Noise stability of functions with low influences: invariance and optimality (Extended Abstract)

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

Abstract

In this paper we study functions with low influences on product probability spaces. 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 [29] from theoretical computer science, which was the original motivation for this work, and the "It Ain't Over Till It's Over" conjecture [27] from social choice theory. The "Majority Is Stablest" conjecture and its generalizations proven here, in conjunction with the "Unique Games Conjecture" and its variants, imply a number of (optimal) inapproximability results for graph problems.

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" [24] 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 [38, 18]):

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 [38, 39, 18, 17, 6, 3, 7, 32, 10].

As in the KKL paper, some of the more refined problems studied in recent years have involved restricting attention to functions with low influences [3, 6, 10] (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. For example, in PCP-based hardness of approximation of results one often needs to analyze so-called "Long Code tests"; this involves distinguishing between dictator functions ("long codes") and functions that are far from being dictators — i.e., functions in which all variables have small influence. 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., [13, 28, 11, 30, 29]. As another example, in the theory of social choice from economics, 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. [26].

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 principle 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., [1]), 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 [29]) *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

$$\mathbb{S}_\rho(f) = \sum_S \rho^{|S|} \hat{f}(S)^2 \tag{2}$$

for the *noise stability* of f. This quantity measures how correlated $f(x)$ and $f(y)$ are when x and y are ρ-correlated random strings. Specifically, let T_ρ denote the operator on functions $f : \{-1,1\}^n \to \mathbb{R}$ defined by

$$(T_\rho f)(x) = \mathop{\mathbf{E}}_y [f(y)], \tag{3}$$

where y is a random string in $\{-1,1\}^n$ chosen so that $\mathbf{E}[x_i y_i] = \rho$ for each i independently. Then it holds that

$$\mathbb{S}_\rho(f) = \mathop{\mathbf{E}}_x [f(x)(T_\rho f)(x)] = \mathop{\mathbf{E}}_{x,y}[f(x)f(y)],$$

where in these expectations x is chosen uniformly at random from $\{-1,1\}^n$.

"Majority Is Stablest" and its generalizations proven in the full version of this paper [34] imply the following hardness of approximation consequences. Assuming Khot's Unique Games Conjecture [28] (UGC) we have:

- MAX-2LIN(2) and MAX-2SAT have $(1-\epsilon, 1-\Theta(\epsilon^{1/2}))$-hardness. This improves upon [28], where a hardness of $(1-\epsilon, 1-\epsilon^{1/2+o(1)})$ is proven. This follows from our results in conjunction with [28].
- MAX-CUT has .878-hardness, matching the Goemans-Williamson approximation factor. This follow from our result together with [29].
- 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. Again, this follows from our results together with [29].
- The MAX-q-CUT problem, has $(1 - 1/q + q^{2+o(1)})$-hardness factor. This asymptotically matches the approximation factor obtained by Frieze and Jerrum [20]. This follows from our results together with [29].

We would also like to mention that in a recent work [12] building on our results, it is proven that coloring a 3-colorable graph with any number of colors is NP-hard assuming a variant of UGC.

The second conjecture we prove using our invariance principle was made by E. Friedgut and G. Kalai [27] in 2001:

Conjecture 1.2 ("It Ain't Over Till It's Over") *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 our invariance principle for the special case of the uniform measure on the discrete cube, a discussion of the two conjectures, and the consequences of their being true. In Section 3 we give mostly complete proof sketches of the main invariance principle and the Majority Is Stablest conjecture. Proof details and extensions can be found in the full version of this paper [34]. Finally, Section 4 of this abstract briefly describes some additional results appearing in [34]

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 [33] (see also [16]).

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′ [36]. As well, a contemporary manuscript of Sourav Chatterjee [9] with an invariance principle of similar flavor has come to our attention. What is common to our work and to [36, 9] 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 [36] 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 [9] 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 [36].

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. [23].

2 Our results

2.1 The invariance principle

In this subsection we present a simplified version of our invariance principle in the special case of the uniform measure on the discrete cube. Suppose that $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)$ is an n-bit string drawn from the uniform measure on $\{-1,1\}^n$, so that $\boldsymbol{X}_1,\dots,\boldsymbol{X}_n$ are independent $p=1/2$ Bernoulli random variables. Let $Q(x_1,\dots,x_n)=\sum_{i=1}^n c_i x_i$ be a linear form where $\sum c_i^2 = 1$. The Berry-Esseen Central Limit Theorem implies that

$$\sup_t \Big|\mathbf{P}[Q(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)\le t]-\mathbf{P}[\boldsymbol{G}\le t]\Big| \le O\big(\textstyle\sum_{i=1}^n |c_i|^3\big),$$

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

$$\sup_t \Big|\mathbf{P}[Q(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)\le t]- \mathbf{P}[Q(\boldsymbol{G}_1,\dots,\boldsymbol{G}_n)\le t]\Big| \le O\big(\max_i |c_i|\big). \quad (4)$$

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

$$Q(x_1,\dots,x_n)=\sum_{S\subseteq[n]} c_S \prod_{i\in S} x_i, \quad (5)$$

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 (4), 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 (4) 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)).

Theorem 2.1 *Let $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ be an n-bit string drawn from the uniform measure on $\{-1,1\}^n$ and let Q be a degree d multilinear polynomial as in (5) with*

$$\mathbf{Var}[Q] := \sum_{|S|>0} c_S^2 = 1,$$

$$\mathrm{Inf}_i(Q) := \sum_{S \ni i} c_S^2 \leq \tau \quad \textit{for all } i.$$

Then

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

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

Note that if d is fixed then the above bound tends to 0 with τ. In fact the same result holds for a much wider family of random variables. For example the condition that $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is drawn from the uniform measure on $\{-1,1\}^n$ may be replaced by the condition that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are i.i.d. random variables with $\mathbf{E}[\boldsymbol{X}_i] = 0$, $\mathbf{E}[\boldsymbol{X}_i^2] = 1$ and either

- $\mathbf{E}[\boldsymbol{X}_i^3] \leq \beta < \infty$ for all i, in which case the error is bounded by $O(d\beta^{1/3}\tau^{1/8d})$ or
- $\boldsymbol{X}_i$ obtains only finitely many values, all with probability at least $\alpha > 0$. In this case the error is bounded by $O(d\,\alpha^{-1/6}\,\tau^{1/8d})$.

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.

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 with a dependence on the smoothness. For "noise stability"-type problems such as ours, this smoothing is essentially harmless.

2.2 Majority Is Stablest

2.2.1 About the problem

The Majority Is Stablest conjecture, Conjecture 1.1, was first formally stated in [29]. However the notion of Hamming balls having the highest noise stability in various senses has always been widespread among the community studying discrete Fourier analysis. Indeed, already in KKL's 1998 paper [24] there is the suggestion that Hamming balls and subcubes should maximize a certain noise stability-like quantity. In [3], it was shown that every "asymptotically noise stable" function is correlated with a weighted majority function; also, in [35] 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 [25] on social choice and one by Khot [28] 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 [25] 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 $1 \leq i < j \leq n$ 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 $3/4 + (3/4)\mathbb{S}_{1/3}(f)$ 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 [22] 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 [28] 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, the analysis of MAX-2LIN(2) in Khot's paper [28] 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 [7] from 2001:

Theorem 2.2 (Bourgain [7]) *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 2.3 *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 enables 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 [21] achieves $1-O(\sqrt{\epsilon})$). As an aside, we note that Khot and Vishnoi [31] recently used Corollary 2.3 to prove that negative type metrics do not embed into ℓ_1 with constant distortion.

Another example of this comes from the work of [29]. Among other things, [29] 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, taking $\rho \approx .69$ implies MAX-CUT is .878-hard to approximate, matching the groundbreaking algorithm of Goemans and Williamson [21].

2.2.2 Consequences of confirming the conjecture

We confirm a generalization of the Majority Is Stablest conjecture. In particular:

Theorem 2.4 *Let $f : \{-1,1\}^n \to [-1,1]$ and 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 $1-\rho$.

Two remarks: First, the original Majority Is Stablest conjecture was concerned with the case $\mu = 0$; in this case, $\mathrm{Thr}_n^{(0)}$ is simply the Majority function Maj_n and the formula $\lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{Maj}_n) = \frac{2}{\pi}\arcsin\rho$ is well known [37]. Second, Theorem 2.4 in fact only needs f to have small "low-degree influences", a distinction crucial for PCP applications.

We now give some consequences of this theorem:

Theorem 2.5 *In the setting of Kalai [25], any odd, balanced social choice function f with $o_n(1)$ influences 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 2.3.

Corollary 2.6 *If $f : \{-1,1\}^n \to \{-1,1\}$ satisfies* $\mathbf{E}[f] = 0$ *and* $\mathrm{Inf}_i(f) \le \epsilon^{-O(1/\epsilon)}$ *for all $i \in [n]$, then*

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

Using Corollary 2.6 instead of Corollary 2.3 in Khot [28] we obtain

Corollary 2.7 *Assuming UGC, MAX-2LIN(2) and MAX-2SAT have $(1-\epsilon, 1-O(\epsilon^{1/2}))$-hardness.*

More generally, [29] now implies

Corollary 2.8 *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 the generalization of "Majority Is Stablest" we prove here. The details of the construction are given in [29]:

Theorem 2.9 *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 2.10 *The MAX-q-CUT problem, has $(1 - 1/q + q^{2+o(1)})$-hardness factor, assuming UGC only. This asymptotically matches the approximation factor obtained by Frieze and Jerrum [20].*

2.3 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 [27] in studying social indeterminacy [19, 26]. The setting here is similar to the setting of Arrow's Theorem from Section 2.2.1 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.

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. In particular,

Theorem 2.11 *The Ain't Over Till It's Over Conecture, Conjecture 1.2 is true.*

3 Proof Sketches

3.1 The invariance principle

The proof of the invariance principle follows Lindeberg's proof of the CLT [33] (see also [16]). Below we will denote by $\boldsymbol{\mathcal{X}}$ the vector $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ drawn uniformly from $\{-1,1\}^n$ and by $\boldsymbol{\mathcal{G}}$ a vector $(\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n)$ of n independent Gaussian random variables. One standard method in probability theory to show that two distributions are close is the following: show that for any sufficiently smooth function, its expectation under the first distribution is close to its expectation under the second distribution. (One then applies this for smoothed versions of indicator functions of intervals.) In particular, in order to prove Theorem 2.1 it is sufficient to prove the following (complete details of the proof of Theorem 2.1 are omitted in this extended abstract):

Theorem 3.1 *Let Q be a mutlilinear polynomial as in (5). Assume further that Q satisfies* $\mathbf{Var}[Q] \le 1$, $\deg(Q) \le d$, *and* $\mathrm{Inf}_i(Q) \le \tau$ *for all i. Let $\Psi : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^4$ function with $|\Psi^{(4)}| \le B$ uniformly. Then*

$$\Big|\mathbf{E}\big[\Psi(Q(\boldsymbol{\mathcal{X}}))\big] - \mathbf{E}\big[\Psi(Q(\boldsymbol{\mathcal{G}}))\big]\Big| \le B\, d\, 9^d\, \tau.$$

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

$$\Big|\mathbf{E}\big[\Psi(\boldsymbol{Q}^{(i-1)})\big] - \mathbf{E}\big[\Psi(\boldsymbol{Q}^{(i)})\big]\Big| \le B\, 9^d\, \mathrm{Inf}_i(Q)^2 \qquad (6)$$

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

$$\sum_{i=1}^{n} \mathrm{Inf}_i(Q)^2 \le \tau \cdot \sum_{i=1}^{n} \mathrm{Inf}_i(Q) \le d\tau.$$

Let us fix a particular $i \in [n]$ and proceed to prove (6). Write

$$\begin{aligned} \boldsymbol{R} &= \sum_{S: i \notin S} c_S \prod_{j \in S} \boldsymbol{\mathcal{Z}}_j^{(i)}, \\ \boldsymbol{S} &= \sum_{S: i \in S} c_S \prod_{j \in S \setminus \{i\}} \boldsymbol{\mathcal{Z}}_j^{(i)}. \end{aligned}$$

Note that $\boldsymbol{R}$ and $\boldsymbol{S}$ are independent of the variables $\boldsymbol{X}_i$ and $\boldsymbol{G}_i$, and that $\boldsymbol{Q}^{(i-1)} = \boldsymbol{R} + \boldsymbol{X}_i \boldsymbol{S}$ and $\boldsymbol{Q}^{(i)} = \boldsymbol{R} + \boldsymbol{G}_i \boldsymbol{S}$.

To bound the left side of (6) — i.e.,

$$\Big|\mathbf{E}[\Psi(\boldsymbol{R} + \boldsymbol{X}_i \boldsymbol{S}) - \Psi(\boldsymbol{R} + \boldsymbol{G}_i \boldsymbol{S})]\Big|$$

— we use Taylor's theorem: for all $x, y \in \mathbb{R}$,

$$\Big|\Psi(x+y) - \sum_{k=0}^{3} \frac{\Psi^{(k)}(x)\, y^k}{k!}\Big| \le \frac{B}{24}\, y^4.$$

In particular,

$$\Big|\mathbf{E}[\Psi(\boldsymbol{R} + \boldsymbol{X}_i \boldsymbol{S})] - \sum_{k=0}^{3} \mathbf{E}\Big[\frac{\Psi^{(k)}(\boldsymbol{R})\, \boldsymbol{X}_i^k\, \boldsymbol{S}^k}{k!}\Big]\Big| \le \frac{B}{24}\, \mathbf{E}\big[\boldsymbol{X}_i^4 \boldsymbol{S}^4\big] = \frac{B}{24}\, \mathbf{E}\big[\boldsymbol{S}^4\big], \quad (7)$$

(where we used independence of $\boldsymbol{X}_i$ from $\boldsymbol{S}$ and also $\mathbf{E}[\boldsymbol{X}_i^4] = 1$ in the last step), and similarly,

$$\Big|\mathbf{E}[\Psi(\boldsymbol{R} + \boldsymbol{G}_i \boldsymbol{S})] - \sum_{k=0}^{3} \mathbf{E}\Big[\frac{\Psi^{(k)}(\boldsymbol{R})\, \boldsymbol{G}_i^k\, \boldsymbol{S}^k}{k!}\Big]\Big| \le \frac{B}{24}\, \mathbf{E}\big[\boldsymbol{G}_i^4 \boldsymbol{S}^4\big] = \frac{B}{24} \cdot 3\mathbf{E}\big[\boldsymbol{S}^4\big]. \quad (8)$$

Since $\boldsymbol{X}_i$ and $\boldsymbol{G}_i$ are independent of $\boldsymbol{R}$ and $\boldsymbol{S}$ and since the first 3 moments of $\boldsymbol{X}_i$ equal those of $\boldsymbol{G}_i$ it follows that for $k = 0, 1, 2, 3$:

$$\begin{aligned} \mathbf{E}[\Psi^{(k)}(\boldsymbol{R})\, \boldsymbol{X}_i^k\, \boldsymbol{S}^k] &= \mathbf{E}[\Psi^{(k)}(\boldsymbol{R})\, \boldsymbol{S}^k] \cdot \mathbf{E}[\boldsymbol{X}_i^k] \\ &= \mathbf{E}[\Psi^{(k)}(\boldsymbol{R})\, \boldsymbol{S}^k] \cdot \mathbf{E}[\boldsymbol{G}_i^k] \\ &= \mathbf{E}[\Psi^{(k)}(\boldsymbol{R})\, \boldsymbol{G}_i^k\, \boldsymbol{S}^k]. \quad (9) \end{aligned}$$

From (7), (8) and (9) it follows that

$$\Big|\mathbf{E}[\Psi(\boldsymbol{R} + \boldsymbol{X}_i \boldsymbol{S}) - \Psi(\boldsymbol{R} + \boldsymbol{G}_i \boldsymbol{S})]\Big| \le \frac{B}{24}(1+3)\mathbf{E}[\boldsymbol{S}^4] \le B\, \mathbf{E}[\boldsymbol{S}^4]. \quad (10)$$

We now use hypercontractivity. By Lemma 3.2 below,

$$\mathbf{E}[\boldsymbol{S}^4] \le 9^d \mathbf{E}[\boldsymbol{S}^2]^2. \quad (11)$$

But using orthonormality of $\boldsymbol{\mathcal{Z}}_j$'s we have

$$\mathbf{E}[\boldsymbol{S}^2] = \sum_{S : i \in S} c_i^2 = \mathrm{Inf}_i(Q). \quad (12)$$

Combining (10), (11) and (12) it follows that

$$\Big|\mathbf{E}[\Psi(\boldsymbol{R} + \boldsymbol{X}_i \boldsymbol{S}) - \Psi(\boldsymbol{R} + \boldsymbol{G}_i \boldsymbol{S})]\Big| \le B\, 9^d\, \mathrm{Inf}_i(Q)^2,$$

confirming (6) and completing the proof. □

The hypercontractivity result we needed is a special case of the Bonami-Beckner theorem [4, 2]. We include a very short proof for completeness.

Lemma 3.2 *Let Q be a multilinear polynomial as in (5) of degree d. Let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ be independent real random variables with $\mathbf{E}[\boldsymbol{Z}_i] = E[\boldsymbol{Z}_i^3] = 0$, $\mathbf{E}[\boldsymbol{Z}_i^2] = 1$ and $\mathbf{E}[\boldsymbol{Z}_i^4] \le 9$ for $i = 1, 2, \ldots, n$. Let $\boldsymbol{Q} = Q(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$. Then $\mathbf{E}[\boldsymbol{Q}^4] \le 9^d \mathbf{E}[\boldsymbol{Q}^2]^2$.*

Proof: The proof is by induction on the number of variables n. The case $n = 0$ is trivial, as Q is just a constant. So assume $n > 0$. We can express $Q(x_1, \ldots, x_n)$ as $R(x_1, \ldots, x_{n-1}) + x_n S(x_1, x_2, \ldots, x_{n-1})$ where R and S are multilinear polynomials in at most $n-1$ variables, $\deg(R) \le d$ and $\deg(S) \le d-1$. Let

$$\boldsymbol{R} = R(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{n-1})$$

and

$$\boldsymbol{S} = S(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{n-1}).$$

Clearly, $\boldsymbol{R}$ and $\boldsymbol{S}$ are independent of $\boldsymbol{Z}_n$. We have

$$\begin{aligned} \mathbf{E}[\boldsymbol{Q}^4] &= \mathbf{E}[(\boldsymbol{R} + \boldsymbol{Z}_n \boldsymbol{S})^4] \\ &= \mathbf{E}[\boldsymbol{R}^4] + 4\mathbf{E}[\boldsymbol{Z}_n] \cdot \mathbf{E}[\boldsymbol{R}^3 \boldsymbol{S}] + \\ &\quad\; 6\mathbf{E}[\boldsymbol{Z}_n^2] \cdot \mathbf{E}[\boldsymbol{R}^2 \boldsymbol{S}^2] \\ &\quad\; +4\mathbf{E}[\boldsymbol{Z}_n^3] \cdot \mathbf{E}[\boldsymbol{R}\boldsymbol{S}^3] + \mathbf{E}[\boldsymbol{Z}_n^4] \cdot \mathbf{E}[\boldsymbol{S}^4] \\ &\le \mathbf{E}[\boldsymbol{R}^4] + 6\mathbf{E}[\boldsymbol{R}^2 \boldsymbol{S}^2] + 9\mathbf{E}[\boldsymbol{S}^4] \\ &\le \left(\sqrt{\mathbf{E}[\boldsymbol{R}^4]} + 3\sqrt{\mathbf{E}[\boldsymbol{S}^4]}\right)^2 \\ &\le \left(3^d \mathbf{E}[\boldsymbol{R}^2] + 3 \cdot 3^{d-1} \mathbf{E}[\boldsymbol{S}^2]\right)^2 \\ &= 9^d \left(\mathbf{E}[\boldsymbol{R}^2] + \mathbf{E}[\boldsymbol{S}^2]\right)^2 \\ &= 9^d \mathbf{E}[\boldsymbol{Q}^2]^2. \end{aligned}$$

where the first inequality uses the conditions on $\boldsymbol{Z}_n$'s moments, the second inequality uses Cauchy-Schwartz, the third inequality used the induction hypothesis for $\boldsymbol{R}$ and $\boldsymbol{S}$, and the final equality uses $\mathbf{E}[\boldsymbol{Z}_n] = 0$ and $\mathbf{E}[\boldsymbol{Z}_n^2] = 1$. □

3.2 Majority Is Stablest

In this section we sketch the proof of the Majority Is Stablest conjecture. The essential idea is that continuous analogues of the conjecture, on the n-dimensional sphere and in n-dimensional Gaussian space, are already known to be true (without any assumption or notion of "influences"). For example, in studying the Goemans-Williamson semidefinite programming algorithm for MAX-CUT, Feige and Schechtman [15] produced an optimal semidefinite programming gap essentially by showing the Majority Is Stablest analogue on the Euclidean sphere. Namely, they showed that among subsets of half of the sphere, the sets with the most self-correlation under ρ-perturbation are hemispheres — i.e., intersections of halfspaces with the sphere. Even more relevantly for us, the analogous (and roughly equivalent) result in Gaussian space was proved in the '80s by

C. Borell [5]: halfspaces have maximum "ρ-noise stability", and this maximal value is $\frac{2}{\pi}\arcsin\rho$, the noise stability of Majority.

Both of the above-mentioned results have conceptually simple proofs: Given a subset of the sphere or of Gaussian space, one shows that symmetrizing it across lower-dimensional subspaces improves its stability under ρ-snoise. Then sufficiently many symmetrizations bring it close to a halfspace. (Note that all halfspaces have equal noise stability by spherical symmetry.) One would like to use a similar argument in the discrete setting of $\{-1,1\}^n$ — i.e., show that some symmetrization operation both improves a function's noise stability and brings it closer to the Majority function (the intersection of $\{-1,1\}^n$ with a "generic" halfspace). However the presence of n "special" directions in $\{-1,1\}^n$ make this impossible — indeed, the halfspaces oriented in these directions, the dictator functions $f(x) = \pm x_i$, have maximal noise stability. This is where our invariance principle comes into play. Under the assumptions of the Majority Is Stablest conjecture, our initial function f on $\{-1,1\}^n$ has small influences and is thus not particularly aligned with any special direction. The invariance principle now allows us to replace f by a highly similar function on n-dimensional Gaussian space, and then Borell's result implies that its noise stability is bounded above by that of Mmjority.

Before quoting Borell's result, let us generalize the notion of noise stability to the setting used in our invariance principle, Theorem 3.1. Let $\boldsymbol{\mathcal{Y}} = (\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n)$ be a sequence of i.i.d. random variables with $\mathbf{E}[\boldsymbol{Y}_i] = 0$ and $\mathbf{E}[\boldsymbol{Y}_i^2] = 1$ (think of the $\boldsymbol{Y}_i$'s as either random ± 1 variables or standard Gaussians). Given $0 \le \rho \le 1$, and a multilinear polynomial Q as in (5) we define

$$(T_\rho Q)(x_1,\dots,x_n) = \sum_S c_S \rho^{|S|} \prod_{i\in S} x_i,$$

and

$$\mathbb{S}_\rho(Q) = \sum_{S\subseteq[n]} \rho^{|S|} c_S^2 = \mathbf{E}[Q(\boldsymbol{\mathcal{Y}})\cdot(T_\rho Q)(\boldsymbol{\mathcal{Y}})]. \qquad (13)$$

Note that the value of $\mathbb{S}_\rho(Q)$ does not depend on the sequence $\boldsymbol{\mathcal{Y}}$.

We now describe Borell's result. Let γ_n be the n-dimensional Gaussian measure. When $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{G}} = (\boldsymbol{G}_1,\dots,\boldsymbol{G}_n)$ is a sequence of independent Gaussians, the operator T_ρ is identical with the Ornstein-Uhlenbeck operator U_ρ which acts on $L^2(\mathbb{R}^n,\gamma_n)$ by

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

where the expectation on y is with respect to γ_n. The operator U_ρ is central in the study of the heat equation and Brownian motion in n dimensions. Since it extends the definition of T_ρ, it is natural to define $\mathbb{S}_\rho(f) = \mathbf{E}[fU_\rho f]$ for all $f \in L^2(\mathbb{R}^n,\gamma_n)$.

Borell [5] (see also Ledoux's lecture notes [14]) showed that for all $\rho \in [0,1]$ and all $f \in L^2(\mathbb{R}^n,\gamma_n)$, $\mathbb{S}_\rho(f)$ is at most $\mathbb{S}_\rho(f^*)$, where f^* denotes the so-called spherical rearrangement of f. In particular, if f's range is $\{-1,1\}$ then f^* is simply the $\{-1,1\}$-valued indicator function of a halfspace with the same measure as f. A straightforward convexity argument extends the same bound to functions $f : (\mathbb{R}^n,\gamma_n) \to [-1,1]$ and thus we have the following corollary of Borell's result:

Theorem 3.3 *Let $f : \mathbb{R}^n \to [-1,1]$ be a measurable function on Gaussian space with $\mathbf{E}[f] = \mu$. Then for all $0 \le \rho \le 1$ we have $\mathbb{S}_\rho(f) \le \mathbb{S}_\rho(\chi_\mu)$, where $\chi_\mu : \mathbb{R} \to \{-1,1\}$ is the indicator function of the interval $(-\infty,t]$, where t is chosen so that $\mathbf{E}[\chi_\mu] = \mu$.*

The proof of Majority Is Stablest goes by using our invariance principle to reduce to the above theorem. There is one small technical difficulty to overcome: namely, the invariance principle can only be applied to multilinear polynomials of bounded degree, whereas a general function $f : \{-1,1\}^n \to [-1,1]$ can have degree as high as n when represented as a multilinear (Fourier) polynomial. As mentioned in Section 2.1, we get around this by giving an invariance principle for "smoothed" multilinear polynomials of arbitrary degree. Specifically, this smoothed invariance principle is very similar to Theorems 2.1 and 3.1, with the following distinction: it holds for polynomials Q of arbitrary degree; however it only shows that $(T_{1-\gamma}Q)(\boldsymbol{\mathcal{X}})$ and $(T_{1-\gamma}Q)(\boldsymbol{\mathcal{G}})$ are close in distribution. The closeness bound we show is $\tau^{\Omega(\gamma)}$ (i.e., it's as if Q had degree $\Theta(1/\gamma)$) and the proof involves straightforward degree-truncation arguments. Since the Majority Is Stablest problem involves studying $T_\rho f$ anyway, we can essentially assume that our functions are already smoothed and thus use the smoothed invariance principle.

We now give a mostly complete sketch of the proof of Majority Is Stablest, Theorem 2.4:

Proof: (of Theorem 2.4) Express f as a multilinear polynomial Q (i.e., in its Fourier representation) over the random ± 1 variables $\boldsymbol{\mathcal{X}} = (\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)$. As usual, write also $\boldsymbol{\mathcal{G}} = (\boldsymbol{G}_1,\dots,\boldsymbol{G}_n)$ for a sequence of independent Gaussians. We now express $\rho = \rho'\cdot(1-\gamma)^2$, where $0 < \gamma \ll 1-\rho$ is a tiny quantity to be chosen later. From (13) we see that

$$\begin{aligned}\mathbb{S}_\rho(f) = \mathbb{S}_\rho(Q(\boldsymbol{\mathcal{X}})) &= \sum_S (\rho'\cdot(1-\gamma)^2)^{|S|} c_S^2 \\ &= \mathbb{S}_{\rho'}((T_{1-\gamma}Q)(\boldsymbol{\mathcal{G}})). \qquad (14)\end{aligned}$$

Let us write $\boldsymbol{R}$ for the random variable $(T_{1-\gamma}Q)(\boldsymbol{\mathcal{X}})$ and $\boldsymbol{S}$ for the random variable $(T_{1-\gamma}Q)(\boldsymbol{\mathcal{G}})$. By the smoothed invariance principle, we know that these two random variables are very close in distribution. In particular, let $\zeta : \mathbb{R} \to \mathbb{R}$ denote the function with $\zeta(t) = 0$ if $t \in [-1,1]$ and $\zeta(t) = (|t|-1)^2$ otherwise. (I.e. ζ measures the L_2^2-distance of a number from being in $[-1,1]$.) Then smoothed invariance tells us that

$$\Big|\mathbf{E}[\zeta(\boldsymbol{R})] - \mathbf{E}[\zeta(\boldsymbol{S})]\Big| \le \tau^{\Omega(\gamma)} \tag{15}$$

(cf. the statement of Theorem 3.1; although ζ is not $\mathcal{C}^4$ it can be very closely approximated by $\mathcal{C}^4$ functions). By assumption, $f(\boldsymbol{\mathcal{X}}) = Q(\boldsymbol{\mathcal{X}})$ is always in the interval $[-1,1]$, and since T_ρ is an averaging operator (see Equation (3)) the same is true of $\boldsymbol{R} = (T_{1-\gamma}Q)(\boldsymbol{\mathcal{X}})$; in other words, $\zeta(\boldsymbol{R}) = 0$. Hence by (15) we have $|\zeta(\boldsymbol{S})| \le \tau^{\Omega(\gamma)}$. Another way to express this is to say that $\|\boldsymbol{S} - \boldsymbol{S}'\|_2^2 \le \tau^{\Omega(\gamma)}$, where $\boldsymbol{S}'$ is the random variable on n-dimensional Gaussian space, dependent on $\boldsymbol{S}$, defined to be the truncation of $\boldsymbol{S}$ to the interval $[-1,1]$. We now have that

$$\begin{aligned}
& |\mathbb{S}_{\rho'}(\boldsymbol{S}) - \mathbb{S}_{\rho'}(\boldsymbol{S}')| \\
= \;& |\mathbf{E}[\boldsymbol{S}\cdot U_{\rho'}\boldsymbol{S}] - \mathbf{E}[\boldsymbol{S}'\cdot U_{\rho'}\boldsymbol{S}']| \\
\le \;& |\mathbf{E}[\boldsymbol{S}\cdot U_{\rho'}\boldsymbol{S}] - \mathbf{E}[\boldsymbol{S}'\cdot U_{\rho'}\boldsymbol{S}]| \\
& \qquad + |\mathbf{E}[\boldsymbol{S}'\cdot U_{\rho'}\boldsymbol{S}] - \mathbf{E}[\boldsymbol{S}'\cdot U_{\rho'}\boldsymbol{S}']| \\
\le \;& (\|\boldsymbol{S}\|_2 + \|\boldsymbol{S}'\|_2)\|\boldsymbol{S} - \boldsymbol{S}'\|_2 \le \tau^{\Omega(\gamma)}, \qquad (16)
\end{aligned}$$

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

Now $\mu = \mathbf{E}[f] = c_\emptyset = \mathbf{E}[Q(\boldsymbol{\mathcal{G}})] = \mathbf{E}[\boldsymbol{S}]$, and writing $\mu' = \mathbf{E}[\boldsymbol{S}']$ it follows from $\|\boldsymbol{S}-\boldsymbol{S}'\|_2 \le \tau^{\Omega(\gamma)}$ and Cauchy-Schwartz that $|\mu - \mu'| \le \tau^{\Omega(\gamma)}$. We now apply the Borell result Theorem 3.3 to conclude that $\mathbb{S}_{\rho'}(\boldsymbol{S}') \le \mathbb{S}_{\rho'}(\chi_{\mu'})$. Combining this with (14) and (16) we get

$$\mathbb{S}_\rho(f) \le \mathbb{S}_{\rho'}(\chi_{\mu'}) + \tau^{\Omega(\gamma)}.$$

An elementary argument can be used to show that $\rho' \approx \rho$ and $\mu' \approx \mu$ imply that

$$|\mathbb{S}_{\rho'}(\chi_{\mu'}) - \mathbb{S}_\rho(\chi_\mu)| \le O(\gamma/(1-\rho)).$$

Further, it's straightforward to see that $\mathbb{S}_\rho(\chi_\mu)$ is exactly $\lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{Thr}_n^{(\mu)})$. Thus we have

$$\mathbb{S}_\rho(f) \le \lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{Thr}_n^{(\mu)}) + \tau^{\Omega(\gamma)} + O(\gamma/(1-\rho)).$$

Optimizing over γ completes the proof. □

4 Other results omitted from this abstract

This short section summarizes some of the additional material appearing in the full version [34] of this extended abstract.

It Ain't Over Till It's Over. The proof of this conjecture is similar in outline to that of the Majority Is Stablest conjecture. We replace the use of Borell's result on Gaussian space as follows: We first use the invariance principle to convert the Ain't problem on $\{-1,1\}^n$ to a related problem on Gaussian space. This related problem can be solved using the results of Borell; however, to achieve sharp bounds most easily, we use the invariance principle again to bring the related problem *back* into $\{-1,1\}^n$! We then use the solution to this related problem on $\{-1,1\}^n$ given by [35], a paper on the topic of "non-interactive correlation distillation".

Generalized domains. As alluded to in the introduction, the invariance principle we prove in [34] holds in the full generality of functions whose domain is any discrete product probability space. The proof is not much more difficult although the notation becomes cumbersome. Using this generalized invariance principle, we extend the Majority Is Stablest result to functions $f : [q]^n \to [-1,1]$ and the It Ain't Over Till It's Over result to multiparty elections. We use the Majority Is Stablest extension to prove the hardness of approximation results for MAX-2LIN(q) and MAX-q-CUT; the paper [12] uses it to prove hardness of approximation for coloring problems.

A counterexample to a conjecture of Kalai. From [29] it is known that among functions with low influences, Majority maximizes the Fourier weight at level 1. From the present work we also know that Majority maximizes $\sum_S \hat{f}(S)^2 \rho^{|S|}$. Kalai [25] made the natural conjecture that among all transitive functions, Majority maximizes $\sum_{|S|\le d} \hat{f}(S)^2$. Surprisingly, we give an explicit counterexample showing this conjecture is false.

5 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 [36] and Olivier Guédon for referring us to [8].

References

[1] 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.

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

[3] 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.

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

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

[6] 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.

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

[8] 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.

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

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

[11] 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.

[12] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. Preprint, 2005.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

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

[19] 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.

[20] 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.

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

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

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

[24] 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.

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

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

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

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

[29] 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.

[30] 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.

[31] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into l1. Preprint. Extended abstract to appear in Proceedings of 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005).

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

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

[34] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. Submitted. Availibe at the Arxiv as math.PR/0503503, 2005.

[35] 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 Journal of Matheamtics, 2005.

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

[37] 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.

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

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