Noise stability of functions with low influences: invariance and optimality

Elchanan Mossel[*]
U.C. Berkeley
mossel@stat.berkeley.edu

Ryan O'Donnell
Microsoft Research
odonnell@microsoft.com

Krzysztof Oleszkiewicz[†]
Warsaw University
koles@mimuw.edu.pl

June 29, 2005

Abstract

In this paper we study functions with low influences on product probability spaces. These are functions $f : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ that have $\mathbf{E}[\mathbf{Var}_{\Omega_i}[f]]$ small compared to $\mathbf{Var}[f]$ for each i. The analysis of boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ with low influences has become a central problem in discrete Fourier analysis. It is motivated by fundamental questions arising from the construction of probabilistically checkable proofs in theoretical computer science and from problems in the theory of social choice in economics.

We prove an invariance principle for multilinear polynomials with low influences and bounded degree; it shows that under mild conditions the distribution of such polynomials is essentially invariant for all product spaces. Ours is one of the very few known non-linear invariance principles. It has the advantage that its proof is simple and that the error bounds are explicit. We also show that the assumption of bounded degree can be eliminated if the polynomials are slightly "smoothed"; this extension is essential for our applications to "noise stability"-type problems.

In particular, as applications of the invariance principle we prove two conjectures: the "Majority Is Stablest" conjecture [35] from theoretical computer science, which was the original motivation for this work, and the "It Ain't Over Till It's Over" conjecture [33] from social choice theory.

1 Introduction

1.1 Harmonic analysis of boolean functions

The motivation for this paper is the study of *boolean functions* $f : \{-1,1\}^n \to \{-1,1\}$, where $\{-1,1\}^n$ is equipped with the uniform probability measure. This topic is of significant interest in theoretical computer science; it also arises in other diverse areas of mathematics including combinatorics (e.g., sizes of set systems, additive combinatorics), economics (e.g., social choice), metric spaces (e.g., non-embeddability of metrics), geometry in Gaussian space (e.g., isoperimetric inequalities), and statistical physics (e.g., percolation, spin glasses).

Beginning with Kahn, Kalai, and Linial's landmark paper "The Influence Of Variables On Boolean Functions" [30] there has been much success in analyzing questions about boolean functions using methods of harmonic analysis. Recall that KKL essentially shows the following (see also [49, 24]):

[*]Supported by a Miller fellowship in Statistics and Computer Science, U.C. Berkeley and by a Sloan fellowship in Mathematics

[†]Partially supported by Polish KBN Grant 2 P03A 027 22.

KKL Theorem: If $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\mathbf{E}[f] = 0$ and $\mathrm{Inf}_i(f) \le \tau$ for all i, then $\sum_{i=1}^n \mathrm{Inf}_i(f) \ge \Omega(\log(1/\tau))$.

We have used here the notation $\mathrm{Inf}_i(f)$ for the *influence of the ith coordinate on* f,

$$\mathrm{Inf}_i(f) = \mathop{\mathbf{E}}_{x}[\mathop{\mathbf{Var}}_{x_i}[f(x)]] = \sum_{S \ni i} \hat{f}(S)^2. \tag{1}$$

Although an intuitive understanding of the analytic properties of boolean functions is emerging, results in this area have used increasingly elaborate methods, combining random restriction arguments, applications of the Bonami-Beckner inequality, and classical tools from probability theory. See for example [49, 50, 24, 23, 10, 5, 11, 38, 15].

As in the KKL paper, some of the more refined problems studied in recent years have involved restricting attention to functions with low influences [5, 10, 15] (or, relatedly, "non-juntas"). There are several reasons for this. The first is that large-influence functions such as "dictators" — i.e., functions $f(x_1, \dots, x_n) = \pm x_i$ — frequently trivially maximize or minimize quantities studied in boolean analysis. However this tends to obscure the truth about extremal behaviors among functions that are "genuinely" functions of n bits. Another reason for analyzing only low-influence functions is that this subclass is often precisely what is interesting or necessary for applications. In particular, the analysis of low-influence boolean functions is crucial for proving hardness of approximation results in theoretical computer science and is also very natural for the study of social choice. Let us describe these two settings briefly.

In the economic theory of social choice, boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ often represent voting schemes, mapping n votes between two candidates into a winner. In this case, it is very natural to exclude voting schemes that give any voter an undue amount of influence; see e.g. [32]. In the study of hardness of approximation and probabilistically checkable proofs (PCPs), the sharpest results often involve the following paradigm: One considers a problem that requires labeling the vertices of a graph using the label set $[n]$; then one relaxes this to the problem of labeling the vertices by functions $f : \{-1,1\}^n \to \{-1,1\}$. In the relaxation one thinks of f as "weakly labeling" a vertex by the *set* of coordinates that have large influence on f. It then becomes important to understand the combinatorial properties of functions that weakly label with the empty set. There are by now quite a few results in hardness of approximation that use results on low-influence functions or require conjectured such results; e.g., [17, 34, 16, 36, 35].

In this paper we give a new framework for studying functions on product probability spaces with low influences. Our main tool is a simple invariance principle for low-influence polynomials; this theorem lets us take an optimization problem for functions on one product space and pass freely to other product spaces, such as Gaussian space. In these other settings the problem sometimes becomes simpler to solve. It is interesting to note that while in the theory of hypercontractivity and isoperimetry it is common to prove results in the Gaussian setting by proving them first in the $\{-1,1\}^n$ setting (see, e.g., [2]), here the invariance principle is actually used to go the other way around.

As applications of our invariance principle we prove two previously unconnected conjectures from boolean harmonic analysis; the first was motivated by hardness of approximation in computer science, the second by the theory of social choice from economics:

Conjecture 1.1 ("Majority Is Stablest" conjecture [35]) *Let $0 \le \rho \le 1$ and $\epsilon > 0$ be given. Then there exists $\tau > 0$ such that if $f : \{-1,1\}^n \to [-1,1]$ satisfies $\mathbf{E}[f] = 0$ and $\mathrm{Inf}_i(f) \le \tau$ for all i, then*

$$\mathbb{S}_\rho(f) \le \tfrac{2}{\pi}\arcsin\rho + \epsilon.$$

Here we have used the notation $\mathbb{S}_\rho(f)$ for $\sum_S \rho^{|S|}\hat{f}(S)^2$, the *noise stability* of f. This equals $\mathbf{E}[f(x)f(y)]$ when $(x,y) \in \{-1,1\}^n \times \{-1,1\}^n$ is chosen so that $(x_i, y_i) \in \{-1,1\}^2$ are independent random variables with $\mathbf{E}[x_i] = \mathbf{E}[y_i] = 0$ and $\mathbf{E}[x_i y_i] = \rho$.

Conjecture 1.2 ("It Ain't Over Till It's Over" conjecture [33]) *Let $0 \le \rho < 1$ and $\epsilon > 0$ be given. Then there exists $\delta > 0$ and $\tau > 0$ such that if $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\mathbf{E}[f] = 0$ and $\mathrm{Inf}_i(f) \le \tau$ for all i, then f has the following property: If V is a random subset of $[n]$ in which each i is included independently with probability ρ, and if the bits $(x_i)_{i\in V}$ are chosen uniformly at random, then*

$$\mathop{\mathbf{P}}_{V,\,(x_i)_{i\in V}}\Big[\big|\mathbf{E}[f \mid (x_i)_{i\in V}]\big| > 1-\delta\Big] \le \epsilon.$$

(In words, the conjecture states that even if a random ρ fraction of voters' votes are revealed, with high probability the election is still slightly undecided, provided f has low influences.)

The truth of these results gives illustration to a recurring theme in the harmonic analysis of boolean functions: the extremal role played the Majority function. It seems this theme becomes especially prominent when low-influence functions are studied. To explain the connection of Majority to our applications: In the former case the quantity $\frac{2}{\pi}\arcsin\rho$ is precisely $\lim_{n\to\infty}\mathbb{S}_\rho(\mathrm{Maj}_n)$; this explains the name of the Majority Is Stablest conjecture. In the latter case, we show that δ can be taken to be on the order of $\epsilon^{\rho/(1-\rho)}$ (up to $o(1)$ in the exponent), which is the same asymptotics one gets if f is Majority on a large number of inputs.

1.2 Outline of the paper

We begin in Section 2 with an overview of the invariance principle, the two applications, and some of their consequences. We prove the invariance principle in Section 3. Our proofs of the two conjectures are in Section 4. Finally, we show in Section 5 that a conjecture closely related to Majority Is Stablest is false. Some minor proofs from throughout the paper appear in appendices.

1.3 Related work

Our multilinear invariance principle has some antecedents. For degree 1 polynomials it reduces to a version of the Berry-Esseen Central Limit Theorems. Indeed, our proof follows the same outlines as Lindeberg's proof of the CLT [42] (see also [21]).

Since presenting our proof of the invariance principle, we have been informed by Oded Regev that related results were proved in the past by V. I. Rotar′ [46]. As well, a contemporary manuscript of Sourav Chatterjee [13] with an invariance principle of similar flavor has come to our attention. What is common to our work and to [46, 13] is a generalization of Lindeberg's argument to the non-linear case. The result of Rotar′ is an invariance principle similar to ours where the condition on the influences generalizes Lindeberg's condition. The setup is not quite the same, however, and the proof in [46] is of a rather qualitative nature. It seems that even after appropriate modification the bounds it gives would be weaker and less useful for our type of applications. (This is quite understandable; in a similar way Lindeberg's CLT can be less precise than the Berry-Esseen inequality even though — indeed, because — it works under weaker assumptions.) The paper [13] is by contrast very clear and explicit. However it does not seem to be appropriate for many applications since it requires low "worst-case" influences, instead of the "average-case" influences used by this work and [46].

Finally, we would like to mention that some chaos-decomposition limit theorems have been proved before in various settings. Among these are limit theorems for U and V statistics and limit theorems for random graphs; see, e.g. [29].

1.4 Acknowledgments

We are grateful to Keith Ball for suggesting a collaboration among the authors. We would also like to thank Oded Regev for referring us to [46] and Olivier Guédon for referring us to [12].

2 Our results

2.1 The invariance principle

In this subsection we present a simplified version of our invariance principle.

Suppose $\boldsymbol{X}$ is a random variable with $\mathbf{E}[\boldsymbol{X}] = 0$ and $\mathbf{E}[\boldsymbol{X}^2] = 1$ and $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independent copies of $\boldsymbol{X}$. Let $Q(x_1, \ldots, x_n) = \sum_{i=1}^n c_i x_i$ be a linear form and assume $\sum c_i^2 = 1$. The Berry-Esseen CLT states that under mild conditions on the distribution of $\boldsymbol{X}$, say $\mathbf{E}[|\boldsymbol{X}|^3] \leq A < \infty$, it holds that

$$\sup_t \bigl|\mathbf{P}[Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \leq t] - \mathbf{P}[\boldsymbol{G} \leq t]\bigr| \leq O\bigl(A \cdot \textstyle\sum_{i=1}^n |c_i|^3\bigr),$$

where $\boldsymbol{G}$ denotes a standard normal random variable. Note that a simple corollary of the above is

$$\sup_t \bigl|\mathbf{P}[Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \leq t] - \mathbf{P}[Q(\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n) \leq t]\bigr| \leq O\bigl(A \cdot \max_i |c_i|\bigr). \tag{2}$$

Here the $\boldsymbol{G}_i$'s denote independent standard normals. We have upper-bounded the sum of $|c_i|^3$ here by a maximum, for simplicity; more importantly though, we have suggestively replaced $\boldsymbol{G}$ by $\sum_i c_i \boldsymbol{G}_i$, which of course has the same distribution.

We would like to generalize (2) to *multilinear polynomials* in the $\boldsymbol{X}_i$'s; i.e., functions of the form

$$Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} \boldsymbol{X}_i, \tag{3}$$

where the real constants c_S satisfy $\sum c_S^2 = 1$. Let $d = \max_{c_S \neq 0} |S|$ denote the degree of Q. Unlike in the $d = 1$ case of the CLT, there is no single random variable $\boldsymbol{G}$ which always provides a limiting distribution. However one can still hope to prove, in light of (2), that the distribution of the polynomial applied to the variables $\boldsymbol{X}_i$ is close to the distribution of the polynomial applied to independent Gaussian random variables. This is indeed what our invariance principle shows.

It turns out that the appropriate generalization of the Berry-Esseen theorem (2) is to control the error by a function of d and of $\max_i \sum_{S \ni i} c_S^2$ — i.e., the maximum of the *influences* of Q (as in (1)). Naturally, we also need some conditions in addition to second moments. In our formulation we impose the condition that the variable $\boldsymbol{X}$ is *hypercontractive*; i.e., there is some $\eta > 0$ such that for all $a \in \mathbb{R}$,

$$\|a + \eta \boldsymbol{X}\|_3 \leq \|a + \boldsymbol{X}\|_2.$$

This condition is satisfied whenever $\mathbf{E}[\boldsymbol{X}] = 0$ and $\mathbf{E}[|\boldsymbol{X}|^3] < \infty$; in particular, it holds for any mean-zero random variable $\boldsymbol{X}$ taking on only finitely many values. Using hypercontractivity, we get a simply proved invariance principle with explicit error bounds. The following theorem (a simplification of Theorem 3.19, bound (21)) is an example of what we prove:

Theorem 2.1 *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random variables satisfying $\mathbf{E}[\boldsymbol{X}_i] = 0$, $\mathbf{E}[\boldsymbol{X}_i^2] = 1$, and $\mathbf{E}[|\boldsymbol{X}_i|^3] \leq \beta$. Let Q be a degree d multilinear polynomial as in (3) with*

$$\sum_{|S|>0} c_S^2 = 1, \qquad \sum_{S \ni i} c_S^2 \leq \tau \quad \textit{for all } i.$$

Then

$$\sup_t \left| \mathbf{P}[Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \leq t] - \mathbf{P}[Q(\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n) \leq t] \right| \leq O(d \beta^{1/3} \tau^{1/8d}),$$

where $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n$ are independent standard Gaussians.

If, instead of assuming $\mathbf{E}[|\boldsymbol{X}_i|^3] \leq \beta$, we assume that each $\boldsymbol{X}_i$ takes only on finitely many values, and that for all i and all $x \in \mathbb{R}$ either $\mathbf{P}[\boldsymbol{X}_i = x] = 0$ or $\mathbf{P}[\boldsymbol{X}_i = x] \geq \alpha$, then

$$\sup_t \left| \mathbf{P}[Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \leq t] - \mathbf{P}[Q(\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n) \leq t] \right| \leq O(d\, \alpha^{-1/6}\, \tau^{1/8d}).$$

Note that if d, β, and α are fixed then the above bound tends to 0 with τ. We call this theorem an "invariance principle" because it shows that $Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ has essentially the same distribution no matter what the $\boldsymbol{X}_i$'s are. Usually we will not push for the optimal constants; instead we will try to keep our approach as simple as possible while still giving explicit bounds useful for our applications.

An unavoidable deficiency of this sort of invariance principle is the dependence on d in the error bound. In applications such as Majority Is Stablest and It Ain't Over Till It's Over, the functions f may well have arbitrarily large degree. To overcome this, we introduce a supplement to the invariance principle: We show that if the polynomial Q is "smoothed" slightly then the dependence on d in the error bound can be eliminated and replaced with a dependence on the smoothness. For "noise stability"-type problems such as ours, this smoothing is essentially harmless.

In fact, the techniques we use are strong enough to obtain Berry-Esseen estimates under Lyapunov-type assumptions. In particular, we believe that the following theorem is new even in the case of sums of independent random variables.

Theorem 2.2 *Let $q \in (2, 3]$. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random variables satisfying $\mathbf{E}[\boldsymbol{X}_i] = 0$, $\mathbf{E}[\boldsymbol{X}_i^2] = 1$, and $\mathbf{E}[|\boldsymbol{X}_i|^q] \leq \beta$. Let Q be a degree d multilinear polynomial as in (3) with*

$$\sum_{|S|>0} c_S^2 = 1, \qquad \sum_{S \ni i} c_S^2 \leq \tau \quad \textit{for all } i.$$

Then

$$\sup_t \left| \mathbf{P}[Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \leq t] - \mathbf{P}[Q(\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n) \leq t] \right| \leq$$
$$O(d \beta^{\frac{d}{qd+1}} (\sum_i (\sum_{S \ni i} c_S^2)^{q/2})^{\frac{1}{qd+1}}) \leq O(d \beta^{\frac{d}{qd+1}} \tau^{\frac{q-2}{2qd+2}}),$$

where $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n$ are independent standard Gaussians.

2.2 Influences and noise stability in product spaces

Our proofs of the Majority Is Stablest and It Ain't Over Till It's Over conjectures hold not just for functions on the uniform-distribution discrete cube, but for functions on arbitrary finite product probability spaces. Harmonic analysis results on influences have often considered the biased product distribution on the discrete cube (see, e.g., [49, 24, 23, 10]); and, some recent works involving influences and noise stability have considered functions on product sets $[q]^n$ endowed with the uniform distribution (e.g., [1, 35]). But since there doesn't appear to be a unified treatment for the general case in the literature, we give the necessary definitions here.

Let $(\Omega_1, \mu_1), \ldots, (\Omega_n, \mu_n)$ be probability spaces and let (Ω, μ) denote the product probability space. Let

$$f : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$$

be any real-valued function on Ω.

Definition 2.3 *The* influence of the ith coordinate on f *is*

$$\mathrm{Inf}_i(f) = \mathop{\mathbf{E}}_{\mu}[\mathop{\mathbf{Var}}_{\mu_i}[f]].$$

Note that for boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ this agrees with the classical notion of influences introduced to computer science by Ben-Or and Linial [4]. When the domain $\{-1,1\}^n$ has a p-biased distribution, our notion differs from that of, say, [22] by a multiplicative factor of $4p(1-p)$. We believe the above definition is more natural, and in any case it is easy to pass between the two.

To define noise stability, we first define the T_ρ operator on the space of functions f:

Definition 2.4 *For any $0 \le \rho \le 1$, the operator T_ρ is defined by*

$$(T_\rho f)(\omega_1, \ldots, \omega_n) = \mathbf{E}[f(\omega'_1, \ldots, \omega'_n)], \tag{4}$$

where each ω'_i is an independent random variable defined to equal ω_i with probability ρ and to be randomly drawn from μ_i with probability $1-\rho$.

We remark that this definition agrees with that of the "Bonami-Beckner operator" introduced in the context of boolean functions by KKL [30] and also with its generalization to $[q]^n$ from [35]. For more on this operator, see Wolff [51]. With this definition in place, we can define noise stability:

Definition 2.5 *The* noise stability of f at $\rho \in [0,1]$ *is*

$$\mathbb{S}_\rho(f) = \mathop{\mathbf{E}}_{\mu}[f \cdot T_\rho f].$$

For the It Ain't Over Till It's Over problem, we introduce a new operator V_ρ:

Definition 2.6 *For any $\rho \in [0,1]$, the operator V_ρ is defined as follows. The operator takes a function $f : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ to a function $g : \Omega_1 \times \cdots \times \Omega_n \times \{0,1\}^n \to \mathbb{R}$, where $\{0,1\}^n$ is equipped with the $(1-\rho, \rho)^{\otimes n}$ measure. It is defined as follows:*

$$(V_\rho f)(\omega_1, \ldots, \omega_n, x_1, \ldots, x_n) = \mathop{\mathbf{E}}_{\omega'}\left[f\left(x_1\omega_1 + (1-x_1)\omega'_1, \ldots, x_n\omega_n + (1-x_n)\omega'_n\right)\right].$$

Finally, we would like to note that our definitions are valid for functions f into the reals, although our motivation is usually $\{-1,1\}$-valued functions. Our proofs of the Majority Is Stablest and It Ain't Over Till It's Over conjectures will hold in the setting of functions $f : \Omega_1 \times \cdots \times \Omega_n \to [-1,1]$ (note that Conjecture 1.1 *requires* this generalized range). For notational simplicity, though, we will give our proofs for functions into $[0,1]$; the reader can easily convert such results to the $[-1,1]$ case by the linear transformation $f \mapsto 2f-1$, which interacts in a simple way with the definitions of Inf_i, $\mathbb{S}_\rho$ and V_ρ.

2.3 Majority Is Stablest

2.3.1 About the problem

The Majority Is Stablest conjecture, Conjecture 1.1, was first formally stated in [35]. However the notion of Hamming balls having the highest noise stability in various senses has been widely spread among the community studying discrete Fourier analysis. Indeed, already in KKL's 1998 paper [30] there is the suggestion that Hamming balls and subcubes should maximize a certain noise stability-like quantity. In [5], it was shown that every 'asymptotically noise stable" function is correlated with a weighted majority function; also, in [43] it was shown that the majority function asymptotically maximizes a high-norm analog of $\mathbb{S}_\rho$.

More concretely, strong motivation for getting sharp bounds on the noise stability of low-influence functions came from two 2002 papers, one by Kalai [31] on social choice and one by Khot [34] on PCPs and hardness of approximation. We briefly discuss these two papers below.

Kalai '02 — Arrow's Impossibility Theorem: Suppose n voters rank three candidates, A, B, and C, and a *social choice* function $f : \{-1,1\}^n \to \{-1,1\}$ is used to aggregate the rankings, as follows: f is applied to the n A-vs.-B preferences to determine whether A or B is globally preferred; then the same happens for A-vs.-C and B-vs.-C. The outcome is termed "non-rational" if the global ranking has A preferable to B preferable to C preferable to A (or if the other cyclic possibility occurs). Arrow's Impossibility Theorem from the theory of social choice states that under some mild restrictions on f (such as f being odd; i.e., $f(-x) = -f(x)$), the only functions which never admit non-rational outcomes given rational voters are the dictator functions $f(x) = \pm x_i$.

Kalai [31] studied the *probability* of a rational outcome given that the n voters vote independently and at random from the 6 possible rational rankings. He showed that the probability of a rational outcome in this case is precisely $3/4 + (3/4)\mathbb{S}_{1/3}(f)$. Thus it is natural to ask which function f with small influences is most likely to produce a rational outcome. Instead of considering small influences, Kalai considered the essentially stronger assumption that f is "transitive-symmetric"; i.e., that for all $1 \le i < j \le n$ there exists a permutation σ on $[n]$ with $\sigma(i) = j$ such that $f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for all $(x_1, \ldots, x_n)$. Kalai conjectured that Majority was the transitive-symmetric function that maximized $3/4 + (3/4)\mathbb{S}_{1/3}(f)$ (in fact, he made a stronger conjecture, but this conjecture is false; see Section 5). He further observed that this would imply that in any transitive-symmetric scheme the probability of a rational outcome is at most $3/4 + (3/2\pi)\arcsin(1/3) + o_n(1) \approx .9123$; however, Kalai could only prove the weaker bound .9192.

Khot '02 — Unique Games and hardness of approximating 2-CSPs: In computer science, many combinatorial optimization problems are NP-hard, meaning it is unlikely there are efficient algorithms that always find the optimal solution. Hence there has been extensive interest in understanding the complexity of *approximating* the optimal solution. Consider for example "k-variable constraint satisfaction problems" (k-CSPs) in which the input is a set of variables over a finite domain, along with some constraints on k-sets of the variables, restricting what sets of values they can simultaneously take. We say a problem has "(c, s)-hardness" if it is NP-hard, given a k-CSP instance in which the optimal assignment satisfies a c-fraction of the constrains, for an algorithm to find an assignment that satisfies an s-fraction of the constraints. In this case we also say that the problem is "s/c-hard to approximate".

The PCP and Parallel Repetition theorems have led to many impressive results showing that it is NP-hard even to give α-approximations for various problems, especially k-CSPs for $k \geq 3$. For example, letting MAX-kLIN(q) denote the problem of satisfying k-variable linear equations over $\mathbf{Z}_q$, it is known [28] that MAX-kLIN(q) has $(1-\epsilon, 1/q+\epsilon)$-hardness for all $k \geq 3$, and this is sharp. However it seems that current PCP theorems are not strong enough to give sharp hardness of approximation results for 2-CSPs (e.g., constraint satisfaction problems on graphs). The influential paper of Khot [34] introduced the "Unique Games Conjecture" (UGC) in order to make progress on 2-CSPs; UGC states that a certain 2-CSP over a large domain has $(1-\epsilon, \epsilon)$-hardness.

Interestingly, it seems that using UGC to prove hardness results for other 2-CSPs typically crucially requires strong results about influences and noise stability of boolean functions. For example, [34]'s analysis of MAX-2LIN(2) required an upper bound on $\mathbb{S}_{1-\epsilon}(f)$ for small ϵ among balanced functions $f : \{-1,1\}^n \to \{-1,1\}$ with small influences; to get this, Khot used the following deep result of Bourgain [11] from 2001:

Theorem 2.7 (Bourgain [11]) *If $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\mathbf{E}[f] = 0$ and $\mathrm{Inf}_i(f) \leq 10^{-d}$ for all $i \in [n]$, then*

$$\sum_{|S|>d} \hat{f}(S)^2 \geq d^{-1/2-O(\sqrt{\log\log d/\log d})} = d^{-1/2-o(1)}.$$

Note that Bourgain's theorem has the following easy corollary:

Corollary 2.8 *If $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\mathbf{E}[f] = 0$ and $\mathrm{Inf}_i(f) \leq 2^{-O(1/\epsilon)}$ for all $i \in [n]$, then*

$$\mathbb{S}_{1-\epsilon}(f) \leq 1 - \epsilon^{1/2+o(1)}.$$

This corollary enabled Khot to show $(1-\epsilon, 1-\epsilon^{1/2+o(1)})$-hardness for MAX-2LIN(2), which is close to sharp (the algorithm of Goemans-Williamson [27] achieves $1 - O(\sqrt{\epsilon})$). As an aside, we note that Khot and Vishnoi [37] recently used Corollary 2.8 to prove that negative type metrics do not embed into ℓ_1 with constant distortion.

Another example of this comes from the work of [35]. Among other things, [35] studied the MAX-CUT problem: Given an undirected graph, partition the vertices into two parts so as to maximize the number of edges with endpoints in different parts. The paper introduced the Majority Is Stablest Conjecture 1.1 and showed that together with UGC it implied $(\frac{1}{2}+\frac{1}{2}\rho-\epsilon, \frac{1}{2}+\frac{1}{\pi}\arcsin\rho+\epsilon)$-hardness for MAX-CUT. In particular, optimizing over ρ (taking $\rho \approx .69$) implies MAX-CUT is .878-hard to approximate, matching the groundbreaking algorithm of Goemans and Williamson [27].

2.3.2 Consequences of confirming the conjecture

In Theorem 4.4 we confirm a generalization of the Majority Is Stablest conjecture. We give a slightly simplified statement of this theorem here:

Theorem 4.4 *Let $f : \Omega_1 \times \cdots \times \Omega_n \to [0,1]$ be a function on a discrete product probability space and assume that for each i the minimum probability of any atom in Ω_i is at least $\alpha \leq 1/2$. Further* *assume that* $\mathrm{Inf}_i(f) \leq \tau$ *for all* i. *Let* $\mu = \mathbf{E}[f]$. *Then for any* $0 \leq \rho < 1$,

$$\mathbb{S}_\rho(f) \leq \lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{Thr}_n^{(\mu)}) + O\Big(\tfrac{\log\log(1/\tau)}{\log(1/\tau)}\Big),$$

where $\mathrm{Thr}_n^{(\mu)} : \{-1,1\}^n \to \{0,1\}$ *denotes the symmetric threshold function with expectation closest to* μ, *and the* $O(\cdot)$ *hides a constant depending only on* α *and* $1-\rho$.

We now give some consequences of this theorem:

Theorem 2.9 *In the terminology of Kalai [31], any odd, balanced social choice function* f *with either*

- $o_n(1)$ *influences or*
- *such that* f *is transitive*

has probability at most $3/4 + (3/2\pi)\arcsin(1/3) + o_n(1) \approx .9123$ *of producing a rational outcome. The majority function on* n *inputs achieves this bound,* $3/4 + (3/2\pi)\arcsin(1/3) + o_n(1)$.

By looking at the series expansion of $\frac{2}{\pi}\arcsin(1-\epsilon)$ we obtain the following strengthening of Corollary 2.8.

Corollary 2.10 *If* $f : \{-1,1\}^n \to \{-1,1\}$ *satisfies* $\mathbf{E}[f] = 0$ *and* $\mathrm{Inf}_i(f) \leq \epsilon^{-O(1/\epsilon)}$ *for all* $i \in [n]$, *then*

$$\mathbb{S}_{1-\epsilon}(f) \leq 1 - (\tfrac{\sqrt{8}}{\pi} - o(1))\epsilon^{1/2}.$$

Using Corollary 2.10 instead of Corollary 2.8 in Khot [34] we obtain

Corollary 2.11 *MAX-2LIN(2) and MAX-2SAT have* $(1-\epsilon, 1-O(\epsilon^{1/2}))$*-hardness.*

More generally, [35] now implies

Corollary 2.12 *MAX-CUT has* $(\frac{1}{2} + \frac{1}{2}\rho - \epsilon, \frac{1}{2} + \frac{1}{\pi}\arcsin\rho + \epsilon)$*-hardness for each* ρ *and all* $\epsilon > 0$, *assuming UGC only. In particular, the Goemans-Williamson .878-approximation algorithm is best possible, assuming UGC only.*

The following two results are consequences of a generalization of "Majority is Stablest" as shown in [35]:

Theorem 2.13 *UGC implies that for each* $\epsilon > 0$ *there exists* $q = q(\epsilon)$ *such that MAX-2LIN(q) has* $(1-\epsilon, \epsilon)$*-hardness. Indeed, this statement is* equivalent *to UGC.*

Theorem 2.14 *The MAX-q-CUT problem, i.e. Approximate q-Coloring, has* $(1 - 1/q + q^{2+o(1)})$*-hardness factor, assuming UGC only. This asymptotically matches the approximation factor obtained by Frieze and Jerrum [26].*

2.4 It Ain't Over Till It's Over

The It Ain't Over Till It's Over conjecture was originally made by Kalai and Friedgut [33] in studying social indeterminacy [25, 32]. The setting here is similar to the setting of Arrow's Theorem from Section 2.3.1 except that there are an arbitrary finite number of candidates. Let R denote the (asymmetric) relation given on the candidates when the *monotone* social choice function f is used. Kalai showed that if f has small influences, then the It Ain't Over Till It's Over Conjecture implies that *every* possible relation R is achieved with probability bounded away from 0. Since its introduction in 2001, the It Ain't Over Till It's Over problem has circulated widely in the community studying harmonic analysis of boolean functions. The conjecture was given as one of the top unsolved problems in the field at a workshop at Yale in late 2004.

In Theorem 4.9 we confirm the It Ain't Over Till It's Over conjecture and generalize it to functions on arbitrary finite product probability spaces with means bounded away from 0 and 1. Further, the asymptotics we give show that symmetric threshold functions (e.g., Majority in the case of mean 1/2) are the "worst" examples. We give a slightly simplified statement of Theorem 4.9 here:

Theorem 4.9 *Let $0 < \rho < 1$ and let $f : \Omega_1 \times \cdots \times \Omega_n \to [0,1]$ be a function on a discrete product probability space; assume that for each i the minimum probability of any atom in Ω_i is at least $\alpha \le 1/2$. Then there exists $\epsilon(\rho,\mu) > 0$ such that if $\epsilon < \epsilon(\rho,\mu)$ and $\mathrm{Inf}_i(f) \le \epsilon^{O(\sqrt{\log(1/\epsilon)})}$ for all i and $\mu = \mathbf{E}[f]$ then*

$$\mathbf{P}[V_\rho f > 1-\delta] \le \epsilon$$

and

$$\mathbf{P}[V_\rho f < \delta] \le \epsilon$$

provided

$$\delta < \epsilon^{\rho/(1-\rho)+O(1/\sqrt{\log(1/\epsilon)})},$$

where the $O(\cdot)$ hides a constant depending only on α, $1-\mu$, ρ, and $1-\rho$.

3 The invariance principle

3.1 Setup and notation

In this section we will describe the setup and notation necessary for our invariance principle. Recall that we are interested in functions on finite product probability spaces, $f : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$. For each i, the space of all functions $\Omega_i \to \mathbb{R}$ can be expressed as the span of a finite set of orthonormal random variables, $\boldsymbol{X}_{i,0} = 1, \boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, \boldsymbol{X}_{i,3}, \ldots$; then f can be written as a multilinear polynomial in the $\boldsymbol{X}_{i,j}$'s. In fact, it will be convenient for us to mostly disregard the Ω_i's and work directly with sets of orthonormal random variables; in this case, we can even drop the restriction of finiteness. We thus begin with the following definition:

Definition 3.1 *We call a collection of finitely many orthonormal real random variables, one of which is the constant 1, an* orthonormal ensemble. *We will write a typical* sequence *of n orthonormal ensembles as $\boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_n)$, where $\boldsymbol{\mathcal{X}}_i = \{\boldsymbol{X}_{i,0} = 1, \boldsymbol{X}_{i,1}, \ldots, \boldsymbol{X}_{i,m_i}\}$. We call a sequence of orthonormal ensembles $\boldsymbol{\mathcal{X}}$* independent *if the ensembles are independent families of random variables.*

We will henceforth be concerned only with independent sequences of orthonormal ensembles, and we will call these sequences of ensembles*, for brevity.*

Remark 3.2 *Given a sequence of independent random variables $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ with $\mathbf{E}[\boldsymbol{X}_i] = 0$ and $\mathbf{E}[\boldsymbol{X}_i^2] = 1$ (as in Theorem 2.1), we can view them as a sequence of ensembles $\boldsymbol{\mathcal{X}}$ by renaming $\boldsymbol{X}_i = \boldsymbol{X}_{i,1}$ and setting $\boldsymbol{X}_{i,0} = 1$ as required.*

Definition 3.3 *We denote by $\boldsymbol{\mathcal{G}}$ the* Gaussian sequence of ensembles*, in which $\boldsymbol{\mathcal{G}}_i = \{\boldsymbol{G}_{i,0} = 1, \boldsymbol{G}_{i,1}, \boldsymbol{G}_{i,2}, \dots\}$ and all $\boldsymbol{G}_{i,j}$'s with $j \geq 1$ are independent standard Gaussians.*

As mentioned, we will be interested in *multilinear polynomials* over sequences of ensembles. By this we mean sums of products of the random variables, where each product is obtained by multiplying one random variable from each ensemble.

Definition 3.4 *A* multi-index *$\boldsymbol{\sigma}$ is a sequence $(\sigma_1, \dots, \sigma_n)$ in $\mathbb{N}^n$; the* degree *of $\boldsymbol{\sigma}$, denoted $|\boldsymbol{\sigma}|$, is $|\{i \in [n] : \sigma_i > 0\}|$. Given a doubly-indexed set of indeterminates $\{x_{i,j}\}_{i \in [n], j \in \mathbb{N}}$, we write $x_{\boldsymbol{\sigma}}$ for the monomial $\prod_{i=1}^n x_{i,\sigma_i}$. We now define a* multilinear polynomial *over such a set of indeterminates to be any expression*

$$Q(x) = \sum_{\boldsymbol{\sigma}} c_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}} \tag{5}$$

where the $c_{\boldsymbol{\sigma}}$'s are real constants, all but finitely many of which are zero. The degree *of $Q(x)$ is $\max\{|\boldsymbol{\sigma}| : c_{\boldsymbol{\sigma}} \neq 0\}$, at most n. We also use the notation*

$$Q^{\leq d}(x) = \sum_{|\boldsymbol{\sigma}| \leq d} c_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}$$

and the analogous $Q^{=d}(x)$ and $Q^{>d}(x)$.

Naturally, we will consider applying multilinear polynomials Q to sequences of ensembles $\boldsymbol{\mathcal{X}}$; the distribution of these random variables $Q(\boldsymbol{\mathcal{X}})$ is the subject of our invariance principle. Since $Q(\boldsymbol{\mathcal{X}})$ can be thought of as a function on a product space $\Omega_1 \times \cdots \times \Omega_n$ as described at the beginning of this section, there is a consistent way to define the notions of influences, T_ρ, and noise stability from Section 2.2. For example, the "influence of the ith ensemble on Q" is

$$\mathrm{Inf}_i(Q(\boldsymbol{\mathcal{X}})) = \mathbf{E}[\mathbf{Var}[Q(\boldsymbol{\mathcal{X}}) \mid \boldsymbol{\mathcal{X}}_1, \dots, \boldsymbol{\mathcal{X}}_{i-1}, \boldsymbol{\mathcal{X}}_{i+1}, \dots, \boldsymbol{\mathcal{X}}_n]].$$

Using independence and orthonormality, it is easy to show the following formulas, familiar from harmonic analysis of boolean functions:

Proposition 3.5 *Let $\boldsymbol{\mathcal{X}}$ be a sequence of ensembles and Q a multilinear polynomial as in (5). Then*

$$\mathbf{E}[Q(\boldsymbol{\mathcal{X}})] = c_{\mathbf{0}}; \qquad \mathbf{E}[Q(\boldsymbol{\mathcal{X}})^2] = \sum_{\boldsymbol{\sigma}} c_{\boldsymbol{\sigma}}^2; \qquad \mathbf{Var}[Q(\boldsymbol{\mathcal{X}})] = \sum_{|\boldsymbol{\sigma}|>0} c_{\boldsymbol{\sigma}}^2;$$

$$\mathrm{Inf}_i(Q(\boldsymbol{\mathcal{X}})) = \sum_{\boldsymbol{\sigma}:\sigma_i>0} c_{\boldsymbol{\sigma}}^2; \qquad T_\rho Q(\boldsymbol{\mathcal{X}}) = \sum_{\boldsymbol{\sigma}} \rho^{|\boldsymbol{\sigma}|} c_{\boldsymbol{\sigma}} \boldsymbol{\mathcal{X}}_{\boldsymbol{\sigma}}; \qquad \mathbb{S}_\rho(Q(\boldsymbol{\mathcal{X}})) = \sum_{\boldsymbol{\sigma}} \rho^{|\boldsymbol{\sigma}|} c_{\boldsymbol{\sigma}}^2.$$

Note that in each case above, the formula does not depend on the sequence of ensembles $\boldsymbol{\mathcal{X}}$; it only depends on Q. Thus we are justified in henceforth writing $\mathbf{E}[Q]$, $\mathbf{E}[Q^2]$, $\mathbf{Var}[Q]$, $\mathrm{Inf}_i(Q)$, and $\mathbb{S}_\rho(Q)$, and in treating T_ρ as a formal operator on multilinear polynomials:

Definition 3.6 *For $\rho \in [0,1]$ we define the operator T_ρ as acting formally on multilinear polynomials $Q(x)$ as in (5) by*

$$(T_\eta Q)(x) = \sum_{\boldsymbol{\sigma}} \rho^{|\boldsymbol{\sigma}|} c_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}.$$

Note that for every sequence of ensembles, we have that Definition 3.6 agrees with Definition 2.4.

We end this section with a short discussion of "low-degree influences", a notion that has proven crucial in the analysis of PCPs (see, e.g., [35]).

Definition 3.7 *The d-*low-degree influence of the *i*th ensemble on *$Q(\boldsymbol{\mathcal{X}})$ is*

$$\mathrm{Inf}_i^{\leq d}(Q(\boldsymbol{\mathcal{X}})) = \mathrm{Inf}_i^{\leq d}(Q) = \sum_{\boldsymbol{\sigma}:|\boldsymbol{\sigma}|\leq d, \sigma_i > 0} c_{\boldsymbol{\sigma}}^2.$$

Note that this gives a way to define low-degree influences $\mathrm{Inf}_i^{\leq d}(f)$ for functions $f : \Omega_1 \times \cdots \Omega_n \to \mathbb{R}$ on finite product spaces.

There isn't an especially natural interpretation of $\mathrm{Inf}_i^{\leq d}(f)$. However, the notion is important for PCPs due to the fact that a function with variance 1 cannot have too many coordinates with substantial low-degree influence; this is reflected in the following easy proposition:

Proposition 3.8 *Suppose Q is multilinear polynomial as in (5). Then*

$$\sum_i \mathrm{Inf}_i^{\leq d}(Q) \leq d \cdot \mathbf{Var}[Q].$$

3.2 Hypercontractivity

As mentioned in Section 2.1, our invariance principle requires that the ensembles involved to be hypercontractive in a certain sense. Recall that a random variable $\boldsymbol{Y}$ is said to be "(p,q,η)-hypercontractive" for $1 \leq p \leq q < \infty$ and $0 < \eta < 1$ if

$$\|a + \eta \boldsymbol{Y}\|_q \leq \|a + \boldsymbol{Y}\|_p \tag{6}$$

for all $a \in \mathbb{R}$. This type of hypercontractivity was introduced (with slightly different notation) in [39]. Some basic facts about hypercontractivity are explained in Appendix A; much more can be found in [40]. Here we just note that for $q > 2$ a random variable $\boldsymbol{Y}$ is $(2,q,\eta)$-hypercontractive with some $\eta \in (0,1)$ if and only if $\mathbf{E}[\boldsymbol{Y}] = 0$ and $\mathbf{E}[|\boldsymbol{Y}|^q] < \infty$. Also, if $\boldsymbol{Y}$ is $(2,q,\eta)$-hypercontractive then $\eta \leq (q-1)^{-1/2}$.

We now define our extension of the notion of hypercontractivity to sequences of ensembles:

Definition 3.9 *Let $\boldsymbol{\mathcal{X}}$ be a sequence of ensembles. For $1 \leq p \leq q < \infty$ and $0 < \eta < 1$ we say that $\boldsymbol{\mathcal{X}}$ is (p,q,η)-*hypercontractive *if*

$$\|(T_\eta Q)(\boldsymbol{\mathcal{X}})\|_q \leq \|Q(\boldsymbol{\mathcal{X}})\|_p$$

for every multilinear polynomial Q over $\boldsymbol{\mathcal{X}}$.

Since T_η is a contractive semi-group, we have

Remark 3.10 *If $\boldsymbol{\mathcal{X}}$ is (p,q,η)-hypercontractive then it is (p,q,η')-hypercontractive for any $0 < \eta' \le \eta$.*

There is a related notion of hypercontractivity for *sets* of random variables which considers all polynomials in the variables, not just multilinear polynomials; see, e.g., Janson [29]. Several of the properties of this notion of hypercontractivity carry over to our setting of sequences of ensembles. In particular, the following facts can easily be proved by repeating the analogous proofs in [29]; for completeness, we give the proofs in Appendix A.

Proposition 3.11 *Suppose $\boldsymbol{\mathcal{X}}$ is a sequence of n_1 ensembles and $\boldsymbol{\mathcal{Y}}$ is an independent sequence of n_2 ensembles. Assume both are (p,q,η)-hypercontractive. Then the sequence of ensembles $\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}} = (\boldsymbol{\mathcal{X}}_1, \dots, \boldsymbol{\mathcal{X}}_{n_1}, \boldsymbol{\mathcal{Y}}_1, \dots, \boldsymbol{\mathcal{Y}}_{n_2})$ is also (p,q,η)-hypercontractive.*

Proposition 3.12 *Let $\boldsymbol{\mathcal{X}}$ be a $(2,q,\eta)$-hypercontractive sequence of ensembles and Q a multilinear polynomial over $\boldsymbol{\mathcal{X}}$ of degree d. Then*

$$\|Q(\boldsymbol{\mathcal{X}})\|_q \le \eta^{-d}\, \|Q(\boldsymbol{\mathcal{X}})\|_2.$$

In light of Proposition 3.11, to check that a sequence of ensembles is (p,q,η)-hypercontractive it is enough to check that each ensemble individually is (p,q,η)-hypercontractive (as a "sequence" of length 1); in turn, it is easy to see that this is equivalent to checking that for each i, all linear combinations of the random variables $\boldsymbol{X}_{i,1}, \dots, \boldsymbol{X}_{i,m_i}$ are hypercontractive in the traditional sense of (6).

We end this section by recording the optimal hypercontractivity constants for the ensembles we consider. The result for ± 1 Rademacher variables is well known and due originally to Bonami [7] and independently Beckner [3]; the same result for Gaussian and uniform random variables is also well known and in fact follows easily from the Rademacher case. The optimal hypercontractivity constants for general finite spaces was recently determined by Wolff [51] (see also [44]):

Theorem 3.13 *Let $\boldsymbol{X}$ denote either a uniformly random ± 1 bit, a standard one-dimensional Gaussian, or a random variable uniform on $[-\sqrt{3}, \sqrt{3}]$. Then $\boldsymbol{X}$ is $(2,q,(q-1)^{-1/2})$-hypercontractive.*

Theorem 3.14 *(Wolff) Let $\boldsymbol{X}$ be any mean-zero random variable on a finite probability space in which the minimum nonzero probability of any atom is $\alpha \le 1/2$. Then $\boldsymbol{X}$ is $(2,q,\eta_q(\alpha))$-hypercontractive, where*

$$\eta_q(\alpha) = \left(\frac{A^{1/q'} - A^{-1/q'}}{A^{1/q} - A^{-1/q}}\right)^{-1/2}$$

$$\text{with} \quad A = \frac{1-\alpha}{\alpha}, \quad 1/q + 1/q' = 1.$$

Note the following special case:

Proposition 3.15

$$\eta_3(\alpha) = \left(A^{1/3} + A^{-1/3}\right)^{-1/2} \overset{\alpha \to 0}{\sim} \alpha^{1/6},$$

and also

$$\tfrac{1}{2}\alpha^{1/6} \le \eta_3(\alpha) \le 2^{-1/2},$$

for all $\alpha \in [0, 1/2]$.

For general random variables with bounded moments we have the following results, proved in Appendix A:

Proposition 3.16 *Let $\boldsymbol{X}$ be a mean-zero random variable satisfying $\mathbf{E}[|\boldsymbol{X}|^q] < \infty$. Then $\boldsymbol{X}$ is $(2, q, \eta_q)$-hypercontractive with $\eta_q = \frac{\|\boldsymbol{X}\|_2}{2\sqrt{q-1}\|\boldsymbol{X}\|_q}$.*

In particular, when $\mathbf{E}[\boldsymbol{X}] = 0$, $\mathbf{E}[\boldsymbol{X}^2] = 1$, and $\mathbf{E}[|\boldsymbol{X}|^3] \leq \beta$, we have that $\boldsymbol{X}$ is $(2, 3, 2^{-3/2}\beta^{-1/3})$-hypercontractive.

Proposition 3.17 *Let $\boldsymbol{X}$ be a mean-zero random variable satisfying $\mathbf{E}[|\boldsymbol{X}|^q] < \infty$ and let $\boldsymbol{V}$ be a random variable independent of $\boldsymbol{X}$ with $\mathbf{P}[\boldsymbol{V} = 0] = 1 - \rho$ and $\mathbf{P}[\boldsymbol{V} = 1] = \rho$. Then $\boldsymbol{V}\boldsymbol{X}$ is $(2, q, \xi_q)$-hypercontractive with $\xi_q = \frac{\|\boldsymbol{X}\|_2}{2\sqrt{q-1}\|\boldsymbol{X}\|_q} \cdot \rho^{\frac{1}{2}-\frac{1}{q}}$.*

3.3 Hypotheses for invariance theorems — some families of ensembles

All of the variants of our invariance principle that we prove in this section will have similar hypotheses. Specifically, they will be concerned with a multilinear polynomial Q over two hypercontractive sequences of ensembles, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$; furthermore, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ will be assumed to have satisfy a "matching moments" condition, as described below. We will now lay out four hypotheses — $\boldsymbol{H1}, \boldsymbol{H2}, \boldsymbol{H3}$, and $\boldsymbol{H4}$ that will be used in the theorems of this section. As can easily be seen (using Theorems 3.13 and 3.14 and Proposition 3.15; see also Appendix A), the hypothesis $\boldsymbol{H1}$ generalizes $\boldsymbol{H2}, \boldsymbol{H3}$, and $\boldsymbol{H4}$; hence all proofs will be carried out only in the setting of $\boldsymbol{H1}$. However the amount of notation and number of parameters under $\boldsymbol{H1}$ is quite cumbersome, and the reader who is interested mainly in functions on finite product spaces ($\boldsymbol{H3}$) or just boolean functions where $\{-1, 1\}^n$ has the uniform distribution ($\boldsymbol{H4}$) may find it easier to proceed through the proofs and results in the restricted cases.

Herewith our hypotheses:

$\boldsymbol{H1}$ Let $r \geq 3$ be an integer and let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be independent sequences of n ensembles which are $(2, r, \eta)$-hypercontractive; recall that $\eta \leq (r-1)^{-1/2}$. Assume furthermore that for all $1 \leq i \leq n$ and all sets $\Sigma \subset \mathbb{N}$ with $|\Sigma| < r$, the sequences $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ satisfy the "matching moments" condition

$$\mathbf{E}\left[\prod_{\sigma \in \Sigma} \boldsymbol{X}_{i,\sigma}\right] = \mathbf{E}\left[\prod_{\sigma \in \Sigma} \boldsymbol{Y}_{i,\sigma}\right]. \tag{7}$$

Finally, let Q be a multilinear polynomial as in (5).

We remark that in $\boldsymbol{H1}$, if $r = 3$ then the matching moment conditions hold automatically since the sequences are orthonormal. We also remark that we have added the condition $\eta \leq (r-1)^{-1/2}$ so that we can take $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{G}}$, the Gaussian sequence of ensembles (see Theorem 3.13).

$\boldsymbol{H2}$ Let $r = 3$. Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be independent sequences of ensembles in which each ensemble has only two random variables, $\boldsymbol{X}_{i,0} = 1$ and $\boldsymbol{X}_{i,1} = \boldsymbol{X}_i$ (respectively, $\boldsymbol{Y}_{i,0} = 1$, $\boldsymbol{Y}_{i,1} = \boldsymbol{Y}_i$), as in Remark 3.2. Further assume that each $\boldsymbol{X}_i$ (respectively $\boldsymbol{Y}_i$) satisfies $\mathbf{E}[\boldsymbol{X}_i] = 0$, $\mathbf{E}[\boldsymbol{X}_i^2] = 1$ and $\mathbf{E}[|\boldsymbol{X}_i|^3] \leq \beta$. Put $\eta = 2^{-3/2}\beta^{-1/3}$, so $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are $(2, 3, \eta)$-hypercontractive. Finally, let Q be a multilinear polynomial as in (5).

The hypothesis $\boldsymbol{H2}$ is used to derive the multilinear version of the Berry-Esseen inequality given in Theorem 2.1.

$H3$ Let $r = 3$ and let $\boldsymbol{\mathcal{X}}$ be a sequence of n ensembles in which the random variables in each ensemble $\boldsymbol{\mathcal{X}}_i$ form a basis for the real-valued functions on some finite probability space Ω_i. Further assume that the least nonzero probability of any atom in any Ω_i is $\alpha \le 1/2$, and let $\eta = \frac{1}{2}\alpha^{1/6}$. Let $\boldsymbol{\mathcal{Y}}$ be any independent $(2, 3, \eta)$-hypercontractive sequence of ensembles. Finally, let Q be a multilinear polynomial as in (5).

We remark that $Q(\boldsymbol{\mathcal{X}})$ in $H3$ encompasses *all* real-valued functions f on finite product spaces, including the familiar cases of the p-biased discrete cube (for which $\alpha = \min\{p, 1-p\}$) and the set $[q]^n$ with uniform measure (for which $\alpha = 1/q$). Note also that $\eta \le 2^{-1/2}$ so we may take $\boldsymbol{\mathcal{Y}}$ to be the Gaussian sequence of ensembles.

$H4$ Let $r = 4$ and $\eta = 3^{-1/2}$. Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be independent sequences of ensembles in which each ensemble has only two random variables, $\boldsymbol{X}_{i,0} = 1$ and $\boldsymbol{X}_{i,1} = \boldsymbol{X}_i$ (respectively, $\boldsymbol{Y}_{i,0} = 1$, $\boldsymbol{Y}_{i,1} = \boldsymbol{Y}_i$), as in Remark 3.2. Further assume that each $\boldsymbol{X}_i$ (respectively $\boldsymbol{Y}_i$) is either a) a uniformly random ± 1 bit; b) a standard one-dimensional Gaussian; or c) uniform on $[-3^{1/2}, 3^{1/2}]$. Hence $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are $(2, 4, \eta)$-hypercontractive. Finally, let Q be a multilinear polynomial as in (5).

Note that this simplest of all hypotheses allows for arbitrary real-valued functions on the uniform-measure discrete cube $f : \{-1, 1\}^n \to \mathbb{R}$. Also, under $H4$, Q is just a multilinear polynomial in the usual sense over the $\boldsymbol{X}_i$'s or $\boldsymbol{Y}_i$'s; in particular, if $f : \{-1, 1\}^n \to \mathbb{R}$ then Q is the "Fourier expansion" of f. Finally, note that the matching moments condition (7) holds in $H4$ since it requires $\mathbf{E}[X_t^3] = \mathbf{E}[Y_t^3]$ for each t, and this is true since both equal 0.

3.4 Basic invariance principle, $\mathcal{C}^r$ functional version

The essence of our invariance principle is that if Q is of bounded degree and has low influences then the random variables $Q(\boldsymbol{\mathcal{X}})$ and $Q(\boldsymbol{\mathcal{Y}})$ are close in distribution. The simplest way to formulate this conclusion is to say that if $\Psi : \mathbb{R} \to \mathbb{R}$ is a sufficiently nice "test function" then $\Psi(Q(\boldsymbol{\mathcal{X}}))$ and $\Psi(Q(\boldsymbol{\mathcal{Y}}))$ are close in expectation.

Theorem 3.18 *Assume hypothesis $H1, H2, H3,$ or $H4$. Further assume* $\mathbf{Var}[Q] \le 1$, $\deg(Q) \le d$, *and* $\mathrm{Inf}_i(Q) \le \tau$ *for all i. Let $\Psi : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^r$ function with $|\Psi^{(r)}| \le B$ uniformly. Then*

$$\Big|\mathbf{E}\big[\Psi(Q(\boldsymbol{\mathcal{X}}))\big] - \mathbf{E}\big[\Psi(Q(\boldsymbol{\mathcal{Y}}))\big]\Big| \le \epsilon,$$

where

$$\epsilon = \begin{cases} (2B/r!)\, d\, \eta^{-rd}\, \tau^{r/2-1} & \textit{under } H1, \\ B\, 30^d\, \beta^d\, \tau^{1/2} & \textit{under } H2, \\ B\, (10\alpha^{-1/2})^d\, \tau^{1/2} & \textit{under } H3, \\ B\, 10^d\, \tau & \textit{under } H4. \end{cases}$$

As will be the case in all of our theorems, the results under $H2, H3$ and $H4$ are immediate corollaries of the result under $H1$; one only needs to substitute in $r = 3$, $\eta = 2^{-3/2}\beta^{-1/3}$ or $r = 3$, $\eta = \frac{1}{2}\alpha^{1/6}$ or $r = 4$, $\eta = 3^{-1/2}$ (we have also here used that $(1/3)\, d\, 2^{9d/2}$ is at most 30^d and that $(1/3)\, d\, 8^d$ and $(1/12)\, d\, 9^d$ are at most 10^d). Thus it will suffice for us to carry out the proof under $H1$.

Proof: We begin by defining intermediate sequences between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$. For $i = 0, 1, \ldots, n$, let $\boldsymbol{\mathcal{Z}}^{(i)}$ denote the sequence of n ensembles $(\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_{i+1}, \ldots, \boldsymbol{\mathcal{X}}_n)$ and let $\boldsymbol{Q}^{(i)} = Q(\boldsymbol{\mathcal{Z}}^{(i)})$. Our goal will be to show

$$\left|\mathbf{E}\big[\Psi(\boldsymbol{Q}^{(i-1)})\big] - \mathbf{E}\big[\Psi(\boldsymbol{Q}^{(i)})\big]\right| \leq \left(\frac{2B}{r!}\,\eta^{-rd}\right) \cdot \mathrm{Inf}_i(Q)^{r/2} \tag{8}$$

for each $i \in [n]$. Summing this over i will complete the proof since $\boldsymbol{\mathcal{Z}}^{(0)} = \boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Z}}^{(n)} = \boldsymbol{\mathcal{Y}}$, and

$$\sum_{i=1}^{n} \mathrm{Inf}_i(Q)^{r/2} \leq \tau^{r/2-1} \cdot \sum_{i=1}^{n} \mathrm{Inf}_i(Q) = \tau^{r/2-1} \cdot \sum_{i=1}^{n} \mathrm{Inf}_i^{\leq d}(Q) \leq d\tau^{r/2-1},$$

where we used Proposition 3.8 and $\mathbf{Var}[Q] \leq 1$.

Let us fix a particular $i \in [n]$ and proceed to prove (8). Given a multi-index $\boldsymbol{\sigma}$, write $\boldsymbol{\sigma} \setminus i$ for the same multi-index except with $\sigma_i = 0$. Now write

$$\begin{aligned}
\tilde{\boldsymbol{Q}} &= \sum_{\boldsymbol{\sigma}:\sigma_i = 0} c_{\boldsymbol{\sigma}} \boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}}, \\
\boldsymbol{R} &= \sum_{\boldsymbol{\sigma}:\sigma_i > 0} c_{\boldsymbol{\sigma}} \boldsymbol{X}_{i,\sigma_i} \cdot \boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}\setminus i}, \\
\boldsymbol{S} &= \sum_{\boldsymbol{\sigma}:\sigma_i > 0} c_{\boldsymbol{\sigma}} \boldsymbol{Y}_{i,\sigma_i} \cdot \boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}\setminus i}.
\end{aligned}$$

Note that $\tilde{\boldsymbol{Q}}$ and the variables $\boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}\setminus i}$ are independent of the variables in $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{Y}}_i$ and that $\boldsymbol{Q}^{(i-1)} = \tilde{\boldsymbol{Q}} + \boldsymbol{R}$ and $\boldsymbol{Q}^{(i)} = \tilde{\boldsymbol{Q}} + \boldsymbol{S}$.

To bound the left side of (8) — i.e., $|\mathbf{E}[\Psi(\tilde{\boldsymbol{Q}} + \boldsymbol{R}) - \Psi(\tilde{\boldsymbol{Q}} + \boldsymbol{S})]|$ — we use Taylor's theorem: for all $x, y \in \mathbb{R}$,

$$\left|\Psi(x+y) - \sum_{k=0}^{r-1} \frac{\Psi^{(k}(x)\; y^k}{k!}\right| \leq \frac{B}{r!}\,|y|^r.$$

In particular,

$$\left|\mathbf{E}[\Psi(\tilde{\boldsymbol{Q}} + \boldsymbol{R})] - \sum_{k=0}^{r-1} \mathbf{E}\Big[\frac{\Psi^{(k)}(\tilde{\boldsymbol{Q}})\; \boldsymbol{R}^k}{k!}\Big]\right| \leq \frac{B}{r!}\,\mathbf{E}\big[|\boldsymbol{R}|^r\big] \tag{9}$$

and similarly,

$$\left|\mathbf{E}[\Psi(\tilde{\boldsymbol{Q}} + \boldsymbol{S})] - \sum_{k=0}^{r-1} \mathbf{E}\Big[\frac{\Psi^{(k)}(\tilde{\boldsymbol{Q}})\; \boldsymbol{S}^k}{k!}\Big]\right| \leq \frac{B}{r!}\,\mathbf{E}\big[|\boldsymbol{S}|^r\big]. \tag{10}$$

We will see below that that $\boldsymbol{R}$ and $\boldsymbol{S}$ have finite r moments. Moreover, for $0 \leq k \leq r$ it holds that $|\Psi^{(k)}(\tilde{\boldsymbol{Q}})\,\boldsymbol{R}^k| \leq |k!\,B\,\tilde{\boldsymbol{Q}}^{r-k}\,\boldsymbol{R}^k|$ (and similarly for $\boldsymbol{S}$). Thus all moments above are finite. We now claim that for all $0 \leq k < r$ it holds that

$$\mathbf{E}[\Psi^{(k)}(\tilde{\boldsymbol{Q}})\,\boldsymbol{R}^k] = \mathbf{E}[\Psi^{(k)}(\tilde{\boldsymbol{Q}})\,\boldsymbol{S}^k]. \tag{11}$$

Indeed,

$$
\begin{aligned}
\mathbf{E}[\Psi^{(k)}(\tilde{\boldsymbol{Q}})\,\boldsymbol{R}^k] &= \mathbf{E}\Big[\Psi^{(k)}(\tilde{\boldsymbol{Q}}) \sum_{\substack{(\boldsymbol{\sigma}^1,\ldots,\boldsymbol{\sigma}^k)\\ \text{s.t. } \forall t,\ \sigma_i^t>0}} \prod_{t=1}^k c_{\boldsymbol{\sigma}^t} \prod_{t=1}^k \boldsymbol{X}_{i,\sigma_i^t} \prod_{t=1}^k \boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}^t\setminus i}\Big] \\
&= \sum_{\substack{(\boldsymbol{\sigma}^1,\ldots,\boldsymbol{\sigma}^k)\\ \text{s.t. } \forall t,\ \sigma_i^t>0}} \prod_{t=1}^k c_{\boldsymbol{\sigma}^i} \cdot \mathbf{E}\Big[\Psi^{(k)}(\tilde{\boldsymbol{Q}}) \prod_{t=1}^k \boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}^t\setminus i}\Big] \cdot \mathbf{E}\Big[\prod_{t=1}^k \boldsymbol{X}_{i,\sigma_i^t}\Big] \qquad (12) \\
&= \sum_{\substack{(\boldsymbol{\sigma}^1,\ldots,\boldsymbol{\sigma}^k)\\ \text{s.t. } \forall t,\ \sigma_i^t>0}} \prod_{t=1}^k c_{\boldsymbol{\sigma}^i} \cdot \mathbf{E}\Big[\Psi^{(k)}(\tilde{\boldsymbol{Q}}) \prod_{t=1}^k \boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}^t\setminus i}\Big] \cdot \mathbf{E}\Big[\prod_{t=1}^k \boldsymbol{Y}_{i,\sigma_i^t}\Big] \qquad (13) \\
&= \mathbf{E}\left[\Psi^{(k)}(\tilde{\boldsymbol{Q}})\,\boldsymbol{S}^k\right].
\end{aligned}
$$

The equality in (12) follows since $\boldsymbol{\mathcal{Z}}^{(i)}_{\boldsymbol{\sigma}^t\setminus i}$ and $\tilde{\boldsymbol{Q}}$ are independent of the variables in $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{Y}}_i$. The equality in (13) follows from the matching moments condition (7).

From (9), (10) and (11) it follows that

$$|\mathbf{E}[\Psi(\tilde{\boldsymbol{Q}}+\boldsymbol{R}) - \Psi(\tilde{\boldsymbol{Q}}+\boldsymbol{S})]| \le \frac{B}{r!}\,(\mathbf{E}[|\boldsymbol{R}|^r] + \mathbf{E}[|\boldsymbol{S}|^r]). \qquad (14)$$

We now use hypercontractivity. By Proposition 3.11 each $\boldsymbol{\mathcal{Z}}^{(i)}$ is $(2,r,\eta)$-hypercontractive. Thus by Proposition 3.12,

$$\mathbf{E}[|\boldsymbol{R}|^r] \le \eta^{-rd}\mathbf{E}[\boldsymbol{R}^2]^{r/2}, \quad \mathbf{E}[|\boldsymbol{S}|^r] \le \eta^{-rd}\mathbf{E}[\boldsymbol{S}^2]^{r/2}. \qquad (15)$$

However,

$$\mathbf{E}[\boldsymbol{S}^2] = \mathbf{E}[\boldsymbol{R}^2] = \sum_{\boldsymbol{\sigma}:\sigma_i>0} c_{\boldsymbol{\sigma}}^2 = \mathrm{Inf}_i(Q). \qquad (16)$$

Combining (14), (15) and (16) it follows that

$$|\mathbf{E}[\Psi(\tilde{\boldsymbol{Q}}+\boldsymbol{R}) - \Psi(\tilde{\boldsymbol{Q}}+\boldsymbol{S})]| \le \left(\frac{2B}{r!}\,\eta^{-rd}\right)\cdot \mathrm{Inf}_i(Q)^{r/2}$$

confirming (8) and completing the proof. $\Box$

3.5 Invariance principle — other functionals, and smoothed version

Our basic invariance principle shows that $\mathbf{E}[\Psi(Q(\boldsymbol{\mathcal{X}}))]$ and $\mathbf{E}[\Psi(Q(\boldsymbol{\mathcal{Y}}))]$ are close if Ψ is a $\mathcal{C}^r$ functional with bounded rth derivative. To show that the distributions of $Q(\boldsymbol{\mathcal{X}})$ and $Q(\boldsymbol{\mathcal{Y}})$ are close in other senses we need the invariance principle for less smooth functionals. This we can obtain using straightforward approximation arguments; we defer the proof of Theorem 3.19 which follows to Section 3.6.

Theorem 3.19 shows closeness of distribution in two senses. The first is closeness in *Lévy's metric*; recall that the distance between two random variables $\boldsymbol{R}$ and $\boldsymbol{S}$ in Lévy's metric is

$$d_L(\boldsymbol{R},\boldsymbol{S}) = \inf\{\lambda>0: \quad \forall t\in\mathbb{R},\ \ \mathbf{P}[\boldsymbol{S}\le t-\lambda]-\lambda \le \mathbf{P}[\boldsymbol{R}\le t] \le \mathbf{P}[\boldsymbol{S}\le t+\lambda]+\lambda\}.$$

We also show the distributions are close in the usual sense with a weaker bound; the proof of this goes by comparing the distributions of $Q(\boldsymbol{\mathcal{X}})$ and $Q(\boldsymbol{\mathcal{Y}})$ to $Q(\boldsymbol{\mathcal{G}})$ and noting that bounded-degree Gaussian polynomials are known to have low "small ball probabilities". Finally, Theorem 3.19 also shows L^1 closeness and, as a technical necessity for applications, shows closeness under the functional $\zeta : \mathbb{R} \to \mathbb{R}$ defined by

$$\zeta(x) = \begin{cases} x^2 & \text{if } x \leq 0, \\ 0 & \text{if } x \in [0,1], \\ (x-1)^2 & \text{if } x \geq 1; \end{cases} \tag{17}$$

this functional gives the squared distance to the interval $[0,1]$.

Theorem 3.19 *Assume Hypothesis* $\boldsymbol{H1}, \boldsymbol{H2}, \boldsymbol{H3}$*, or* $\boldsymbol{H4}$*. Further assume* $\mathbf{Var}[Q] \leq 1$*,* $\deg(Q) \leq d$ *and* $\mathrm{Inf}_i(Q) \leq \tau$ *for all* i*. Then*

$$\begin{aligned} \Big| \|Q(\boldsymbol{\mathcal{X}})\|_1 - \|Q(\boldsymbol{\mathcal{Y}})\|_1 \Big| &\leq O(\epsilon^{1/r}), & (18) \\ d_L(Q(\boldsymbol{\mathcal{X}}), Q(\boldsymbol{\mathcal{Y}})) &\leq O(\epsilon^{1/(r+1)}), & (19) \\ \Big| \mathbf{E}\big[\zeta(Q(\boldsymbol{\mathcal{X}}))\big] - \mathbf{E}\big[\zeta(Q(\boldsymbol{\mathcal{Y}}))\big] \Big| &\leq O(\epsilon^{2/r}), & (20) \end{aligned}$$

where $O(\cdot)$ *hides a constant depending only on* r*, and*

$$\epsilon = \begin{cases} d\,\eta^{-rd}\,\tau^{r/2-1} & \textit{under } \boldsymbol{H1}, \\ 30^d \beta^d\, \tau^{1/2} & \textit{under } \boldsymbol{H2}, \\ (10\alpha^{-1/2})^d\, \tau^{1/2} & \textit{under } \boldsymbol{H3}, \\ 10^d\, \tau & \textit{under } \boldsymbol{H4}. \end{cases}$$

If in addition $\mathbf{Var}[Q] = 1$ *then*

$$\sup_t \Big| \mathbf{P}\big[Q(\boldsymbol{\mathcal{X}}) \leq t\big] - \mathbf{P}\big[Q(\boldsymbol{\mathcal{Y}}) \leq t\big] \Big| \leq O\big(d\,\epsilon^{1/(rd+1)}\big). \tag{21}$$

As discussed in Section 2.1, Theorem 3.19 has the unavoidable deficiency of having error bounds depending on the degree d of Q. This can be overcome if we first "smooth" Q by applying $T_{1-\gamma}$ to it, for some $0 < \gamma < 1$. Theorem 3.20 which follows will be our main tool for applications; its proof is a straightforward degree truncation argument which we also defer to Section 3.6. As an additional benefit of this argument, we will show that Q need only have small *low-degree influences*, $\mathrm{Inf}_i^{\leq d}(Q)$, as opposed to small influences. As discussed at the end of Section 3.1, this feature has proven essential for applications involving PCPs.

Theorem 3.20 *Assume hypothesis* $\boldsymbol{H1}, \boldsymbol{H3}$*,or* $\boldsymbol{H4}$*. Further assume* $\mathbf{Var}[Q] \leq 1$ *and* $\mathrm{Inf}_i^{\leq \log(1/\tau)/K}(Q) \leq \tau \leq$ *for all* i*, where*

$$K = \begin{cases} \log(1/\eta) & \textit{under } \boldsymbol{H1}, \\ \log(1/\alpha) & \textit{under } \boldsymbol{H3}, \\ 1 & \textit{under } \boldsymbol{H4}. \end{cases}$$

Given $0 < \gamma < 1$*, write* $\boldsymbol{R} = (T_{1-\gamma}Q)(\boldsymbol{\mathcal{X}})$ *and* $\boldsymbol{S} = (T_{1-\gamma}Q)(\boldsymbol{\mathcal{Y}})$*. Then*

$$\begin{aligned} d_L(\boldsymbol{R}, \boldsymbol{S}) &\leq \tau^{\Omega(\gamma/K)}, \\ \Big| \mathbf{E}\big[\zeta(\boldsymbol{R})\big] - \mathbf{E}\big[\zeta(\boldsymbol{S})\big] \Big| &\leq \tau^{\Omega(\gamma/K)}, \end{aligned}$$

where the $\Omega(\cdot)$ hides a constant depending only on r.

More generally the statement of the theorem holds for $\boldsymbol{R} = Q(\boldsymbol{\mathcal{X}}), \boldsymbol{S} = Q(\boldsymbol{\mathcal{Y}})$ if $\mathbf{Var}[Q^{>d}] \leq (1-\gamma)^{2d}$ for all d.

3.6 Proofs of extensions of the invariance principle

In this section we will prove Theorems 3.19 and 3.20 under hypothesis $\boldsymbol{H1}$. The results under $\boldsymbol{H2}, \boldsymbol{H3}$, and $\boldsymbol{H4}$ are corollaries.

3.6.1 Invariance principle for some C^0 and C^1 functionals

In this section we prove (18), (19), (20) of Theorem 3.19. We do it by approximating the following functions in the sup norm by smooth functions:

$$\ell_1(x) = |x|; \quad \Delta_{s,t}(x) = \begin{cases} 1 & \text{if } x \leq t-s, \\ \frac{t-x+s}{2s} & \text{if } x \in [t-s, t+s], \\ 0 & \text{if } x \geq t+s; \end{cases} \quad \zeta(x) = \begin{cases} x^2 & \text{if } x \leq 0, \\ 0 & \text{if } x \in [0,1], \\ (x-1)^2 & \text{if } x \geq 1. \end{cases}$$

Lemma 3.21 *Let $r \geq 2$ be an integer. Then there exist constant B_r for which the following holds. For all $0 < \lambda \leq 1/2$ there exist $\mathcal{C}^\infty$ functions ℓ_1^λ, $\Delta_{\lambda,t}^\lambda$ and ζ^λ satisfying the following:*

- *$\|\ell_1^\lambda - \ell_1\|_\infty \leq 2\lambda$; and, $\|(\ell_1^\lambda)^{(r)}\|_\infty \leq 4B_r\,\lambda^{1-r}$.*
- *$\Delta_{\lambda,t}^\lambda$ agrees with $\Delta_{\lambda,t}$ outside the interval $(t-2\lambda, t+2\lambda)$, and is otherwise in $[0,1]$; and, $\|(\Delta_{\lambda,t}^\lambda)^{(r)}\|_\infty \leq B_r\,\lambda^{-r}$.*
- *$\|\zeta^\lambda - \zeta\|_\infty \leq 2\lambda^2$; and, $\|(\zeta^\lambda)^{(r)}\|_\infty \leq 2B_{r-1}\,\lambda^{2-r}$.*

Proof: Let $f(x) = x1_{\{x \geq 0\}}$. We will show that for all $\lambda > 0$ there is a $\mathcal{C}^\infty$ function f_λ satisfying the following:

- f_λ and f agree on $(-\infty, -\lambda]$ and $[\lambda, \infty)$;
- $0 \leq f_\lambda(x) \leq f(x) + \lambda$ on $(-\lambda, \lambda)$; and,
- $\|f_\lambda^{(r)}\|_\infty \leq 2B_r\,\lambda^{1-r}$.

The construction of f easily gives the construction of the other functionals by letting $\ell_1^\lambda(x) = f_\lambda(x) + f_\lambda(-x)$ and

$$\Delta_{\lambda,t}^\lambda(x) = \begin{cases} \frac{1}{2\lambda} f_\lambda(t-x+\lambda) & \text{if } x \geq t, \\ 1 - \frac{1}{2\lambda} f_\lambda(x-t+\lambda) & \text{if } x \leq t; \end{cases} \qquad \zeta^\lambda(x) = \begin{cases} \int_{-\infty}^{x-1} f_\lambda(t)dt & \text{if } x \geq 1/2, \\ \int_{-\infty}^{1-x} f_\lambda(t)dt & \text{if } x \leq 0. \end{cases} \tag{22}$$

To construct f, first let ψ be a nonnegative $\mathcal{C}^\infty$ function satisfying the following: ψ is 0 outside $(-1,1)$, $\int_{-1}^1 \psi(x)\,dx = 1$, and $\int_{-1}^1 x\psi(x)\,dx = 0$. It is well known that such functions ψ exist. Define the constant B_r to be $\|\psi^{(r)}\|_\infty$.

Next, write $\psi_\lambda(x) = \psi(x/\lambda)/\lambda$, so ψ_λ satisfies the same three properties as ψ with respect to the interval $(-\lambda, \lambda)$ rather than $(-1,1)$. Note that $\|\psi_\lambda^{(r)}\|_\infty = B_r\,\lambda^{-1-r}$.

Finally, take $f_\lambda = f * \psi_\lambda$, which is $\mathcal{C}^\infty$. The first two properties demanded of f follow easily. To see the third, first note that $f_\lambda^{(r)}$ is identically 0 outside $(-\lambda, \lambda)$ and then observe that for $|x| < \lambda$,

$$|f_\lambda^{(r)}(x)| = |(f * \psi_\lambda)^{(r)}(x)| = |(f * \psi_\lambda^{(r)})(x)| \leq \|\psi_\lambda^{(r)}\|_\infty \cdot \int_{x-\lambda}^{x+\lambda} |f| \leq 2B_r \lambda^{1-r}.$$

This completes the proof. □

We now prove (18), (19) and (20).

Proof: Note that the properties of $\Delta_{\lambda,t}^{\lambda}$ imply that

$$\mathbf{P}[\boldsymbol{R} \leq t - 2\lambda] \leq \mathbf{E}[\Delta_{\lambda,t}^{\lambda}(\boldsymbol{R})] \leq \mathbf{P}[\boldsymbol{R} \leq t + 2\lambda] \tag{23}$$

holds for every random variable $\boldsymbol{R}$ and every t and $0 < \lambda \leq 1/2$.

Let us first prove (18), with

$$\epsilon = d\,\eta^{-rd}\,\tau^{r/2-1}$$

since we assume $\boldsymbol{H1}$. Taking $\Psi = \ell_1^\lambda$ in Theorem 3.18 we obtain

$$\begin{aligned}\Big|\mathbf{E}\big[\ell_1(Q(\boldsymbol{\mathcal{X}}))\big] - E\big[\ell_1(Q(\boldsymbol{\mathcal{Y}}))\big]\Big| &\leq \Big|\mathbf{E}\big[\ell_1^\lambda(Q(\boldsymbol{\mathcal{X}}))\big] - E\big[\ell_1^\lambda(Q(\boldsymbol{\mathcal{Y}}))\big]\Big| + 4\lambda \\ &\leq (4B_r\,\lambda^{1-r}/r!)\,d\,\eta^{-rd}\,\tau^{r/2-1} + 4\lambda = O(\epsilon\,\lambda^{1-r}) + 4\lambda.\end{aligned}$$

Taking $\lambda = \epsilon^{1/r}$, gives the bound (18). Next, using (23) and applying Theorem 3.18 with $\Psi = \Delta_{\lambda,t}^\lambda$ we obtain

$$\begin{aligned}d_L(Q(\boldsymbol{\mathcal{X}}), Q(\boldsymbol{\mathcal{Y}})) &\leq \max\left\{4\lambda, \sup_t \Big|\mathbf{E}\big[\Delta_{\lambda,t}^\lambda(Q(\boldsymbol{\mathcal{X}}))\big] - \mathbf{E}\big[\Delta_{\lambda,t}^\lambda(Q(\boldsymbol{\mathcal{Y}}))\big]\Big|\right\} \\ &\leq \max\left\{(B_r\,\lambda^{-r}/r!)\,d\,\eta^{-rd}\,\tau^{r/2-1}, 4\lambda\right\} = \max\{O(\epsilon\,\lambda^{-r}), 4\lambda\}.\end{aligned}$$

Again taking $\lambda = \epsilon^{1/(r+1)}$ we achieve (19). Finally, using $\Psi = \zeta^\lambda$ we get

$$\begin{aligned}\Big|\mathbf{E}\big[\zeta(Q(\boldsymbol{\mathcal{X}}))\big] - \mathbf{E}\big[\zeta(Q(\boldsymbol{\mathcal{Y}}))\big]\Big| &\leq \Big|\mathbf{E}\big[\zeta^\lambda(Q(\boldsymbol{\mathcal{X}}))\big] - \mathbf{E}\big[\zeta^\lambda(Q(\boldsymbol{\mathcal{Y}}))\big]\Big| + 4\lambda^2 \\ &\leq (2B_{r-1}\,\lambda^{2-r}/r!)\,d\,\eta^{-rd}\,\tau^{r/2-1} + 4\lambda^2 = O(\epsilon\,\lambda^{2-r}) + 4\lambda^2,\end{aligned}$$

and taking $\lambda = \epsilon^{1/r}$ we get (20). This concludes the proof of the first three bounds in Theorem 3.19. □

3.6.2 Closeness in distribution

We proceed to prove (21) from Theorem 3.19. By losing constant factors it will suffice to prove the bound in the case that $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{G}}$, the sequence of independent Gaussian ensembles. As mentioned, we will use the fact that bounded-degree multilinear polynomials over $\boldsymbol{\mathcal{G}}$ have low "small ball probabilities". Specifically, the following theorem is an immediate consequence of Theorem 8 in [12] (taking $q = 2d$ in their notation):

Theorem 3.22 *There exists a universal constant C such that for all multilinear polynomials Q of degree d over $\boldsymbol{\mathcal{G}}$ and all $\epsilon > 0$,*

$$\mathbf{P}[|Q(\boldsymbol{\mathcal{G}})| \le \epsilon] \le C\, d\, (\epsilon/\|Q(\boldsymbol{\mathcal{G}})\|_2)^{1/d}.$$

Thus we have the following:

Corollary 3.23 *For all multilinear polynomials Q of degree d over $\boldsymbol{\mathcal{G}}$ with $\mathbf{Var}[Q] = 1$ and for all $t \in \mathbb{R}$ and $\epsilon > 0$,*

$$\mathbf{P}[|Q(\boldsymbol{\mathcal{G}}) - t| \le \epsilon] \le O(d\, \epsilon^{1/d}).$$

We now prove (21).

Proof: We will use Theorem 3.18 with $\Psi = \Delta^{\lambda}_{\lambda,t}$, where λ will be chosen later. Writing $\Delta_t = \Delta^{\lambda}_{\lambda,t}$ for brevity and using fact (23) twice, we have

$$\begin{aligned}
\mathbf{P}[Q(\boldsymbol{\mathcal{X}}) \le t] &\le \mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{X}})] \\
&\le \mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{G}})] + |\mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{X}})] - \mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{G}})]| \\
&\le \mathbf{P}[Q(\boldsymbol{\mathcal{G}}) \le t + 4\lambda] + |\mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{X}})] - \mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{G}})]| \\
&= \mathbf{P}[Q(\boldsymbol{\mathcal{G}}) \le t] + \mathbf{P}[t < Q(\boldsymbol{\mathcal{G}}) \le t + 4\lambda] + |\mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{X}})] - \mathbf{E}[\Delta_{t+2\lambda}(\boldsymbol{\mathcal{G}})]|. \quad (24)
\end{aligned}$$

The second quantity in (24) is at most $O(d\,(4\lambda)^{1/d})$ by Corollary 3.23; the third quantity in (24) is at most $O(\epsilon\,\lambda^{-r})$ by Lemma 3.21 and Theorem 3.18. Thus we conclude

$$\mathbf{P}[Q(\boldsymbol{\mathcal{X}}) \le t] \le \mathbf{P}[Q(\boldsymbol{\mathcal{G}}) \le t] + O(d\,\lambda^{1/d}) + O(\epsilon\,\lambda^{-r}),$$

independently of t. Similarly it follows that

$$\mathbf{P}[Q(\boldsymbol{\mathcal{X}}) \le t] \ge \mathbf{P}[Q(\boldsymbol{\mathcal{G}}) \le t] - O(d\,\lambda^{1/d}) - O(\epsilon\,\lambda^{-r}).$$

independently of t. Choosing $\lambda = \epsilon^{d/(rd+1)}$ we get

$$\Big|\mathbf{P}\big[Q(\boldsymbol{\mathcal{X}}) \le t\big] - \mathbf{P}\big[Q(\boldsymbol{\mathcal{G}}) \le t\big]\Big| \le O(d\,\epsilon^{1/(rd+1)}),$$

as required. □

The proof of Theorem 3.19 is now complete.

3.6.3 Invariance principle for smoothed functions

The proof of Theorem 3.20 is by truncating at degree $d = c\log(1/\tau)/\log(1/\eta)$, where $c > 0$ is a sufficiently small constant to be chosen later. Let $L(\boldsymbol{R}) = (T_{1-\gamma}Q)^{\le d}(\boldsymbol{\mathcal{X}})$, $H(\boldsymbol{R}) = (T_{1-\gamma}Q)^{>d}(\boldsymbol{\mathcal{X}})$, and define $L(\boldsymbol{S})$, and $H(\boldsymbol{S})$ analogously for $\boldsymbol{\mathcal{Y}}$. Note that the low-degree influences of $T_{1-\gamma}Q$ are no more than those of Q.

We first prove the upper bound on $d_L(\boldsymbol{R}, \boldsymbol{S})$. By Theorem 3.19 we have

$$d_L(L(\boldsymbol{R}), L(\boldsymbol{S})) \le d^{\Theta(1)}\, \eta^{-\Theta(d)}\, \tau^{\Theta(1)} = \eta^{-\Theta(d)}\, \tau^{\Theta(1)}. \quad (25)$$

As for $H(\boldsymbol{R})$ and $H(\boldsymbol{S})$ we have $\mathbf{E}[H(\boldsymbol{R})] = \mathbf{E}[H(\boldsymbol{S})] = 0$ and $\mathbf{E}[H(\boldsymbol{R})^2] = \mathbf{E}[H(\boldsymbol{S})^2] \le (1-\gamma)^{2d}$ (since $\mathbf{Var}[Q] \le 1$). Thus by Chebyshev's inequality it follows that for all λ,

$$\mathbf{P}[|H(\boldsymbol{R})| \ge \lambda] \le (1-\gamma)^{2d}/\lambda^2, \qquad \mathbf{P}[|H(\boldsymbol{S})| \ge \lambda] \le (1-\gamma)^{2d}/\lambda^2. \quad (26)$$

Combining (25) and (26) and taking $\lambda = (1-\gamma)^{2d/3}$ we conclude that the Lévy distance between $\boldsymbol{R}$ and $\boldsymbol{S}$ is at most

$$\eta^{-\Theta(d)}\,\tau^{\Theta(1)} + 4(1-\gamma)^{2d/3} \le \eta^{-\Theta(d)}\,\tau^{\Theta(1)} + \exp\bigl(-\gamma\,\Theta(d)\bigr). \tag{27}$$

Our choice of d, with c taken sufficiently small so that the second term above dominates, completes the proof of the upper bound on $d_L(\boldsymbol{R}, \boldsymbol{S})$.

To prove the claim about ζ we need the following simple lemma:

Lemma 3.24 *For all $a, b \in \mathbb{R}$, $|\zeta(a+b) - \zeta(a)| \le 2|ab| + 2b^2$.*

Proof: We have

$$|\zeta(a+b) - \zeta(a)| \le |b| \sup_{x\in[a,a+b]} |\zeta'(x)|.$$

The claim follows since $\zeta'(x) = 0$ for $|x| \le 1$ and $|\zeta'(x)| = 2||x|-1| \le 2|x|$ for $|x| \ge 1$. □

By (20) in Theorem 3.19 we get the upper bound of $\eta^{-\Theta(d)}\,\tau^{\Theta(1)}$ for $|\mathbf{E}[\zeta(L(\boldsymbol{R})) - \zeta(L(\boldsymbol{S}))]|$. The Lemma above and Cauchy-Schwartz imply

$$\begin{aligned}\mathbf{E}\Bigl[\bigl|\zeta(\boldsymbol{R})) - \zeta(L(\boldsymbol{R}))\bigr|\Bigr] &= \mathbf{E}\Bigl[\bigl|\zeta(L(\boldsymbol{R}) + H(\boldsymbol{R})) - \zeta(L(\boldsymbol{R}))\bigr|\Bigr] \le 2\mathbf{E}\bigl[|L(\boldsymbol{R})H(\boldsymbol{R})|\bigr] + \mathbf{E}\bigl[H(\boldsymbol{R})^2\bigr] \\ &\le 2\sqrt{\mathbf{E}[H(\boldsymbol{R})^2]} + \mathbf{E}[H(\boldsymbol{R})^2] \le 2(1-\gamma)^d + (1-\gamma)^{2d} \le \exp\bigl(-\gamma\,\Theta(d)\bigr),\end{aligned}$$

and similarly for $\boldsymbol{S}$. Thus

$$|\mathbf{E}[\zeta(\boldsymbol{R})] - \mathbf{E}[\zeta(\boldsymbol{S})]| \le \eta^{-\Theta(d)}\,\tau^{\Theta(1)} + \exp\bigl(-\gamma\,\Theta(d)\bigr)$$

as in (27) and we get the same upper bound.

Finally, it is easy to see that the second statement of the theorem also holds as the only property of $\boldsymbol{R}$ we have used is that $\mathbf{Var}[Q^{>d}] \le (1-\gamma)^{2d}$ for all d.

3.7 Invariance principle under Lyapunov conditions

Here we sketch a proof of Theorem 2.2.

Proof: (sketch) Let $\Delta : \mathbb{R} \to [0,1]$ be a nondecreasing smooth function with $\Delta(0) = 0$, $\Delta(1) = 1$ and $A := \sup_{x\in\mathbb{R}} |\Delta'''(x)| < \infty$. Then $\sup_{x\in\mathbb{R}} |\Delta''(x)| \le A/2$ and therefore for $x, y \in \mathbb{R}$ we have

$$|\Delta''(x) - \Delta''(y)| \le A^{3-q}|\Delta''(x) - \Delta''(y)|^{q-2} \le A^{3-q}(A|x-y|)^{q-2} = A|x-y|^{q-2}.$$

For $s > 0$ let $\Delta_s(x) = \Delta(x/s)$, so that $|\Delta_s''(x) - \Delta_s''(y)| \le As^{-q}|x-y|^{q-2}$ for all $x, y \in \mathbb{R}$. Let $\boldsymbol{Y}$ and $\boldsymbol{Z}$ be random variables with $\mathbf{E}[\boldsymbol{Y}] = \mathbf{E}[\boldsymbol{Z}]$, $\mathbf{E}[\boldsymbol{Y}^2] = \mathbf{E}[\boldsymbol{Z}^2]$ and $\mathbf{E}[|\boldsymbol{Y}|^q], \mathbf{E}[|\boldsymbol{Z}|^q] < \infty$. Then $|\mathbf{E}[\Delta_s(x+\boldsymbol{Y})] - \mathbf{E}[\Delta_s(x+\boldsymbol{Z})]| \le As^{-q}(\mathbf{E}[|\boldsymbol{Y}|^q] + \mathbf{E}[|\boldsymbol{Z}|^q])$ for all $x \in \mathbb{R}$. Indeed, for $u \in [0,1]$ let $\varphi(u) = \mathbf{E}[\Delta_s(x+u\boldsymbol{Y})] - \mathbf{E}[\Delta_s(x+u\boldsymbol{Z})]$. Then $\varphi(0) = \varphi'(0) = 0$ and

$$|\varphi''(u)| = |\mathbf{E}[\boldsymbol{Y}^2(\Delta_s''(x+u\boldsymbol{Y}) - \Delta_s''(x))] - \mathbf{E}[\boldsymbol{Z}^2(\Delta_s''(x+u\boldsymbol{Z}) - \Delta_s''(x))]| \le As^{-q}u^{q-2}(\mathbf{E}[|\boldsymbol{Y}|^q] + \mathbf{E}[|\boldsymbol{Z}|^q]),$$

so that $|\varphi(1)| \le As^{-q}(\mathbf{E}[|\boldsymbol{Y}|^q] + \mathbf{E}[|\boldsymbol{Z}|^q])$. Now, using the above estimate and the fact that both $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{G}}$ are $(2, q, \eta)$-hypercontractive with $\eta = \frac{\beta^{-1/q}}{2\sqrt{q-1}}$ one arrives at

$$|\mathbf{E}[\Delta_s(Q(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n))] - \mathbf{E}[\Delta_s(Q(\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n))]| \le O(s^{-q}\eta^{-qd}\sum_i (\sum_{S\ni i} c_S^2)^{q/2}).$$

Replacing Q by $Q+t$ and using the arguments of subsection 3.6.2 yields

$$\sup_t \left| \mathbf{P}[Q(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \le t] - \mathbf{P}[Q(\boldsymbol{G}_1,\ldots,\boldsymbol{G}_n) \le t] \right| \le$$

$$O(ds^{1/d}) + O(s^{-q}\eta^{-qd}\sum_i (\sum_{S \ni i} c_S^2)^{q/2}).$$

Optimizing over s ends the proof. We skip some elementary calculations. □

4 Proofs of the conjectures

Our applications of the invariance principle have the following character: We wish to study certain noise stability properties of low-influence functions on finite product probability spaces. By using the invariance principle for slightly smoothed functions, Theorem 3.20, we can essentially analyze the properties in the product space of our choosing. And as it happens, the necessary result for Majority Is Stablest is already known in Gaussian space [9] and the necessary result for It Ain't Over Till It's Over is already known on the uniform-measure discrete cube [43].

In the case of the Majority Is Stablest problem, one needs to find a set of prescribed Gaussian measure which maximizes the probability that the Ornstein-Uhlenbeck process (started at the Gaussian measure) will belong to the set at times 0 and time t for some fixed time t. This problem was solved by Borell in [9] using symmetrization arguments. It should also be noted that the analogous result for the sphere has been proven in more than one place, including a paper of Feige and Schechtman [20]. It fact, one can deduce Borell's result and Majority is Stablest from the spherical result using the proximity of spherical and Gaussian measures in high dimensions and the invariance principle proven here.

In the case of the It Ain't Over Till It's Over problem, the necessary result on the discrete cube $\{-1,1\}^n$ was essentially proven in the recent paper [43] using the reverse Bonami-Beckner inequality (which is also due to Borell [8]). This paper did not solve the conjecture though (nor did that paper note the relevance), even when the conjecture is set on $\{-1,1\}^n$; the reason is that reduction of the problem to a question about T_ρ already involves transferring to a different product domain (e.g., $\{-1,0,1\}^n$ with biased measure) and so the invariance principle is required.

Note that in both cases the necessary auxiliary result is valid without any assumptions about low influences. This should not be surprising in the Gaussian case, since given a multilinear polynomial Q over Gaussians it is easy to define another multilinear polynomial $\tilde{Q}$ over Gaussians with exactly the same distribution and arbitrarily low influences, by letting

$$\tilde{Q}(x_{1,1},\ldots,x_{1,N},\ \ldots\ ,x_{n,1},\ldots,x_{n,N}) = Q\Big(\frac{x_{1,1}+\cdots+x_{1,N}}{N^{1/2}},\ \ldots\ ,\frac{x_{n,1}+\cdots+x_{n,N}}{N^{1/2}}\Big).$$

The fact that low influences are not required for the the results of [43] is perhaps more surprising.

4.1 Noise stability in Gaussian space

We begin by recalling some definitions and results relevant for "Gaussian noise stability". Throughout this section we consider $\mathbb{R}^n$ to have the standard n-dimensional Gaussian distribution, and our probabilities and expectations are over this distribution.

Let U_ρ denote the Ornstein-Uhlenbeck operator acting on $L^2(\mathbb{R}^n)$ by

$$(U_\rho f)(x) = \mathop{\mathbf{E}}_{y}[f(\rho x + \sqrt{1-\rho^2}\, y)],$$

where y is a random standard n-dimensional Gaussian. It is easy to see that if $f(x)$ is expressible as a *multilinear* polynomial in its n independent Gaussian inputs,

$$f(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i,$$

then $U_\rho f$ is the following multilinear polynomial:

$$(U_\rho f)(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} \rho^{|S|} c_S \prod_{i \in S} x_i.$$

Thus U_ρ acts identically to T_ρ for multilinear polynomials Q over $\boldsymbol{\mathcal{G}}$, the Gaussian sequence of ensembles.

Next, given any function $f : \mathbb{R}^n \to \mathbb{R}$, recall that its *(Gaussian) nonincreasing spherical rearrangement* is defined to be the upper semicontinuous nondecreasing function $f^* : \mathbb{R} \to \mathbb{R}$ which is equimeasurable with f; i.e., for all $t \in \mathbb{R}$, f^* satisfies $\mathbf{P}[f > t] = \mathbf{P}[f^* > t]$ under Gaussian measure.

We now state a result of Borell concerning the Ornstein-Uhlenbeck operator U_ρ (see also Ledoux's Saint-Flour lecture notes [18]). Borell uses Ehrhard symmetrization to show the following:

Theorem 4.1 *(Borell [9]) Let $f, g \in L^2(\mathbb{R}^n)$. Then for all $0 \le \rho \le 1$ and all $q \ge 1$,*

$$\mathbf{E}[(U_\rho f)^q \cdot g] \le \mathbf{E}[(U_\rho f^*)^q \cdot g^*].$$

Borell's result is more general and is stated for Lipschitz functions, but standard density arguments immediately imply the validity of the statement above. One immediate consequence of the theorem is that $\mathbb{S}_\rho(f) \le \mathbb{S}_\rho(f^*)$, where we define

$$\mathbb{S}_\rho(f) = \mathbf{E}[f \cdot U_\rho f] = \mathbf{E}[(U_{\sqrt{\rho}} f)^2]. \tag{28}$$

One can think of this quantity as the "(Gaussian) noise stability of f at ρ"; again, it is compatible with our earlier definition of $\mathbb{S}_\rho$ if f is a multilinear polynomial over $\boldsymbol{\mathcal{G}}$.

Note that the latter equality in (28) and the fact that $U_{\sqrt{\rho}}$ is positivity-preserving and linear imply that $\sqrt{\mathbb{S}_\rho}$ defines an L^2 norm on $L^2(\mathbb{R}^n)$, dominated by the usual L^2 norm, so that it is a continuous convex functional on $L^2(\mathbb{R}^n)$. The set of all $[0,1]$-valued functions from $L^2(\mathbb{R}^n)$ having the same mean as f is closed and bounded in the standard L^2 norm and one can easily check that its extremal points are indicator functions; hence by the Edgar-Choquet theorem (see [19]; clearly $L^2(\mathbb{R}^n)$ is separable and it has the Radon-Nikodym property since it is a Hilbert space):

$$\sqrt{\mathbb{S}_\rho}(f) \le \sup_{\chi} \sqrt{\mathbb{S}_\rho}(\chi),$$

where the supremum is taken over all functions $\chi : \mathbb{R}^n \to \{0,1\}$ with $\mathbf{E}[\chi] = \mathbf{E}[f]$. Since by Borell's result $\mathbb{S}_\rho(\chi) \le \mathbb{S}_\rho(\chi^*)$, we have $\mathbb{S}_\rho(f) \le \mathbb{S}_\rho(\chi_\mu)$ where $\chi_\mu : \mathbb{R} \to \{0,1\}$ is the indicator function of a halfline with measure $\mu = \mathbf{E}[f]$.

Let us introduce some notation:

Definition 4.2 *Given $\mu \in [0,1]$, define $\chi_\mu : \mathbb{R} \to \{0,1\}$ to be the indicator function of the interval $(-\infty, t]$, where t is chosen so that $\mathbf{E}[\chi_\mu] = \mu$. Explicitly, $t = \Phi^{-1}(\mu)$, where Φ denotes the distribution function of a standard Gaussian. Furthermore, define*

$$\Gamma_\rho(\mu) = \mathbb{S}_\rho(\chi_\mu) = \mathbf{P}[\boldsymbol{X} \le t, \boldsymbol{Y} \le t],$$

where $(\boldsymbol{X}, \boldsymbol{Y})$ is a two dimensional Gaussian vector with covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$.

Summarizing the above discussion, we obtain:

Corollary 4.3 *Let $f : \mathbb{R}^n \to [0,1]$ be a measurable function on Gaussian space with $\mathbf{E}[f] = \mu$. Then for all $0 \le \rho \le 1$ we have $\mathbb{S}_\rho(f) \le \Gamma_\rho(\mu)$.*

This is the result we will use to prove the Majority Is Stablest conjecture. We note that in general there is no closed form for $\Gamma_\rho(\mu)$; however, some asymptotics are known: For balanced functions we have Sheppard's formula $\Gamma_\rho(1/2) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$. Some other properties of $\Gamma_\rho(\mu)$ are given in Appendix B.

4.2 Majority Is Stablest

In this section we prove a strengthened form of the Majority Is Stablest conjecture. The implications of this result were discussed in Section 2.3.

Theorem 4.4 *Let $f : \Omega_1 \times \cdots \times \Omega_n \to [0,1]$ be a function on a finite product probability space and assume that for each i the minimum probability of any atom in Ω_i is at least $\alpha \le 1/2$. Write $K = \log(1/\alpha)$. Further assume that there is a $0 < \tau < 1/2$ such that $\mathrm{Inf}_i^{\le \log(1/\tau)/K}(f) \le \tau$ for all i. (See Definition 3.7 for the definition of low-degree influence.) Let $\mu = \mathbf{E}[f]$. Then for any $0 \le \rho < 1$,*

$$\mathbb{S}_\rho(f) \le \Gamma_\rho(\mu) + \epsilon,$$

where

$$\epsilon = O\Big(\frac{K}{1-\rho}\Big) \cdot \frac{\log\log(1/\tau)}{\log(1/\tau)}.$$

For the reader's convenience we record here two facts from Appendix B:

$$\begin{aligned} \Gamma_\rho(\tfrac{1}{2}) &= \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho \\ \Gamma_\rho(\mu) &\sim \mu^{2/(1+\rho)} \, (4\pi \ln(1/\mu))^{-\rho/(1+\rho)} \, \frac{(1+\rho)^{3/2}}{(1-\rho)^{1/2}} \qquad \text{as } \mu \to 0. \end{aligned}$$

Proof: As discussed in Section 3.1, let $\boldsymbol{\mathcal{X}}$ be the sequence of ensembles such that $\boldsymbol{\mathcal{X}}_i$ spans the functions on Ω_i, and express f as the multilinear polynomial Q. We use the invariance principle under hypothesis $\boldsymbol{H2}$. Express $\rho = \rho' \cdot (1-\gamma)^2$, where $0 < \gamma \ll 1 - \rho$ will be chosen later. Writing $Q(x) = \sum c_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}$ (with $c_{\mathbf{0}} = \mu$) we see that

$$\mathbb{S}_\rho(Q(\boldsymbol{\mathcal{X}})) = \sum (\rho' \cdot (1-\gamma)^2)^{|\boldsymbol{\sigma}|} c_{\boldsymbol{\sigma}}^2 = \mathbb{S}_{\rho'}((T_{1-\gamma}Q)(\boldsymbol{\mathcal{G}})),$$

where $\boldsymbol{\mathcal{G}}$ is the sequence of independent Gaussian ensembles.

Since $Q(\boldsymbol{\mathcal{X}})$ is bounded in $[0,1]$ the same is true of $\boldsymbol{R} = (T_{1-\gamma}Q)(\boldsymbol{\mathcal{X}})$. In other words, $\mathbf{E}[\zeta(\boldsymbol{R})] = 0$, where ζ is the function from (17). Writing $\boldsymbol{S} = (T_{1-\gamma}Q)(\boldsymbol{\mathcal{G}})$, we conclude from Theorem 3.20

that $\mathbf{E}[\zeta(\boldsymbol{S})] \le \tau^{\Omega(\gamma/K)}$. That is, $\|\boldsymbol{S} - \boldsymbol{S}'\|_2^2 \le \tau^{\Omega(\gamma/K)}$, where $\boldsymbol{S}'$ is the random variable depending on $\boldsymbol{S}$ defined by

$$\boldsymbol{S}' = \begin{cases} 0 & \text{if } \boldsymbol{S} \le 0, \\ \boldsymbol{S} & \text{if } \boldsymbol{S} \in [0,1], \\ 1 & \text{if } \boldsymbol{S} \ge 1. \end{cases}$$

Then

$$\begin{aligned} |\mathbb{S}_{\rho'}(\boldsymbol{S}) - \mathbb{S}_{\rho'}(\boldsymbol{S}')| &= |\mathbf{E}[\boldsymbol{S} \cdot U_{\rho'}\boldsymbol{S}] - \mathbf{E}[\boldsymbol{S}' \cdot U_{\rho'}\boldsymbol{S}']| \\ &\le |\mathbf{E}[\boldsymbol{S} \cdot U_{\rho'}\boldsymbol{S}] - \mathbf{E}[\boldsymbol{S}' \cdot U_{\rho'}\boldsymbol{S}]| + |\mathbf{E}[\boldsymbol{S}' \cdot U_{\rho'}\boldsymbol{S}] - \mathbf{E}[\boldsymbol{S}' \cdot U_{\rho'}\boldsymbol{S}']| \\ &\le (\|\boldsymbol{S}\|_2 + \|\boldsymbol{S}'\|_2)\|\boldsymbol{S} - \boldsymbol{S}'\|_2 \le \tau^{\Omega(\gamma/K)}, \end{aligned}$$

where we have used the fact that $U_{\rho'}$ is a contraction on L^2.

Writing $\mu' = \mathbf{E}[\boldsymbol{S}']$ it follows from Cauchy-Schwartz that $|\mu - \mu'| \le \tau^{\Omega(\gamma/K)}$. Since $\boldsymbol{S}'$ takes values in $[0,1]$ it follows from Corollary 4.3 that $\mathbb{S}_{\rho'}(\boldsymbol{S}') \le \Gamma_{\rho'}(\mu')$. We thus conclude

$$\mathbb{S}_{\rho}(Q(\boldsymbol{\mathcal{X}})) = \mathbb{S}_{\rho'}(\boldsymbol{S}) \le \mathbb{S}_{\rho'}(\boldsymbol{S}') + \tau^{\Omega(\gamma/K)} \le \Gamma_{\rho'}(\mu') + \tau^{\Omega(\gamma/K)}.$$

We can now bound the difference $|\Gamma_\rho(\mu) - \Gamma_{\rho'}(\mu')|$ using Lemmas B.3 and Corollary B.5 in Appendix B. We get a contribution of $2|\mu - \mu'| \le \tau^{\Omega(\gamma/K)}$ from the difference in the μ's and a contribution of at most $O(\gamma/(1-\rho))$ from the difference in the ρ's. Thus we have

$$\mathbb{S}_{\rho}(Q(\boldsymbol{\mathcal{X}})) \le \Gamma_\rho(\mu) + \tau^{\Omega(\gamma/K)} + O(\gamma/(1-\rho)).$$

Taking

$$\gamma = C \cdot K \cdot \frac{\log\log(1/\tau)}{\log(1/\tau)}$$

for some large enough constant C and this gives the claimed bound. □

4.3 It Ain't Over Till It's Over

As mentioned, our proof of the It Ain't Over Till It's Over conjecture will use a result due essentially to [43]:

Theorem 4.5 *Let $f : \{-1,1\}^n \to [0,1]$ have $\mathbf{E}[f] = \mu$ (with respect to uniform measure on $\{-1,1\}^n$). Then for any $0 < \rho < 1$ and any $0 < \epsilon \le 1 - \mu$ we have*

$$\mathbf{P}[T_\rho f > 1 - \delta] < \epsilon$$

provided

$$\delta < \epsilon^{\rho^2/(1-\rho^2)+O(\kappa)},$$

where

$$\kappa = \frac{\sqrt{c(\mu)}}{1-\rho} \cdot \frac{1}{\sqrt{\log(1/\epsilon)}}, \qquad c(\mu) = \mu \log(e/(1-\mu)).$$

This theorem follows from the proof of Theorem 4.1 in [43]; for completeness we give an explicit derivation in Appendix C.

Remark 4.6 *Since the only fact about $\{-1,1\}^n$ used in the proof of Theorem 4.5 is the reverse Bonami-Beckner inequality, and since this inequality also holds in Gaussian space, we conclude that Theorem 4.5 also holds for measurable functions on Gaussian space $f : \mathbb{R}^n \to [0,1]$. In this setting the result can be proved using Borell's Corollary 4.3 instead of using the reverse Bonami-Beckner inequality.*

The first step of the proof of It Ain't Over Till It's Over is to extend Theorem 4.5 to functions on arbitrary product probability spaces. Note that if we only want to solve the problem for functions on $\{-1,1\}^n$ with the uniform measure, this step is unnecessary. The proof of the extension is very similar to the proof of Theorem 4.4. In order to state the theorem it would be helpful to let $u > 0$ be a constant such that Theorem 3.20 holds with the bound $\tau^{u\gamma/K}$.

Theorem 4.7 *Let $f : \Omega_1 \times \cdots \times \Omega_n \to [0,1]$ be a function on a finite product probability space and assume that for each i the minimum probability of any atom in Ω_i is at least $\alpha \leq 1/2$. Let $K \geq \log(1/\alpha)$. Further assume that there is a $\tau > 0$ such that $\mathrm{Inf}_i^{\leq \log(1/\tau)/K}(f) \leq \tau$ for all i (recall Definition 3.7). Let $\mu = \mathbf{E}[f]$. Then for any $0 < \rho < 1$ there exists $\epsilon(\mu, \rho)$ such that if $0 < \epsilon < \epsilon(\mu, \rho)$ we have*

$$\mathbf{P}[T_\rho f > 1 - \delta] \leq \epsilon$$

provided

$$\delta < \epsilon^{\rho^2/(1-\rho^2)+C\kappa}, \qquad \tau \leq \epsilon^{(100K/u(1-\rho))(1/(1-\rho)^3+C\kappa)}$$

where

$$\kappa = \frac{\sqrt{c(\mu)}}{1-\rho} \cdot \frac{1}{\sqrt{\log(1/\epsilon)}}, \qquad c(\mu) = \mu \log(e/(1-\mu)) + \epsilon$$

and $C > 0$ is some constant.

Proof: Without loss of generality we assume that $\delta = \epsilon^{\rho^2/(1-\rho^2)+C\kappa}$ as taking a smaller δ yields a smaller tail probability. We can also assume $\epsilon(\mu, \rho) < 1/10$. Let $\boldsymbol{\mathcal{X}}$ and Q be as in the proof of Theorem 4.4 and this time decompose $\rho = \rho' \cdot (1-\gamma)$ where we take $\gamma = \kappa \cdot (1-\rho)^2$. Note that taking $\epsilon(\mu, \rho)$ sufficiently small we have $\kappa < 1, \gamma < 0.1$ and $(1-\rho)/(1-\rho') \leq 2$. Let $\boldsymbol{R} = (T_{1-\gamma}Q)(\boldsymbol{\mathcal{X}})$ as before, and let $\boldsymbol{S} = (T_{1-\gamma}Q)(\boldsymbol{\mathcal{Y}})$, where $\boldsymbol{\mathcal{Y}}$ denotes the Rademacher sequence of ensembles ($\boldsymbol{Y}_{i,0} = 1$, $\boldsymbol{Y}_{i,1} = \pm 1$ independently and uniformly random). Since $\mathbf{E}[\zeta(\boldsymbol{R})] = 0$ as before, we conclude from Theorem 3.20 that we have $\mathbf{E}[\zeta(\boldsymbol{S})] \leq \tau^{u\gamma/K} \leq \epsilon^{10/(1-\rho)+2C\kappa}$; i.e.,

$$\|\boldsymbol{S} - \boldsymbol{S}'\|_2^2 \leq \epsilon^{10/(1-\rho)+2C\kappa} \tag{29}$$

where $\boldsymbol{S}'$ is the truncated version of $\boldsymbol{S}$ as in the proof of Theorem 4.4. Now $\boldsymbol{S}'$ is a function $\{-1,1\}^n \to [0,1]$ with mean μ' differing from μ by at most ϵ^5 (using Cauchy-Schwartz, as before). This implies that $c(\mu') \leq O(c(\mu))$.

Furthermore, our assumed upper bound on δ also holds with ρ' in place of ρ. This is because

$$\frac{\rho'^2}{1-\rho'^2} - \frac{\rho^2}{1-\rho^2} = \frac{1}{1-\rho'^2} - \frac{1}{1-\rho^2} \leq (\rho'^2 - \rho^2)\frac{1}{(1-\rho'^2)^2} \leq \frac{2\gamma}{(1-\rho')^2} \leq \frac{8\gamma}{(1-\rho)^2} = 8\kappa.$$

Thus Theorem 4.5 implies that if C is sufficiently large then

$$\mathbf{P}[T_{\rho'}\boldsymbol{S}' > 1 - 4\delta] < \epsilon/2.$$

This, in turn implies that

$$\mathbf{P}[T_{\rho'}\boldsymbol{S} > 1 - 2\delta] < 3\epsilon/4.$$

This follows by (29) since,

$$\mathbf{P}[T_{\rho'}\boldsymbol{S} > 1 - 4\delta] - \mathbf{P}[T_{\rho'}\boldsymbol{S}' > 1 - 2\delta] \leq \delta^{-2}\|T_{\rho'}\boldsymbol{S} - T_{\rho'}\boldsymbol{S}'\|_2^2 \leq \delta^{-2}\|\boldsymbol{S} - \boldsymbol{S}'\|_2^2.$$

We now use Theorem 3.20 again, bounding the Lévy distance of $(T_\rho Q)(\boldsymbol{\mathcal{Y}})$ and $(T_\rho Q)(\boldsymbol{\mathcal{X}})$ by $\tau^{u(1-\rho)/K}$, which is smaller than δ and $\epsilon/8$. Thus

$$\mathbf{P}[(T_\rho Q)(\boldsymbol{\mathcal{X}}) > 1-\delta] \le \mathbf{P}[T_\rho f > 1-2\delta] + \epsilon/8 < \epsilon,$$

as needed. $\square$

The second step of the proof of It Ain't Over Till It's Over is to use the invariance principle to show that the random variable $V_\rho f$ (recall Definition 2.6) has essentially the same distribution as $T_{\sqrt{\rho}} f$.

Theorem 4.8 *Let $0 < \rho < 1$ and let $f : \Omega_1 \times \cdots \times \Omega_n \to [0,1]$ be a function on a finite product probability space; assume that for each i the minimum probability of any atom in Ω_i is at least $\alpha \le 1/2$. Further assume that there is a $0 < \tau < 1/2$ such that $\mathrm{Inf}_i^{\le \log(1/\tau)/K'} \le \tau$ for all i, where $K' = \log(1/(\alpha\rho(1-\rho)))$. Then*

$$d_L(V_\rho f, T_{\sqrt{\rho}} f) \le \tau^{\Omega((1-\rho)/K')}.$$

Proof: Introduce $\boldsymbol{\mathcal{X}}$ and Q as in the proof of Theorems 4.4 and 4.7. We now define a new independent sequence of orthonormal ensembles $\boldsymbol{\mathcal{X}}^{(\rho)}$ as follows. Let $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n$ be independent random variables, each of which is 1 with probability ρ and 0 with probability $1-\rho$. Now define $\boldsymbol{\mathcal{X}}^{(\rho)} = (\boldsymbol{\mathcal{X}}_1^{(\rho)}, \ldots, \boldsymbol{\mathcal{X}}_n^{(\rho)})$ by $\boldsymbol{X}_{i,0}^{(\rho)} = 1$ for each i, and $\boldsymbol{X}_{i,j}^{(\rho)} = \rho^{-1/2} \boldsymbol{V}_i \boldsymbol{X}_{i,j}$ for each i and $j > 0$. It is easy to verify that $\boldsymbol{\mathcal{X}}^{(\rho)}$ is indeed an independent sequence of orthonormal ensembles. We will also use the fact that each atom in the ensemble $\boldsymbol{\mathcal{X}}_i^{(\rho)}$ has weight at least $\alpha' = \alpha \cdot \min\{\rho, 1-\rho\} \ge \alpha\rho(1-\rho)$. (one can also use Proposition 3.17 to get a bit better estimate on K').

The crucial observation is now simply that the random variable $V_\rho f$ has precisely the same distribution as the random variable $(T_{\sqrt{\rho}} Q)(\boldsymbol{\mathcal{X}}^{(\rho)})$. The reason is that when the randomness in the $\boldsymbol{V}_i = 1$ ensembles is fixed, the expectation of the restricted function is given by substituting 0 for all other random variables $\boldsymbol{X}_{i,j}$. The $T_{\sqrt{\rho}}$ serves to cancel the factors $\rho^{-1/2}$ introduced in the definition of $\boldsymbol{X}_{i,j}^{(\rho)}$ to ensure orthonormality.

It now simply remains to use Theorem 3.20 to bound the Lévy distance of $(T_{\sqrt{\rho}} Q)(\boldsymbol{\mathcal{X}}^{(\rho)})$ and $(T_{\sqrt{\rho}} Q)(\boldsymbol{\mathcal{X}})$, where here $\boldsymbol{\mathcal{X}}$ denotes a copy of this sequence of ensembles. We use hypothesis ***H3*** and get a bound of $\tau^{\Omega((1-\sqrt{\rho})/K')} = \tau^{\Omega((1-\rho)/K')}$, as required. $\square$

Our generalization of It Ain't Over Till It's Over is now simply a corollary of Theorems 4.7 and 4.8; by taking K' instead of K in the upper bound on τ and taking δ to have its maximum possible value, we make the error of

$$\tau^{u((1-\rho)/K')} \le \epsilon^{(100/(1-\rho))(1/(1-\rho)^3 + C\kappa)}$$

from Theorem 4.8 which is negligible compared to both ϵ and δ below.

Theorem 4.9 *Let $0 < \rho < 1$ and let $f : \Omega_1 \times \cdots \times \Omega_n \to [0,1]$ be a function on a finite product probability space; assume that for each i the minimum probability of any atom in Ω_i is at least $\alpha \le 1/2$. Further assume that there is a $0 < \tau < 1/2$ such that $\mathrm{Inf}_i^{\le \log(1/\tau)/K'} \le \tau$ for all i, where $K' = \log(1/(\alpha\rho(1-\rho)))$. Let $\mu = \mathbf{E}[f]$. Then there exists an $\epsilon(\rho, \mu) > 0$ such that if $\epsilon \le \epsilon(\rho, \mu)$ then*

$$\mathbf{P}[V_\rho f > 1-\delta] \le \epsilon$$

provided

$$\delta < \epsilon^{\rho^2/(1-\rho^2)+C\kappa}, \qquad \tau \le \epsilon^{(100K'/u(1-\rho))(1/(1-\rho)^3+C\kappa)}$$

where

$$\kappa = \frac{\sqrt{c(\mu)}}{1-\rho} \cdot \frac{1}{\sqrt{\log(1/\epsilon)}}, \qquad c(\mu) = \mu \log(e/(1-\mu)) + \epsilon,$$

where $C > 0$ is some finite constant.

Remark 4.10 *To get $V_\rho f$ bounded away from both 0 and 1 as desired in Conjecture 1.2, simply use Theorem 4.9 twice, once with f, once with $1-f$.*

5 Weight at low levels — a counterexample

The simplest version of the Majority Is Stablest result states roughly that among all balanced functions $f : \{-1,1\}^n \to \{-1,1\}$ with small influences, the Majority function maximizes $\sum_S \rho^S \hat{f}(S)^2$ for each ρ. One might conjecture that more is true; specifically, that Majority maximizes $\sum_{|S|\le d} \hat{f}(S)^2$ for each $d = 1, 2, 3, \ldots$. This is known to be the case for $d = 1$ ([35]) and is somewhat suggested by the theorem of Bourgain [11] which says that $\sum_{|S|\le d} \hat{f}(S)^2 \le 1 - d^{-1/2-o(1)}$ for functions with low influences. An essentially weaker conjecture was made Kalai [31]:

Conjecture 5.1 *Let $d \ge 1$ and let C_n denote the collection of all functions $f : \{-1,1\}^n \to \{-1,1\}$ which are odd and transitive-symmetric (see Section 2.3.1's discussion of [31]). Then*

$$\limsup_{n\to\infty} \sup_{f \in C_n} \sum_{|S|\le d} \hat{f}(S)^2 = \lim_{n \text{ odd } \to\infty} \sum_{|S|\le d} \widehat{\mathrm{Maj}_n}(S)^2.$$

We now show that these conjectures are false: We construct a sequence (f_n) of completely symmetric odd functions with small influences that have more weight on levels 1, 2, and 3 than Majority has. By "completely symmetric" we mean that $f_n(x)$ depends only on $\sum_{i=1}^n x_i$; because of this symmetry our counterexample is more naturally viewed in terms of the Hermite expansions of functions $f : \mathbb{R} \to \{-1,1\}$ on one-dimensional Gaussian space.

There are several normalizations of the Hermite polynomials in the literature. We will follow [41] and define them to be the orthonormal polynomials with respect to the one-dimensional Gaussian density function $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$. Specifically, we define the Hermite polynomials $h_d(x)$ for $d \in \mathbb{N}$ by

$$\exp(\lambda x - \lambda^2/2) = \sum_{d=0}^{\infty} \frac{\lambda^d}{\sqrt{d!}}\, h_d(x).$$

The first few such polynomials are $h_0(x) = 1$, $h_1(x) = x$, $h_2(x) = (x^2-1)/\sqrt{2}$, and $h_3(x) = (x^3-3x)/\sqrt{6}$. The orthonormality condition these polynomials satisfy is

$$\int_{\mathbb{R}} h_d(x) h_{d'}(x) \varphi(x)\, dx = \begin{cases} 1 & \text{if } d = d', \\ 0 & \text{else.} \end{cases}$$

We will actually henceforth consider functions whose domain is $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$, for simplicity; the value of a function at a single point makes no difference to its Hermite expansion. Given a function $f : \mathbb{R}^* \to \mathbb{R}$ we write $\hat{f}(d)$ for $\int h_d(x) f(x) \varphi(x)\, dx$. Let us also use the notation Maj for the function which is 1 on $(0,\infty)$ and -1 on $(-\infty, 0)$.

Theorem 5.2 *There is an odd function $f : \mathbb{R}^* \to \{-1, 1\}$ with*

$$\sum_{d\le 3} \hat{f}(d)^2 \ge .75913 > \frac{2}{\pi} + \frac{1}{3\pi} = \sum_{d\le 3} \widehat{\mathrm{Maj}}(d)^2.$$

Proof: Let $t > 0$ be a parameter to be chosen later, and let f be the function which is 1 on $(-\infty, -t]$ and $(0, t)$, and -1 on $(-t, 0)$ and $[t, \infty)$. Since f is odd, $\hat{f}(0) = \hat{f}(2) = 0$. Elementary integration gives

$$F_1(t) = \int h_1(x)\varphi(x)\,dx = -e^{-t^2/2}/\sqrt{2\pi}, \qquad F_3(t) = \int h_3(x)\varphi(x)\,dx = (1-t^2)e^{-t^2/2}/\sqrt{12\pi};$$

thus

$$\begin{aligned}
\hat{f}(1) &= 2(F_1(t) + F_1(-t) - F_1(0)) - F_1(\infty) - F_1(-\infty) = \sqrt{2/\pi}\,(1 - 2e^{-t^2/2}),\\
\hat{f}(3) &= 2(F_1(t) + F_1(-t) - F_1(0)) - F_1(\infty) - F_1(-\infty) = -\sqrt{1/3\pi}\,(1 - 2(1-t^2)e^{-t^2/2}).
\end{aligned}$$

We conclude

$$\sum_{d\le 3} \hat{f}(d)^2 = \frac{2}{\pi}\Big(1 - 2e^{-t^2/2}\Big)^2 + \frac{1}{3\pi}\Big(1 - 2(1-t^2)e^{-t^2/2}\Big)^2. \tag{30}$$

As $t \to 0$ or ∞ we recover the fact, well known in the boolean regime (see, e.g., [6]), that $\sum_{d\le 3} \widehat{\mathrm{Maj}}(d)^2 = 2/\pi + 1/3\pi$. But the above expression is not maximized for these t; rather, it is maximized at $t = 2.69647$, where the expression becomes roughly .75913. Fixing this particular t completes the proof. $\Box$

It is now clear how to construct the sequence of completely symmetric odd functions $f_n : \{-1,1\}^n \to \{-1,1\}$ with the same property — take $f_n(x) = f((x_1 + \cdots + x_n)/\sqrt{n})$. The proof that the property holds follows essentially from the fact that the limits of Kravchuk polynomials are Hermite polynomials. For completeness, give a direct proof of Corollary 5.3 in Appendix D.

Corollary 5.3 *For n odd there is a sequence of completely symmetric odd functions $f_n : \{-1,1\}^n \to \{-1,1\}$ satisfying $\mathrm{Inf}_i(f_n) \le O(1/\sqrt{n})$ for each i, and*

$$\lim_{n \text{ odd} \to\infty} \sum_{|S|\le 3} \widehat{f_n}(S)^2 \ge 0.75913 > \frac{2}{\pi} + \frac{1}{3\pi} = \lim_{n \text{ odd} \to\infty} \sum_{|S|\le 3} \widehat{\mathrm{Maj}_n}(S)^2.$$

In light of this counterexample, it seems we can only hope to sharpen Bourgain's Theorem 2.7 in the asymptotic setting; one might ask whether its upper bound can be improved to

$$1 - (1 - o(1))(2/\pi)^{3/2}\, d^{-1/2},$$

the asymptotics for Majority.

References

[1] N. Alon, I. Dinur, E. Friedgut, and B. Sudakov. Graph products, Fourier analysis and spectral techniques. *GAFA*, 14(5):913–940, 2004.

[2] D. Bakry, R. D. Gill, and S. A. Molchanov. *Lectures on probability theory*, volume 1581 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1994. Lectures from the Twenty-second Saint-Flour Summer School held July 9–25, 1992, edited by P. Bernard.

[3] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.

[4] M. Ben-Or and N. Linial. Collective coin flipping. In S. Micali, editor, *Randomness and Computation*. Academic Press, New York, 1990.

[5] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[6] A. Bernasconi. *Mathematical techniques for the analysis of Boolean functions*. PhD thesis, Institute for Computational Mathematics, Pisa, 1998.

[7] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335–402 (1971), 1970.

[8] C. Borell. Positivity improving operators and hypercontractivity. *Math. Zeitschrift*, 180(2):225–234, 1982.

[9] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985.

[10] J. Bourgain. An appendix to *Sharp thresholds of graph properties, and the k-SAT problem*, by E. Friedgut. *J. American Math. Soc.*, 12(4):1017–1054, 1999.

[11] J. Bourgain. On the distribution of the Fourier spectrum of boolean functions. *Israel Journal of Mathematics*, 131:269–276, 2002.

[12] A. Carbery and J. Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Math. Res. Lett.*, 8(3):233–248, 2001.

[13] S. Chatterjee. A simple invariance principle. Submitted, 2005.

[14] E. de Klerk, D. Pasechnik, and J. Warners. Approximate graph colouring and MAX-k-CUT algorithms based on the theta function. To appear in *Journal of Combinatorial Optimization*, 2005.

[15] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell. On the Fourier tails of bounded functions over the discrete cube. Manuscript.

[16] I. Dinur, V. Guruswami, S. Khot, and O. Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 595–601, 2003.

[17] I. Dinur and S. Safra. The importance of being biased. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing*, pages 33–42, 2002.

[18] R. Dobrushin, P. Groeneboom, and M. Ledoux. *Lectures on probability theory and statistics*, volume 1648 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1996. Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994, edited by P. Bernard.

[19] G. A. Edgar. A noncompact Choquet theorem. *Proc. Amer. Math. Soc.*, 49:354–358, 1975.

[20] U. Feige and G. Schechtman. On the optimality of the random hyperplane rounding technique for MAX-CUT. *Random Structures and Algorithms*, 20:403–440, 2002.

[21] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[22] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):474–483, 1998.

[23] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[24] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124:2993–3002, 1996.

[25] E. Friedgut, G. Kalai, and A. Naor. Boolean functions whose Fourier transform is concentrated on the first two levels. *Adv. in Appl. Math.*, 29(3), 2002.

[26] A. Frieze and M. Jerrum. Improved approximation algorithms for MAX-k-CUT and MAX-BISECTION. In E. Balas and J. Clausen, editors, *Integer Programming and Combinatorial Optimization*, volume 920, pages 1–13. Springer, 1995.

[27] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *JACM*, 42:1115–1145, 1995.

[28] J. Håstad. Some optimal inapproximability results. *J. ACM*, 48:798–869, 2001.

[29] S. Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[30] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[31] G. Kalai. A Fourier-theoretic perspective on the Concordet paradox and Arrow's theorem. *Adv. in Appl. Math.*, 29(3):412–426, 2002.

[32] G. Kalai. Social Indeterminacy. *Econometrica*, 72:1565–1581, 2004.

[33] G. Kalai and E. Friedgut. It ain't over till it's over. Private communication, 2001.

[34] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th Ann. STOC*, pages 767–775, 2002.

[35] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 146–154. IEEE, 2004.

[36] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2-\varepsilon$. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity*, pages 379–388. IEEE, 2003.

[37] S. Khot and N. Vishnoi. On embeddability of negative type metrics into ℓ_1. In preparation.

[38] G. Kindler and S. Safra. Noise-resistant boolean functions are juntas. Manuscript, 2002.

[39] W. Krakowiak and J. Szulga. Hypercontraction principle and random multilinear forms. *Probab. Theory Related Fields*, 77(3):325–342, 1988.

[40] S. Kwapień and W. A. Woyczyński. *Random series and stochastic integrals: single and multiple*. Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, 1992.

[41] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer, 1991.

[42] J. W. Lindeberg. Eine neue herleitung des exponential-gesetzes in der wahrscheinlichkeit srechnung. *Math. Zeit.*, 15:211–235, 1922.

[43] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov. Non-interactive correlation distillation, inhomegenous markov chains and the Reverse Bonami-Beckner inequality. To appear in *Israel J. Math.*, 2005.

[44] K. Oleszkiewicz. On a nonsymmetric version of the Khinchine-Kahane inequality. *Progress In Probability*, 56:156–168, 2003.

[45] Y. Rinott and V. Rotar′. A remark on quadrant normal probabilities in high dimensions. *Statistics and Probability Letters*, 51(1):47–51, 2001.

[46] V. I. Rotar′. Limit theorems for polylinear forms. *J. Multivariate Anal.*, 9(4):511–530, 1979.

[47] W. Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Phil. Trans. Royal Soc. London*, 192:101–168, 1899.

[48] J. Szulga. *Introduction to random chaos*. Chapman & Hall, London, 1998.

[49] M. Talagrand. On Russo's approximate 0-1 law. *Annals of Probability*, 22:1576–1587, 1994.

[50] M. Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

[51] P. Wolff. Hypercontractivity of simple random variables. submitted, 2005.

A Hypercontractivity of sequences of ensembles

We now give the proofs of Propositions 3.11 and 3.12 that were omitted. As mentioned, the proofs are completely straightforward adaptations of the analogous proofs in [29].

Proof: (of Proposition 3.11) Let Q be a multilinear polynomial over $\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}}$. Note that we can write $Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})$ as $\sum_j c_j \boldsymbol{\mathcal{X}}_{\boldsymbol{\sigma}_j} \boldsymbol{\mathcal{Y}}_{\boldsymbol{\upsilon}_j}$, where the $\boldsymbol{\sigma}$'s are multi-indexes for $\boldsymbol{\mathcal{X}}$, the $\boldsymbol{\upsilon}$'s are multi-indexes for $\boldsymbol{\mathcal{Y}}$, and the c_j's are constants. Then

$$\begin{aligned}
\|(T_\eta Q)(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_q &= \|\sum_j \eta^{|\sigma_j|+|\upsilon_j|} c_j \boldsymbol{\mathcal{X}}_{\sigma_j} \boldsymbol{\mathcal{Y}}_{\upsilon_j}\|_q \\
&= \|\|T_\eta^{(\boldsymbol{\mathcal{Y}})}\Big(\sum_j (\eta^{|\sigma_j|} c_j \boldsymbol{\mathcal{X}}_{\sigma_j}) \boldsymbol{\mathcal{Y}}_{\upsilon_j}\Big)\|_{L^q(\boldsymbol{\mathcal{Y}})}\|_{L^q(\boldsymbol{\mathcal{X}})} \\
&\le \|\|\sum_j (\eta^{|\sigma_j|} c_j \boldsymbol{\mathcal{X}}_{\sigma_j}) \boldsymbol{\mathcal{Y}}_{\upsilon_j}\|_{L^p(\boldsymbol{\mathcal{Y}})}\|_{L^q(\boldsymbol{\mathcal{X}})} \\
&\le \|\|\sum_j (\eta^{|\sigma_j|} c_j \boldsymbol{\mathcal{X}}_{\sigma_j}) \boldsymbol{\mathcal{Y}}_{\upsilon_j}\|_{L^q(\boldsymbol{\mathcal{X}})}\|_{L^p(\boldsymbol{\mathcal{Y}})} \\
&= \|\|T_\eta^{(\boldsymbol{\mathcal{X}})}\Big(\sum_j (c_j \boldsymbol{\mathcal{Y}}_{\upsilon_j}) \boldsymbol{\mathcal{X}}_{\sigma_j}\Big)\|_{L^q(\boldsymbol{\mathcal{X}})}\|_{L^p(\boldsymbol{\mathcal{Y}})} \\
&\le \|\sum_j c_j \boldsymbol{\mathcal{Y}}_{\upsilon_j} \boldsymbol{\mathcal{X}}_{\sigma_j}\|_{L^p(\boldsymbol{\mathcal{X}})}\|_{L^p(\boldsymbol{\mathcal{Y}})} \\
&= \|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_p,
\end{aligned}$$

where the second inequality used a simple consequence of the integral version of Minkowski's inequality and $p \le q$ (see [29, Prop. C.4]). $\square$

Proof: (of Proposition 3.12) Note that if $Q = Q^{=d}$ then we obviously have equality. In the general case, write $Q = \sum_{i=0}^d Q^{=i}$, and note that $\mathbf{E}[Q^{=i}(\boldsymbol{\mathcal{X}})Q^{=j}(\boldsymbol{\mathcal{X}})] = 0$ for $i \neq j$ is easy to check. Thus

$$\begin{aligned}
\|Q(\boldsymbol{\mathcal{X}})\|_q &= \|T_\eta\Big(\sum_{i=0}^d \eta^{-i} Q^{=i}(\boldsymbol{\mathcal{X}})\Big)\|_q \;\le\; \|\sum_{i=0}^d \eta^{-i} Q^{=i}(\boldsymbol{\mathcal{X}})\|_2 \\
&= \left(\sum_{i=0}^d \eta^{-2i} \|Q^{=i}(\boldsymbol{\mathcal{X}})\|_2^2\right)^{1/2} \;\le\; \eta^{-d}\|Q(\boldsymbol{\mathcal{X}})\|_2.
\end{aligned}$$

$\square$

Let us also mention some standard facts about the $(2, q, \eta)$-hypercontractivity of random variables. Let $q > 2$. Clearly, if we want $\boldsymbol{X}$ to be $(2, q, \eta)$-hypercontractive, we must assume that $\mathbf{E}[|\boldsymbol{X}|^q] < \infty$. If $\boldsymbol{X}$ is $(2, q, \eta)$-hypercontractive for some $\eta \in (0, 1)$ then $\mathbf{E}[\boldsymbol{X}] = 0$ and $\eta \le (q-1)^{-1/2}$. Indeed, it suffices to consider the first and second order Taylor expansions in both sides of the inequality $\|1 + \eta b\boldsymbol{X}\|_q \le \|1 + b\boldsymbol{X}\|_2$ as $b \to 0$. We leave details to the reader.

We now give the proofs of Proposition 3.16 and Proposition 3.17, which follow the argument of Szulga [48, Prop. 2.20]:

Proof: (of Proposition 3.16) Let $\boldsymbol{X}'$ be an independent copy of $\boldsymbol{X}$ and put $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{X}'$. By the triangle inequality $\|\boldsymbol{Y}\|_q \le 2\|\boldsymbol{X}\|_q$. Let ϵ be a symmetric ± 1 Bernoulli random variable independent of $\boldsymbol{Y}$. Note that $\boldsymbol{Y}$ is symmetric, so it has the same distribution as $\epsilon\boldsymbol{Y}$. Now by Jensen's inequality, Fubini's theorem, $(2, q, (q-1)^{-1/2})$-hypercontractivity of ϵ, and Minkowski's inequality we get

$$\|a + \eta_q\boldsymbol{X}\|_q \le \|a + \eta_q\boldsymbol{Y}\|_q = \|a + \eta_q\epsilon\boldsymbol{Y}\|_q \le (\mathbf{E}_{\mathbf{Y}}[(\mathbf{E}_\epsilon[|a + (q-1)^{1/2}\eta_q\epsilon\boldsymbol{Y}|^2])^{q/2}])^{1/q} =$$

$$(\mathbf{E}[(a^2 + (q-1)\eta_q^2\boldsymbol{Y}^2)^{q/2}])^{1/q} = \|a^2 + (q-1)\eta_q^2\boldsymbol{Y}^2\|_{q/2}^{1/2} \le (a^2 + (q-1)\eta_q^2\|\boldsymbol{Y}^2\|_{q/2})^{1/2} =$$

$$(a^2 + (\frac{\|\boldsymbol{Y}\|_q}{2\|\boldsymbol{X}\|_q})^2 \cdot \mathbf{E}[\boldsymbol{X}^2])^{1/2} \leq (a^2 + \mathbf{E}[\boldsymbol{X}^2])^{1/2} = \|a + \boldsymbol{X}\|_2.$$

□

Proof: (of Proposition 3.17) Let $(\boldsymbol{X}', \boldsymbol{V}')$ be an independent copy of $(\boldsymbol{X}, \boldsymbol{V})$ and put $\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{X} - \boldsymbol{V}'\boldsymbol{X}'$. Then $\|\boldsymbol{Y}\|_q \leq 2\|\boldsymbol{V}\|_q\|\boldsymbol{X}\|_q = 2\rho^{1/q}\|\boldsymbol{X}\|_q$ and as in the previous proof we get

$$\|a+\xi_q\boldsymbol{V}\boldsymbol{X}\|_q \leq \|a+\xi_q\boldsymbol{Y}\|_q \leq (a^2+(q-1)\xi_q^2\|\boldsymbol{Y}\|_q^2)^{1/2} \leq (a^2+4(q-1)\xi_q^2\rho^{2/q}\|\boldsymbol{X}\|_q^2)^{1/2} = \|a+\boldsymbol{V}\boldsymbol{X}\|_2.$$

□

If $\boldsymbol{X}$ is defined on a finite probability space in which probability of all atoms is at least α then obviously $\mathbf{E}[\boldsymbol{X}^2] \geq \alpha\|\boldsymbol{X}\|_\infty^2$, so that $\mathbf{E}[|\boldsymbol{X}|^q] \leq \|\boldsymbol{X}\|_\infty^{q-2} \cdot \mathbf{E}[\boldsymbol{X}^2] \leq (\mathbf{E}[\boldsymbol{X}^2])^{q/2}\alpha^{1-\frac{q}{2}}$. In particular, if $q = 3$ then $\|\boldsymbol{X}\|_3/\|\boldsymbol{X}\|_2 \leq \alpha^{-1/6}$, so that $\boldsymbol{V}\boldsymbol{X}$ is $(2, 3, \xi_3)$-hypercontractive with $\xi_3 = 2^{-3/2}\alpha^{1/6}\rho^{1/6}$.

Let us also point out that if $\mathbf{E}[\boldsymbol{X}^4] < \infty$ and $\boldsymbol{X}$ is symmetric then a direct and elementary calculation shows that $\boldsymbol{X}$ is $(2, 4, \eta_4)$-hypercontractive with $\eta_4 = \min(3^{-1/2}, \|\boldsymbol{X}\|_2/\|\boldsymbol{X}\|_4)$ and the constant is optimal. Therefore the random variable $\boldsymbol{X}_{i,j}^{(\rho)}$ which appears in the proof of Theorem 4.8 is $(2, 4, \min(\rho^{1/4}, 3^{-1/2}))$-hypercontractive if $\boldsymbol{\mathcal{X}}$ is the ± 1 Rademacher ensemble; this can be used to get a smaller value for K' if ρ is close to 1.

B Properties of $\Gamma_\rho(\mu)$

Sheppard's Formula [47] gives the value of $\Gamma_\rho(1/2)$:

Theorem B.1 *$\Gamma_\rho(1/2) = \frac{1}{4} + \frac{1}{2\pi}\arcsin\rho$.*

For fixed ρ, the asymptotics of $\Gamma_\rho(\mu)$ as $\mu \to 0$ can be determined precisely; calculations of this nature appear in [45, 14].

Theorem B.2 *As $\mu \to 0$,*

$$\Gamma_\rho(\mu) \sim \mu^{2/(1+\rho)}\,(4\pi\ln(1/\mu))^{-\rho/(1+\rho)}\,\frac{(1+\rho)^{3/2}}{(1-\rho)^{1/2}}.$$

Proof: This follows from, e.g., Lemma 11.1 of [14]; although we have $\rho > 0$ as opposed to $\rho < 0$ as in [14], the formula there can still be seen to hold when $x = y$ (in their notation). □

Lemma B.3 *For all $0 \leq \rho \leq 1$ and all $0 \leq \mu_1 \leq \mu_2 \leq 1$,*

$$\Gamma_\rho(\mu_2) - \Gamma_\rho(\mu_1) \leq 2(\mu_2 - \mu_1).$$

Proof: Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be ρ-correlated Gaussians and write $t_i = \Phi^{-1}(\mu_i)$. Then

$$\begin{aligned}\Gamma_\rho(\mu_2) - \Gamma_\rho(\mu_1) &= \mathbf{P}[\boldsymbol{X} \leq t_2, \boldsymbol{Y} \leq t_2] - \mathbf{P}[\boldsymbol{X} \leq t_1, \boldsymbol{Y} \leq t_1] \\ &\leq 2\mathbf{P}[t_1 \leq \boldsymbol{X} \leq t_2] \;=\; 2(\mu_2 - \mu_1).\end{aligned}$$

□

Lemma B.4 *Let $0 < \mu < 1$ and $0 < \rho_1 < \rho_2 < 1$, and write $I_2 = (\Gamma_{\rho_2}(\mu) - \mu^2)/\rho_2$. Then $I_2 \leq \mu$ and*

$$\Gamma_{\rho_2}(\mu) - \Gamma_{\rho_1}(\mu) \leq 4 \cdot \frac{1 + \ln(\mu/I_2)}{1 - \rho_2} \cdot I_2 \cdot (\rho_2 - \rho_1)\,.$$

Proof: Let

$$d = \frac{1 + \ln(\mu/I_2)}{1 - \rho_2}.$$

The proof will rely on the fact that $\Gamma_\rho(\mu)$ is a convex function of ρ. This implies in particular that $I_2 \le \mu$. Moreover, by the Mean Value Theorem it suffices to show that the derivative at ρ_2 is at most $4dI_2$. If we write the Hermite polynomial expansion of χ_μ as $\chi_\mu(x) = \sum_i c_i H_i(x)$, then $\Gamma_\rho(\mu) = \sum_i c_i^2 \rho^i$, and thus

$$\frac{\partial}{\partial \rho}\Gamma_\rho(\mu)\bigg|_{\rho=\rho_2} = \sum_{i\ge 1} i c_i^2 \rho_2^{i-1} \le \sum_{1\le i\le d+1} i c_i^2 \rho_2^{i-1} + \sum_{i \ge d+1} i c_i^2 \rho_2^{i-1}. \tag{31}$$

We will complete the proof by showing that both terms in (31) are at most $2dI_2$. The first term is visibly at most $(d+1)I_2 \le 2dI_2$. As for the second term, the quantity $i\rho_2^{i-1}$ decreases for $i \ge \rho_2/(1-\rho_2)$. Since $d + 1 \ge (2-\rho)/(1-\rho) \ge \rho_2/(1-\rho_2)$ the second term is therefore at most $(d+1)\rho_2^d I_2 \le (d+1)\rho_2^d \mu$. But

$$\rho_2^d \le \rho_2^{\ln(\mu/I_2)/(1-\rho_2)} \le I_2/\mu$$

since $\rho_2^{1/(1-\rho_2)} \le 1/e$ for all ρ_2. Thus the second term of (31) is also at most $(d+1)I_2 \le 2dI_2$, as needed. $\square$

Using the fact that $-I_2 \ln I_2$ is a bounded quantity we obtain:

Corollary B.5 *For all $0 < \mu < 1$ and $0 < \rho < 1$, if $0 < \delta < (1-\rho)/2$ then*

$$\Gamma_{\rho+\delta}(\mu) - \Gamma_\rho(\mu) \le \frac{O(1)}{1-\rho} \cdot \delta.$$

C Proof of Theorem 4.5

Proof: The proof is essentially the same as the proof of the "upper bound" part of the proof of Theorem 4.1 in [43]. By way of contradiction, suppose the upper bound on δ holds and yet $\mathbf{P}[T_\rho f > 1 - \delta\}] \ge \epsilon$. Let g be the indicator function of a subset of $\{x : T_\rho f(x) > 1 - \delta\}$ whose measure is ϵ.

Let $h = \mathbf{1}_{\{f \le b\}}$, where $b = 1/2 + \mu/2$. By a Markov argument,

$$\mu = \mathbf{E}[f] \ge (1 - \mathbf{E}[h])b \implies \mathbf{E}[h] \ge 1 - \mathbf{E}[f]/b = \frac{1-\mu}{1+\mu}.$$

By another Markov argument, whenever $g(x) = 1$ we have

$$T_\rho(1-f) < \delta \implies T_\rho(h(1-b)) < \delta \implies T_\rho h < \frac{\delta}{1-b}.$$

Thus

$$\mathbf{E}[gT_\rho h] \le \frac{2\epsilon\delta}{1-\mu}. \tag{32}$$

But by Corollary 3.5 in [43] (itself a simple corollary of the reverse Bonami-Beckner inequality),

$$\mathbf{E}[gT_\rho h] \ge \epsilon \cdot \epsilon^{(\sqrt{\alpha}+\rho)^2/(1-\rho^2)}, \tag{33}$$

where $\alpha = \log(1/\mathbf{E}[h])/\log(1/\epsilon)$. (In Gaussian space, this fact can also be proven using Borell's Corollary 4.3.) Note that since $\mathbf{E}[h] \ge (1-\mu)/(1+\mu)$ we get $\alpha \le O(c(\mu)/\log(1/\epsilon))$, which is also at most 1 since we assume $\epsilon \le 1-\mu$. Therefore the exponent $(\sqrt{\alpha}+\rho)^2$ is $\rho^2 + O(\sqrt{\alpha})$. Now (33) implies

$$\mathbf{E}[gT_\rho h] \ge \epsilon \cdot \epsilon^{\rho^2/(1-\rho^2)} \cdot \epsilon^{O(\sqrt{c(\mu)/\log(1/\epsilon)}/(1-\rho))} = \epsilon \cdot \epsilon^{\rho^2/(1-\rho^2)+O(\kappa)}. \tag{34}$$

Combining (32) and (34) yields

$$\delta \ge (\tfrac{1-\mu}{2}) \cdot \epsilon^{\rho^2/(1-\rho^2)+O(\kappa)} = \epsilon^{\log(2/(1-\mu))/\log(1/\epsilon)} \epsilon^{\rho^2/(1-\rho^2)+O(\kappa)} = \epsilon^{\rho^2/(1-\rho^2)+O(\kappa)},$$

a contradiction. □

D Proof of Corollary 5.3

Proof: We define f_n by setting $1 \le u < n$ to be the odd integer nearest to $t\sqrt{n}$ (where t is the number chosen in Theorem 5.3) and then taking

$$f_n(x) = \begin{cases} 1 & \text{if } |x| \in [1,u] & \text{or} \quad |x| \in [-n, -(u+2)], \\ -1 & \text{if } |x| \in [u+2, n] & \text{or} \quad |x| \in [-u, -1], \end{cases}$$

where $|x|$ denotes $\sum_{i=1}^n x_i$. This is clearly a completely symmetric odd function. It is well known that for any boolean function, $\sum_{i=1}^n \mathrm{Inf}_i(f_n)$ equals the expected number of pivotal bits for f_n in a random input. One can easily see that this is $O(\sqrt{n})$. Thus each of f_n's coordinates has influence $O(1/\sqrt{n})$, by symmetry.

Let $p(n,s)$ denote the probability that the sum of n independent ± 1 Bernoulli random variables is exactly s, so

$$p(n,s) = 2^{-n}\binom{n}{\frac{1}{2}n + \frac{1}{2}s},$$

and for a set S of integers let $p(n,S)$ denote $\sum_{s \in S} p(n,s)$.

By symmetry all of f_n's Fourier coefficients at level d have the same value; we will write $\widehat{f_n}(d)$ for this quantity. Since f_n is odd, $\widehat{f_n}(0) = \widehat{f_n}(2) = 0$. By explicit calculation, we have

$$\widehat{f_n}(1) = p(n-1, 0) - 2p(n-1, u+1)$$

and

$$\begin{aligned} \widehat{f_n}(3) &= \frac{1}{4}\Big(p(n-3, \{-2, 2\}) - 2p(n-3, 0)\Big) \\ &\qquad -\frac{1}{4}\Big(p(n-3, \{\pm(u-1), \pm(u+3)\}) - 2p(n-3, \{\pm(u+1)\})\Big) \\ &= -\frac{1}{n-1}\, p(n-3,0) + 2\,\frac{(n-1)-(u+1)^2}{(n-1)^2-(u+1)^2}\, p(n-3, u+1), \end{aligned}$$

where the last equality is by explicit conversion to factorials and simplification. Using $p(n, t\sqrt{n}) = (1+o(1))\sqrt{2/\pi}e^{-t^2/2}\, n^{-1/2}$ as $n \to \infty$, we conclude

$$\widehat{f_n}(1) \sim \sqrt{2/\pi}(1 - 2e^{-u^2/2}), \qquad \widehat{f_n}(3) \sim \sqrt{2/\pi}(-1 + 2(1-u^2)e^{-u^2/2})n^{-3/2}.$$

But the weight of f_n at level 1 is $n \cdot \widehat{f_n}(1)^2$ and the weight of f_n at level 3 is $\binom{n}{3} \cdot \widehat{f_n}(3)^2 \sim (n^3/6)\widehat{f_n}(3)^2$; thus the above imply (30) from Theorem 5.2 in the limit $n \to \infty$ and the proof is complete. □