taking  $X \sim \nu_\ell$  and choosing  $Y \supset X$  uniformly among the  $n - \ell$  choices; note that  $Y \sim \nu_{\ell+1}$ . Since  $(X, Y)$  is a uniform mixture of the distributions  $(X^i, Y^i)$ , we deduce

$$\begin{aligned} \mathbb{E}[(f(X) - f(Y))^2] &\leq O_p \left( \frac{d}{n} \mathbb{V}[f]_{\nu_{pn}} + \frac{1}{n} \text{Inf}^s[f]_{\nu_{pn}} \right) \\ &\leq O_p \left( \frac{d}{n} \mathbb{V}[f]_{\nu_{pn}} \right), \end{aligned}$$

using the Poincaré inequality  $\text{Inf}^s[f] \leq d \mathbb{V}[f]$  (see Section 2.2). The lemma now follows along the lines of Lemma 2.1.  $\square$

We now apply a variant of the invariance principle for Lipschitz functionals due to Isaksson and Mossel.

**Proposition 4.4** ([IM12, Theorem 3.4]). *Let  $Q_1, \dots, Q_k$  be  $n$ -variate multilinear polynomials of degree at most  $d$  such that with respect to  $\mu_p$ ,  $\mathbb{V}[F_i] \leq 1$  and  $\text{Inf}_j^c[F_i] \leq \tau$  for all  $i \in [k]$  and  $j \in [n]$ . For any  $C$ -Lipschitz functional  $\Psi: \mathbb{R}^k \rightarrow \mathbb{R}$  (i.e., a function satisfying  $|\Psi(x) - \Psi(y)| \leq C\|x - y\|_2$ ),*

$$\left| \mathbb{E}_{\mu_p}[\Psi(Q_1, \dots, Q_k)] - \mathbb{E}_{\mathcal{G}_p}[\Psi(Q_1, \dots, Q_k)] \right| = O_k(C\rho_p^d \tau^{1/6}),$$

for some (explicit) constant  $\rho_p \geq 1$ .

**Lemma 4.5.** *Denote*

$$S = \frac{\sum_{i=1}^n x_i - np}{\sqrt{p(1-p)n}}.$$

*Let  $f$  be a harmonic multilinear polynomial of low degree  $d \geq 1$  such that with respect to  $\mu_p$ ,  $\mathbb{E}[f] = 0$ ,  $\mathbb{V}[f] \leq 1$  and  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$ . Suppose that  $\tau \leq R_p^{-d}$  and  $n \geq R_p^d$ , for some constant  $R_p$ . For any  $C$ -Lipschitz functional  $\psi$  such that  $\psi(0) = 0$  and  $B$ -Lipschitz functional  $\phi$  (where  $B \geq 1$ ) satisfying  $\|\phi\|_\infty \leq 1$ ,*

$$\left| \mathbb{E}_{\mu_p}[\psi(f)\phi(S)] - \mathbb{E}_{\mathcal{G}_p}[\psi(f)\phi(S)] \right| = CO_p \left( \sqrt{B}\rho_p^{d/2} \left( \tau + \frac{d}{n} \right)^{1/12} \right).$$

*The condition  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$  can be replaced by the condition  $\text{Inf}_i^c[f]_{\mu_p} \leq \tau$  for all  $i \in [n]$ .*

*Proof.* For  $M$  to be chosen later, define

$$\tilde{\psi}(x) = \begin{cases} -M & \text{if } \psi(x) \leq -M, \\ \psi(x) & \text{if } -M \leq \psi(x) \leq M, \\ M & \text{if } M \leq \psi(x). \end{cases}$$

It is not hard to check that  $\tilde{\psi}$  is also  $C$ -Lipschitz.

We are going to apply Proposition 4.4 with  $Q_1 = f$ ,  $Q_2 = S/\sqrt{p(1-p)n}$ , and  $\Psi(y_1, y_2) = \tilde{\psi}(y_1)\phi(y_2)$ . With respect to  $\mu_p$ ,  $\mathbb{V}[Q_2] = 1$  and  $\text{Inf}_i^c[Q_2] = 1/(p(1-p)n)$  for all  $i \in [n]$ . Lemma 4.2 shows that  $\text{Inf}_i^c[f] = O_p(\frac{d}{n} + \tau)$ , and so the cube-influences of  $Q_1, Q_2$  are bounded by  $O_p(\tau + \frac{d}{n})$ . Since

$$|\Psi(y_1, y_2) - \Psi(z_1, z_2)| \leq |\Psi(y_1, y_2) - \Psi(y_1, z_2)| + |\Psi(y_1, z_2) - \Psi(z_1, z_2)| \leq MB|y_2 - z_2| + C|y_1 - z_1|,$$

we see that  $\Psi$  is  $(MB + C)$ -Lipschitz. Therefore

$$|\mathbb{E}_{\mu_p}[\tilde{\psi}(f)\phi(S)] - \mathbb{E}_{\mathcal{G}_p}[\tilde{\psi}(f)\phi(S)]| = O_p\left((MB + C)\rho_p^d\left(\tau + \frac{d}{n}\right)^{1/6}\right).$$

Next, we want to replace  $\tilde{\psi}$  with  $\psi$ . For  $\pi \in \{\mu_p, \mathcal{G}_p\}$  we have

$$|\mathbb{E}_{\pi}[\tilde{\psi}(f)\phi(S)] - \mathbb{E}_{\pi}[\psi(f)\phi(S)]| \leq \mathbb{E}_{\pi}[|\psi(f)||\phi(S)|\mathbf{1}_{|\psi(f)| \geq M}] \leq C \mathbb{E}_{\pi}[|f|\mathbf{1}_{|f| \geq M/C}] \leq \frac{C^2}{M} \mathbb{E}_{\pi}[|f|^2] \leq \frac{C^2}{M}.$$

Therefore

$$|\mathbb{E}_{\mu_p}[\psi(f)\phi(S)] - \mathbb{E}_{\mathcal{G}_p}[\psi(f)\phi(S)]| = O_p\left((MB + C)\rho_p^d\left(\tau + \frac{d}{n}\right)^{1/6} + \frac{C^2}{M}\right).$$

Choosing  $M = C/\sqrt{B\rho_p^d(\tau + \frac{d}{n})^{1/6}}$  completes the proof. The conditions on  $\tau, n$  guarantee that  $\rho_p^d(\tau + \frac{d}{n})^{1/6} \leq 1$ , and so  $B \geq 1$  allows us to obtain the stated error bound.  $\square$

In order to finish the proof, we combine Lemma 4.5 with Lemma 4.3.

**Theorem 4.6.** *Let  $f$  be a harmonic multilinear polynomial of degree  $d$  such that with respect to  $\nu_{pn}$ ,  $\mathbb{V}[f] \leq 1$  and  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$ . Suppose that  $\tau \leq I_p^{-d}\delta^K$  and  $n \geq I_p^d/\delta^K$ , for some constants  $I_p, K$ . For any  $C$ -Lipschitz functional  $\psi$  and for  $\pi \in \{\mathcal{G}_p, \mu_p\}$ ,*

$$|\mathbb{E}_{\nu_{pn}}[\psi(f)] - \mathbb{E}_{\pi}[\psi(f)]| = O_p(C\delta).$$

The condition  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$  can be replaced by the condition  $\text{Inf}_i^c[f]_{\mu_p} \leq \tau$  for all  $i \in [n]$ .

*Proof.* We prove the theorem for  $\pi = \mathcal{G}_p$ . The version for  $\mu_p$  then follows from the classical invariance principle, using Lemma 4.2.

Replacing  $f$  with  $f - \mathbb{E}[f]$  (recall that the expectation of  $f$  is the same with respect to both  $\mu_p$  and  $\pi$ ) doesn't change the variance and influences of  $f$ , so we can assume without loss of generality that  $\mathbb{E}[f] = 0$ . Similarly, we can replace  $\psi$  with  $\psi - \psi(0)$  without affecting the quantity  $\mathbb{E}_{\nu_{pn}}[\psi(f)] - \mathbb{E}_{\mu_p}[\psi(f)]$ , and so we can assume without loss of generality that  $\psi(0) = 0$ .

For a parameter  $\sigma \leq 1$  to be chosen later, define a function  $\phi$  supported on  $[-\sigma, \sigma]$  by

$$\phi(x) = \begin{cases} 1 + x/\sigma & \text{if } -\sigma \leq x \leq 0, \\ 1 - x/\sigma & \text{if } 0 \leq x \leq \sigma. \end{cases}$$

Note that  $\|\phi\|_{\infty} = 1$  and that  $\phi$  is  $(1/\sigma)$ -Lipschitz. Lemma 4.5 (together with Lemma 4.1) shows that

$$|\mathbb{E}_{\mu_p}[\psi(f)\phi(S)] - \mathbb{E}_{\mathcal{G}_p}[\psi(f)\phi(S)]| = CO_p\left(\sigma^{-1/2}\rho_p^{d/2}\left(\tau + \frac{d}{n}\right)^{1/12}\right) = CO_p(\sigma^{-1/2}\rho_p^{d/2}(\tau^{1/12} + n^{-1/24})),$$

assuming  $\tau \leq R_p^{-d}$  and  $n \geq R_p^d$  (the condition on  $n$  implies that  $d$  is low degree).

Let  $\alpha$  be the distribution of  $x_1 + \dots + x_n$  under  $\mathcal{G}_p$ . Lemma 2.2 and Lemma 2.7 show that

$$\mathbb{E}_{\mathcal{G}_p}[\psi(f)\phi(S)] = \mathbb{E}_{q \sim \alpha} \left[ \mathbb{E}_{\gamma_{p,q}}[\psi(f)]\phi\left(\frac{q-np}{\sqrt{p(1-p)n}}\right) \right] = \mathbb{E}_{\mathcal{G}_p}[\psi(f)] \mathbb{E}_{q \sim \alpha} \left[ \phi\left(\frac{q-np}{\sqrt{p(1-p)n}}\right) \right] = \mathbb{E}_{\mathcal{G}_p}[\psi(f)] \mathbb{E}_{\mathcal{G}_p}[\phi(S)].$$



Similarly, Lemma 4.3 shows that

$$\begin{aligned}
& \left| \mathbb{E}_{\mu_p} [\psi(f)\phi(S)] - \mathbb{E}_{\nu_{np}} [\psi(f)] \mathbb{E}_{\mu_p} [\phi(S)] \right| \\
& \leq \sum_{|k-np| \leq \sigma \sqrt{p(1-p)n}} \frac{\Pr[S = \frac{k-np}{\sqrt{p(1-p)n}}] \mu_p}{\sqrt{p(1-p)n}} \left| \phi\left(\frac{k-np}{\sqrt{p(1-p)n}}\right) \mathbb{E}_{\nu_k} [\psi(f)] - \mathbb{E}_{\nu_{pn}} [\psi(f)] \right| \\
& \leq \sum_{|k-np| \leq \sigma \sqrt{p(1-p)n}} \frac{\Pr[S = \frac{k-np}{\sqrt{p(1-p)n}}] \mu_p}{\sqrt{p(1-p)n}} |k-np| O_p \left( C \sqrt{\frac{d}{n}} \right) \\
& \leq \mathbb{E}_{\mu_p} [\phi(S)] O_p(C\sigma\sqrt{d}).
\end{aligned}$$

Therefore

$$\left| \mathbb{E}_{\mathcal{G}_p} [\psi(f)] \mathbb{E}_{\mathcal{G}_p} [\phi(S)] - \mathbb{E}_{\nu_{np}} [\psi(f)] \mathbb{E}_{\mu_p} [\phi(S)] \right| \leq O_p(C\sigma^{-1/2} \rho_p^{d/2} (\tau^{1/12} + n^{-1/24})) + O_p(C\sigma\sqrt{d} \mathbb{E}_{\mu_p} [\phi(S)]).$$

Proposition 4.4 shows that

$$\left| \mathbb{E}_{\mathcal{G}_p} [\phi(S)] - \mathbb{E}_{\mu_p} [\phi(S)] \right| = O_p(\sigma^{-1/2} n^{-1/6}).$$

Moreover,  $\mathbb{E}_{\mathcal{G}_p} [\psi(f)] \leq C \mathbb{E}_{\mathcal{G}_p} [|f|] = O_p(C)$ . It follows that

$$\left| \mathbb{E}_{\mu_p} [\phi(S)] \mathbb{E}_{\mathcal{G}_p} [\psi(f)] - \mathbb{E}_{\nu_{np}} [\psi(f)] \mathbb{E}_{\mu_p} [\phi(S)] \right| \leq O_p(C\sigma^{-1/2} \rho_p^{d/2} (\tau^{1/12} + n^{-1/24})) + O_p(C\sigma\sqrt{d} \mathbb{E}_{\mu_p} [\phi(S)]).$$

It is not hard to check that  $\mathbb{E}_{\mathcal{G}_p} [\phi(S)] = \Theta_p(\sigma)$ , and so for  $n \geq A_p \sigma^{-9}$  we have  $\mathbb{E}_{\mu_p} [\phi(S)] = \Theta_p(\sigma)$ , implying

$$\left| \mathbb{E}_{\mathcal{G}_p} [\psi(f)] - \mathbb{E}_{\nu_{np}} [\psi(f)] \right| \leq O_p(C\sigma^{-3/2} \rho_p^{d/2} (\tau^{1/12} + n^{-1/24})) + C\sigma\sqrt{d}.$$

Choosing  $\sigma = \rho_p^{d/5} (\tau^{1/12} + n^{-1/24})^{2/5} / d^{1/5}$ , we obtain

$$\left| \mathbb{E}_{\mathcal{G}_p} [\psi(f)] - \mathbb{E}_{\nu_{np}} [\psi(f)] \right| \leq O_p(C\rho_p^{d/5} (\tau^{1/30} + n^{-1/60}) d^{3/10}).$$

It is not hard to check that if  $d \leq B_p n^\beta$  and  $n \geq M_p$  then  $n \geq A_p \sigma^{-9}$ , and that if  $\tau, n^{-1} \leq \rho_p^{-\gamma d}$  then  $\sigma \leq 1$ ; these are the conditions necessary for our estimate to hold. In fact, for an appropriate choice of  $\gamma_p \geq \gamma$ , the condition  $n \geq \rho_p^{\gamma_p d}$  implies the condition  $d \leq B_p n^\beta$ , and furthermore allows us to estimate  $n^{-1/60} d^{3/10} = O_p(n^{-1/70})$  (say), and to control the other error term similarly. This completes the proof of the theorem.  $\square$

## 4.2 Corollaries

Theorem 4.6 allows us to bound the Lévy distance between the distribution of a low degree polynomial with respect to  $\nu_{pn}$  and the distribution of the same polynomial with respect to  $\mathcal{G}_p$  or  $\mu_p$ . This is the analog of [MOO10, Theorem 3.19].

**Corollary 4.7.** *Let  $f$  be a harmonic multilinear polynomial of degree  $d$  such that with respect to  $\nu_{pn}$ ,  $\mathbb{V}[f] \leq 1$  and  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$ . There are parameters  $X_p, X$  such that for any  $0 < \epsilon < 1/2$ , if  $\tau \leq X_p^{-d} \epsilon^X$  and  $n \geq X_p^d / \epsilon^X$  then the Lévy distance between  $f(\nu_{pn})$  and  $f(\pi)$  is at most  $\epsilon$ , for  $\pi \in \{\mathcal{G}_p, \mu_p\}$ . In other words, for all  $\sigma$ ,*

$$\Pr_{\nu_{pn}} [f \leq \sigma - \epsilon] - \epsilon \leq \Pr_{\pi} [f \leq \sigma] \leq \Pr_{\nu_{pn}} [f \leq \sigma + \epsilon] + \epsilon.$$

*Proof.* Given  $\sigma$  and  $\epsilon$ , define a function  $\psi$  by

$$\psi(x) = \begin{cases} 0 & \text{if } x \leq \sigma, \\ \frac{x-\sigma}{\epsilon} & \text{if } \sigma \leq x \leq \sigma + \epsilon, \\ 1 & \text{if } x \geq \sigma + \epsilon. \end{cases}$$

Note that  $\psi$  is  $(1/\eta)$ -Lipschitz. Theorem 4.6 shows that if  $\tau \leq I_p^{-d}\delta^K$  and  $n \geq I_p^d/\delta^k$ ,

$$\Pr_{\pi}[f \leq \sigma] - \Pr_{\nu_{pn}}[f \leq \sigma + \epsilon] \leq \mathbb{E}_{\pi}[\psi(f)] - \mathbb{E}_{\nu_{pn}}[\psi(f)] = O_p(\delta/\epsilon).$$

We can similarly get a bound in the other direction. To complete the proof, choose  $\delta = c_p\epsilon^2$  for an appropriate  $c_p$ .  $\square$

Using the Carbery–Wright theorem, we can bound the actual CDF distance. This is the analog of [MOO10, Theorem 3.19(30)].

**Proposition 4.8** (Carbery–Wright). *Let  $f$  be a polynomial of degree at most  $d$  such that  $\mathbb{V}[f]_{\mathcal{G}_p} = 1$ . Then for all  $\epsilon > 0$  and all  $x$ ,*

$$\Pr_{\mathcal{G}_p}[|f - x| \leq \epsilon] = O(d\epsilon^{1/d}).$$

**Corollary 4.9.** *Let  $f$  be a harmonic multilinear polynomial of degree  $d$  such that with respect to  $\nu_{pn}$ ,  $\mathbb{V}[f] = 1$  and  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$ . There are parameters  $Y_p, Y$  such that for any  $0 < \epsilon < 1/2$ , if  $\tau \leq (Y_p d)^{-d}\epsilon^{Yd}$  and  $n \geq (Y_p d)^d/\epsilon^{Yd}$  then the CDF distance between  $f(\nu_{pn})$  and  $f(\pi)$  is at most  $\epsilon$ , for  $\pi \in \{\mathcal{G}_p, \mu_p\}$ . In other words, for all  $\sigma$ ,*

$$|\Pr_{\nu_{pn}}[f \leq \sigma] - \Pr_{\pi}[f \leq \sigma]| \leq \epsilon.$$

*Proof.* It is enough to prove the corollary for  $\pi = \mathcal{G}_p$ , the other case following from the corresponding result in the classical setting. Corollary 4.7 and the Carbery–Wright theorem show that for  $\tau \leq X_p^{-d}\eta^X$  and  $n \geq X_p^d/\eta^X$  we have

$$\Pr_{\nu_{pn}}[f \leq \sigma] \leq \Pr_{\mathcal{G}_p}[f \leq \sigma + \eta] + \eta \leq \Pr_{\mathcal{G}_p}[f \leq \sigma] + O_p(d\eta^{1/d}).$$

We can similarly obtain a bound from the other direction. To complete the proof, choose  $\eta = c_p(\epsilon/d)^d$  for an appropriate  $c_p$ .  $\square$

All bounds we have considered so far apply only to low degree functions. We can get around this restriction by applying a small amount of noise to the functions before applying the invariance principle itself. This is the analog of [MOO10, Theorem 3.20] Even though the natural noise operator to apply on the slice is  $H_\rho$ , from the point of view of applications it is more natural to use  $U_\rho$  (which we apply syntactically). Lemma 2.14 shows that the difference between the two noise operators is small.

**Corollary 4.10.** *Let  $f$  be a harmonic multilinear polynomial such that with respect to  $\nu_{pn}$ ,  $\mathbb{V}[f] \leq 1$  and  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$ . There is a parameter  $Z_p$  such that for any  $0 < \epsilon < 1/2$  and  $0 < \delta < 1/2$ , if  $\tau \leq \epsilon^{Z_p/\delta}$  and  $n \geq 1/\epsilon^{Z_p/\delta}$  then for  $\pi \in \{\mathcal{G}_p, \mu_p\}$ ,*

$$|\mathbb{E}_{\nu_{pn}}[\psi(U_{1-\delta}f)] - \mathbb{E}_{\pi}[\psi(U_{1-\delta}f)]| = O_p(C\epsilon).$$

*Proof.* Let  $g = U_{1-\delta}f$ . Let  $d$  be a low degree to be decided, and split  $g = g^{\leq d} + g^{> d}$ . With respect to  $\nu_{pn}$ ,  $\|g^{> d}\|^2 = \sum_{t>d} (1-\delta)^{2t} \|f^t\|^2 \leq (1-\delta)^{2d}$ . On the other hand, Theorem 4.6 shows that if  $\tau \leq I_p^{-d}\epsilon^K$  and  $n \geq I_p^d/\epsilon^K$  then

$$|\mathbb{E}_{\nu_{pn}}[\psi(g^{\leq d})] - \mathbb{E}_{\pi}[\psi(g^{\leq d})]| = O_p(C\epsilon).$$

Since  $\|g - g^{\leq d}\| = \|g^{> d}\|$ , Lemma 2.1 and Lemma 4.1 show that as long as the degree  $d$  is low,

$$|\mathbb{E}_{\nu_{pn}}[\psi(g)] - \mathbb{E}_{\pi}[\psi(g)]| = O_p(C\epsilon + C(1 - \delta)^d) = O_p(C\epsilon + Ce^{-\delta d}).$$

Choosing  $d = \log(1/\epsilon)/\delta$ , the resulting error is  $O_p(C\epsilon)$ . This degree is low if  $\log(1/\epsilon)/\delta \leq S_p\sqrt{n}$ , a condition which is implied by the stated condition on  $n$ .  $\square$

## 5 Majority is stablest

Recall Borell's theorem.

**Theorem 5.1** (Borell [Bor85]). *Let  $f: \mathbb{R}^n \rightarrow [0, 1]$  have expectation  $\mu$  with respect to  $N(0, 1)^n$ . Then  $\mathbb{S}_\rho^c[f] \leq \Gamma_\rho(\mu)$ , where  $\Gamma_\rho(\mu)$  is the probability that two  $\rho$ -correlated Gaussians be at most  $\Phi^{-1}(\mu)$ .*

Borell's theorem remains true if we replace the standard Gaussian with  $\mathcal{G}_p$ . Indeed, given a function  $f$ , define a new function  $g$  by  $g(x) = f(\sqrt{p(1-p)}x + p)$ . If  $x \sim N(0, 1)$  then  $\sqrt{p(1-p)}x + p \sim \mathcal{G}_p$ , and so  $\mathbb{E}[f]_{\mathcal{G}_p} = \mathbb{E}[g]_{N(0,1)}$ . Similarly,  $\mathbb{S}_\rho^c[f]_{\mathcal{G}_p} = \mathbb{S}_\rho^c[g]_{N(0,1)}$ . Indeed, if  $y = \rho x + \sqrt{1 - \rho^2} N(0, 1)$  then

$$\begin{aligned} \sqrt{p(1-p)}y + p &= p + \rho\sqrt{p(1-p)}x + \sqrt{1 - \rho^2} N(0, p(1-p)) \\ &= (1 - \rho)p + \rho(\sqrt{p(1-p)}x + p) + \sqrt{1 - \rho^2} N(0, p(1-p)). \end{aligned}$$

Therefore Borell's theorem for  $f$  and  $\mathcal{G}_p$  follows from the theorem for  $g$  and  $N(0, 1)$ .

*Majority is stablest* states that a similar bound essentially holds for all low influence functions on the slice. This result was originally proved using the invariance principle in [MOO10]. An alternative inductive proof appears in [DMN13].

It is known (see for example [MOO10]) that the bound  $\Phi^{-1}(\mu)$  is achieved by threshold functions. Corollary 2.10 together with Lemma 2.14 shows that threshold functions achieve the bound also on the slice. Indeed, take a threshold function  $f$  on  $d$  variables such that with respect to  $\mu_p$ ,  $\mathbb{E}[f] = \mu$  and  $\mathbb{S}_\rho^c[f] \geq \Gamma_\rho(\mu) - \epsilon$ . Let  $\tilde{f}$  be the restriction of  $f$  to the slice  $\binom{[n]}{pn}$ . Corollary 2.10 shows that  $\mathbb{E}[\tilde{f}] = \mu \pm o_n(1)$  and  $\mathbb{S}_\rho^c[\tilde{f}] = \mathbb{S}_\rho^c[f] \pm o_n(1)$ . Lemma 2.14 shows that  $\mathbb{S}_\rho^s[\tilde{f}] = \mathbb{S}_\rho^c[f] \pm o_n(1)$ . Therefore for large  $n$ ,  $\mathbb{E}[\tilde{f}] \approx \mu$  and  $\mathbb{S}_\rho^c[\tilde{f}] \geq \Gamma_\rho(\mu) - 2\epsilon$ .

Our proof of majority is stablest closely follows the proof of [MOO10, Theorem 4.4] presented in [O'D14, §11.7]. We need an auxiliary result on  $\Gamma_\rho$ .

**Proposition 5.2** ([MOO10, Lemma B.6]). *For each  $\rho$ , the function  $\Gamma_\rho$  defined in Theorem 5.1 is 2-Lipschitz.*

**Theorem 5.3.** *Let  $f: \binom{[n]}{pn} \rightarrow [0, 1]$  have expectation  $\mu$  and satisfy  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$ . For any  $0 < \rho < 1$ , we have*

$$\mathbb{S}_\rho^s[f] \leq \Gamma_\rho(\mu) + O_{p,\rho} \left( \frac{\log \log \frac{1}{\alpha}}{\log \frac{1}{\alpha}} \right) + O_\rho \left( \frac{1}{n} \right), \text{ where } \alpha = \min(\tau, \frac{1}{n}).$$

*The condition  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$  can be replaced by the condition  $\text{Inf}_i^c[f]_{\mu_p} \leq \tau$  for all  $i \in [n]$ .*

*Proof.* We identify  $f$  with the unique harmonic multilinear polynomial agreeing with it on  $\binom{[n]}{pn}$ . For a parameter  $0 < \delta < 1/2$  to be chosen later, let  $g = H_{1-\delta}f$ . Note that the range of  $g$  on  $\binom{[n]}{pn}$  is included in  $[0, 1]$  as well, since  $H_{1-\delta}$  is an averaging operator. We have

$$\mathbb{S}_\rho^s[f] - \mathbb{S}_\rho^s[g] = \mathbb{S}_\rho^s[f] - \mathbb{S}_{\rho(1-\delta)^2}^s[f] = \sum_{d=0}^{pn} \rho^{d(1-(d-1)/n)} (1 - (1-\delta)^{2d(1-(d-1)/n)}) \|f^{=d}\|^2.$$

Since  $d \leq n/2$ , we have

$$\rho^{d(1-(d-1)/n)}(1 - (1 - \delta)^{2d(1-(d-1)/n)}) \leq \rho^{d/2}(1 - (1 - \delta)^{2d}) \leq 2\delta d\rho^{d/2}.$$

The expansion  $x/(1-x)^2 = \sum_d dx^d$  shows that  $d\rho^{d/2} \leq \sqrt{\rho}/(1-\sqrt{\rho})^2$ , and so

$$|\mathbb{S}_\rho^c[f] - \mathbb{S}_\rho^c[g]| \leq 2\delta \frac{\sqrt{\rho}}{(1-\sqrt{\rho})^2}. \quad (1)$$

From now on we concentrate on estimating  $\mathbb{S}_\rho^c[g]$ .

Define the clumped square function Sq by

$$\text{Sq}(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x^2 & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } x \geq 1. \end{cases}$$

It is not difficult to check that Sq is 2-Lipschitz. Corollary 4.10 together with Lemma 2.14 shows that for all  $\epsilon > 0$ , if  $\tau, \frac{1}{n} \leq \epsilon^{Z_p/\delta}$  then

$$|\mathbb{S}_\rho^s[g] - \mathbb{S}_\rho^c[g]| = |\mathbb{E}_{\nu_{pn}}[\text{Sq}(H_{\sqrt{\rho}}g)] - \mathbb{E}_{\mathcal{G}_p}[\text{Sq}(U_{\sqrt{\rho}}g)]| = O_p(\epsilon) + O\left(\frac{(1-\sqrt{\rho})^{-2}}{n}\right). \quad (2)$$

We would like to apply Borell's theorem in order to bound  $\mathbb{S}_\rho^c[g]$ , but  $g$  is not necessarily bounded by  $[0, 1]$  on  $\mathbb{R}^n$ . In order to handle this, we define the function  $\tilde{g} = \max(0, \min(1, g))$ , which is bounded by  $[0, 1]$ . Let  $\text{dist}_{[0,1]}$  be the function which measures the distance of a point  $x$  to the interval  $[0, 1]$ . The function  $\text{dist}_{[0,1]}$  is clearly 1-Lipschitz, and so Corollary 4.10 implies that under the stated assumptions on  $\tau, \frac{1}{n}$ , we have

$$\mathbb{E}_{\mathcal{G}_p}[|g - \tilde{g}|] = \mathbb{E}_{\mathcal{G}_p}[\text{dist}_{[0,1]}(g)] = |\mathbb{E}_{\nu_{pn}}[\text{dist}_{[0,1]}(g)] - \mathbb{E}_{\mathcal{G}_p}[\text{dist}_{[0,1]}(g)]| = O_p(\epsilon).$$

Since  $U_{\sqrt{\rho}}$  is an averaging operator and Sq is 2-Lipschitz, we conclude that

$$|\mathbb{S}_\rho^c[g] - \mathbb{S}_\rho^c[\tilde{g}]| = |\mathbb{E}_{\mathcal{G}_p}[\text{Sq}(U_{\sqrt{\rho}}g)] - \mathbb{E}_{\mathcal{G}_p}[\text{Sq}(U_{\sqrt{\rho}}\tilde{g})]| = O_p(\epsilon). \quad (3)$$

Lemma 2.6 shows that  $\mathbb{E}_{\mathcal{G}_p}[g] = \mathbb{E}_{\nu_{pn}}[g] = \mathbb{E}_{\nu_{pn}}[f] = \mu$ , and so  $|\mathbb{E}_{\mathcal{G}_p}[g] - \mu| = O_p(\epsilon)$ . Proposition 5.2 implies that  $\Gamma_\rho(\mathbb{E}[\tilde{g}]) \leq \Gamma_\rho(\mu) + O_p(\epsilon)$ . Applying Borell's theorem (Theorem 5.1), we deduce that

$$\mathbb{S}_\rho^c[\tilde{g}] \leq \Gamma_\rho(\mathbb{E}[\tilde{g}]) \leq \Gamma_\rho(\mu) + O_p(\epsilon). \quad (4)$$

Putting (1),(2),(3),(4) together, we conclude that

$$\mathbb{S}_\rho^s[f] \leq \Gamma_\rho(\mu) + O_p(\epsilon) + O\left(\frac{(1-\sqrt{\rho})^{-2}}{n}\right) + 2\delta \frac{\sqrt{\rho}}{(1-\sqrt{\rho})^2}.$$

Taking  $\delta = \epsilon$ , we obtain

$$\mathbb{S}_\rho^s[f] \leq \Gamma_\rho(\mu) + O_{p,\rho}(\epsilon) + O_p\left(\frac{1}{n}\right).$$

The bounds on  $\tau, \frac{1}{n}$  now become  $\tau, \frac{1}{n} \leq \epsilon^{Z_p/\epsilon}$ , from which we can extract the theorem.  $\square$

## 6 Bourgain's theorem

Bourgain's theorem in Gaussian space gives a lower bound on the tails of Boolean functions (in this section, *Boolean* means that the range of the function is  $\{\pm 1\}$ ). We quote its version from [KKO14, Theorem 2.11].

**Theorem 6.1** (Bourgain). *Let  $f: \mathbb{R}^n \rightarrow \{\pm 1\}$ . For any  $k \geq 1$  we have, with respect to Gaussian measure  $N(0, 1)$ ,*

$$\|f^{>k}\|^2 = \Omega\left(\frac{\mathbb{V}[f]}{\sqrt{k}}\right).$$

While the theorem is stated for  $N(0, 1)$ , it holds for  $\mathcal{G}_p$  as well. Indeed, given a function  $f$ , define a new function  $g$  by  $g(x) = f(\sqrt{p(1-p)}x + p)$ . If  $x \sim N(0, 1)$  then  $\sqrt{p(1-p)}x + p \sim \mathcal{G}_p$ , and so  $\mathbb{V}[f]_{\mathcal{G}_p} = \mathbb{V}[g]_{N(0,1)}$ . Our definition of  $f^{=i}$  for  $\mathcal{G}_p$  makes it clear that  $g^{=i}(x) = f^{=i}(\sqrt{p(1-p)}x + p)$ , where  $g^{=i}$  is the degree  $i$  homogeneous part of  $g$ . This implies that  $\|f^{>k}\|_{\mathcal{G}_p}^2 = \|g^{>k}\|_{N(0,1)}^2$ . Therefore Bourgain's theorem for  $f$  and  $\mathcal{G}_p$  follows from the theorem for  $g$  and  $N(0, 1)$ .

Following closely the proof of [KKO14, Theorem 3.1], we can prove a similar result for the slice.

**Theorem 6.2.** *Fix  $k \geq 2$ . Let  $f: \binom{[n]}{pn} \rightarrow \{\pm 1\}$  satisfy  $\text{Inf}_i^s[f^{\leq k}] \leq \tau$  for all  $i \in [n]$ . For some constants  $W_{p,k}, C$ , if  $\tau \leq W_{p,k}^{-1} \mathbb{V}[f]^C$  and  $n \geq W_{p,k}/\mathbb{V}[f]^C$  then*

$$\|f^{>k}\|^2 = \Omega\left(\frac{\mathbb{V}[f]}{\sqrt{k}}\right).$$

*The condition  $\text{Inf}_i^s[f] \leq \tau$  for all  $i \in [n]$  can be replaced by the condition  $\text{Inf}_i^c[f]_{\mu_p} \leq \tau$  for all  $i \in [n]$ .*

*Proof.* We treat  $f$  as a harmonic multilinear polynomial. Since  $f$  is Boolean, working over  $\nu_{pn}$  we have

$$\|f^{>k}\|^2 = \|f - f^{\leq k}\|^2 \geq \|f^{\leq k} - \text{sgn}(f^{\leq k})\|^2 = \|f^{\leq k}\|^2 + 1 - 2\mathbb{E}[|f^{\leq k}|].$$

Lemma 4.1 shows that  $\|f^{\leq k}\|_{\mathcal{G}_p}^2 = \|f^{\leq k}\|^2(1 \pm O_p(k^2/n))$ . Since the absolute value function is 1-Lipschitz, Theorem 4.6 applied to the parameter  $\delta > 0$  shows that if  $\tau \leq I_p^{-k} \delta^K$  and  $n \geq I_p^k / \delta^K$  then then  $|\mathbb{E}_{\nu_{pn}}[|f^{\leq k}|] - \mathbb{E}_{\mathcal{G}_p}[|f^{\leq k}|]| = O_p(\delta)$ . This shows that

$$\|f^{>k}\|_{\nu_{pn}}^2 \geq \|f^{\leq k} - \text{sgn}(f^{\leq k})\|_{\mathcal{G}_p}^2 - O_p\left(\delta + \frac{k^2}{n}\right).$$

Let  $g = \text{sgn}(f^{\leq k})$ . With respect to Gaussian measure  $\mathcal{G}_p$ ,  $\|f^{\leq k} - g\|^2 \geq \|g^{>k}\|^2 = \Omega(\mathbb{V}[g]/\sqrt{k})$ , using Bourgain's theorem (Theorem 6.1). Putting everything together, we conclude that

$$\|f^{>k}\|_{\nu_{pn}}^2 \geq \Omega\left(\frac{\mathbb{V}[g]_{\mathcal{G}_p}}{\sqrt{k}}\right) - O_p\left(\delta + \frac{k^2}{n}\right). \quad (5)$$

It remains to lower bound  $\mathbb{V}[g]_{\mathcal{G}_p}$ . Note first that over  $\nu_{pn}$ ,  $\mathbb{V}[f] = 4\Pr[f = 1]\Pr[f = -1]$ , and so  $\Pr[f = 1], \Pr[f = -1] \geq \mathbb{V}[f]/4$ . We can furthermore assume that

$$\Pr_{\nu_{pn}}[f^{\leq k} \geq \frac{2}{3}], \Pr_{\nu_{pn}}[f^{\leq k} \leq -\frac{2}{3}] \geq \frac{\mathbb{V}[f]}{8},$$

since if for example  $\Pr[f^{\leq k} \geq \frac{2}{3}] \leq \mathbb{V}[f]/8$  then with probability at least  $\mathbb{V}[f]/8$  we have  $f = 1$  and  $f^{\leq k} < \frac{2}{3}$ , and so  $\|f^{>k}\|^2 = \|f - f^{\leq k}\|^2 \geq \frac{1}{9} \cdot \mathbb{V}[f]/8 = \Omega(\mathbb{V}[f])$ . Corollary 4.7 applied with  $\epsilon = \mathbb{V}[f]/16 \leq 1/3$  shows that for an appropriate  $c$ , if  $\tau \leq X_p^{-k}/c$  and  $n \geq cX_p^k$  then

$$\Pr_{\mathcal{G}_p}[f^{\leq k} \geq \frac{1}{3}], \Pr_{\mathcal{G}_p}[f^{\leq k} \leq -\frac{1}{3}] \geq \frac{\mathbb{V}[f]}{16},$$

and so  $\mathbb{V}_{\mathcal{G}_p}[g] \geq 4\mathbb{V}_{\nu_{pn}}[f]/16(1 - \mathbb{V}_{\nu_{pn}}[f]/16) = \Omega(\mathbb{V}_{\nu_{pn}}[f])$ . Combining this with (5) shows that under  $\nu_{pn}$ ,

$$\|f^{>k}\|^2 \geq \Omega\left(\frac{\mathbb{V}[f]}{\sqrt{k}}\right) - O_p\left(\delta + \frac{k^2}{n}\right). \quad (6)$$

Choosing  $\delta = c_p \mathbb{V}[f]/\sqrt{k}$  for an appropriate  $c_p$  completes the proof.  $\square$

We do not attempt to match here [KKO14, Theorem 3.2], which has the best constant in front of  $\mathbb{V}[f]/\sqrt{k}$ .

## 7 Kindler–Safra theorem

Theorem 6.2 implies a version of the Kindler–Safra theorem [KS04, Kin02], Theorem 7.5 below.

We start by proving a structure theorem for almost degree  $k$  functions. We start with a hypercontractive estimate due to Lee and Yau [LY98] (see for example [Fil14, Proposition 6.2]).

**Proposition 7.1.** *For every  $p$  there exists a constant  $r_p$  such that for all functions  $f: \binom{[n]}{pn} \rightarrow \mathbb{R}$ ,  $\|H_{r_p} f\|_2 \leq \|f\|_{4/3}$ .*

This implies the following dichotomy result.

**Lemma 7.2.** *Fix parameters  $p$  and  $k$ , and let  $f: \binom{[n]}{pn} \rightarrow \{\pm 1\}$  satisfy  $\|f^{>k}\|^2 = \epsilon$ . For any  $i, j \in [n]$ , either  $\text{Inf}_{ij}^s[f] \leq \epsilon/2$  or  $\text{Inf}_{ij}^s[f] \geq J_{p,k}$ , for some constant  $J_{p,k}$ .*

*Proof.* Let  $r = r_p$  be the parameter in Proposition 7.1. Let  $g = (f - f^{(ij)})/2$ , so that  $\text{Inf}_{ij}^s[f] = 4\|g\|^2$ . Since  $g(x) \in \{0, \pm 1\}$ ,  $\|g\|_{4/3}^2 = \|g\|^2 = 4\text{Inf}_{ij}^s[f]$ . Proposition 7.1 therefore implies that

$$(4\text{Inf}_{ij}^s[f])^{3/2} = \|g\|_{4/3}^2 \geq \|H_r g\|_2^2 \geq \|H_r g^{\leq k}\|_2^2 \geq r^k \|g^{\leq k}\|^2.$$

Since  $g = (f - f^{(ij)})/2$ , we can bound  $\|g^{>k}\|^2 \leq \|f^{>k}\|^2 = \epsilon$ . Therefore

$$(4\text{Inf}_{ij}^s[f])^{3/2} \geq r^k (\|g\|^2 - \epsilon) = r^k (4\text{Inf}_{ij}^s[f] - \epsilon).$$

If  $4\text{Inf}_{ij}^s[f] > 2\epsilon$  then  $4\text{Inf}_{ij}^s[f] - \epsilon > 4\text{Inf}_{ij}^s[f]/2$  and so  $4\text{Inf}_{ij}^s[f] \geq r^{2k}/4$ .  $\square$

We need the following result, due to Wimmer [Wim14, Proposition 5.3].

**Lemma 7.3** ([Wim14, Proposition 5.3], [Fil14, Lemma 5.2]). *Let  $f: \binom{[n]}{pn} \rightarrow \mathbb{R}$ . For every  $\tau > 0$  there is a set  $J \subseteq [n]$  of size  $O(\text{Inf}^s[f]/\tau)$  such that  $\text{Inf}_{ij}^s[f] < \tau$  whenever  $i, j \notin J$ .*

Combining Lemma 7.2 and Lemma 7.3, we deduce that bounded degree functions depend on a constant number of coordinates, the analog of [NS94, Theorem 1].

**Corollary 7.4.** *Fix parameters  $p$  and  $k$ . If  $f: \binom{[n]}{pn} \rightarrow \{\pm 1\}$  has degree  $k$  then  $f$  depends on  $O_{p,k}(1)$  coordinates (that is,  $f$  is invariant under permutations of all other coordinates).*

*Proof.* Apply Lemma 7.3 with  $\tau = J_{p,k}$  to obtain a set  $J$  of size  $O(k/J_{p,k})$ . Lemma 7.2 with  $\epsilon = 0$  shows that for  $i, j \notin J$  we have  $\text{Inf}_{ij}^s[f] = 0$ , and so  $f$  is invariant under permutations of coordinates outside of  $J$ .  $\square$

Using Bourgain’s tail bound, we can deduce a stability version of Corollary 7.4, namely a Kindler–Safra theorem for the slice.

**Theorem 7.5.** *Fix the parameter  $k \geq 2$ . Let  $f: \binom{[n]}{pn} \rightarrow \{\pm 1\}$  satisfy  $\|f^{>k}\|^2 = \epsilon$ . There exists a function  $h: \binom{[n]}{pn} \rightarrow \{\pm 1\}$  of degree  $k$  depending on  $O_{k,p}(1)$  coordinates (that is, invariant under permutations of all other coordinates) such that*

$$\|f - h\|^2 = O_{p,k} \left( \epsilon^{1/C} + \frac{1}{n^{1/C}} \right),$$

for some constant  $C$ .

*Proof.* Let  $F = f^{\leq k}$ . We can assume that  $2\epsilon < J_{p,k}/2$ , since otherwise the theorem is trivial. Apply Lemma 7.3 to  $F$  with parameter  $\tau = J_{p,k} - 2\epsilon > J_{p,k}/2$ , obtaining a set  $J$  of size  $O(k/\tau) = O_{p,k}(1)$ . It is not hard to check that

$$\text{Inf}_{ij}^s[F] \leq \text{Inf}_{ij}^s[f] \leq \text{Inf}_{ij}^s[F] + 2\|f^{>k}\|^2 = \text{Inf}_{ij}^s[F] + 2\epsilon.$$

Therefore if  $i, j \notin J$  then  $\text{Inf}_{ij}^s[f] < \tau + 2\epsilon = J_{k,p}$ , and so Lemma 7.2 shows that  $\text{Inf}_{ij}^s[F] \leq \text{Inf}_{ij}^s[f] = O(\epsilon)$ .

For  $x \in \{0, 1\}^J$ , let  $G_x$  and  $g_x$  result from  $F$  and  $f$  (respectively) by restricting the coordinates in  $J$  to the value  $x$ . It is not hard to check that  $\Pr_{S \sim \nu_{pn}}[S|_J = x] \geq (p - O_p(|J|/n))^{|J|} = \Omega_{p,k}(1)$ , as long as  $n \geq N_{p,k}$  for some constant  $N_{p,k}$ ; if  $n \leq N_{p,k}$  then the theorem is trivial. We conclude that  $\text{Inf}_{ij}^s[G_x] = O_{p,k}(\epsilon)$  for all  $i, j \notin J$  and  $\|G_x - g_x\|^2 = \|g_x^{>k}\|^2 = O_{p,k}(\epsilon)$ . Together these imply that  $\text{Inf}_{ij}^s[g_x] = O_{p,k}(\epsilon)$  for all  $i, j \notin J$ , and so  $\text{Inf}_i^s[g_x] = O_{p,k}(\epsilon)$  for all  $i \notin J$ .

We can assume that  $n - |J| \geq n/2$  (otherwise the theorem is trivial) and that the skew  $p_x$  of the slice on which  $G_x, g_x$  are defined satisfies  $p_x = p \pm O_p(|J|/n) = \Theta(p)$ , and so Theorem 6.2 implies that either  $\max_i \text{Inf}_i^s[g_x] > W_{p,k}^{-1} \mathbb{V}[g_x]^C$ , or  $n < 2W_{p,k}/\mathbb{V}[g_x]^C$ , or  $\mathbb{V}[g_x] = O(\sqrt{k}\|g_x^{>k}\|^2) = O_{p,k}(\epsilon)$ . Since  $\max_i \text{Inf}_i^s[g_x] = O_{p,k}(\epsilon)$ , we conclude that

$$\mathbb{V}[g_x] = O_{p,k} \left( \epsilon^{1/C} + \frac{1}{n^{1/C}} \right).$$

Define a function  $g$  by  $g(S) = \mathbb{E}[g_{S|_J}]$ . The bound on  $\mathbb{V}[g_x]$  implies

$$\|f - g\|^2 = O_{p,k} \left( \epsilon^{1/C} + \frac{1}{n^{1/C}} \right).$$

If we let  $h = \text{sgn } g$  then we obtain the desired bound  $\|f - h\|^2 \leq 4\|f - g\|^2$ .

It remains to show that  $h$  has degree  $k$  if  $\epsilon$  is small enough and  $n$  is large enough. We can assume without loss of generality that  $J = [M]$ , where  $M$  is the bound on  $|J|$ . We have  $\|f - h\|^2 \geq \|f^{>k} - h^{>k}\|^2 \geq (\|h^{>k}\| - \sqrt{\epsilon})^2$ . Therefore

$$\|h^{>k}\| \leq \sqrt{\epsilon} + O_{p,k} \left( \epsilon^{1/2C} + \frac{1}{n^{1/2C}} \right).$$

On the other hand, we can write  $h$  as a Boolean function  $H$  of  $x_1, \dots, x_M$ . Lemma 3.1 shows that  $\deg h \leq \deg H$ , and so  $\deg h > k$  implies that  $\deg H > k$ . Corollary 3.4(5) implies that for large enough  $n$ ,  $\|h^{>k}\| = \Omega_{p,H}(1)$ . Since there are only finitely many Boolean functions on  $x_1, \dots, x_M$  which can play the role of  $H$ , we conclude that if  $\epsilon$  is small enough and  $n$  is large enough then  $\deg h \leq k$ .  $\square$

We conjecture that Theorem 7.5 holds with an error bound of  $O_{p,k}(\epsilon)$  rather than  $O_{p,k}(\epsilon^{1/C} + 1/n^{1/C})$ .

## 8 $t$ -Intersecting families

As an application of Theorem 7.5, we prove a stability result for the  $t$ -intersecting Erdős–Ko–Rado theorem, along the lines of Friedgut [Fri08]. We start by stating the  $t$ -intersecting Erdős–Ko–Rado theorem, which was first proved by Wilson [Wil84].

**Theorem 8.1** ([Wil84]). *Let  $t \geq 1$ ,  $k \geq t$ , and  $n \geq (t+1)(k-t+1)$ . Suppose that the family  $\mathcal{F} \subseteq \binom{[n]}{k}$  is  $t$ -intersecting: every two sets in  $\mathcal{F}$  have at least  $t$  points in common. Then:*

(a)  $|\mathcal{F}| \leq \binom{n-t}{k-t}$ .

(b) If  $n > (t+1)(k-t+1)$  and  $|\mathcal{F}| = \binom{n-t}{k-t}$  then  $\mathcal{F}$  is a  $t$ -star: a family of the form

$$\mathcal{F} = \left\{ A \in \binom{[n]}{k} : S \subseteq A \right\}, \quad |S| = t.$$

(c) If  $t \geq 2$ ,  $n = (t+1)(k-t+1)$  and  $|\mathcal{F}| = \binom{n-t}{k-t}$  then  $\mathcal{F}$  is either a  $t$ -star or a  $(t, 1)$ -Frankl family:

$$\mathcal{F} = \left\{ A \in \binom{[n]}{k} : |A \cap S| \geq t+1 \right\}, \quad |S| = t+2.$$

The case  $t = 1$  is the original Erdős–Ko–Rado theorem [EKR61]. Ahlswede and Khachatrian [AK97, AK99] found the optimal  $t$ -intersecting families for all values of  $n, k, t$ .

A stability version of Theorem 8.1 would state that if  $|\mathcal{F}| \approx \binom{n-t}{k-t}$  then  $\mathcal{F}$  is close to a  $t$ -star. Frankl [Fra87] proved an optimal such result for the case  $t = 1$ . Friedgut [Fri08] proved a stability result for all  $t$  assuming that  $k/n$  is bounded away from  $1/(t+1)$ .

**Theorem 8.2** ([Fri08]). *Let  $t \geq 1$ ,  $k \geq t$ ,  $\lambda, \zeta > 0$ , and  $\lambda n < k < (\frac{1}{t+1} - \zeta)n$ . Suppose  $\mathcal{F} \subseteq \binom{[n]}{k}$  is a  $t$ -intersecting family of measure  $|\mathcal{F}| = \binom{n-t}{k-t} - \epsilon \binom{n}{k}$ . Then there exists a family  $\mathcal{G}$  which is a  $t$ -star such that*

$$\frac{|\mathcal{F} \Delta \mathcal{G}|}{\binom{n}{k}} = O_{t,\lambda,\zeta}(\epsilon).$$

Careful inspection of Friedgut’s proof shows that it is meaningful even for sub-constant  $\zeta$ , but only as long as  $\zeta = \omega(1/\sqrt{n})$ . We prove a stability version of Theorem 8.1 which works all the way up to  $\zeta = 0$ .

**Theorem 8.3.** *Let  $t \geq 2$ ,  $k \geq t+1$  and  $n = (t+1)(k-t+1) + r$ , where  $r > 0$ . Suppose that  $k/n \geq \lambda$  for some  $\lambda > 0$ . Suppose  $\mathcal{F} \subseteq \binom{[n]}{k}$  is a  $t$ -intersecting family of measure  $|\mathcal{F}| = \binom{n-t}{k-t} - \epsilon \binom{n}{k}$ . Then there exists a family  $\mathcal{G}$  which is a  $t$ -star or a  $(t, 1)$ -Frankl family such that*

$$\frac{|\mathcal{F} \Delta \mathcal{G}|}{\binom{n}{k}} = O_{t,\lambda} \left( \max \left( \left( \frac{k}{r} \right)^{1/C}, 1 \right) \epsilon^{1/C} + \frac{1}{n^{1/C}} \right),$$

for some constant  $C$ .

Furthermore, there is a constant  $A_{t,\lambda}$  such that  $\epsilon \leq A_{t,\lambda} \min(r/k, 1)^{C+1}$  implies that  $\mathcal{G}$  is a  $t$ -star.

We do not know whether the error bound we obtain is optimal. We conjecture that Theorem 8.3 should hold with an error bound of  $O_{t,\lambda}(\max(k/r, 1)\epsilon)$ .

Friedgut’s approach proceeds through the  $\mu_p$  version of Theorem 8.1, first proved by Dinur and Safra [DS05] as a simple consequence of the work of Ahlswede and Khachatrian. The special case  $p = 1/d$  (where  $d \geq 3$ ) also follows from earlier work of Ahlswede and Khachatrian [AK98], who found the optimal  $t$ -agreeing families in  $\mathbb{Z}_d^n$ .

**Theorem 8.4** ([DS05],[Fri08],[Fil13]). *Let  $t \geq 1$  and  $p \leq 1/(t+1)$ . Suppose that  $\mathcal{F} \subseteq \{0, 1\}^n$  is  $t$ -intersecting. Then:*

- (a)  $\mu_p(\mathcal{F}) \leq p^t$  [DS05].
- (b) If  $p < 1/(t+1)$  and  $\mu_p(\mathcal{F}) = p^t$  then  $\mathcal{F}$  is a  $t$ -star [Fri08].
- (c) If  $t \geq 2$ ,  $p = 1/(t+1)$  and  $\mu_p(\mathcal{F}) = p^t$  then  $\mathcal{F}$  is either a  $t$ -star or a  $(t, 1)$ -Frankl family [Fil13].

Friedgut [Fri08] deduces his stability version of Theorem 8.1 from a stability version of Theorem 8.4. While Friedgut’s stability version of Theorem 8.4 is meaningful for all  $p < 1/(t+1)$ , his stability version of Theorem 8.1 is meaningful only for  $k/n < 1/(t+1) - \omega(1/\sqrt{n})$ . A more recent stability result for compressed cross- $t$ -intersecting families due to Frankl, Lee, Siggers and Tokushige [FLST14], using completely different techniques, also requires  $k/n$  to be bounded away from  $1/(t+1)$ .

Friedgut’s argument combines a spectral approach essentially due to Lovász [Lov79] with the Kindler–Safra theorem [KS04, Kin02]. Using Theorem 7.5 instead of the Kindler–Safra theorem, we are able to obtain a stability result for the entire range of parameters of Theorem 8.1. We restrict ourselves to the case  $t \geq 2$ .

Our starting point is a calculation due to Wilson [Wil84].

**Theorem 8.5** ([Wil84]). *Let  $t \geq 2$ ,  $k \geq t+1$ , and  $n \geq (t+1)(k-t+1)$ . There exists an  $\binom{[n]}{k} \times \binom{[n]}{k}$  symmetric matrix  $A$  such that  $A_{SS} = 1$  for all  $S \in \binom{[n]}{k}$ ,  $A_{ST} = 0$  for all  $S \neq T \in \binom{[n]}{k}$  satisfying  $|S \cap T| \geq t$ , and for all functions  $f: \binom{[n]}{k} \rightarrow \mathbb{R}$ ,*

$$Af = \sum_{e=0}^k \lambda_e f^{=e}, \quad \lambda_e = 1 + (-1)^{t-1-e} \sum_{i=0}^{t-1} (-1)^i \binom{k-1-i}{k-t} \binom{k-e}{i} \binom{n-k-e+i}{k-e} \binom{n-k-t+i}{k-t}^{-1}.$$



The eigenvalues  $\lambda_e$  satisfy the following properties:

- (a)  $\lambda_0 = \binom{n}{k} \binom{n-t}{k-t}^{-1}$ .
- (b)  $\lambda_1 = \dots = \lambda_t = 0$ .
- (c)  $\lambda_{t+2} \geq 0$ , with equality if and only if  $n = (t+1)(k-t+1)$ .
- (d)  $\lambda_{t+1} > \lambda_{t+2}$  and  $\lambda_e > \lambda_{t+2}$  for  $e > t+2$ .

Wilson's result actually needs  $n \geq 2k$ , but this is implied by our stronger assumption  $k \geq t+1$  (Wilson only assumes that  $k \geq t$ ) since  $(t+1)(k-t+1) - 2k = (t-1)(k-(t+1)) \geq 0$ .

We need to know exact asymptotics of  $\lambda_{t+2}$ .

**Lemma 8.6.** *Let  $t \geq 2$ ,  $k \geq t+1$  and  $n = (t+1+\rho)(k-t+1)$ , where  $\rho > 0$ . Let  $\lambda = \lambda_{t+2}$  be the quantity defined in Theorem 8.5. Then*

$$\lambda = \Omega_t(\min(\rho, 1)), \quad \lim_{\rho \rightarrow \infty} \lambda = 1.$$

*Proof.* Wilson [Wil84, (4.5)] gives the following alternative formula for  $\lambda$ :

$$\lambda = 1 - \binom{t+1}{2} \sum_{i=0}^{t-1} \frac{2}{i+2} \binom{t-1}{i} \frac{\binom{k-t}{i+2}}{\binom{n-k-t+i}{i+2}}.$$

Algebraic manipulation shows that

$$\lambda = 1 - \sum_{i=0}^{t-1} (i+1) \binom{t+1}{i+2} \frac{\binom{k-t}{i+2}}{\binom{n-k-t+i}{i+2}}.$$

Calculation shows that  $n - k - t = (t+\rho)(k-t+1) - 2t + 1$ . Therefore

$$\lambda = 1 - \sum_{i=0}^{t-1} (i+1) \binom{t+1}{i+2} \frac{\binom{k-t}{i+2}}{\binom{(t+\rho)(k-t+1)-2t+1+i}{i+2}}.$$

This formula makes it clear that  $\lim_{\rho \rightarrow \infty} \lambda = 1$ , and that  $\lambda$  is an increasing function of  $\rho$ .

Assume now that  $\rho \leq 1$ . Then

$$\lambda = 1 - \sum_{i=0}^{t-1} (i+1) \binom{t+1}{i+2} (t+\rho)^{-i-2} \left( 1 \pm O_t \left( \frac{1}{k} \right) \right).$$

Let us focus on the main term. Setting  $\alpha = 1/(t+\rho)$ , we have

$$\begin{aligned} \sum_{i=0}^{t-1} (i+1) \binom{t+1}{i+2} \alpha^{i+2} &= \sum_{i=0}^{t-1} (i+2) \binom{t+1}{i+2} \alpha^{i+2} - \sum_{i=0}^{t-1} \binom{t+1}{i+2} \alpha^{i+2} \\ &= (t+1) \sum_{i=0}^{t-1} \binom{t}{i+1} \alpha^{i+2} - \sum_{i=0}^{t-1} \binom{t+1}{i+2} \alpha^{i+2} \\ &= (t+1) \alpha \sum_{i=1}^t \binom{t}{i} \alpha^i - \sum_{i=2}^{t+1} \binom{t+1}{i} \alpha^i \\ &= (t+1) \alpha ((1+\alpha)^t - 1) - ((1+\alpha)^{t+1} - 1 - (t+1)\alpha) \\ &= 1 - (1+\alpha)^t (1 - t\alpha). \end{aligned}$$

Substituting  $\alpha = 1/(t + \rho)$ , we obtain

$$\sum_{i=0}^{t-1} (i+1) \binom{t+1}{i+2} (t+\rho)^{-i-2} = 1 - \frac{(t+\rho+1)^t}{(t+\rho)^t} \frac{\rho}{t+\rho} = 1 - \frac{\rho(t+1+\rho)^t}{(t+\rho)^{t+1}}.$$

Therefore when  $\rho \leq 1$ ,

$$\lambda = \frac{\rho(t+1+\rho)^t}{(t+\rho)^{t+1}} \pm O_t\left(\frac{1}{k}\right).$$

In particular, we can find some constant  $C_t$  such that

$$\lambda \geq \frac{(t+1+\rho)^t}{(t+\rho)^{t+1}} \left(\rho - \frac{C_t}{k}\right).$$

Therefore for  $2C_t/k \leq \rho \leq 1$ , we have  $\lambda = \Omega_t(\rho)$ . Since  $\lambda$  is an increasing function of  $\rho$ , this shows that for  $\rho \geq 2C_t/k$ , we have  $\lambda = \Omega_t(\min(\rho, 1))$ .

In order to finish the proof, we handle the case  $\rho \leq C_t/k$ . Consider  $n = (t+1)(k-t+1) + 1$ . The value of  $1 - \lambda$  in this case is

$$\begin{aligned} 1 - \lambda &= \sum_{i=0}^{t-1} \binom{t+1}{i+2} \binom{t-1}{i} \frac{\binom{k-t}{i+2}}{\binom{t(k-t+1)-2t+2+i}{i+2}} \\ &= \sum_{i=0}^{t-1} \binom{t+1}{i+2} \binom{t-1}{i} \frac{\binom{k-t}{i+2}}{\binom{t(k-t+1)-2t+1+i}{i+2}} \left(1 - \frac{i+2}{t(k-t+1)-2t+2+i}\right). \end{aligned}$$

The value of the last expression without the correction term  $1 - \frac{i+2}{t(k-t+1)-2t+2+i}$  is exactly 1 by Theorem 8.5, and so

$$\lambda \geq \frac{2}{t(k-t+1)-2t+2} = \Omega_t\left(\frac{1}{k}\right).$$

Since  $\lambda$  is increasing in  $\rho$ , this shows that for all  $\rho > 0$  we have  $\lambda = \Omega_t(1/k)$ . If also  $\rho \leq C_t/k$  then this implies that  $\lambda = \Omega_t(\rho)$ , finishing the proof.  $\square$

We need a similar result comparing the measures of  $t$ -stars and  $(t, 1)$ -Frankl families.

**Lemma 8.7.** *Let  $t \geq 2$ ,  $k \geq t+1$  and  $n = (t+1+\rho)(k-t) + t+1$ , where  $\rho > 0$ . Let  $m$  be the measure of a  $t$ -star, and let  $m_1$  be the measure of a  $(t, 1)$ -Frankl family. Then*

$$\frac{m - m_1}{m} = \Omega_t(\min(\rho, 1)), \quad \lim_{\rho \rightarrow \infty} \frac{m - m_1}{m} = 1.$$

*Proof.* We have

$$m = \binom{n-t}{k-t}, \quad m_1 = (t+2) \binom{n-t-2}{k-t-1} + \binom{n-t-2}{k-t-2}.$$

Computation shows that

$$\frac{m - m_1}{m_1} = 1 - \frac{(t+2)(n-k)(k-t) + (k-t)(k-t-1)}{(n-t)(n-t-1)}.$$

If  $n = (t+1)(k-t+1) + r$  then calculation shows that

$$\frac{m - m_1}{m_1} = \frac{r(r+t(k-t)+1)}{(n-t)(n-t-1)} \geq \frac{n-t-1}{n-t} \frac{r(r+t(k-t))}{(n-t-1)^2}.$$

Substituting  $r = (k-t)\rho$ , we obtain

$$\frac{m - m_1}{m_1} \geq \frac{n-t-1}{n-t} \frac{(k-t)^2 \rho(\rho+t)}{(k-t)^2 (t+1+\rho)^2} = \frac{n-t-1}{n-t} \frac{\rho(t+\rho)}{(t+1+\rho)^2}.$$

This shows that  $\lim_{\rho \rightarrow \infty} (m - m_1)/m_1 = 1$ . Since  $n \geq (t+1)(k-t+1) \geq t+2$  implies  $(n-t-1)/(n-t) \geq 1/2$ , we also get

$$\frac{m - m_1}{m_1} \geq \frac{\rho(t + \rho)}{2(t + 1 + \rho)^2}.$$

As  $\rho \rightarrow \infty$ , the lower bound tends to  $1/2$ , and in particular, we can find  $c_t$  such that for  $\rho \geq c_t$  we have  $(m - m_1)/m_1 \geq 1/3$ . When  $\rho \leq c_t$ , we clearly have  $(m - m_1)/m_1 = \Omega_t(\rho)$ , completing the proof.  $\square$

The method of Lovász [Lov79] as refined by Friedgut [Fri08] allows us to deduce an upper bound on  $\|f^{>t}\|^2$  for the characteristic function of a  $t$ -intersecting family.

**Lemma 8.8.** *Let  $t \geq 2$ ,  $k \geq t+1$  and  $n = (t+1)(k-t+1) + r$ , where  $r > 0$ . Let  $\mathcal{F} \subseteq \binom{[n]}{k}$  be a  $t$ -intersecting family, and  $f$  its characteristic function. Then*

$$\|f^{>t}\|^2 = O\left(\max\left(\frac{k}{r}, 1\right)\right) \cdot (m - \mathbb{E}[f]), \text{ where } m = \frac{\binom{n-t}{k-t}}{\binom{n}{k}}.$$

*Proof.* Let  $A$  be the matrix from Theorem 8.5. Since  $\|f^{=0}\| = \mathbb{E}[f]$ ,

$$\mathbb{E}[f] = \langle f, Af \rangle \geq \lambda_0 \mathbb{E}[f]^2 + \lambda_{t+2} \|f^{>t}\|^2.$$

This already implies that  $\mathbb{E}[f] \leq \lambda_0^{-1} = m$ . Since  $\lambda_0 = m^{-1}$  and  $\mathbb{E}[f] \leq m$ , we conclude that

$$\|f^{>t}\|^2 \leq \frac{\mathbb{E}[f] - m^{-1} \mathbb{E}[f]^2}{\lambda_{t+2}} = \frac{\mathbb{E}[f](1 - m^{-1} \mathbb{E}[f])}{\lambda_{t+2}} \leq \frac{m - \mathbb{E}[f]}{\lambda_{t+2}}.$$

Lemma 8.6 completes the proof.  $\square$

In order to prove our stability result, we need a result on cross-intersecting families.

**Theorem 8.9** ([FT92]). *Let  $\mathcal{F} \subseteq \binom{[n]}{a}$  and  $\mathcal{G} \subseteq \binom{[n]}{b}$  be cross-intersecting families: every set in  $\mathcal{F}$  intersects every set in  $\mathcal{G}$ . If  $n \geq a + b$  and  $b \geq a$  then*

$$|\mathcal{F}| + |\mathcal{G}| \leq \binom{n}{b} - \binom{n-a}{b} + 1 \leq \binom{n}{b}.$$

We can now prove our stability result.

*Proof of Theorem 8.3.* In what follows, all big  $O$  notations depend on  $t$  and  $\lambda$ . We can assume that  $n$  is large enough (as a function of  $t$  and  $\lambda$ ), since otherwise the theorem is trivial. We use the parameter  $p = (k-t+1)/n$  which satisfies  $\lambda/2 < p < 1/(t+1)$ .

Let  $f$  be the characteristic function of  $\mathcal{F}$ , so that  $\mathbb{E}[f] = m - \epsilon$ , where  $m = \binom{n-t}{k-t}/\binom{n}{k}$ . Lemma 8.8 shows that  $\|f^{>t}\|^2 = O(\max(k/r, 1))\epsilon$ , and so Theorem 7.5 shows that  $\|f - g\|^2 \leq \delta$  for the characteristic function  $g$  of some family  $\mathcal{G}$  depending on  $J = J_t$  coordinates, for some constant  $J_t$ , where  $\delta = O(\max((k/r)^{1/C}, 1)\epsilon^{1/C} + 1/n^{1/C})$ ; here we use the fact that  $\lambda \leq k/n \leq 1/2$ . We want to show that if  $\delta$  is small enough (as a function of  $t$ ) then  $\mathcal{G}$  must be a  $t$ -star or a  $(t, 1)$ -Frankl family; if  $\delta$  is large then the theorem becomes trivial.

We start by showing that if  $\delta$  is small enough then  $\mathcal{G}$  must be  $t$ -intersecting. Suppose without loss of generality that  $\mathcal{G}$  depends only on the first  $J$  coordinates. We will show that  $\mathcal{J} = \mathcal{G}|_{[J]} \subseteq \{0, 1\}^J$  must be  $t$ -intersecting. If  $\mathcal{J}$  is not  $t$ -intersecting, then pick  $A, B \in \mathcal{J}$  which are not  $t$ -intersecting, with  $|A| \geq |B|$ . Let  $\mathcal{A} = \{S \in \binom{[n] \setminus [J]}{k-|A|} : A \cup S \in \mathcal{F}\}$  and  $\mathcal{B} = \{S \in \binom{[n] \setminus [J]}{k-|B|} : B \cup S \in \mathcal{F}\}$ . Since  $n \geq (t+1)k - (t^2 - 1)$  and  $k \geq \lambda n$ , if  $n$  is large enough then  $(k - |A|) + (k - |B|) \leq n - 2J$ , and so Theorem 8.9 shows that  $|\mathcal{A}| + |\mathcal{B}| \leq \binom{n-J}{k-|B|}$ . Therefore

$$\|f - g\|^2 = \frac{|\mathcal{F} \Delta \mathcal{G}|}{\binom{n}{k}} \geq \frac{\binom{n-J}{k-|A|}}{\binom{n}{k}} = p^{|A|} (1-p)^{J-|A|} \left(1 \pm O\left(\frac{1}{p(1-p)n}\right)\right) = \Omega(1),$$

using Lemma 2.4 (for large enough  $n$ ) and the fact that  $p > \lambda/2$ . We conclude that if  $\delta$  is small enough,  $\mathcal{J} = \mathcal{G}|_{[J]}$  must be  $t$ -intersecting.

Next, we show that if  $\delta$  is small enough then  $\mathcal{G}$  must be either a  $t$ -star or a  $(t, 1)$ -Frankl family. If  $\mathcal{G}$  is neither then  $\mu_p(\mathcal{J}) < p^t$  for all  $0 < p \leq 1/(t+1)$  by Theorem 8.4, and in particular, since  $p > \lambda/2$ ,  $\mu_p(\mathcal{J}) \leq p^t - \gamma$  for some  $\gamma > 0$ ; here we use the fact that there are finitely many  $t$ -intersecting families on  $J$  points. Since  $\nu_k(\mathcal{J}) = \mu_p(\mathcal{J})(1 \pm O(1/n))$  due to Lemma 2.4, for large enough  $n$  and small enough  $\epsilon$  we have

$$\|f - g\|^2 \geq (\mathbb{E}[f] - \mathbb{E}[g])^2 \geq (\gamma(1 \pm O(1/n)) - \epsilon)^2 = \Omega(1).$$

We deduce that if  $n$  is large enough and  $\epsilon$  is small enough then  $\mathcal{G}$  is either a  $t$ -star or a  $(t, 1)$ -Frankl family.

It remains to show that if  $\epsilon \leq A_{t,\lambda} \min(r/k, 1)^{C+1}$  then  $\mathcal{G}$  cannot be a  $(t, 1)$ -Frankl family. Define  $\tau = \min(r/k, 1)$ . Let  $m_1$  be the measure of a  $(t, 1)$ -Frankl family. Lemma 8.7 shows that  $m - m_1 = \Omega(\tau)$  (since  $p > \lambda/2$  implies  $m = \Omega(1)$ ). Therefore if  $\mathcal{G}$  is a  $(t, 1)$ -Frankl family then  $\mathbb{E}[g] \leq m - \Omega(\tau)$ . On the other hand,  $\mathbb{E}[g] \geq \mathbb{E}[f] - \delta = m - \epsilon - O((\epsilon/\tau)^{1/C} + 1/n^{1/C})$ . Put together, we obtain

$$\Omega(\tau) \leq \epsilon + O((\epsilon/\tau)^{1/C} + 1/n^{1/C}).$$

Choose a constant  $c$  so that  $\epsilon \leq c\tau$  implies

$$\Omega(\tau) \leq O((\epsilon/\tau)^{1/C} + 1/n^{1/C});$$

if  $\epsilon > c\tau$  then the theorem becomes trivial. The inequality implies that  $\tau^C = O(\epsilon/\tau)$  and so  $\tau^{C+1} = O(\epsilon)$ , contradicting our assumption on  $\epsilon$  for an appropriate choice of  $A_{t,\lambda}$ .  $\square$

Our conjecture on the optimal error bound in Theorem 7.5 implies an error bound of  $O_{t,\lambda}(\max(k/r, 1)\epsilon)$  in Theorem 8.3.

## 9 Open problems

Our work gives rise to several open questions.

1. Prove (or refute) an invariance principle comparing  $\nu_{pn}$  and  $\mu_p$  directly without the condition on the influences.
2. Prove (or refute) an invariance principle comparing  $\nu_{pn}$  and  $\gamma_{p,p}$  for arbitrary (non-harmonic) multilinear polynomials.
3. Prove a tight version of the Kindler–Safra theorem on the slice (Theorem 7.5).
4. The uniform distribution on the slice is an example of a negatively associated vector of random variables. Generalize the invariance principle to this setting.
5. The slice  $\binom{[n]}{k}$  can be thought of as a 2-coloring of  $[n]$  with a given histogram. Generalize the invariance principle to  $c$ -colorings with given histogram.
6. The slice  $\binom{[n]}{k}$  has a  $q$ -analog: all  $k$ -dimensional subspaces of  $\mathbb{F}_q^n$  for some prime power  $q$ . The analog of the Boolean cube consists of all subspaces of  $\mathbb{F}_q^n$  weighted according to their dimension. Generalize the invariance principle to the  $q$ -analog, and determine the analog of Gaussian space.

## Acknowledgement

This paper started when all authors were members of a semester long program on “Real Analysis in Computer Science” at the Simons Institute for Theory of Computing at U.C. Berkeley. The authors would like to thank the institute for enabling this work.

Y.F. would like to mention that this material is based upon work supported by the National Science Foundation under agreement No. DMS-1128155. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the National Science Foundation.

E.M. would like to acknowledge the support of the following grants: NSF grants DMS 1106999 and CCF 1320105, DOD ONR grant N00014-14-1-0823, and grant 328025 from the Simons Foundation.

K.W. would like to acknowledge the support of NSF grant CCF 1117079.

## References

- [AK97] Rudolf Ahlswede and Levon H. Khachatrian. The complete intersection theorem for systems of finite sets. *Eur. J. Comb.*, 18(2):125–136, 1997. [24](#)
- [AK98] Rudolf Ahlswede and Levon H. Khachatrian. The diametric theorem in Hamming spaces—optimal anticodes. *Adv. Appl. Math.*, 20(4):429–449, 1998. [24](#)
- [AK99] Rudolf Ahlswede and Levon H. Khachatrian. A pushing-pulling method: New proofs of intersection theorems. *Combinatorica*, 19(1):1–15, 1999. [24](#)
- [BI84] Eiichi Bannai and Tatsuro Ito. *Algebraic Combinatorics I: Association schemes*. Benjamin/Cummings Pub. Co., 1984. [8](#)
- [Bor85] C. Borell. Geometric bounds on the Ornstein–Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985. [19](#)
- [DDG<sup>+</sup>15] Roei David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. In *ITCS 2015*, 2015. [2](#)
- [DMN13] Anindya De, Elchanan Mossel, and Joe Neeman. Majority is stable: discrete and SoS. In *45th ACM Symposium on Theory of Computing*, pages 477–486, 2013. [19](#)
- [DS05] Irit Dinur and Shmuel Safra. On the hardness of approximating minimum vertex cover. *Ann. Math.*, 162(1):439–485, 2005. [24](#)
- [Dun76] Charles F. Dunkl. A Krawtchouk polynomial addition theorem and wreath products of symmetric groups. *Indiana Univ. Math. J.*, 25:335–358, 1976. [2](#)
- [Dun79] Charles F. Dunkl. Orthogonal functions on some permutation groups. In *Relations between combinatorics and other parts of mathematics*, volume 34 of *Proc. Sump. Pure Math.*, pages 129–147, Providence, RI, 1979. Amer. Math. Soc. [2](#)
- [EKR61] Paul Erdős, Chao Ko, and Richard Rado. Intersection theorems for systems of finite sets. *Quart. J. Math. Oxford*, 12(2):313–320, 1961. [24](#)
- [Fil13] Yuval Filmus. *Spectral methods in extremal combinatorics*. PhD thesis, University of Toronto, 2013. [24](#)
- [Fil14] Yuval Filmus. Orthogonal basis for functions over a slice of the boolean hypercube, 2014. Submitted. [2](#), [4](#), [5](#), [6](#), [7](#), [22](#)

- [FLST14] Peter Frankl, Sang June Lee, Mark Siggers, and Norihide Tokushige. An Erdős–Ko–Rado theorem for cross- $t$ -intersecting families. *J. Combin. Th., Ser. A*, 128:207–249, 2014. [24](#)
- [Fra87] Péter Frankl. Erdős-Ko-Rado theorem with conditions on the maximal degree. *J. Comp. Theory A*, 46:252–263, 1987. [24](#)
- [Fri08] Ehud Friedgut. On the measure of intersecting families, uniqueness and stability. *Combinatorica*, 28(5):503–528, 2008. [2](#), [23](#), [24](#), [27](#)
- [FT92] Peter Frankl and Norihide Tokushige. Some best-possible inequalities concerning cross-intersecting families. *J. Combin. Th., Ser. A*, 61:87–97, 1992. [27](#)
- [IM12] Marcus Isaksson and Elchanan Mossel. Maximally stable Gaussian partitions with discrete applications. *Israel J. Math.*, 189(1):347–396, 2012. [15](#)
- [Kho10] Subhash Khot. Inapproximability of NP-complete problems, discrete fourier analysis, and geometry. In *Proceedings of the International Congress of Mathematicians*, Hyderabad, India, 2010. [1](#)
- [Kin02] Guy Kindler. *Property testing, PCP and Juntas*. PhD thesis, Tel-Aviv University, 2002. [22](#), [24](#)
- [KKO14] Guy Kindler, Naomi Kirshner, and Ryan O’Donnell. Gaussian noise sensitivity and Fourier tails, 2014. Manuscript. [20](#), [21](#)
- [KS04] Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas, 2004. Unpublished manuscript. [22](#), [24](#)
- [Lov79] László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1–7, 1979. [24](#), [27](#)
- [LY98] Tzong-Yau Lee and Horng-Tzer Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Prob.*, 26(4):1855–1873, 1998. [22](#)
- [MOO10] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Ann. Math.*, 171:295–341, 2010. [1](#), [17](#), [18](#), [19](#)
- [NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Comp. Comp.*, 4(4):301–313, 1994. [22](#)
- [O’D14] Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. [6](#), [19](#)
- [QZ07] Li Qiu and Xingzhi Zhan. On the span of Hadamard products of vectors. *Linear Algebra Appl.*, (422):304–307, 2007. [8](#)
- [Sri11] 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. [4](#)
- [Tan09] Hajime Tanaka. A note on the span of Hadamard products of vectors. *Linear Algebra Appl.*, (430):865–867, 2009. [8](#)
- [Wil84] Richard M. Wilson. The exact bound in the Erdős-Ko-Rado theorem. *Combinatorica*, 4:247–257, 1984. [2](#), [23](#), [24](#), [25](#)
- [Wim14] Karl Wimmer. Low influence functions over slices of the Boolean hypercube depend on few coordinates. In *CCC*, 2014. [22](#)