University of Birmingham
Research at Birmingham

Hypercontractivity for global functions and sharp thresholds

Keevash, Peter; Lifshitz, Noam; Long, Eoin; Minzer, Dor

DOI:
10.1090/jams/1027

License:
None: All rights reserved

Document Version
Peer reviewed version

Citation for published version (Harvard):
Keevash, P, Lifshitz, N, Long, E & Minzer, D 2023, 'Hypercontractivity for global functions and sharp thresholds', *Journal of the American Mathematical Society*. https://doi.org/10.1090/jams/1027

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 09. May. 2024

Hypercontractivity for global functions and sharp thresholds

Peter Keevash[*] Noam Lifshitz [†] Eoin Long [‡] Dor Minzer[§]

Abstract

The classical hypercontractive inequality for the noise operator on the discrete cube plays a crucial role in many of the fundamental results in the Analysis of Boolean functions, such as the KKL (Kahn-Kalai-Linial) theorem, Friedgut's junta theorem and the invariance principle of Mossel, O'Donnell and Oleszkiewicz. In these results the cube is equipped with the uniform ($1/2$-biased) measure, but it is desirable, particularly for applications to the theory of sharp thresholds, to also obtain such results for general p-biased measures. However, simple examples show that when p is small there is no hypercontractive inequality that is strong enough for such applications.

In this paper, we establish an effective hypercontractivity inequality for general p that applies to 'global functions', i.e. functions that are not significantly affected by a restriction of a small set of co-ordinates. This class of functions appears naturally, e.g. in Bourgain's sharp threshold theorem, which states that such functions exhibit a sharp threshold. We demonstrate the power of our tool by strengthening Bourgain's theorem, making progress[1] on two conjectures of Kahn and Kalai, and proving a p-biased analogue of the seminal invariance principle of Mossel, O'Donnell, and Oleszkiewicz.

In this 2023 version of our paper we will also survey many further applications of our results that have been obtained by various authors since we arXived the first version in 2019.

1 Introduction

The field of analysis of Boolean functions is centered around the study of functions on the discrete cube $\{0,1\}^n$, via their Fourier–Walsh expansion, often using the classical hypercontractive inequality for the noise operator, obtained independently by Bonami [17], Gross [35] and Beckner [10]. In particular, the fundamental KKL theorem of Kahn, Kalai and Linial [43] applies hypercontractivity to obtain structural information on Boolean valued functions with small 'total influence' / 'edge boundary' (see Section 1.2); such functions cannot be 'global': they must have a co-ordinate with large influence.

The theory of sharp thresholds is closely connected (see Section 1.3) to the structure of Boolean functions of small total influence, not only in the KKL setting of uniform measure on the cube, but also in the general p-biased setting. However, we will see below that the hypercontractivity theorem is ineffective for small p. This led Friedgut [29], Bourgain [29, appendix], and Hatami [38] to develop new ideas for proving p-biased analogues of the KKL theorem. The theme of these works can be roughly summarised by the statement: an effective analog of the KKL theorem holds for a certain class of 'global' functions. However, these theorems were incomplete in two important respects:

[*]Mathematical Institute, University of Oxford, UK. Research supported in part by ERC Consolidator Grant 647678.

[†]Einstein Institute of Mathematics, Hebrew University, Jerusalem, Israel.

[‡]University of Birmingham, Birmingham, UK.

[§]Department of Mathematics, Massachusetts Institute of Technology. Research was done while the author was a postdoctoral fellow at the Institute for Advanced Study, Princeton and supported by an NSF grant CCF-1412958 and Rothschild Fellowship.

[1]Both these conjectures were open when we arXived this paper in 2019. One of them was solved in 2022; the other is still open.

- *Sharpness:* Unlike the KKL theorem, they are not sharp up to constant factors.
- *Applicability:* They are only effective in the 'dense setting' when $\mu_p(f)$ is bounded away from 0 and 1, whereas the 'sparse setting' $\mu_p(f) = o(1)$ is needed for many important open problems.

Main result

The fundamental new contribution of this paper is a hypercontractive theorem for functions that are 'global' (in a sense made precise below). This has many applications, of which the following are included in this paper (subsequent applications will be discussed in Section 1.6).

- We strengthen Bourgain's Theorem by obtaining an analogue of the KKL theorem that is both quantitively tight and applicable in the sparse regime.
- We prove a variant form of the Kahn-Kalai Isoperimetric Conjecture concerning the structure of functions that are close to optimal for the edge-isoperimetric inequality.
- We obtain a sharp threshold result for global monotone functions in the spirit of the Kahn-Kalai Threshold Conjecture.[2]
- We obtain a p-biased generalisation of the seminal invariance principle of Mossel, O'Donnell and Oleszkiewicz [63] (itself a generalisation of the Berry-Esseen theorem from linear functions to polynomials of bounded degree), thus opening the door to p-biased versions of its many striking applications in Hardness of Approximation and Social Choice Theory (see O'Donnell [64, Section 11.5]) and Extremal Combinatorics (see Dinur–Friedgut–Regev [21]).

1.1 Hypercontractivity of global functions

Before formally stating our main theorem, we start by recalling (the p-biased version of) the classical hypercontractive inequality. Let[3] $p \in \left(0, \frac{1}{2}\right]$. For $r \geq 1$ we write $\| \cdot \|_r$ (suppressing p from our notation) for the norm on $L^r(\{0,1\}^n, \mu_p)$.

Definition 1.1 (Noise operator)**.** For $x \in \{0,1\}^n$ we define the ρ-correlated distribution $N_\rho(x)$ on $\{0,1\}^n$: a sample $\mathbf{y} \sim N_\rho(x)$ is obtained by, independently for each i setting $\mathbf{y}_i = x_i$ with probability ρ, or otherwise (with probability $1-\rho$) we resample $\mathbf{y}_i$ with $\mathbb{P}(\mathbf{y}_i = 1) = p$. We define the noise operator T_ρ on $L^2(\{0,1\}^n, \mu_p)$ by

$$\mathrm{T}_\rho\left(f\right)\left(x\right) = \mathbb{E}_{\mathbf{y} \sim N_\rho(x)}\left[f\left(\mathbf{y}\right)\right].$$

Hölder's inequality gives $\|f\|_r \leq \|f\|_s$ whenever $r \leq s$. The hypercontractivity theorem gives an inequality in the other direction after applying noise to f; for example, for $p = 1/2$, $r = 2$ and $s = 4$ we have

$$\|\mathrm{T}_\rho f\|_4 \leq \|f\|_2$$

for any $\rho \leq \frac{1}{\sqrt{3}}$. A similar inequality also holds when $p = o(1)$, but the correlation ρ has to be so small that it is not useful in applications; e.g. if $f(x) = x_1$ (the 'dictator' or 'half cube'), then $\|f\|_2 = \sqrt{\mu_p(f)} = \sqrt{p}$

[2] Subsequent to the 2019 version of this paper, the Kahn-Kalai Threshold Conjecture has been proved, see [5, 27, 65]. The proof techniques are completely different to ours, in that no Fourier Analysis is involved, although one can identify in their proof strategy some elements of our notion of globality.

[3] The case where $p > \frac{1}{2}$ is similar.

and $\mathrm{T}_\rho f(x) = \mathbb{E}_{\mathbf{y}\sim N_\rho(x)}\mathbf{y}_1 = \rho x_1 + (1-\rho)p$, so $\|\mathrm{T}_\rho f\|_4 > (\mathbb{E}[\rho^4 x_1^4])^{1/4} = \rho p^{1/4}$. Thus we need $\rho = O(p^{1/4})$ to obtain any hypercontractive inequality for general f.

Local and global functions

To resolve this issue, we note that the tight examples for the hypercontractive inequality are *local*, in the sense that a small number of coordinates can significantly influence the output of the function. On the other hand, many functions of interest are *global*, in the sense that a small number of coordinates can change the output of the function only with a negligible probability; such global functions appear naturally in Random Graph Theory [2], Theoretical Computer Science [29] and Number Theory [30]. Our hypercontractive inequality will show that constant noise suffices for functions that are global in a sense captured by *generalised influences*, which we will now define.

Let $f\colon \{0,1\}^n \to \mathbb{R}$. For $S \subset [n]$ and $x \in \{0,1\}^S$, we write $f_{S\to x}$ for the function obtained from f by restricting the coordinates of S according to x (if $S = \{i\}$ is a singleton we simplify notation to $f_{i\to x}$). We write $|x|$ for the number of ones in x. For $i \in [n]$, the ith influence is $\mathrm{I}_i(f) = \|f_{i\to 1} - f_{i\to 0}\|_2^2$, where the norm is with respect to the implicit measure μ_p. In general, we define the influence with respect to any $S \subset [n]$ by sequentially applying the operators $f \mapsto f_{i\to 1} - f_{i\to 0}$ for all $i \in S$, as follows.

Definition 1.2. For $f\colon \{0,1\}^n \to \mathbb{R}$ and $S \subset [n]$ we let (suppressing p in the notation)

$$\mathrm{I}_S(f) = \mathbb{E}_{\mu_p}\Big[\Big(\sum_{x\in\{0,1\}^S} (-1)^{|S|-|x|} f_{S\to x}\Big)^2\Big].$$

We say f has β-*small generalised influences* if $\mathrm{I}_S[f] \le \beta\,\mathbb{E}[f^2]$ for all $S \subseteq [n]$.

The reader familiar with the KKL theorem and the invariance principle may wonder why it is necessary to introduce generalised influences rather than only considering influences (of singletons). The reason is that under the uniform measure the properties of having small influences or small generalised influences are qualitatively equivalent, but this is no longer true in the p-biased setting for small p (consider $f(x) = \frac{x_1x_2+\cdots+x_{n-1}x_n}{\|x_1x_2+\cdots+x_{n-1}x_n\|}$).

We are now ready to state our main theorem, which shows that global[4] functions are hypercontractive for a noise operator with a constant rate. Moreover, our result applies to general L^r norms and product spaces (see Section 3), but for simplicity here we just highlight the case of $(4,2)$-hypercontractivity in the cube.

Theorem 1.3. *Let $p \in \left(0, \frac{1}{2}\right]$. Suppose $f \in L^2\left(\{0,1\}^n, \mu_p\right)$ has β-small generalised influences (for p). Then $\|\mathrm{T}_{1/5} f\|_4 \le \beta^{1/4}\|f\|_2$.*

We now move on to demonstrate the power of global hypercontractivity in the contexts of isoperimetry, noise sensitivity, sharp thresholds, and invariance. We emphasise that Theorem 1.3 is the only new ingredient required for these applications, so we expect that it will have many further applications to generalising results proved via usual hypercontractivity on the cube with uniform measure.

[4]Strictly speaking, our assumption is stronger than the most natural notion of global functions: we require all generalised influences to be small, whereas a function should be considered global if it has small generalised influences $I_S(f)$ for small sets S. However, in practice, the hypercontractivity Theorem is typically applied to low-degree truncations of Boolean functions (see Section 3.1) , when there is no difference between these notions, as $I_S(f) = 0$ for large S.

1.2 Isoperimetry and influence

Stability of isoperimetric problems is a prominent open problem at the interface of Geometry, Analysis and Combinatorics. This meta-problem is to characterise sets whose boundary is close to the minimum possible given their volume; there are many specific problems obtained by giving this a precise meaning. Such results in Geometry were obtained for the classical setting of Euclidean Space by Fusco, Maggi and Pratelli [33] and for Gaussian Space by Mossel and Neeman [62].

The relevant setting for our paper is that of the cube $\{0,1\}^n$, endowed with the p-biased measure μ_p. We refer to this problem as the (p-biased) edge-isoperimetric stability problem. We identify any subset of $\{0,1\}^n$ with its characteristic Boolean function $f:\{0,1\}^n \to \{0,1\}$, and define its 'boundary' as the (total) influence[5]

$$\mathrm{I}[f] = \sum_{i=1}^{n} \mathrm{I}_i[f], \text{ where each } \mathrm{I}_i[f] = \Pr_{\mathbf{x}\sim\mu_p}[f(\mathbf{x}\oplus e_i) \neq f(\mathbf{x})],$$

i.e. the *ith influence* $\mathrm{I}_i[f]$ of f is the probability that f depends on bit i at a random input according to μ_p. (The notion of influence for real-valued functions, given in Section 1.1, coincides with this notion for Boolean-valued functions). When $p = 1/2$ the total influence corresponds to the classical combinatorial notion of edge-boundary[6].

The KKL theorem of Kahn, Kalai and Linial [43] concerns the structure of functions $f:\{0,1\}^n \to \{0,1\}$, considering the cube under the uniform measure, with variance bounded away from 0 and 1 and with total influence is upper bounded by some number K. It states that f has a coordinate with influence at least $e^{-O(K)}$. The tribes example of Ben-Or and Linial [11] shows that this is sharp.

p-biased versions

The p-biased edge-isoperimetric stability problem is somewhat understood in the *dense regime* (where $\mu_p(f)$ is bounded away from 0 and 1) especially for Boolean functions f that are *monotone* (satisfy $f(x) \leq f(y)$ whenever all $x_i \leq y_i$). Roughly speaking, most edge-isoperimetric stability results in the dense regime say that Boolean functions of small influence have some 'local' behaviour (see the seminal works of Friedgut–Kalai [31], Friedgut [28, 29], Bourgain [29, Appendix], and Hatami [38]). In particular, Bourgain (see also [64, Chapter 10]) showed that for any monotone Boolean function f with $\mu_p(f)$ bounded away from 0 and 1 and $pI[f] \leq K$ there is a set J of $O(K)$ coordinates such that $\mu_p(f_{J\to 1}) \geq \mu_p(f) + e^{-O(K^2)}$. This result is often interpreted as 'almost isoperimetric (dense) subsets of the p-biased cube must be local' or on the contrapositive as 'global functions have large total influence'. Indeed, if a restriction of a small set of coordinates can significantly boost the p-biased measure of a function, then this intuitively means that it is of a local nature.

For monotone functions, the conclusion in Bourgain's theorem is equivalent (see Section 4) to having some set J of size $O(K)$ with $\mathrm{I}_J(f) \geq e^{-O(K^2)}$. Thus Bourgain's theorem can be viewed as a p-biased analog of the KKL theorem, where influences are replaced by generalised influences. However, unlike the KKL Theorem, Bourgain's result is not sharp, and the anti-tribes example of Ben-Or and Linial only shows that the K^2 term in the exponent cannot drop below K.

As a first application of our hypercontractivity theorem we replace the term $e^{-O(K^2)}$ by the term $e^{-O(K)}$, which is sharp by Ben-Or and Linial's example, see Section 4.

[5]Everything depends on p, which we fix and suppress in our notation.

[6]For the vertex boundary, stability results showing that approximately isoperimetric sets are close to Hamming balls were obtained independently by Keevash and Long [49] and by Przykucki and Roberts [66].

Theorem 1.4. *Let $p \in \left(0, \frac{1}{2}\right]$, and let $f\colon \{0,1\}^n \to \{0,1\}$ be a monotone Boolean function with $\mu_p(f)$ bounded away from 0 and 1 and $\mathrm{I}[f] \le \frac{K}{p}$. Then there is a set J of $O(K)$ coordinates such that $\mu_p(f_{J\to 1}) \ge \mu_p(f) + e^{-O(K)}$.*

For general functions we prove a similar result, where the conclusion $\mu_p(f_{J\to 1}) \ge \mu_p(f) + e^{-O(K)}$ is replaced with $\mathrm{I}_J(f) \ge e^{-O(K)}$.

The sparse regime

On the other hand, the *sparse regime* (where we allow any value of $\mu_p(f)$) seemed out of reach of previous methods in the literature. Here Russo [67], and independently Kahn and Kalai [42], gave a proof of the p-biased isoperimetric inequality: $p\mathrm{I}[f] \ge \mu_p(f)\log_p(\mu_p(f))$ for every f. They also showed that equality holds only for the monotone sub-cubes. Kahn and Kalai posed the problem of determining the structure of monotone Boolean functions f that they called *d-optimal*, meaning that $p\mathrm{I}[f] \le d\mu_p(f)\log_p(\mu_p(f))$, i.e. functions with total influence within a certain multiplicative factor of the minimal value guaranteed by the isoperimetric inequality. The (see [42, Conjecture 4.1(a)]) states that for any constant $C > 0$ there are constants $K, \delta > 0$ such that if f is $C\log(1/p)$-optimal then there is a set J of $\le K\log\frac{1}{\mu_p(f)}$ coordinates such that $\mu_p(f_{J\to 1}) \ge (1+\delta)\mu_p(f)$.

Our variant form of the Kahn–Kalai Isoperimetric Conjecture applies to functions satisfying the slightly stronger assumption of $C\log(1/p)$-optimality with C sufficiently small; the conjecture requires an arbitrary constant C, although we note that the conjecture was previously open even for C-optimal functions! Furthermore, we compensate for our stronger hypothesis in the following result by obtaining a much stronger conclusion than that asked for by Kahn and Kalai, which is sharp up to the constant factor C.

Theorem 1.5. *Let $p \in \left(0, \frac{1}{2}\right]$, $K \ge 1$ and let f be a Boolean function with $p\mathrm{I}[f] < K\mu_p(f)$. Then there is a set J of at most CK coordinates, where C is an absolute constant, such that $\mu_p(f_{J\to 1}) \ge e^{-CK}$.*

Note that if f is $\frac{\log(1/p)}{100C}$-optimal then Theorem 1.5 applies with $K = \frac{\log_{1/p}(1/\mu_p(f))}{100C}\log(1/p) = \frac{\log(1/\mu_p(f))}{100C}$, giving a set J of size at most CK such that

$$\mu_p(f_{J\to 1}) \ge e^{-CK} = e^{-\log(1/\mu_p(f))/100} = \mu_p(f)^{0.01}.$$

1.3 Sharp thresholds

The results of Friedgut and Bourgain mentioned above also had the striking consequence that any 'global' Boolean function has a sharp threshold, which was a breakthrough in the understanding of this phenomenon, as it superceded many results for specific functions.

The sharp threshold phenomenon concerns the behaviour of $\mu_p(f_n)$ for p around the critical probability, defined as follows. Consider any sequence $f_n\colon \{0,1\}^n \to \{0,1\}$ of monotone Boolean functions. For $t \in [0,1]$ let $p_n(t) = \inf\{p : \mu_p(f_n) \ge t\}$. In particular, $p_n^c := p_n(1/2)$ is commonly known as the 'critical probability' (which we think of as small in this paper). A classical theorem of Bollobás and Thomason [16] shows that for any $\varepsilon > 0$ there is $C > 0$ such that $p_n(1-\varepsilon) \le Cp_n(\varepsilon)$. This motivates the following definition: we say that the sequence (f_n) has a *coarse threshold* if for each $\varepsilon > 0$ the length of the interval $[p_n(\varepsilon), p_n(1-\varepsilon)]$ is $\Theta(p_n^c)$, otherwise we say that it has a *sharp threshold.*

The classical approach for understanding sharp thresholds is based on the Margulis–Russo formula $\frac{d\mu_p(f)}{dp} = \mathrm{I}_{\mu_p}(f)$, see [58] and [67]. Here we note that if f has a coarse threshold, then by the Mean Value

Theorem there is a constant $\epsilon > 0$, some p with $\mu_p(f) \in (\epsilon, 1-\epsilon)$ and $p\mathrm{I}_{\mu_p}(f) = \Theta(1)$, so one can apply various results mentioned in Section 1.2. Thus Bourgain's Theorem implies that there is a set J of $O(K)$ coordinates such that $\mu_{p'}(f_{J\to 1}) \ge \mu_{p'}(f) + e^{-O(K^2)}$. While this approach is useful for studying the behaviour of f around the critical probability, it rarely gives any information regarding the location of the critical probability. Indeed, many significant papers are devoted to locating the critical probability of specific interesting functions, see e.g. the breakthroughs of Johansson, Kahn and Vu [41] and Montgomery [60].

A general result was conjectured by Kahn and Kalai for the class of Boolean functions of the form $f_n\colon \{0,1\}^{\binom{[n]}{2}} \to \{0,1\}$, whose input is a graph G and whose output is 1 if G contains a certain fixed graph H. For such functions there is a natural 'expectation heuristic' p_n^E for the critical probability, namely the least value of p such that the expected number of copies of any subgraph of H in $G(n,p)$ is at least $1/2$. Markov's inequality implies $p_n^c \ge p_n^E$. The hope of the Kahn–Kalai Threshold Conjecture (see [42, Conjecture 2.1]) is that there is a corresponding upper bound up to some multiplicative factor, i.e. that $p_n^c = O\left(p_n^E \log n\right)$. They also outlined a proof strategy based on their Isoperimetric Conjecture discussed above, which was partly the motivation for our work.

While the Isoperimetric Conjecture remains open, the Threshold Conjecture has now been resolved by Park and Pham [65], building on advances on the sunflower conjecture [5] and Talagrand's fractional version of the Threshold Conjecture [27]. Nevertheless, even given the Threshold Conjecture, it remains a challenging task to determine some specific thresholds for which it is hard to estimate the expectation threshold. For example, it is open to determine the thresholds for designs in random hypergraphs (except for the recent solution for Latin Squares and Steiner Triple Systems independently in [45, 40]).

Thus there are still potential applications for Theorem 1.5 for estimating thresholds in cases where one lacks techniques for estimating the expectation threshold. In particular, we note the following consequence, after combining with Russo's Lemma. Let f be a monotone Boolean function. We say that f is *M-global* in an interval I if for each set J of size $\le M$ and each $p \in I$ we have $\mu_p(f_{J\to 1}) \le \mu_p(f)^{0.01}$.

Theorem 1.6. *There exists an absolute constant C such that the following holds for any monotone Boolean function f with critical probability p_c and $p \le p_c$. Suppose for some $M > 0$ that f is M-global in the interval $[p, p_c]$ and that $\mu_p(f) \ge e^{-M/C}$. Then $p_c \le M^C p$.*

To see the utility of Theorem 1.6, imagine that one wants to bound the critical probability as $p_n^c \le p$, but instead of showing $\mu_p(f_n) \ge \frac{1}{2}$ one can only obtain a weaker lower bound $\mu_p(f) \ge e^{-M/C}$, where f is M-global; then one can still bound the critical probability as $p_n^c \le M^{O(1)}p$.

1.4 Noise sensitivity

Studying the effect of 'noise' on a Boolean function is a fundamental paradigm in various contexts, including hypercontractivity (as in Section 1.1) and Gaussian isoperimetry (via the invariance principle, see Section 8). Roughly speaking, a function f is 'noise sensitive' if $f(x)$ and $f(y)$ are approximately independent for a random input x and a random small perturbation y of x; an equivalent formulation (which we adopt below) is that the 'noise stability' of f is small (compared to $\mu_p(f)$). Formally, we use the following definition.

Definition 1.7. The noise stability $\mathrm{Stab}_\rho(f)$ of $f \in L^2(\{0,1\}^n, \mu_p)$ is defined by

$$\mathrm{Stab}_\rho(f) = \langle f, \mathrm{T}_\rho f\rangle = \mathbb{E}_{\mathbf{x}\sim\mu_p}\left[f(\mathbf{x})\,\mathrm{T}_\rho f(\mathbf{x})\right].$$

A sequence f_n of Boolean functions is said to be noise sensitive if for each fixed ρ we have $\mathrm{Stab}_\rho(f_n) = \mu_p(f_n)^2 + o(\mu_p(f_n))$.

Note that everything depends on p, but this will be clear from the context, so we suppress p from the notation Stab_ρ. Kahn, Kalai, and Linial [43] (see also [64, Section 9]) showed that sparse subsets of the uniform cube are noise sensitive, where we recall that the sequence (f_n) is *sparse* if $\mu_p(f_n) = o(1)$ and *dense* if $\mu_p(f_n) = \Theta(1)$.

The relationship between noise and influence in the cube under the uniform measure was further studied by Benjamini, Kalai, and Schramm [14] (with applications to percolation), who gave a complete characterisation: a sequence (f_n) of monotone dense Boolean functions is noise sensitive if and only if the sum of the squares of the influences of f_n is $o(1)$. Schramm and Steif [68] proved that any dense Boolean function on n variables that can be computed by an algorithm that reads $o(n)$ of the input bits is noise sensitive. Their result had the striking application that the set of exceptional times in dynamical critical site percolation on the triangular lattice, in which an infinite cluster exists, is of Hausdorff dimension in the interval $\left[\frac{1}{6}, \frac{31}{36}\right]$. Ever since, noise sensitivity was considered in many other contexts (see e.g. the recent results and open problems of Lubetzky–Steif [57] and Benjamini-Brieussel [13]).

The p-biased setting

In contrast to the uniform setting, in the p-biased setting for small p it is no longer true that sparse sets are noise sensitive (e.g. consider dictators). Our main contribution to the theory of noise sensitivity is showing that 'global' sparse sets are noise sensitive. Formally, we say that a sequence f_n of sparse Boolean functions is *weakly global* if for any $\varepsilon, C > 0$ there is $n_0 > 0$ so that $\mu_p\left((f_n)_{J\to 1}\right) < \varepsilon$ for all $n > n_0$ and J of size at most C.

Theorem 1.8. *Any weakly global sequence of Boolean functions is noise sensitive.*

Besides being of interest in its own right, noise sensitivity provides an alternative approach for proving results on sharp thresholds. In particular, we obtain the following consequence which will underpin our combinatorial applications in the companion paper on Extremal Combinatorics.

Theorem 1.9. *For any $\zeta > 0$ there is $C_0 > 1$ so that for any $\varepsilon, p, q \in (0, 1/2)$ with $q \geq (1+\zeta)p$ and $C > C_0$, writing $r = C \log \varepsilon^{-1}$ and $\delta = C^{-r}$, any monotone (r, δ)-global Boolean function f whose p-biased measure is at most δ satisfies $\mu_q(f) \geq \varepsilon^{-1}\mu_p(f)$.*

1.5 The Invariance Principle

Besides the applications of Theorem 1.3 to isoperimetry, sharp thresholds and noise sensitivity discussed above, in Section 8 we will also generalise the Invariance Principle of Mossel, O'Donnell and Oleszkiewicz [63] to the p-biased setting: we show that if a low degree function on the p-biased cube is global (has small generalised influences) then it is close in distribution to a low degree function on Gaussian space.

We defer a precise statement of the Invariance Principle to Section 8; here we instead highlight the following application to a variant of the 'Majority is Stablest' Theorem of Mossel, O'Donnell and Oleszkiewicz [63] (see also [61]). We need the following notation for its statement. The p-biased *α-Hamming ball* is the function $H_\alpha : \{0,1\}^n \to \{0,1\}$ with $H_\alpha(x) = 1$ if and only if $|\{i : x_i = 1\}| \geq t$, where $t \in \mathbb{N}$ is chosen to minimise $|\alpha - \mu_p(H_\alpha)|$.

Corollary 1.10. *For each $\epsilon > 0$, there exists $\delta > 0$, such that the following holds. Let $\rho \in (\epsilon, 1-\epsilon)$, let $n > \delta^{-1}$, and let $f, g \in L^2(\{0,1\}^n, \mu_p)$. Suppose that $\mathrm{I}_S[f] \leq \delta$ and that $\mathrm{I}_S[g] \leq \delta$ for each set S of at most δ^{-1} coordinates. Then*

$$\langle \mathrm{T}_\rho f, g\rangle \leq \left\langle \mathrm{T}_\rho H_{\mu_p(f)}, H_{\mu_p(g)}\right\rangle + \epsilon.$$

1.6 Further applications of global hypercontractivity

Here we survey some additional applications of global hypercontractivity, mostly subsequent to the 2019 version of this paper.

1. *Exotic settings:* Noise sensitivity of sparse sets is related to small-set expansion on graphs, which has found many applications in Computer Science. Here the interpretation of Theorem 1.8 is that although not all small sets in the p-biased cube expand, global small sets do expand. Results of a similar nature were proved for the Grassmann graph [54] and the Johnson graph [53]. The former result was essential in the proof of the 2-to-2 Games Conjecture, a prominent problem in the field of hardness of approximation. In subsequent works [24, 25, 47] hypercontractive results for global functions are proven for various domains by reducing to the p-biased cube and using Theorem 1.3; these methods are strong enough to establish global hypercontractivity results for general product domains, as well as domains that are not product but instead are product-like (such as the Johnson graph, the multi-slice and the symmetric group). In [24], they were also shown to be powerful enough to give an alternative proof for global-hypercontractivity style results over the Grassmann graph [54].

2. *Extremal Combinatorics:* The junta method, introduced by Dinur and Friedgut [20] and further developed by Keller and Lifshitz [51], is a powerful tool for solving problems in Extremal Combinatorics via the sharp threshold phenomenon. Specifically, it is useful for the study of the Turán problem for hypergraphs, where one asks how large can a k-uniform hypergraph on n vertices be if it does not contain a copy of a given hypergraph H. This method was applied in [51] to resolve many such questions for a wide class of hypergraphs called *expanded hypergraphs* in which the edge uniformity can be linear in n, although the number of edges in H is fixed.

   In a companion paper [46], we apply the sharp threshold technology developed in the current paper to the regime where the number of edges of H can grow with n, thus settling many cases of the Huang–Loh–Sudakov conjecture [39] on cross matchings in uniform hypergraphs and the Füredi–Jiang–Seiver conjecture [32] on path expansions.

   In another companion paper [47], we apply the theory of global hypercontractivity to extremal problems for codes with forbidden intersections. What is the largest subset $A \subseteq [m]^n$ in which no two vectors agree on exactly t coordinates? We solve this question for any $m > 2$ and $n > n_0(t)$, thereby strengthening the more classical version of the question regarding intersecting codes due to Ahlswede and Khachatrian [4] and independently by Frankl and Tokushige [26].

3. *Product-free Sets:* A subset A of a group G is called product-free if it contains no solutions to the equation $ab = c$. What is the largest size of a product-free set in a given group G? This question was posed in 1985 by Babai and Sós [6], both for general groups and with specific attention to the alternating group A_n. The problem for A_n has recently been resolved in [48]. Moreover, the structure of families that achieve the maximum size is exactly determined, as well as stability results that give structure for any sizable product-free set in A_n. Besides building on earlier work by Gowers [34] and Eberhard [22], a key component in the proof is a global hypercontractivity result for the symmetric group S_n, established in [25] and directly motivated by the current work.

4. *Error Correcting Codes:* The Reed-Muller code is one of the most useful codes in Theoretical Computer Science. A fundamental question on its local testability concerns the effectiveness of the so-called 'flat tester', which queries a random affine subspace and accepts if the restriction forms a degree d polynomial. Suppose $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$, $d \in \mathbb{N}$, $\varepsilon > 0$ is small and with probability at least $1 - \varepsilon$

for a random affine $(d+1)$-dimensional subspace A the restriction $f|_A$ is a polynomial of degree d. Must f be close to a degree d polynomial?

A new approach for deriving an optimal testing result is derived in [44], using global hypercontractivity over the affine Grassmann graph as a key component. The new method is able to recover the results of [15] and improves upon the results for larger fields [37]. Also, it is more general and thus is more likely to work for a richer class of codes (the technique of [15] is very specific to the Reed-Muller code, and also has tower-type dependency on the field size due to use of the Density Hales-Jewett Theorem).

5. *Hypercontractivity over High Dimensional Expanders:* High-dimensional expanders can be thought of as sparse analogues of the Johnson graph, so morally speaking one expects analogous results in the setting, although these are often much harder to prove. Motivated by some of the ideas herein (and in subsequent works), new hypercontractive estimates have been proved in the setting of high-dimensional expanders [36, 9]. These have potential to be useful in this emerging field of study, e.g. as a new tool to prove mixing results.

6. *Algorithms for Unique-Games over Specialized Instances:* As discussed earlier, global hypercontractivity has its roots in the proof of the 2-to-2-Games Conjecture, wherein a baby form of global hypercontractivity plays a crucial role. Subsequent study regarding the more well-known sibling of 2-to-2-Games, known as the Unique-Games Conjecture, has mostly focused on whether global hypercontractivity can also be used to eliminate certain avenues towards a proof of the Unique-Games Conjecture. In particular, in [7, 8] the authors show that global hypercontractivity can be used in the realm of Sum-of-Squares algorithms to show that Unique-Games instances over specialized graphs, such as the Johnson graph and the Grassmann graph, can be solved efficiently.

Organization The organisation of this paper is as follows. After introducing some background on Fourier analysis on the cube in the next section, we prove Theorem 1.3 in Section 3. In Section 4 we establish the equivalence between the two notions of globalness referred to above, namely control of generalised influences and insensitivitity of the measure under restriction to a small set of coordinates. Section 5 concerns the total influence of global functions, and includes the proofs of our stability results for the isoperimetric inequality (Theorems 1.4 and 1.5) and our first sharp threshold result (Theorem 1.6). In Section 6 we prove our result on noise sensitivity and apply this to deduce an alternative sharp threshold result. Section 7 generalises our hypercontractivity result in two directions: we consider general norms and general product spaces. In Section 8 we prove our p-biased version of the Invariance Principle and sketch its application to a variant of the 'Majority is Stablest' theorem and a sharp threshold result for almost monotone functions. We end with some concluding remarks.

2 Notations

Here we summarise some notation and basic properties of Fourier analysis on the cube. We fix $p \in (0,1)$ and suppress it in much of our notation, i.e. we consider $\{0,1\}^n$ to be equipped with the p-biased measure μ_p, unless otherwise stated. We let $\sigma = \sqrt{p(1-p)}$ (the standard deviation of a p-biased bit). For each $i \in [n]$ we define $\chi_i \colon \{0,1\}^n \to \mathbb{R}$ by $\chi_i(x) = \frac{x_i - p}{\sigma}$ (so χ_i has mean 0 and variance 1). We use the orthonormal Fourier basis $\{\chi_S\}_{S\subset[n]}$ of $L^2(\{0,1\}^n, \mu_p)$, where each $\chi_S := \prod_{i\in S}\chi_i$. Any $f : \{0,1\}^n \to \mathbb{R}$ has a unique expression $f = \sum_{S\subset[n]} \hat{f}(S)\chi_S$ where $\{\hat{f}(S)\}_{S\subset[n]}$ are the *p-biased Fourier coefficients* *of* f. Orthonormality gives the Plancherel identity $\langle f,g\rangle = \sum_{S\subset[n]} \hat{f}(S)\hat{g}(S)$. In particular, we have the Parseval identity $\mathbb{E}[f^2] = \|f\|_2^2 = \langle f,f\rangle = \sum_{S\subset[n]} \hat{f}(S)^2$. For $\mathcal{F}\subset\{0,1\}^n$ we define the $\mathcal{F}$-truncation $f^{\mathcal{F}} = \sum_{S\in\mathcal{F}} \hat{f}(S)\chi_S$. Our truncations will always be according to some degree threshold r, for which we write $f^{\le r} = \sum_{|S|\le r} \hat{f}(S)\chi_S$.

For $i\in[n]$, the *i-derivative* f_i and *i-influence* $\mathrm{I}_i(f)$ of f are

$$f_i = \mathrm{D}_i\left[f\right] = \sigma\big(f_{i\to 1} - f_{i\to 0}\big) = \sum_{S:i\in S} \hat{f}\left(S\right)\chi_{S\setminus\{i\}}, \text{ and}$$
$$\mathrm{I}_i(f) = \|f_{i\to 1} - f_{i\to 0}\|_2^2 = \sigma^{-2}\mathbb{E}[f_i^2] = \tfrac{1}{p(1-p)}\sum_{S:i\in S}\hat{f}(S)^2.$$

The *influence* of f is

$$\mathrm{I}(f) = \sum_i \mathrm{I}_i(f) = (p(1-p))^{-1}\sum_S |S|\hat{f}(S)^2. \tag{2.1}$$

In general, for $S\subset[n]$, the *S-derivative* of f is obtained from f by sequentially applying D_i for each $i\in S$, i.e.

$$\mathrm{D}_S(f) = \sigma^{|S|}\sum_{x\in\{0,1\}^S}(-1)^{|S|-|x|}f_{S\to x} = \sum_{T:S\subset T}\hat{f}(T)\chi_{T\setminus S}.$$

The *S-influence* of f (as in Definition 1.2) is

$$\mathrm{I}_S(f) = \sigma^{-2|S|}\|\mathrm{D}_S\left(f\right)\|_2^2 = \sigma^{-2|S|}\sum_{E:S\subset E}\hat{f}(E)^2. \tag{2.2}$$

Recalling that a function f has α-small generalised influences if $\mathrm{I}_S(f)\le\alpha\mathbb{E}[f^2]$ for all $S\subset[n]$, we see that this is equivalent to $\mathbb{E}[\mathrm{D}_S\left(f\right)^2]\le\alpha\sigma^{2|S|}\mathbb{E}[f^2]$ for all $S\subset[n]$.

3 Hypercontractivity of functions with small generalised influences

In this section we prove our hypercontractive inequality (Theorem 1.3), which is the fundamental result that underpins all of the results in this paper.

The idea of the proof is to reduce hypercontractivity in μ_p to hypercontractivity in $\mu_{1/2}$ via the 'replacement method' (the idea of Lindeberg's proof of the Central Limit Theorem, and of the proof of Mossel, O'Donnell and Oleszkiewicz [63] of the invariance principle). Throughout this section we fix $f:\{0,1\}^n\to\mathbb{R}$ and express f in the p-biased Fourier basis as $\sum_S \hat{f}(S)\chi_S^p$, where $\chi_S^p = \prod_{i\in S}\chi_i^p$ and $\chi_i^p(x) = \frac{x_i-p}{\sigma}$ (the same notation as above, except that we introduce the superscript p to distinguish the p-biased and uniform settings).

For $0\le t\le n$ we define $f_t = \sum_S \hat{f}(S)\chi_S^t$, where

$$\chi_S^t = \prod_{i\in S\cap[t]}\chi_i^{1/2}(x)\prod_{i\in S\setminus[t]}\chi_i^p(x)\in L^2(\Omega_t), \text{ with } \Omega_t = (\{0,1\}^{[t]},\mu_{1/2})\times(\{0,1\}^{[n]\setminus[t]},\mu_p).$$

Thus f_t interpolates from $f_0 = f\in L^2(\{0,1\}^n,\mu_p)$ to $f_n = \sum_S \hat{f}(S)\chi_S^{1/2}\in L^2(\{0,1\}^n,\mu_{1/2})$. As $\{\chi_S^t : S\subset[n]\}$ is an orthonormal basis we have $\|f_t\|_2 = \|f\|_2$ for all t.

We also define noise operators $\mathrm{T}^t_{\rho',\rho}$ on $L^2(\Omega_t)$ by $\mathrm{T}^t_{\rho',\rho}(g)(\mathbf{x}) = \mathbb{E}_{\mathbf{y}\sim N_{\rho',\rho}(\mathbf{x})}[f(\mathbf{y})]$, where to sample $\mathbf{y}$ from $N_{\rho',\rho}(\mathbf{x})$, for $i \le t$ we let $y_i = x_i$ with probability ρ' or otherwise we resample y_i from $\mu_{1/2}$, and for $i > t$ we let $y_i = x_i$ with probability ρ or otherwise we resample y_i from μ_p. Thus $\mathrm{T}^t_{\rho',\rho}$ interpolates from $\mathrm{T}^0_{\rho',\rho} = \mathrm{T}_\rho$ (for μ_p) to $\mathrm{T}^n_{\rho',\rho} = \mathrm{T}_{\rho'}$ (for $\mu_{1/2}$).

We record the following estimate for 4-norms of p-biased characters:

$$\lambda := \mathbb{E}[(\chi^p_i)^4] = \sigma^{-4}(p(1-p)^4 + (1-p)p^4) = \sigma^{-2}((1-p)^3 + p^3) \le \sigma^{-2}.$$

The core of our argument by replacement is the following lemma which controls the evolution of $\mathbb{E}[(\mathrm{T}^t_{2\rho,\rho} f_t)^4] = \|\mathrm{T}^t_{2\rho,\rho} f_t\|_4^4$ for $0 \le t \le n$. Note that expectations are with respect to Ω_{t-1} on the left-hand-side and Ω_t on the right-hand-side.

Lemma 3.1. $\mathbb{E}[(\mathrm{T}^{t-1}_{2\rho,\rho} f_{t-1})^4] \le \mathbb{E}[(\mathrm{T}^t_{2\rho,\rho} f_t)^4] + 3\lambda\rho^4\mathbb{E}[(\mathrm{T}^t_{2\rho,\rho}((\mathrm{D}_t f)_t))^4]$.

Proof. We write

$$\begin{aligned} &f_t = \chi^{1/2}_t g + h \quad\text{and}\quad f_{t-1} = \chi^p_t g + h, \quad\text{where} \\ &g = (\mathrm{D}_t f)_t = \sum_{S:t\in S} \hat{f}(S)\chi^t_{S\setminus\{t\}} = \sum_{S:t\in S} \hat{f}(S)\chi^{t-1}_{S\setminus\{t\}} = (\mathrm{D}_t f)_{t-1}, \quad\text{and} \\ &h = \mathbb{E}_{x_t\sim\mu_{1/2}} f_t = \sum_{S:t\notin S} \hat{f}(S)\chi^t_S = \sum_{S:t\notin S} \hat{f}(S)\chi^{t-1}_S = \mathbb{E}_{x_t\sim\mu_p} f_{t-1}. \end{aligned}$$

We also write

$$\begin{gathered} \mathrm{T}^t_{2\rho,\rho} f_t = 2\rho\chi^{1/2}_t d + e \quad\text{and}\quad \mathrm{T}^{t-1}_{2\rho,\rho} f_{t-1} = \rho\chi^p_t d + e, \quad\text{where} \\ d = \mathrm{T}^t_{2\rho,\rho} g = \mathrm{T}^{t-1}_{2\rho,\rho} g \quad\text{and}\quad e = \mathrm{T}^t_{2\rho,\rho} h = \mathrm{T}^{t-1}_{2\rho,\rho} h. \end{gathered}$$

We can calculate the expectations in the statement of the lemma by conditioning on all coordinates other than x_t, i.e. $\mathbb{E}_{\mathbf{x}}[\cdot] = \mathbb{E}_{\mathbf{x}'}[\mathbb{E}_{x_t}[\cdot \mid \mathbf{x}']]$ where $\mathbf{x}'$ is obtained from $\mathbf{x} = (x_1, \dots, x_n)$ by removing x_t. It therefore suffices to establish the required inequality for each fixed $\mathbf{x}'$ with expectations over the choice of x_t; thus we can treat d and e as constants, and it suffices to show

$$\mathbb{E}_{x_t\sim\mu_p}[(\rho d\chi^p_t(x_t) + e)^4] \le \mathbb{E}_{x_t\sim\mu_{1/2}}[(2\rho d\chi^{1/2}_t(x_t) + e)^4] + 3\lambda\rho^4 d^4. \tag{3.1}$$

As χ^p_t has mean 0, we can expand the left hand side of (3.1) as

$$(\rho d)^4\mathbb{E}[(\chi^p_t)^4] + 4e(\rho d)^3\mathbb{E}[(\chi^p_t)^3] + 6e^2(\rho d)^2\mathbb{E}[(\chi^p_t)^2] + e^4 \le 3\lambda(d\rho)^4 + 8(de\rho)^2 + e^4,$$

where we bound the second term using Cauchy-Schwarz and then AM-GM by

$$4\cdot\mathbb{E}[(d\rho\chi^p_t)^4]^{1/2}\cdot\mathbb{E}[(de\rho\chi^p_t)^2]^{1/2} \le 2\left(\mathbb{E}[(d\rho\chi^p_t)^4] + \mathbb{E}[(de\rho\chi^p_t)^2]\right) = 2(\lambda(d\rho)^4 + (de\rho)^2).$$

Similarly, as $\mathbb{E}[\chi^{1/2}_t] = \mathbb{E}[(\chi^{1/2}_t)^3] = 0$, we can expand the first term on the right hand side of (3.1) as

$$(2\rho d)^4\mathbb{E}[(\chi^{1/2}_t)^4] + 6e^2(2\rho d)^2\mathbb{E}[(\chi^{1/2}_t)^2] + e^4 = (2\rho d)^4 + 6(2\rho de)^2 + e^4 \ge 8(de\rho)^2 + e^4.$$

The lemma follows. $\square$

Now we apply the previous lemma inductively to prove the following estimate.

Lemma 3.2. $\|\mathrm{T}^i_{2\rho,\rho} f_i\|_4^4 \le \sum_{S\subset[n]\setminus[i]} (3\lambda\rho^4)^{|S|} \|\mathrm{T}^n_{2\rho,\rho}((\mathrm{D}_S f)_n)\|_4^4$ *for all* $0 \le i \le n$.

Proof. We prove the inequality by induction on $n-i$ simultaneously for all functions f. If $n=i$ then equality holds trivially. Now suppose that $i<n$. By Lemma 3.1 with $t=i+1$, and the induction hypothesis applied to f and $\mathrm{D}_t f$ with i replaced by t, we have

$$\begin{aligned}\|\mathrm{T}^i_{2\rho,\rho} f_i\|_4^4 &\le \|\mathrm{T}^t_{2\rho,\rho} f_t\|_4^4 + 3\lambda\rho^4 \|\mathrm{T}^t_{2\rho,\rho}((\mathrm{D}_t f)_t)\|_4^4 \\ &\le \sum_{S\subset[n]\setminus[t]} (3\lambda\rho^4)^{|S|} \|\mathrm{T}^n_{2\rho,\rho}((\mathrm{D}_S f)_n)\|_4^4 + 3\lambda\rho^4 \sum_{S\subset[n]\setminus[t]} (3\lambda\rho^4)^{|S|} \|\mathrm{T}^n_{2\rho,\rho}((\mathrm{D}_S \mathrm{D}_t f)_n)\|_4^4 \\ &= \sum_{S\subset[n]\setminus[i]} (3\lambda\rho^4)^{|S|} \|\mathrm{T}^n_{2\rho,\rho}((\mathrm{D}_S f)_n)\|_4^4. \qquad \square\end{aligned}$$

In particular, recalling that $\mathrm{T}^0_{2\rho,\rho} = \mathrm{T}_\rho$ on μ_p and $\mathrm{T}^n_{2\rho,\rho} = \mathrm{T}_{2\rho}$ on $\mu_{1/2}$, the case $i=0$ of Lemma 3.2 is as follows.

Proposition 3.3. $\|\mathrm{T}_\rho f\|_4^4 \le \sum_{S\subset[n]} (3\lambda\rho^4)^{|S|} \|\mathrm{T}_{2\rho}((\mathrm{D}_S f)_n)\|_4^4$.

The 4-norms on the right hand side of Proposition 3.3 are with respect to the uniform measure $\mu_{1/2}$, where we can apply standard hypercontractivity (the 'Beckner-Bonami Lemma') for $\rho \le 1/2\sqrt{3}$ to obtain $\|\mathrm{T}_{2\rho}((\mathrm{D}_S f)_n)\|_4^4 \le \|(\mathrm{D}_S f)_n\|_2^4 = \|\mathrm{D}_S f\|_2^4 = \sigma^{4|S|} \mathrm{I}_S[f]^2$. Recalling that $\lambda \le \sigma^{-2}$, we deduce the following bound for $\|\mathrm{T}_\rho f\|_4^4$ in terms of the generalised influences of f.

Theorem 3.4. *If* $\rho \le 1/\sqrt{12}$ *then* $\|\mathrm{T}_\rho f\|_4^4 \le \sum_{S\subset[n]} (3\lambda\rho^4)^{|S|} \|\mathrm{D}_S f\|_2^4 \le \sum_{S\subset[n]} (3\sigma^2\rho^4)^{|S|} \mathrm{I}_S[f]^2$.

Now we deduce our hypercontractivity inequality. It is convenient to prove the following slightly stronger statement, which implies Theorem 1.3 using $\|\mathrm{D}_S f\|_2^2 = \sigma^{2|S|}\mathrm{I}_S[f] \le \lambda^{-|S|}\mathrm{I}_S[f]$ and $\|\mathrm{T}_{1/5} f\|_4 \le \|\mathrm{T}_{1/\sqrt{24}} f\|_4$ (any T_ρ is a contraction in L^p for any $p \ge 1$).

Theorem 3.5. *Let* $f \in L^2(\{0,1\}^n, \mu_p)$ *with all* $\|\mathrm{D}_S f\|_2^2 \le \beta\lambda^{-|S|}\mathbb{E}[f^2]$. *Then* $\|\mathrm{T}_{1/\sqrt{24}} f\|_4 \le \beta^{1/4}\|f\|_2$.

Proof. By Theorem 3.4 applied to $\mathrm{T}_{1/\sqrt{2}} f$ with $\rho = 1/\sqrt{12}$ we have

$$\|\mathrm{T}_{1/\sqrt{24}} f\|_4^4 \le \sum_{S\subset[n]} (3\lambda\rho^4)^{|S|} \|\mathrm{D}_S \mathrm{T}_{1/\sqrt{2}} f\|_2^4.$$

As $\|\mathrm{D}_S \mathrm{T}_{1/\sqrt{2}} f\|_2^2 = \sum_{E:S\subset E} 2^{-|E|}\hat{f}(E)^2 \le \sum_{E:S\subset E} \hat{f}(E)^2 = \|\mathrm{D}_S f\|_2^2 \le \beta\lambda^{-|S|}\mathbb{E}[f^2]$ we deduce

$$\|\mathrm{T}_{1/\sqrt{24}} f\|_4^4 \le \sum_{S\subset[n]} \sum_{E:S\subset E} \beta\mathbb{E}[f^2] 2^{-|E|}\hat{f}(E)^2 = \beta\mathbb{E}[f^2]\sum_E \hat{f}(E)^2 = \beta\|f\|_2^4. \qquad \square$$

3.1 Hypercontractivity in practice

We will mostly use the following application of the hypercontractivity theorem.

Lemma 3.6. *Let* f *be a function of degree* r. *Suppose that* $\mathrm{I}_S[f] \le \delta$ *for all* $|S| \le r$. *Then*

$$\|f\|_4 \le 5^r \delta^{\frac{1}{4}} \|f\|_2^{1/2}.$$

The proof uses the following lemma, which is immediate from the Fourier expression in (2.2).

Lemma 3.7. $\mathrm{I}_S[f^{\le r}] \le \mathrm{I}_S[f]$ *for all* $S \subset [n]$ *and* $\mathrm{I}_S[f^{\le r}] = 0$ *if* $|S| > r$.

Proof of Lemma 3.6. Write $f = \mathrm{T}_{1/5}(h)$, where $h = \sum_{|T| \le r} 5^{|T|}\hat{f}(T)\chi_T$. We will bound the 4-norm of f by applying Theorem 1.3 to h, so we need to bound the generalised influences of h.

By Lemma 3.7, for $S \subset [n]$ we have $\mathrm{I}_S[h] = 0$ if $|S| > r$. For $|S| \le r$, we have

$$\mathrm{I}_S[h] = \sigma^{-2|S|} \sum_{T: S \subset T, |T| \le r} 5^{2|T|}\hat{f}(T)^2 \le 5^{2r}\mathrm{I}_S[f] \le 5^{2r}\delta = \alpha\|h\|_2^2,$$

where $\alpha = 5^{2r}\delta/\|h\|_2^2$. By Theorem 1.3, we have

$$\|f\|_4 = \|\mathrm{T}_{1/5}h\|_4 \le \alpha^{\frac{1}{4}}\|h\|_2 = 5^{r/2}\delta^{\frac{1}{4}}\sqrt{\|h\|_2} \le 5^r\delta^{\frac{1}{4}}\sqrt{\|f\|_2}.$$

In the final inequality we used $\|h\|_2 \le 5^r\|f\|_2$, which follows from Parseval. $\square$

4 Characterising global functions

Above we have introduced two notions of what it means for a Boolean function f to be global. The first globalness condition, which appears e.g. in Theorem 1.4, is that the measure of f is not sensitive to restrictions to small sets of coordinates. The second condition is a bound on generalised influences $I_S(f)$ for small sets S. In this section we show that we can move freely between these notions for two classes of Boolean functions: namely sparse ones and monotone ones.

Throughout we assume $p \le 1/2$, which does not involve any loss in generality in our main results; indeed, if $p > 1/2$ we can consider the dual $f^*(x) = 1 - f(1-x)$ of any Boolean function f, for which $\mu_{1-p}(f^*) = 1 - \mu_p(f)$ and $\mathrm{I}_{\mu_{1-p}}(f^*) = \mathrm{I}_{\mu_p}(f)$.

We start by formalising our first notion of globalness.

Definition 4.1. We say that a Boolean function f is (r,δ)-*global* if $\mu_p\left(f_{J\to 1}\right) \le \mu_p\left(f\right) + \delta$ for each set J of size at most r.

We remark that Definition 4.1 is a rather weak notion of globalness, so it is quite surprising that it suffices for Theorems 1.5 and 1.8, where one might have expected to need the stricter notion that $\mu_p(f_{J\to 1})$ is close to $\mu_p(f)$.

The following lemma shows that if a sparse Boolean function is global in the sense of Definition 4.1 then it has small generalised influences.

Lemma 4.2. *Suppose that* $f : \{0,1\}^n \to \{0,1\}$ *is an* (r,δ)*-global Boolean function with* $\mu_p(f) \le \delta$. *Then* $\mathrm{I}_S\left(f^{\le r}\right) \le \mathrm{I}_S\left(f\right) \le 8^r\delta$ *for all* $S \subset [n]$ *with* $|S| \le r$.

Proof. The first inequality is from Lemma 3.7. Next, we estimate

$$\sqrt{\mathrm{I}_S\left(f\right)} = \left\| \sum_{x \in \{0,1\}^S} (-1)^{|S|-|x|} f_{S\to x} \right\|_2 \le \sum_{x \in \{0,1\}^S} \|f_{S\to x}\|_2 = \sum_{x \in \{0,1\}^S} \sqrt{\mu_p(f_{S\to x})}. \tag{4.1}$$

Next we fix $x \in \{0,1\}^S$ and claim that $\mu_p(f_{S\to x}) \le 2^r\delta$. By substituting this bound in (4.1) we see that this suffices to complete the proof. Let T be the set of all $i \in S$ such that $x_i = 1$. Since f is nonnegative, we have $\mu_p(f_{T\to 1}) \ge (1-p)^{|S\setminus T|}\mu_p(f_{S\to x})$. As f is (r,δ)-global and $\mu_p(f) \le \delta$, we have $\mu_p\left(f_{T\to 1}\right) \le 2\delta$, so $\mu_p(f_{S\to x}) \le (1-p)^{|T|-r}2\delta \le 2^r\delta$, where for the last inequality we can assume $T \ne \emptyset$, as $\mu_p\left(f_{\emptyset\to 1}\right) = \mu_p(f) \le \delta \le 2^r\delta$. This completes the proof. $\square$

Next we show an analogue of the previous lemma replacing the assumption that f is sparse by the assumption that f is monotone.

Lemma 4.3. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a monotone Boolean (r,δ)-global function. Then $\mathrm{I}_S[f] \le 8^r\delta$ for every nonempty S of size at most r.*

The proof is based on the following lemma showing that globalness is inherited (with weaker parameters) under restriction of a coordinate.

Lemma 4.4. *Suppose that f is a monotone (r,δ)-global function. Then for each i:*

1. *$f_{i\to 1}$ is $(r-1,\delta)$-global,*
2. *$\mu_p(f_{i\to 0}) \ge \mu_p(f) - \frac{p\delta}{1-p}$,*
3. *$f_{i\to 0}$ is $\left(r-1,\frac{\delta}{1-p}\right)$-global.*

Proof. To see (1), note that for any J with $|J| \le r-1$ we have $\mu_p((f_{i\to 1})_{J\to 1}) = \mu_p(f_{J\cup\{i\}\to 1}) \le \mu_p(f)+\delta \le \mu_p(f_{i\to 1})+\delta$, where the last inequality holds as f is monotone. Statement (2) follows from the upper bound $\mu_p(f_{i\to 1}) \le \mu_p(f)+\delta$ and $\mu_p(f_{i\to 0}) = \frac{\mu_p(f)-p\mu_p(f_{i\to 1})}{(1-p)}$.

For (3), we note that by monotonicity $\mu_p\left((f_{i\to 0})_{S\to 1}\right) \le \mu_p\left(f_{\{i\}\cup S\to 1}\right)$. As f is (r,δ)-global,

$$\mu_p\left(f_{S\cup\{i\}\to 1}\right) \le \mu_p(f)+\delta \le \mu_p(f_{i\to 0})+\delta+\frac{p\delta}{1-p} = \mu_p(f_{i\to 0})+\frac{\delta}{1-p},$$

using (2). Hence, $f_{i\to 0}$ is $\left(r,\frac{\delta}{1-p}\right)$-global. □

Proof of Lemma 4.3. We argue by induction on r. In the case where $r=1$, Lemma 4.4 and monotonicity of f imply (using $p \le 1/2$)

$$\mathrm{I}_i(f) = \mu_p(f_{i\to 1}) - \mu_p(f_{i\to 0}) \le \delta + \frac{p\delta}{1-p} \le 2\delta.$$

Now we bound $\mathrm{I}_{S\cup\{i\}}(f)$ for $r>1$ and S of size $r-1$ with $i \notin S$.

Note that $\mathrm{D}_{S\cup\{i\}}(f) = \mathrm{D}_S[\mathrm{D}_i(f)]$. By the triangle inequality, we have

$$\sqrt{\mathrm{I}_{S\cup\{i\}}(f)} = \sigma^{-r}\|\mathrm{D}_{S\cup\{i\}}(f)\|_2 = \sigma^{1-r}\|\mathrm{D}_S(f_{i\to 1}) - \mathrm{D}_S(f_{i\to 0})\|_2 \le \sqrt{\mathrm{I}_S[f_{i\to 1}]} + \sqrt{\mathrm{I}_S[f_{i\to 0}]}.$$

By the induction hypothesis and Lemma 4.4 the right hand side is at most

$$\sqrt{8^{r-1}\delta} + \sqrt{8^{r-1}2\delta} \le \sqrt{8^r\delta}.$$

Taking squares, we obtain $\mathrm{I}_{S\cup\{i\}}(f) \le 8^r\delta$. □

We conclude this section by showing the converse direction of the equivalence between our two notions of globalness, i.e. that if the generalised influences of a function f are small then f is global in the sense of its measure being insensitive to restrictions to small sets. (We will not use the lemma in the sequel but include the proof for completeness.)

Lemma 4.5. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function and let $r > 0$. Suppose that $\mathrm{I}_S[f] \le \delta$ for each nonempty set S of at most r coordinates. Then f is $(r, 9^r\delta)$-global.*

Proof. To facilitate a proof by induction on r we prove the slightly stronger statement that f is $(r, \sum_{i=1}^{r} 9^{i-1}\delta)$-global. Suppose first that $r = 1$. Our goal is to show that if $\mathrm{I}_i[f] \le \delta$, then $\mu_p(f_{i\to 1}) - \mu_p(f_{i\to 0}) \le \delta$, and indeed,

$$\mu_p(f_{i\to 1}) - \mu_p(f_{i\to 0}) \le \Pr[f_{i\to 1} \ne f_{i\to 0}] = \|f_{i\to 1} - f_{i\to 0}\|_2^2 = \mathrm{I}_i[f] \le \delta.$$

Now suppose that $r > 1$ and that the lemma holds with $r-1$ in place of r. The lemma will follow once we show that for all i and all nonempty sets S of size at most $r-1$, we have $\mathrm{I}_S[f_{i\to 1}] \le 9\delta$. Indeed, the induction hypothesis and the $n = 1$ case will imply that for each set S of size at most r and each $i \in S$ we have $\mu_p(f_{S\to 1}) \le \mu_p(f_{i\to 1}) + \sum_{i=1}^{r-1} 9^{i-1}\cdot 9\delta \le \mu_p(f) + \sum_{i=1}^{r} 9^{i-1}\delta$.

We now turn to showing the desired upper bound on the generalised influences of $f_{i\to 1}$. Let S be a set of size at most $r-1$. Recall that $\mathrm{I}_S[f_{i\to 1}] = \sigma^{-2|S|}\|\mathrm{D}_S[f_{i\to 1}]\|_2^2$. We may assume that $i \notin S$ for otherwise the generalised influence $\mathrm{I}_S[f_{i\to 1}]$ is 0. We make two observations. Firstly, we have

$$\mathrm{D}_{S\cup\{i\}}[f] = \sigma(\mathrm{D}_S[f_{i\to 1}] - \mathrm{D}_S[f_{i\to 0}]).$$

Secondly, conditioning on the output of the coordinate i we have

$$\|\mathrm{D}_S[f]\|_2^2 = p\|\mathrm{D}_S[f_{i\to 1}]\|_2^2 + (1-p)\|\mathrm{D}_S[f_{i\to 0}]\|_2^2,$$

which implies $\|\mathrm{D}_S[f_{i\to 0}]\|_2 \le \sqrt{2}\|\mathrm{D}_S[f]\|_2$. We may now apply the triangle inequality on the first observation and use the second observation to obtain

$$\begin{aligned}
\sqrt{\mathrm{I}_S[f_{i\to 1}]} = \sigma^{-|S|}\|\mathrm{D}_S[f_{i\to 1}]\|_2 &= \sigma^{-|S|}\|\sigma^{-1}\mathrm{D}_{S\cup\{i\}}[f] + \mathrm{D}_S[f_{i\to 0}]\|_2 \\
&\le \sigma^{-(|S|+1)}\|\mathrm{D}_{S\cup\{i\}}[f]\|_2 + \sigma^{-|S|}\|\mathrm{D}_S[f_{i\to 0}]\|_2 \\
&\le \sigma^{-(|S|+1)}\|\mathrm{D}_{S\cup\{i\}}[f]\|_2 + \sqrt{2}\sigma^{-|S|}\|\mathrm{D}_S[f]\|_2 \\
&= \sqrt{I_{S\cup\{i\}}[f]} + \sqrt{2I_S[f]} \le (1+\sqrt{2})\sqrt{\delta} \le 3\delta.
\end{aligned}$$

Taking squares, we obtain the desired upper bound on the generalised influences of $f_{i\to 1}$. □

5 Total influence of global functions

In this section we show that our hypercontractive inequality (Theorem 1.3) implies our stability results for the isoperimetric inequality, namely Theorems 1.4 and 1.5. We also deduce our first sharp threshold result, Theorem 1.6.

5.1 The spectrum of sparse global sets

The key step in the proofs of Theorems 1.5 and 1.8 is to show that the Fourier spectrum of global sparse subsets of the p-biased cube is concentrated on the high degrees. We recall first a proof that in the uniform cube (i.e. cube with uniform measure), *all* sparse sets have this behaviour (not just the global ones). Our proof is based on ideas from Talagrand [70] and Bourgain and Kalai [19].

Theorem 5.1. *Let f be a Boolean function on the uniform cube, and let $r > 0$. Then*

$$\left\|f^{\le r}\right\|_2^2 \le 3^r \mu_{1/2}(f)^{1.5}.$$

The idea of the proof is to bound $\left\|f^{\leq r}\right\|_2^2 = \left\langle f^{\leq r}, f\right\rangle$ via Hölder by $\left\|f^{\leq r}\right\|_4 \|f\|_{4/3}$, bound the 4-norm via hypercontractivity and express the $4/3$-norm in terms of the measure of f using the assumption that f is Boolean. For future reference, we decompose the argument into two lemmas, the first of which applies also to the p-biased settting and the second of which requires hypercontractivity, and so is specific to the uniform setting. Theorem 5.1 follows immediately from Lemmas 5.2 and 5.3 below.

In the following lemma we consider $\{-1,0,1\}$-valued functions so that it can be applied to either a Boolean function or its discrete derivative.

Lemma 5.2. *Let $f\colon \{0,1\}^n \to \{0,1,-1\}$, let $\mathcal{F}$ be a family of subsets of $[n]$, and let $g(x) = f^{\mathcal{F}}(x) = \sum_{S\in\mathcal{F}} \hat{f}(S)\chi_S(x)$. Then $\|g\|_2^2 \leq \|g\|_4\|f\|_2^{1.5}$, where the norms can be taken with respect to an arbitrary p-biased measure.*

Proof. By Plancherel and Hölder's inequality, $\mathbb{E}[g^2] = \langle f, g\rangle \leq \|f\|_{4/3}\|g\|_4$, where $\|f\|_{4/3} = \mathbb{E}[f^2]^{3/4} = \|f\|_2^{1.5}$ as f is $\{-1,0,1\}$-valued. □

Applying Lemma 5.2 with $g = f^{\leq r}$, we obtain a lower bound on the 4-norm of g. We now upper bound it by appealing to the Hypercontractivity Theorem.

Lemma 5.3. *Let g be a function of degree r on the uniform cube. Then $\|g\|_4 \leq \sqrt{3}^r \|g\|_2$.*

Proof. Let h be the function, such that $\mathrm{T}_{1/\sqrt{3}}h = g$, i.e. $h = \sum_{|S|\leq r} \sqrt{3}^{|S|}\hat{g}(S)\chi_S$. Then the Hypercontractivity Theorem implies that $\|g\|_4 \leq \|h\|_2$, and by Parseval $\|h\|_2 \leq \sqrt{3}^r\|g\|_2$. □

We shall now adapt the proof of Theorem 5.1 to global functions on the p-biased cube. The only part in the above proof that needs an adjustment is Lemma 5.3, and in fact we have already provided the required adjustment in Section 3 in the form of Lemma 3.6.

Theorem 5.4. *Let $r \geq 1$, and let $f\colon \{0,1\}^n \to \{0,1,-1\}$. Suppose that $\mathrm{I}_S[f^{\leq r}] \leq \delta$ for each set S of size at most r. Then $\mathbb{E}[(f^{\leq r})^2] \leq 5^{4r/3}\delta^{\frac{1}{3}}\mathbb{E}\left[f^2\right]$.*

Proof. Applying Lemma 3.6 with $g = f^{\leq r}$, we obtain the upper bound $\|g\|_4 \leq 5^r\delta^{\frac{1}{4}}\|g\|_2^{0.5}$. Since the function f takes values only in the set $\{0,1,-1\}$, we may apply Lemma 5.2. Combining it with the upper bound on the 4-norm of g, we obtain

$$\|g\|_2^2 \leq \|g\|_4\|f\|_2^{1.5} \leq 5^r\delta^{\frac{1}{4}}\|g\|_2^{0.5}\|f\|_2^{1.5}.$$

Rearranging, and raising everything to the power $\frac{4}{3}$, we obtain $\|g\|_2^2 \leq 5^{4r/3}\delta^{\frac{1}{3}}\|f\|_2^2$. □

Let us say that f is *ϵ-concentrated* above degree r if $\|f^{\leq r}\|_2^2 \leq \epsilon\|f\|_2^2$. The significance of Theorem 5.4 stems from the fact that it implies the following result showing that for each $r, \epsilon > 0$ there exists a $\delta > 0$ such that any sparse (r,δ)-global function is ϵ-concentrated above degree r.

Corollary 5.5. *Let $r \geq 1$. Suppose that f is an (r,δ)-global Boolean function with $\mu_p(f) \leq \delta$. Then $\mathbb{E}[(f^{\leq r})^2] \leq 20^r\delta^{\frac{1}{3}}\mu_p(f)$.*

Proof. By Lemma 4.2, for each S of size r we have $\mathrm{I}_S\left[f^{\leq r}\right] \leq \mathrm{I}_S[f] \leq 8^r\delta$. Then Theorem 5.4 implies $\|f^{\leq r}\|_2^2 \leq 5^{4/3r}(8^r\delta)^{1/3}\|f\|_2^2 \leq 20^r\delta^{\frac{1}{3}}\|f\|_2^2$, where since f is Boolean we have $\|f\|_2^2 = \mu_p(f)$. □

5.2 Isoperimetric stability

We are now ready to prove our variant of the Kahn-Kalai Conjecture and sharp form of Bourgain's Theorem, both of which can be thought of as isoperimetric stability results. Both proofs closely follow existing proofs and substitute our new hypercontractivity inequality for the standard hypercontractivity theorem: for the first we follow a proof of the isoperimetric inequality, and for the second the proof of KKL given by Bourgain and Kalai [19] (their main idea is to apply the argument we gave in Theorem 5.1 for each of the derivatives of f).

Proof of Theorem 1.5. We prove the contrapositive statement that for a sufficiently large absolute constant C, if f is a Boolean function such that $\mu_p(f_{J\to 1}) \leq e^{-CK}$ for all J of size at most CK, then $p\mathrm{I}[f] \geq K\mu_p(f)$. Let f be such a function, and set $\delta = e^{-CK}$. Provided that $C \geq 2$, f is $(2K, \delta)$-global, and has p-biased measure at most δ. By Corollary 5.5, we have

$$\|f^{\leq 2K}\|_2^2 \leq 20^{2K}\delta^{\frac{1}{3}}\mu_p\left(f\right) \leq \mu_p\left(f\right)/2,$$

provided that C is sufficiently large. Hence,

$$\|f^{>2K}\|_2^2 = \|f\|_2^2 - \|f^{\leq 2K}\|_2^2 \geq \mu_p\left(f\right)/2.$$

By (2.1) on page 10 we obtain $p(1-p)\mathrm{I}[f] \geq 2K\|f^{>2K}\|_2^2$, so $p\mathrm{I}[f] \geq K\mu_p(f)$. □

Next we require the following lemma which bounds the norm of a low degree truncation in terms of the total influence.

Lemma 5.6. *Let $r \geq 0$. Suppose that for each nonempty set S of size at most r, $\mathrm{I}_S\left[f^{\leq r}\right] \leq \delta$. Then*

$$\|f^{\leq r}\|_2^2 \leq \mu_p(f)^2 + 10^{r-1}\delta^{\frac{1}{3}}\sigma^2\mathrm{I}[f].$$

Proof. Let $g_i := f_{i\to 1} - f_{i\to 0}$. Then for each S of size at most $r-1$ with $i \notin S$ we have

$$\mathrm{I}_S[g_i^{\leq r-1}] = \mathrm{I}_{S\cup\{i\}}[f^{\leq r}] \leq \delta,$$

and for each S containing i we have $\mathrm{I}_S[(g_i)^{\leq r-1}] = 0$. By Theorem 5.4, $\mathbb{E}[((g_i)^{\leq r-1})^2] \leq 5^{4(r-1)/3}\delta^{\frac{1}{3}}\mathbb{E}[g_i^2]$. The lemma now follows by summing over all i, using $\sum_i \mathbb{E}[g_i^2] = \mathrm{I}[f]$:

$$\begin{aligned}\|f^{\leq r}\|_2^2 &= \sum_{|S|\leq r}\hat{f}(S)^2 \leq \hat{f}(\emptyset)^2 + \sum_{|S|\leq r}|S|\hat{f}(S)^2 \\ &= \mu_p(f)^2 + \sigma^2\sum_i \mathbb{E}[((g_i)^{\leq r-1})^2] \\ &\leq \mu_p(f)^2 + 10^{r-1}\delta^{1/3}\sigma^2\mathrm{I}[f].\end{aligned}$$

□

We now establish a variant of Bourgain's Theorem for general Boolean functions, in which we replace the conclusion on the measure of a restriction by finding a large generalised influence.

Theorem 5.7. *Let $f\colon \{0,1\}^n \to \{0,1\}$. Suppose that $p\mathrm{I}[f] \leq K\mu_p\left(f\right)\left(1-\mu_p(f)\right)$. Then there exists an S of size $2K$, such that $\mathrm{I}_S[f] \geq 10^{-30K}$.*

Proof. Let $r = 2K$ and let $\delta = 10^{-30K}$. Suppose for contradiction that $\mathrm{I}_S[f] \le \delta$ for each set S of size at most r. By Lemma 5.6,

$$\|f^{\le r}\|_2^2 - \mu_p(f)^2 \le 10^{r-1}\delta^{1/3}\sigma^2 I[f] < p\mathrm{I}[f]/2K \le \mu_p(f)(1-\mu_p(f))/2.$$

On the other hand, by Parseval

$$\|f - f^{\le r}\|_2^2 = \sum_{|S|>r} \hat{f}(S)^2 \le r^{-1} \sum_{|S|>r} |S|\hat{f}(S)^2 \le r^{-1}p(1-p)\mathrm{I}[f] \le \mu_p(f)(1-\mu_p(f))/2.$$

However, these bounds contradict the fact that

$$\mu_p(f)(1-\mu_p(f)) = \|f\|_2^2 - \mu_p(f)^2 = \|f^{\le r}\|_2^2 - \mu_p(f)^2 + \|f - f^{\le r}\|_2^2. \qquad \square$$

Proof of Theorem 1.4. The theorem follows immediately from Theorem 5.7 and Lemma 4.3. $\square$

5.3 Sharpness examples

We now give two examples showing sharpness of the theorems in this section, both based on the tribes function of Ben-Or and Linial [11].

Example 5.8. We consider the anti-tribes function $f = f_{s,w} : \{0,1\}^n \to \{0,1\}$ defined by s disjoint sets $T_1, \dots, T_s \subset [n]$ each of size w, where $f(x) = \prod_{j=1}^s \max_{i \in T_j} x_i$, i.e. $f(x) = 1$ if for every j we have $x_i = 1$ for some $i \in T_j$, otherwise $f(x) = 0$. We have $\mu_p(f) = (1-(1-p)^w)^s$ and $\mathrm{I}[f] = \mu_p(f)' = sw(1-p)^{w-1}(1-(1-p)^w)^{s-1}$. We choose s, w with $s(1-p)^w = 1$ (ignoring the rounding to integers) so that $\mu_p(f) = (1-s^{-1})^s$ is bounded away from 0 and 1, and $K = (1-p)p\mathrm{I}[f] = pw(1-s^{-1})^{-1}\mu_p(f) = \Theta(pw)$. Thus $\log s = w \log(1-p)^{-1} = \Theta(K)$. However, for any $J \subset [n]$ with $|J| = t \le s$ we have $\mu_p(f_{J\to 1}) \le (1-s^{-1})^{s-t} \le 2^{t/s}\mu_p(f)$, so to obtain a density bump of $e^{-o(K)}$ we need $t = e^{-o(K)}s = e^{\Omega(K)} \gg K$. Thus Theorem 1.4 is sharp.

Example 5.9. Let $f(x) = f_{s,w}(x)\prod_{i\in T} x_i$ with $f_{s,w}$ as in Example 5.8 and $T \subset [n]$ a set of size t disjoint from $\cup_j T_j$. We have $\mu_p(f) = p^t(1-(1-p)^w)^s$ and $I[f] = \mu_p(f)' = tp^{t-1}(1-(1-p)^w)^s + p^t sw(1-p)^{w-1}(1-(1-p)^w)^{s-1}$. We fix $K > 1$ and choose s, w with $s(1-p)^w = K$, so that $\mu_p(f) = p^t(1-K/s)^s = p^t e^{-\Theta(K)}$ for $s > 2K$ and $p(1-p)I[f] = \mu_p(f)((1-p)t + pwK(1-K/s)^{-1}) = \mu_p(f)\Theta(K)$ if $pw = \Theta(1)$ and $t = O(K)$. For any $J \subset [n]$ with $|J| = t + u \le t + s$ we have $\mu_p(f_{J\to 1}) \le (1-K/s)^{s-u} \le e^{-K(1-u/s)} \le e^{-K/2}$ unless $u > s/2 = \Theta(K)$. Thus Theorem 1.5 is sharp.

5.4 Sharp thresholds: the traditional approach

In this section we deduce Theorem 1.6 from our edge-isoperimetric stability results and the Margulis–Russo Lemma. Recall that a monotone Boolean function is M-global in an interval if $\mu_p(f_{J\to 1}) \le \mu_p(f)^{0.01}$ for each p in the interval and set J of size M. We prove the following slightly stronger version of Theorem 1.6.

Theorem 5.10. *There exists an absolute constant C such that the following holds for any monotone Boolean function f that is M-global in some interval $[p,q]$: if $q \le p_c$ and $\mu_p(f) \ge e^{-M/C}$ then*

$$\mu_q(f) \ge \mu_p(f)^{\left(\frac{p}{q}\right)^{1/C}}. \tag{5.1}$$

In particular, $q \le M^C p$.

Proof. By Theorem 1.5, since f is M-global throughout the interval, there exists a constant C such that $\mathrm{I}_x[f] \geq \frac{\mu_x(f)\log(\frac{1}{\mu_x(f)})}{Cx}$ for all x in the interval $[p,q]$. By the Margulis-Russo lemma,

$$\frac{d}{dx}\log\left(-\log(\mu_x(f))\right) = \frac{\mu_x(f)'}{\mu_x(f)\log(\mu_x(f))} = \frac{I_x[f]}{\mu_x(f)\log(\mu_x(f))} \leq \frac{-1}{Cx}$$

in all of the interval $[p,q]$. Hence,

$$\log\left(-\log(\mu_q(f))\right) \leq \log(-\log(\mu_p(f))) - \frac{\log(\frac{q}{p})}{C}.$$

The first part of the theorem follows by taking exponentials, multiplying by -1 then taking exponentials again. To see the final statement, note that $q \leq p_c$ implies $\mu_q(f) \leq \frac{1}{2}$. We cannot have $q \geq M^c p$, as then the right hand side in (5.1) would be larger than $e^{-\frac{1}{C}} > 1/2$ for large C. To obtain Theorem 1.6 we substitute $q = p_c$. □

6 Noise sensitivity and sharp thresholds

We start this section by showing that sparse global functions are noise sensitive; Theorem 1.8 follows immediately from Theorem 6.1.

Theorem 6.1. *Let $\rho \in (0,1)$, and let $\epsilon > 0$. Let $r = \frac{\log(2/\epsilon)}{\log(1/\rho)}$, and let $\delta = 20^{-3r-1}\epsilon^3$. Suppose that f is an (r,δ)-global Boolean function with $\mu_p(f) < \delta$. Then*

$$\mathrm{Stab}_\rho(f) \leq \epsilon\mu_p(f).$$

Proof. We have

$$\langle \mathrm{T}_\rho f, f\rangle \leq \sum_{|S|\leq r}\hat{f}(S)^2 + \rho^r\sum_{|S|>r}\hat{f}(S)^2 \leq \mathbb{E}\left[\left(f^{\leq r}\right)^2\right] + \frac{\varepsilon}{2}\mu_p(f).$$

The statement now follows from Corollary 5.5, which gives $\mathbb{E}[(f^{\leq r})^2] \leq 20^r\delta^{1/3}\mathbb{E}[f^2] < \varepsilon\mu_p(f)/2$. □

In the remainder of this section, following [56], we deduce sharp thresholds from noise sensitivity via the following *directed noise operator*, which is implicit in the work of Ahlberg, Broman, Griffiths and Morris [3] and later studied in its own right by Abdullah and Venkatasubramanian [1].

Definition 6.2. Let $D(p,q)$ denote the unique distribution on pairs $(\boldsymbol{x},\boldsymbol{y}) \in \{0,1\}^n \times \{0,1\}^n$ such that $\mathbf{x} \sim \mu_p$, $\mathbf{y} \sim \mu_q$, all $\mathbf{x}_i \leq \mathbf{y}_i$ and $\{(\mathbf{x}_i,\mathbf{y}_i) : i \in [n]\}$ are independent. We define a linear operator $\mathrm{T}^{p\to q} : L^2(\{0,1\}^n,\mu_p) \to L^2(\{0,1\}^n,\mu_q)$ by

$$\mathrm{T}^{p\to q}(f)(y) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim D(p,q)}\left[f(\mathbf{x})\,|\,\mathbf{y}=y\right].$$

The directed noise operator $\mathrm{T}^{p\to q}$ is a version of the noise operator where bits can be flipped only from 0 to 1. The associated notion of directed noise stability, i.e. $\langle f, \mathrm{T}^{p\to q}f\rangle_{\mu_q}$, is intuitively a measure of how close a Boolean function f is to being monotone. Indeed, for any $(\mathbf{x},\mathbf{y})$ with all $x_i \leq y_i$ we have $f(\mathbf{x})f(\mathbf{y}) \leq f(\mathbf{x})$, with equality if f is monotone, so

$$\langle f, \mathrm{T}^{p\to q}f\rangle = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim D(p,q)}\left[f(\mathbf{x})f(\mathbf{y})\right] \leq \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim D(p,q)}\left[f(\mathbf{x})\right] = \mu_p(f),$$

with equality if f is monotone[7]. We note that the adjoint operator $(\mathrm{T}^{p\to q})^\star : L^2(\{0,1\}^n,\mu_q)\to L^2(\{0,1\}^n,\mu_p)$ defined by $\langle \mathrm{T}^{p\to q} f, g\rangle = \langle f, (\mathrm{T}^{p\to q})^\star g\rangle$ satisfies $(\mathrm{T}^{p\to q})^\star = \mathrm{T}^{q\to p}$, where

$$\mathrm{T}^{q\to p}(g)(x) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim D(p,q)}\left[g(\mathbf{y}) \,|\, \mathbf{x}=x\right].$$

The following simple calculation relates these operators to the noise operator.

Lemma 6.3. *Let $0<p<q<1$ and $\rho = \frac{p(1-q)}{q(1-p)}$. Then $(\mathrm{T}^{p\to q})^\star \mathrm{T}^{p\to q} = \mathrm{T}_\rho$ on $L^2(\{0,1\}^n,\mu_p)$.*

Proof. We need to show that the following distributions on pairs of p-biased bits $(\mathbf{x},\mathbf{x}')$ are identical: (a) let $\mathbf{x}$ be a p-biased bit, with probability ρ let $\mathbf{x}'=\mathbf{x}$, otherwise let $\mathbf{x}'$ be an independent p-biased bit, (b) let $(\mathbf{x},\mathbf{y})\sim D(p,q)$ and then $(\mathbf{x}',\mathbf{y})\sim D(p,q)\mid y$. It suffices to show $\mathbb{P}(x\neq x')$ is the same in both distributions. We condition on x. Consider $x=1$. In distribution (a) we have $\mathbb{P}(\mathbf{x}'=0)=(1-\rho)(1-p)$. In distribution (b) we have $\mathbb{P}(\mathbf{y}=1)=1$ and then $\mathbb{P}(\mathbf{x}'=0)=1-p/q=(1-\rho)(1-p)$, as required. Now consider $\mathbf{x}=0$. In distribution (a) we have $\mathbb{P}(\mathbf{x}'=1)=(1-\rho)p$. In distribution (b) we have $\mathbb{P}(\mathbf{y}=1)=\frac{q-p}{1-p}$ and then $\mathbb{P}(\mathbf{x}'=1\mid \mathbf{y}=1)=p/q$, so $\mathbb{P}(\mathbf{x}'=1)=\frac{p(q-p)}{q(1-p)}=(1-\rho)p$, as required. □

We now give an alternative way to deduce sharp threshold results, using noise sensitivity, rather than the traditional approach via total influence (as in the proof of Theorem 5.10). Our alternative approach has the following additional nice features, both of which have been found useful in Extremal Combinatorics (see [56]).

1. To deduce a sharp threshold result in an interval $[p,q]$ it is enough to show that f is global only according to the p-biased distribution. This is a milder condition than the one in the traditional approach, that requires globalness throughout the entire interval.

2. The monotonicity requirement may be relaxed to "almost monotonicity".

Proposition 6.4. *Let $f\colon \{0,1\}^n\to\{0,1\}$ be a monotone Boolean function. Let $0<p<q<1$ and $\rho=\frac{p(1-q)}{q(1-p)}$. Then $\mu_q(f)\geq \mu_p(f)^2/\mathrm{Stab}_\rho(f)$.*

Proof. By Cauchy–Schwarz and Lemma 6.3,

$$\mu_p(f)^2 = \langle \mathrm{T}^{p\to q} f, f\rangle_{\mu_q}^2 \leq \langle \mathrm{T}^{p\to q} f, \mathrm{T}^{p\to q} f\rangle_{\mu_q}\langle f,f\rangle_{\mu_q} = \langle \mathrm{T}_\rho f, f\rangle_{\mu_p}\mu_q(f). \qquad \square$$

The above proof works not only for monotone functions, but also for functions where the first equality above is replaced by approximate equality (which is a natural notion for a function to be "almost monotone").

We conclude this section by noting that Theorem 1.9 (sharp thresholds for global functions) is immediate from Theorem 6.1 and Proposition 6.4.

7 General hypercontractivity

In this section we generalise Theorem 1.3 in two different directions. One direction is showing hypercontractivity from general q-norms to the 2-norm (rather than merely treating the case $q=4$); the other is replacing the cube by general product spaces.

[7]The starting point for [56] is the observation that this inequality is close to an equality if f is almost monotone.

7.1 Hypercontractivity with general norms

We start by describing a more convenient general setting in which we replace characters on the cube by arbitrary random variables. To motivate this setting, we remark that one can extend the proof of Theorem 3.4 to any random variable of the form

$$f = \sum_{S \subset [n]} a_S \prod_{i \in S} \mathbf{Z}_i, \tag{7.1}$$

where $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are independent real-valued random variables having expectation 0, variance 1 and 4th moment at most σ^{-2}. To motivate the analogous setting for general integers $q > 2$, we note that the characters χ_i^p satisfy

$$\mathbb{E}[|\chi_i^p|^q] \le \|\chi_i^p\|_\infty^{q-2} \|\chi_i^p\|_2^2 \le \sigma^{2-q}.$$

This suggests replacing the 4th moment condition by $\|\mathbf{Z}_i\|_q^q \le \sigma^{2-q}$. Given f as in (7.1), we define the (generalised) derivatives by substituting the random variables Z_i for the characters χ_i^p in our earlier Fourier formulas, i.e.

$$\mathrm{D}_i[f] = \sum_{S:\, i \in S} a_S \prod_{j \in S \setminus \{i\}} \mathbf{Z}_i \quad \text{and} \quad \mathrm{D}_T(f) = \sum_{S:\, T \subset S} a_S \prod_{j \in S \setminus T} \mathbf{Z}_i,$$

Similarly, we adopt analogous definitions of the generalised influences and noise operator, i.e.

$$\mathrm{I}_S[f] = \|\sigma^{-|S|} \mathrm{D}_S[f]\|_2^2 \quad \text{and} \quad \mathrm{T}_\rho[f] = \sum_S \rho^{|S|} a_S \prod_{i \in S} \mathbf{Z}_i.$$

We prove the following hypercontractive inequality.

Theorem 7.1. *Let $q \ge 2$ be an even integer and $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be independent real-valued random variables satisfying*

$$\mathbb{E}[\mathbf{Z}_i] = 0, \quad \mathbb{E}[\mathbf{Z}_i^2] = 1, \quad \textit{and} \quad \mathbb{E}[|\mathbf{Z}_i|^q] \le \sigma^{2-q}.$$

Let $f = \sum_{S \subset [n]} a_S \prod_{i \in S} \mathbf{Z}_i$ and $\rho < \frac{1}{2q^{1.5}}$. Then

$$\|\mathrm{T}_\rho f\|_q^q \le \sum_{S \subset [n]} \sigma^{(2-q)|S|} \|\mathrm{D}_S(f)\|_2^q.$$

Theorem 7.1 is a qualitative generalisation of Theorem 3.4 (with smaller ρ, which we do not attempt to optimise). The following generalised variant of Theorem 1.3 follows by repeating the proof of Theorem 1.3 in Section 3.

Theorem 7.2. *For $q > 2$ and $\mathbf{Z}_i$'s as in Theorem 7.1, let $f = \sum_{S \subset [n]} a_S \prod_{i \in S} \mathbf{Z}_\mathbf{i}$ let $\delta > 0$, and let $\rho \le (2q)^{-1.5}$. Suppose that $\mathrm{I}_S[f] \le \beta \|f\|_2^2$ for all $S \subset [n]$. Then*

$$\|\mathrm{T}_\rho[f]\|_q \le \beta^{\frac{q-2}{2q}} \|f\|_2.$$

We now begin with the ingredients of the proof of Theorem 7.1, following that of Theorem 3.4. For $0 \le t \le n$ let

$$f_t = \sum_S a_S \chi_S^t, \quad \text{where } \chi_S^t = \prod_{i \in S \cap [t]} \chi_i^{1/2} \prod_{i \in S \setminus [t]} \mathbf{Z}_i.$$

Here, just as in Section 3, the function f_t interpolates from the original function $f_0 = f$ to $f_n = \sum_S a_S \chi_S^{1/2} \in L^2(\{0,1\}^n, \mu_{1/2})$. As $\{\chi_S^t : S \subset [n]\}$ are orthonormal we have $\|f_t\|_2 = \|f\|_2$ for all t.

As before, we define the noise operators $\mathrm{T}^t_{\rho',\rho}$ on a function $f = \sum_S a_S \chi_S^t$ by

$$\mathrm{T}^t[f] = \sum_S \rho'^{|S \cap [t]|} \rho^{|S \setminus [t]|} a_S \chi_S^t.$$

Thus $\mathrm{T}^t_{\rho',\rho}$ interpolates from $\mathrm{T}^0_{\rho',\rho} = \mathrm{T}_\rho$ (for the original function) to $\mathrm{T}^n_{\rho',\rho} = \mathrm{T}_{\rho'}$ (for $\mu_{1/2}$).

Our goal will now be to adjust Lemma 3.1 to the general setting, which is similar in spirit to the 4-norm case, although somewhat trickier. It turns out that the case $n = 1$ poses the main new difficulties, so we start with this in the next lemma.

Lemma 7.3. *Let $q > 2$ be an even integer and $\mathbf{Z}$ be a random variable satisfying $\mathbb{E}[\mathbf{Z}] = 0, \mathbb{E}[\mathbf{Z}^2] = 1, \mathbb{E}[|\mathbf{Z}|^q] \le \sigma^{2-q}$. Let $e, d \in \mathbb{R}$ and $\rho \in (0, \frac{1}{2q}]$. Then $\|e + \rho d\mathbf{Z}\|_q^q \le \|e + d\chi^{\frac{1}{2}}\|_q^q + \sigma^{2-q} d^q$.*

Proof. If $e = 0$ then the lemma is trivial, so we assume $e \neq 0$. If $e < 0$ we can multiply e, d, $\mathbb{Z}$ and $\chi^{\frac{1}{2}}$ all by -1 and get an equivalent statement of this form where now $e > 0$, hence we assume henceforth that $e > 0$ (we used the fact that the distribution of $\chi^{\frac{1}{2}}$ is invariant under multiplication by a sign and the assumptions on $\mathbb{Z}$ are also invariant under multiplication by a sign). By rescaling (d, e) to $(d/e, 1)$ we can also assume that $e = 1$.

It will be convenient to consider both sides of the inequality as functions of d: we write

$$f(d) = \|1 + \rho d\mathbf{Z}\|_q^q \quad \text{and} \quad g(d) = \|1 + d\chi^{\frac{1}{2}}\|_q^q + \sigma^{2-q} d^q.$$

As $f(0) = g(0)$, it suffices to show that $f'(0) = g'(0)$ and $f'' \le g''$ everywhere.

Let us compute the derivatives. We note that the function $x \mapsto |x^q|$ has derivative $q|x|^{q-1}\mathrm{sign}(x)$, which is in turn continuously differentiable for $q > 2$. Thus

$$\begin{aligned} f' &= \mathbb{E}[q\,|1 + \rho d\mathbf{Z}|^{q-1}\,\mathrm{sign}(1 + \rho d\mathbf{Z})\rho\mathbf{Z}] = \rho q \mathbb{E}[|1 + \rho d\mathbf{Z}|^{q-1}\mathrm{sign}(1 + \rho d\mathbf{Z})\mathbf{Z}] \quad \text{and} \\ f'' &= (q-1)q\rho^2 \mathbb{E}[|1 + \rho d\mathbf{Z}|^{q-2}\mathbf{Z}^2]. \end{aligned}$$

Differentiating g we obtain

$$\begin{aligned} g' &= q\mathbb{E}\left[\left|1 + d\chi^{\frac{1}{2}}\right|^{q-1} \mathrm{sign}(1 + d\chi^{\frac{1}{2}})\chi^{\frac{1}{2}}\right] + q\sigma^{2-q}d^{q-1} \quad \text{and} \\ g'' &= q(q-1)\mathbb{E}\left[\left|1 + d\chi^{\frac{1}{2}}\right|^{q-2}\left(\chi^{\frac{1}{2}}\right)^2\right] + q(q-1)d^{q-2}\sigma^{2-q} \ge q(q-1)/2 + q(q-1)d^{q-2}\sigma^{2-q}. \end{aligned}$$

Thus $g'(0) = f'(0) = 0$ and it remains to show $f'' \le g''$ everywhere. Our strategy for bounding f'' is to decompose the expectation over two complementary events E_1 and E_2, where E_1 is the event that $|1 + \rho d\mathbf{Z}| \le |d\mathbf{Z}|$ (and E_2 is its complementary event). We write $f'' = f_1'' + f_2''$, where each

$$f_i'' = (q-1)q\rho^2\mathbb{E}[|1 + \rho d\mathbf{Z}|^{q-2}\mathbf{Z}^2\mathbf{1}_{E_i}].$$

First we note the bound

$$f_1'' \le q(q-1)\rho^2 d^{q-2}\mathbb{E}[|\mathbf{Z}|^q] \le q(q-1)d^{q-2}\sigma^{2-q}.$$

Given the above lower bound on g'', it remains to show $f_2'' \le q(q-1)/2$. On the event E_2 we have

$$|d\mathbf{Z}| \le |1 + \rho d\mathbf{Z}| \le 1 + |\rho d\mathbf{Z}|.$$

Rearranging, we obtain $|\rho d\mathbf{Z}|(\rho^{-1}-1) \leq 1$. Since $\rho^{-1} \geq 2q$, we get

$$1 + |\rho d\mathbf{Z}| \leq 1 + \frac{1}{2q-1}.$$

Using $\mathbb{E}[\mathbf{Z}^2] = 1$ this yields

$$f_2'' \leq q(q-1)\rho^2\Big(1 + \frac{1}{2q-1}\Big)^{q-2} \leq \sqrt{e}\rho^2 q(q-1) \leq q(q-1)/2.$$

Hence $f'' = f_1'' + f_2'' \leq g''$ for any value of d. This completes the proof of the lemma. $\square$

We are now ready to show the replacement step.

Lemma 7.4. $\mathbb{E}[(\mathrm{T}^{t-1}_{2q\rho,\rho} f_{t-1})^q] \leq \mathbb{E}[(\mathrm{T}^{t}_{2q\rho,\rho} f_t)^q] + \sigma^{2-q}\mathbb{E}[(\mathrm{T}^{t}_{2q\rho,\rho}((\mathrm{D}_t f)_t))^q]$.

Proof. We write

$$\begin{aligned} &f_t = \chi_t^{1/2} g + h \quad \text{and} \quad f_{t-1} = \chi_t^p g + h, \quad \text{where} \\ &g = (\mathrm{D}_t f)_t = \sum_{S:t\in S} a_S \chi^t_{S\setminus\{t\}} = \sum_{S:t\in S} a_S \chi^{t-1}_{S\setminus\{t\}} = (\mathrm{D}_t f)_{t-1}, \quad \text{and} \\ &h = \mathbb{E}_{x_t\sim\mu_{1/2}} f_t = \sum_{S:t\notin S} a_S \chi^t_S = \sum_{S:t\notin S} a_S \chi^{t-1}_S = \mathbb{E}_{\mathbf{Z}_t} f_{t-1}. \end{aligned}$$

We also write

$$\begin{gathered} \mathrm{T}^t_{2q\rho,\rho} f_t = 2q\rho\chi_t^{1/2} d + e \quad \text{and} \quad \mathrm{T}^{t-1}_{2q\rho,\rho} f_{t-1} = \rho\mathbf{Z}_t d + e, \quad \text{where} \\ d = \mathrm{T}^t_{2q\rho,\rho} g = \mathrm{T}^{t-1}_{2q\rho,\rho} g \quad \text{and} \quad e = \mathrm{T}^t_{2q\rho,\rho} h = \mathrm{T}^{t-1}_{2q\rho,\rho} h. \end{gathered}$$

Conditioning on all coordinates other than $\mathbf{Z}_t$, we use Lemma 7.3 with $\rho' = \frac{1}{2q}$ and $d' = \frac{\rho}{\rho'} d$ to argue that the left hand side of the lemma is equal to

$$\|\rho'\mathbf{Z}_t \frac{\rho}{\rho'} d + e\|_q^q = \|e + \rho' d' \mathbf{Z}_t\|_q^q \leq \|e + d'\chi_t^{1/2}\|_q^q + \sigma^{2-q} d'^q = \|e + 2q\rho d\chi_t^{1/2}\|_q^q + \sigma^{2-q}(2q\rho d)^q.$$

Taking expectation over the coordinates outside $\mathbf{Z}_t$ and using $2q\rho \leq 1$ concludes the proof. $\square$

From now on, everything is similar to Section 3. We may apply the previous lemma inductively to obtain.

Lemma 7.5. $\|\mathrm{T}^i_{2q\rho,\rho} f_i\|_q^q \leq \sum_{S\subset[n]\setminus[i]} \sigma^{(2-q)|S|}\|\mathrm{T}^n_{2q\rho,\rho}((\mathrm{D}_S f)_n)\|_q^q$ *for all* $0 \leq i \leq n$.

In particular, recalling that $\mathrm{T}^0_{2q\rho,\rho} = \mathrm{T}_\rho$ on the original function and $\mathrm{T}^n_{2q\rho,\rho} = \mathrm{T}_{2q\rho}$ on $\mu_{1/2}$, the case $i = 0$ of Lemma 7.5 is as follows.

Proposition 7.6. $\|\mathrm{T}_\rho f\|_q^q \leq \sum_{S\subset[n]} \sigma^{(2-q)|S|}\|\mathrm{T}_{2q\rho}((\mathrm{D}_S f)_n)\|_q^q$.

The q-norms on the right hand side of Proposition 7.6 are with respect to the uniform measure $\mu_{1/2}$, where we can apply standard hypercontractivity with noise rate $\leq 1/\sqrt{q-1}$ to obtain

$$\|\mathrm{T}_{2q\rho}((\mathrm{D}_S f)_n)\|_q^q \leq \|(\mathrm{D}_S f)_n\|_2^q = \|\mathrm{D}_S f\|_2^q.$$

This completes the proof of Theorem 7.1.

In the case where the $\mathbf{Z}_i$ have different qth moments, the proof can be adjusted to give a better upper bound. We write

$$\mathbb{E}[\mathbf{Z}_i^q] = \sigma_i^{2-q}, \quad \sigma_S = \prod_{i\in S}\sigma_i \quad \text{and} \quad \mathrm{I}_S[f] = \|\frac{1}{\sigma_S}\mathrm{D}_S[f]\|_2^2. \tag{7.2}$$

The proof of Theorem 7.1 yields the following variant of Theorem 3.4.

Theorem 7.7. *Let $q \geq 2$ be an even integer, let $0 < \rho \leq \frac{1}{2q^{1.5}}$, and let $f = \sum a_S \prod_{i\in S}\mathbf{Z}_i$ with Z_i as in* (7.2). *Then*

$$\|\mathrm{T}_\rho f\|_q^q \leq \sum_{S\subset[n]} \sigma_S^{2-q}\|\mathrm{D}_S[f]\|_2^q.$$

The following variant of Theorem 1.3 follows from Theorem 7.7. The proof is similar to the one given in Section 3, where Theorem 1.3 is deduced from Theorem 3.4.

Theorem 7.8. *Let $q > 2$ be an even integer, $\beta > 0$ and $0 < \rho \leq \frac{1}{2q^{1.5}}$. Suppose $f = \sum_{S\subset[n]} a_S \prod_{i\in S}\mathbf{Z}_i$ with Z_i as in* (7.2) *has $\mathrm{I}_S[f] \leq \beta\|f\|_2^2$ for all $S \subset [n]$. Then*

$$\|\mathrm{T}_\rho f\|_q \leq \beta^{\frac{q-2}{2q}}\|f\|_2.$$

Finally, we state the following variant of Lemma 3.6, which is easy to deduce from Theorem 7.8 (mimicking the proof of Lemma 3.6).

Lemma 7.9. *Let $q > 2$ be an even integer and $\delta > 0$. Suppose $f = \sum_{S\subset[n]} a_S \prod_{i\in S}\mathbf{Z}_i$ with Z_i as in* (7.2) *has $\mathrm{I}_S[f] \leq \delta$ for all $|S| \leq r$ and f has degree at most r. Then*

$$\|f\|_q \leq (2q)^{1.5r}\delta^{\frac{q-2}{2q}}\|f\|_2^{\frac{2}{q}}.$$

7.2 A hypercontractive inequality for product spaces

Now we consider the setting of a general discrete product space $(\Omega,\nu) = \prod_{t=1}^n(\Omega_t,\nu_t)$. We assume $p_t = \min_{\omega_t\in\Omega_t}\nu_t(\omega_t) \in (0,1/2)$ for each $t \in [n]$, and we write $p = \min_t p_t$. We recall the projections E_J on $L^2(\Omega,\nu)$ defined by $(\mathrm{E}_J f)(\omega) = \mathbb{E}_{\omega_J}[f(\omega) \mid \omega_{\overline{J}}]$, the generalised Laplacians L_S defined by composing L_t for all $t \in S$, where $\mathrm{L}_t f = f - \mathrm{E}_t f$, and the generalised influences $\mathrm{I}_S[f] = \mathbb{E}[\mathrm{L}_S[f]^2]\prod_{i\in S}\sigma_i^{-2}$, where $\sigma_i^2 = p_i(1-p_i)$.

We will require the theory of orthogonal decompositions in product spaces, which we summarise following the exposition in [64, Section 8.3]. For $f \in L^2(\Omega,\nu)$ and $J, S \subset [n]$ we write $f^{\subset J} = \mathrm{E}_{\overline{J}}f$ and define $f^{=S} = \sum_{J\subset S}(-1)^{|S\setminus J|}f^{\subset J}$ (inclusion-exclusion for $f^{\subset J} = \sum_{S\subset J} f^{=S}$). This decomposition is known as the Efron–Stein decomposition [23]. The key properties of $f^{=S}$ are that it only depends on coordinates in S and it is orthogonal to any function that depends only on some set of coordinates not containing S; in particular, $f^{=S}$ and $f^{=S'}$ are orthogonal for $S \neq S'$. We note that $f = f^{\subset[n]} = \sum_S f^{=S}$. We have similar Plancherel / Parseval relations as for Fourier decompositions, namely $\langle f, g\rangle = \sum_S f^{=S}g^{=S}$, so $\mathbb{E}[f^2] = \sum_S (f^{=S})^2$.

Our goal in this section is to prove an hypercontractive inequality for the Efron–Stein decomposition in the spirit of Theorem 3.4. The noise operator is defined by $\mathrm{T}_\rho[f] = \sum_{S\subset[n]}\rho^{|S|}f^{=S}$. It also has a combinatorial interpretation, which is similar to the usual one on the p-biased setting. Given $x \in \Omega$, a sample $\mathbf{y} \sim N_\rho(x)$ is chosen by independently setting y_i to x_i with probability ρ and resampling it from

(Ω_i, ν_i) with probability $1-\rho$. In the general product space setting there are no good analogues to $\mathrm{D}_i[f]$ and $\mathrm{D}_S[f]$, so we instead work with the Laplacians, which have similar Fourier formulae: $\mathrm{L}_i[f] = \sum_{S\colon i \in S} f^{=S}$ and $\mathrm{L}_T[f] = \sum_{S\colon T \subset S} f^{=S}$. In the special case where $\Omega_i = \{0,1\}$ we have $\|\mathrm{L}_S[f]\|_2 = \|\mathrm{D}_S[f]\|_2$. It will be convenient to write $\sigma_S = \prod_{i\in S} \sigma_i$.

The main result of this section is the following theorem.

Theorem 7.10. *Let $f \in L^2(\Omega, \nu)$, let $q > 2$ be an even integer, and let $\rho \leq \frac{1}{4q^{1.5}}$. Then*

$$\|\mathrm{T}_\rho f\|_q^q \leq \sum_{S \subset [n]} \sigma_S^{2-q} \|\mathrm{L}_S[f]\|_2^q.$$

The idea of the proof is as follows. We encode our function $f \in L^2(\Omega, \nu)$ as a function $\tilde{f} := \sum_S \|f^{=S}\|_2 \chi_S$ for appropriate $\chi_S = \prod_{i\in S} \chi_i$ (in fact, these will be biased characters on the cube). We then bound $\|\mathrm{T}_\rho f\|_q$ by $\|\mathrm{T}_\rho \tilde{f}\|_q$ and use Theorem 7.8 to bound the latter norm.

The main technical component of the theorem is the following proposition.

Proposition 7.11. *Let $q \geq 2$ be an even integer, let $g \in L^2(\Omega, \nu)$ let $\chi_S = \prod_{i\in S} \chi_i$, where χ_i are independent random variables having expectation 0, variance 1, and satisfying $\mathbb{E}[\chi_i^j] \geq \sigma_i^{2-j}$ for each integer $j \in (2, q]$. Let $\tilde{g} = \sum_{S\subset[n]} \|g^{=S}\|_2 \chi_S$. Then*

$$\|g\|_q \leq \|\tilde{g}\|_q.$$

Below, we fix χ_S as in the proposition, and let $\tilde{\circ}$ denote the operator mapping a function $g \in L^2(\Omega, \nu)$ to the function $\sum_{S\subset[n]} g^{=S} \chi_S$.

To prove the proposition, we will expand out $\|g\|_q^q$ and $\|\tilde{g}\|_q^q$ according to their definitions and compare similar terms: namely, we show that a term of the form $\mathbb{E}[\prod_{i=1}^q g^{=S_i}]$ is bounded by the corresponding term in $\|\tilde{g}\|_q^q$, i.e. $\prod_{i=1}^q \|g^{=S_i}\|_2 \mathbb{E}[\prod_{i=1}^q \chi_{S_i}]$. We now establish such a bound.

We begin with identifying cases in which both terms are equal to 0, and for that we use the orthogonality of the decomposition $\{g^{=S}\}_{S\subset[n]}$. Afterwards, we only rely on the fact that $g^{=S}$ depends only on the coordinates in S.

Lemma 7.12. *Let q be some integer, let $g \in L^2(\Omega, \nu)$, and let $S_1, \ldots, S_q \subset [n]$ be some sets. Suppose that some $j \in [n]$ belongs to exactly one of the sets $S_1, \ldots, S_q$. Then*

$$\mathbb{E}\left[\prod_{i=1}^q g^{=S_i}\right] = 0 \quad \textit{and} \quad \mathbb{E}\left[\prod_{i=1}^q \chi_{S_i}\right] = 0.$$

Proof. Assume without loss of generality that $j \in S_1$. The second equality $\mathbb{E}\left[\prod_{i=1}^q \chi_{S_i}\right] = 0$ follows by taking expectation over χ_j, using the independence between the random variables χ_i. For the first equality, observe that the function $\prod_{i=2}^q g^{=S_i}$ depends only on coordinates in $S_2 \cup \cdots \cup S_q \subset [n] \setminus \{j\}$. Hence the properties of the Efron–Stein decomposition imply

$$0 = \left\langle g^{=S_1}, \prod_{i=2}^q g^{=S_i} \right\rangle = \mathbb{E}\left[\prod_{i=1}^q g^{=S_i}\right]. \qquad \square$$

Thus we only need to consider terms corresponding to $S_1, \ldots, S_q$ in which each coordinate appears in at least two sets. To facilitate our inductive proof we work with general functions f_i that depend only on coordinates of S_i (rather than only with the functions of the form $g^{=S_i}$).

Lemma 7.13. *Let $f_1, \dots, f_q \in L^2(\Omega, \nu)$ be functions that depend on sets $S_1, \dots, S_q$ respectively. Let T_i for $i = 3, \dots, q$ be the set of coordinates covered by the sets $S_1, \dots, S_q$ exactly i times. Then*

$$\left| \mathbb{E}\left[\prod_{i=1}^{q} f_i \right] \right| \leq \prod_{i=1}^{q} \|f_i\|_2 \cdot \prod_{j=3}^{q} \sigma_{T_j}^{2-j}.$$

Proof. The proof is by induction on n, simultaneously for all functions. We start with the case $n = 1$, which we prove by reducing to the case that all f_i are eqal.

The case $n = 1$

Here each f_i either depends on a single input or is constant and depends only on the empty set. We may assume that none of the f_i's is constant, as otherwise we may eliminate it from the inequality by dividing by $|f_i|$. By the generalised Hölder inequality we have

$$\left| \mathbb{E}\left[\prod_{i=1}^{q} f_i \right] \right| \leq \prod_{i=1}^{q} \|f_i\|_q.$$

Hence the case $n = 1$ of the lemma will follow once we prove it assuming all the f_i are equal.

The $n = 1$ case with equal f_i's

We show that if (Ω, ν) is a discrete probability space in which any atom has probability at least p, then $\|f\|_q^q \leq \|f\|_2^q \sigma^{2-q}$, where $\sigma = \sqrt{p(1-p)}$.

While the inequality $\|f\|_2 \leq \|f\|_q$ holds in any probability space, the reverse inequality holds in any measure space where each atom has measure at least 1. Accordingly, we consider the measure $\tilde{\nu}$ on Ω defined by $\tilde{\nu}(x) = \nu(x)p^{-1}$. Then

$$\|f\|_{q,\nu}^q = p\|f\|_{q,\tilde{\nu}}^q \leq p\|f\|_{2,\tilde{\nu}}^q = p^{1-\frac{q}{2}}\|f\|_{2,\nu}^q \leq \sigma^{2-q}\|f\|_{2,\nu}^q.$$

This completes the proof of the $n = 1$ case.

The inductive step

Let $f_1, \dots, f_q \in L^2(\Omega, \nu)$ be functions. Let $\mathbf{x} \sim \prod_{i=1}^{n-1}(\Omega_i, \nu_i)$. By the $n = 1$ case we have:

$$\left| \mathbb{E}\left[\prod_{k=1}^{q} f_k \right] \right| = \left| \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}\left[\prod_{k=1}^{q} (f_k)_{[n-1]\to\mathbf{x}} \right] \right] \right| \leq \mathbb{E}_{\mathbf{x}}\left[\prod_{k=1}^{q} \|(f_k)_{[n-1]\to\mathbf{x}}\|_2 \sigma_n^j \right],$$

writing $j = \min\{0, 2 - j'\}$ where $n \in T_{j'}$, noting that at most j' of the functions $(f_k)_{[n-1]\to\mathbf{x}}$ depend on n. The lemma now follows by applying the inductive hypothesis to the functions $f_k^*(\mathbf{x}) = \|(f_k)_{[n-1]\to\mathbf{x}}\|_2$ (depending on coordinates $S_k^* \subset S_k$), using $\left\| \|(f_k)_{[n-1]\to\mathbf{x}}\|_2 \right\|_{2,\mathbf{x}} = \|f_k\|_2$. $\square$

Proof of Proposition 7.11. As $q \geq 2$ is even, we wish to upper bound

$$\|g\|_q^q = \mathbb{E}[|g|^q] = \mathbb{E}[g^q] = \sum_{S_1,\dots,S_q} \mathbb{E}\left[\prod_{i=1}^{q} g^{=S_i} \right]$$

by

$$\sum_{S_1,\dots,S_q} \mathbb{E}\left[\prod_{i=1}^{q} \chi_{S_i}\right] \prod_{i=1}^{q} \|g^{=S_i}\|_2.$$

We upper bound each term participating in the expansion of g^q by the corresponding term in $\tilde{g}^q$. In the case the sets S_i cover some element exactly once, Lemma 7.12 implies that both terms are 0. Otherwise, the sets S_i cover each element either 0 times or at least 2 times; let T_i be the set of elements of $S_1, \dots, S_q$ appearing in exactly i of the sets (as in Lemma 7.13). By the assumption of the proposition, we have $\mathbb{E}\left[\prod_{i=1}^{q} \chi_{S_i}\right] \geq \prod_{i=3}^{q} \sigma_{T_i}^{2-|T_i|}$. The proof is concluded by combining this with the upper bound on $\mathbb{E}\left[\prod_{i=1}^{q} g^{=S_i}\right]$ following from Lemma 7.13 with $f_i = g^{=S_i}$. □

Proof of Theorem 7.10. Let $\sigma_i' = \sqrt{(p_i/4)(1-p_i/4)}$. We choose χ_i to be the $\frac{p_i}{4}$-biased character, $\chi_i = \frac{x_i - p_i/4}{\sigma_i'}$. Clearly χ_i has mean 0 and variance 1. We also claim that $\mathbb{E}\left[\chi_i^j\right] \geq (\sigma_i)^{2-j}$ for all integer $j > 2$. Indeed,

$$\mathbb{E}\left[\chi_i^j\right] \geq \frac{p_i}{4}\left(\frac{1-p_i/4}{\sigma_i'}\right)^j \geq \frac{(1-p_i/4)^{j-1}}{(\sigma_i')^{j-2}} \geq (1-p_i/4)\left(2\sqrt{1-p_i/4}\sqrt{1-p_i}\right)^{j-2}\sigma_i^{2-j},$$

which is at least σ_i^{2-j} as $p_i \leq 1/2$. Hence all of the conditions of Proposition 7.11 hold.

Denote $\sigma_S' = \prod_{i\in S}\sigma_i'$ and set $h = T_{\frac{1}{2}} f$, $g = \mathrm{T}_{\frac{1}{2q^{1.5}}} h$. By Proposition 7.11 and Theorem 7.7 we have

$$\|\mathrm{T}_{\frac{1}{4q^{1.5}}} f\|_q^q = \|g\|_q^q \leq \|\tilde{g}\|_q^q \leq \sum_S (\sigma_S')^{2-q}\|\mathrm{D}_S[\tilde{h}]\|_2^q.$$

We note that by Parseval, the 2-norm of $\tilde{h}$ and its derivatives are equal to the 2-norm of h and its Laplacians, and thus the last sum is equal to

$$\sum_S (\sigma_S')^{2-q}\|\mathrm{L}_S[h]\|_2^q \leq \sum_S (\sigma_S)^{2-q}\|\mathrm{L}_S[f]\|_2^q.$$

In the last inequality we used $\sigma_S' \geq 2^{-|S|}\sigma_S$ and $\|\mathrm{L}_S[h]\|^q \leq 2^{-q|S|}\|\mathrm{L}_S[f]\|_2^q$ (which follows from Parseval). This completes the proof of the theorem. □

8 An invariance principle (for global functions)

Invariance (also known as Universality) is a fundamental paradigm in Probability, describing the phenomenon that many random processes converge to a specific distribution that is the same for many different instances of the process. The prototypical example is the Berry-Esseen Theorem, giving a quantitative version of the Central Limit Theorem (see e.g. [64, Section 11.5]). More sophisticated instances of the phenomenon that have been particularly influential on recent research in several areas of Mathematics include the universality of Wigner's semicircle law for random matrices (see [59]) and of Schramm–Loewner evolution (SLE) e.g. in critical percolation (see [69]).

In the context of the cube, the Invariance Principle is a powerful tool developed by Mossel, O'Donnell and Oleszkiewicz [63] while proving their 'Majority is Stablest' Theorem, which can be viewed as an isoperimetric theorem for the noise operator. Roughly speaking, the result (in a more general form due to Mossel [61]) is that 'majority functions' (characteristic functions of Hamming balls) minimise noise sensitivity among functions that are 'far from being dictators'. The Invariance Principle converts many problems on the cube to equivalent problems in Gaussian Space; in particular, 'Majority is Stablest' is converted into an isoperimetric problem in Gaussian Space which was solved by a classical theorem of Borell [18] (half-spaces are isoperimetric).

In the basic form (see [64, Section 11.6]) of the Invariance Principle, we consider a multilinear real-valued polynomial f of degree $\leq k$ and wish to compare $f(\mathbf{x})$ to $f(\mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are random vectors each having independent coordinates, according to a smooth (to third order) test function ϕ. (Comparison of the cumulative distributions requires ϕ to be a step function, but this can be handled by smooth approximation.) The version of [64, Remark 11.66] shows that if the coordinates x_i have mean 0, variance 1 and are suitably hypercontractive (satisfy $\|a+\rho b x_i\|_3 \leq \|a+bx_i\|_2$ for any $a, b \in \mathbb{R}$), and similarly for y_i, then

$$\left|\mathbb{E}[\phi(f(\mathbf{x}))] - \mathbb{E}[\phi(f(\mathbf{y}))]\right| \leq \tfrac{1}{3}\|\phi'''\|_\infty \rho^{-3k} \sum_{i\in[n]} \mathrm{I}_i(f)^{3/2}. \tag{8.1}$$

The hypercontractivity assumption applies e.g. if the coordinates are standard Gaussians or p-biased bits (renormalised to have mean 0 and variance 1) with p bounded away from 0 or 1, but if $p = o(1)$ then we need $\rho = o(1)$, in which case their theorem becomes ineffective. We will apply our hypercontractivity inequality to obtain an invariance principle that is effective for small probabilities and functions with small generalised influences. We adopt the following setup.

Setup 8.1. Let $\sigma_1, \ldots, \sigma_n > 0$, let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ be random vectors with independent coordinates, where X_i and Y_i are real-valued random variables with mean 0, variance 1, and satisfy $\|X_i\|_3^3 \leq \sigma_i^{-1}$ and $\|Y_i\|_3^3 \leq \sigma_i^{-1}$. Let $f \in \mathbb{R}[v]$ be a multilinear polynomial of degree d in n variables $v = (v_1, \ldots, v_n)$. Let $\phi \in C^3(\mathbb{R})$ be continuously thrice differentiable.

For $S \subset [n]$ we write $\hat{f}(S)$ for the coefficient in f of $v_S = \prod_{i\in S} v_i$. We write $W_S(f) = \sum_{J:S\subset J} \hat{f}(J)^2$ and similarly to Section 7.1 we define the generalised influences by $\mathrm{I}_S[f] = W_S(f)\prod_{i\in S}\sigma_i^{-2}$.

We write $\mathrm{T}_\rho[f] = \sum_{S\subset[n]} \rho^{|S|}\hat{f}(S)v_S$.

Now we state our invariance principle, which compares $f(\mathbf{X})$ to $f(\mathbf{Y})$.

Theorem 8.2. *Under Setup 8.1, if $\mathrm{I}_S[f] \leq \epsilon$ for each nonempty set S, then*

$$|\mathbb{E}[\phi(f(\mathbf{X}))] - \mathbb{E}[\phi(f(\mathbf{Y}))]| \leq 2^{12d}\|\phi'''\|_\infty W_\emptyset(f)\sqrt{\epsilon}.$$

The term $W_\emptyset(f)$ can be replaced by either $\mathbb{E}[f(\mathbf{X})^2]$ or $\mathbb{E}[f(\mathbf{Y})^2]$ as they are all equal.

Theorem 8.2 can be informally interpreted as saying that if a multilinear, low degree polynomial f is global then the distribution of $f(\mathbf{X})$ is essentially independent of the distribution of $\mathbf{X}$ given the mean and variance of each coordinate. In particular, it does not make much difference whether we plug in p-biased characters or uniform characters. A posteriori, this may be seen as an intuitive explanation for Theorem 1.3 given the standard hypercontractivity theorem for the uniform cube.

Next, we set up some notations and preliminary observations for the proof of Theorem 8.2. Throughout we fix $\mathbf{X}$, $\mathbf{Y}$, f, and ϕ as in Setup 8.1. We write $\mathbf{X}_S = \prod_{i\in S}\mathbf{X}_i$, and similarly for $\mathbf{Y}$. Recall that $f = \sum_S \hat{f}(S)v_S$ is a (formal) multilinear polynomial in $\mathbb{R}[v]$ of degree d. Note that $f(\mathbf{X}) = \sum_S \hat{f}(S)\mathbf{X}_S$ has $\mathbb{E}[f(\mathbf{X})^2] = \sum_S \hat{f}(S)^2$, as $\mathbb{E}\mathbf{X}_S^2 = 1$ and $\mathbb{E}[\mathbf{X}_S\mathbf{X}_T] = 0$ for $S \neq T$. The random variable $f(\mathbf{X})$

has the orthogonal decomposition $f = \sum_S f^{=S}$ with each $f^{=S} = \hat{f}(S)\mathbf{X}_S$. Further note that $\mathrm{L}_S f(\mathbf{X}) = \sum_{J:S\subset J} \hat{f}(J)\mathbf{X}_J$ so we have the identities

$$\mathrm{I}_S[f]\prod_{i\in S}\sigma_i^2 = \mathbb{E}[(\mathrm{L}_S f(\mathbf{X}))^2] = \mathbb{E}[(\mathrm{L}_S f(\mathbf{Y}))^2] = \sum_{J:S\subset J}\hat{f}(J)^2 = W_S(f).$$

We apply the replacement method as in Section 3 (and as in the proof of the original invariance principle by Mossel, O'Donnell and Oleszkiewicz [63]). For $0 \le t \le n$, define $\mathbf{Z}^{:t} = (\mathbf{Z}_1^{:t}, \ldots, \mathbf{Z}_n^{:t}) = (\mathbf{Y}_1, ..., \mathbf{Y}_t, \mathbf{X}_{t+1}, ..., \mathbf{X}_n)$, and note that $f(\mathbf{Z}^{:t})$ has the orthogonal decomposition $f(\mathbf{Z}^{:t}) = \sum_S f(\mathbf{Z}^{:t})^{=S}$ with

$$f(\mathbf{Z}^{:t})^{=S} = \hat{f}(S)\mathbf{Z}_S = \hat{f}(S)\mathbf{Y}_{S\cap[t]}\mathbf{X}_{S\setminus[t]}.$$

Proof of Theorem 8.2. We adapt the exposition in [64, Section 11.6]. As $\mathbf{Z}^{:0} = \mathbf{X}$ and $\mathbf{Z}^{:n} = \mathbf{Y}$ we have by telescoping and the triangle inequality

$$|\mathbb{E}[\phi(f(\mathbf{X}))] - \mathbb{E}[\phi(f(\mathbf{Y}))]| \le \sum_{t=1}^{n} |\mathbb{E}[\phi(f(\mathbf{Z}^{:t-1}))] - \mathbb{E}[\phi(f(\mathbf{Z}^{:t}))]|.$$

Consider any $t \in [n]$ and write

$$f(\mathbf{Z}^{:t-1}) = U_t + \Delta_t\mathbf{X}_t \quad\text{and}\quad f(\mathbf{Z}^{:t}) = U_t + \Delta_t\mathbf{Y}_t, \quad\text{where}$$
$$U_t = \mathrm{E}_t f(\mathbf{Z}^{:t-1}) = \mathrm{E}_t f(\mathbf{Z}^{:t}) \quad\text{and}\quad \Delta_t = \mathrm{D}_t f(\mathbf{Z}^{:t-1}) = \mathrm{D}_t f(\mathbf{Z}^{:t}).$$

Both of the functions U_t and Δ_t are independent of the random variables X_t and Y_t.

By Taylor's Theorem,

$$\begin{aligned}\phi(f(\mathbf{Z}^{:t-1})) &= \phi(U_t) + \phi'(U_t)\Delta_t\mathbf{X}_t + \tfrac{1}{2}\phi''(U_t)(\Delta_t\mathbf{X}_t)^2 + \tfrac{1}{6}\phi'''(A)(\Delta_t\mathbf{X}_t)^3, \quad\text{and}\\ \phi(f(\mathbf{Z}^{:t})) &= \phi(U_t) + \phi'(U_t)\Delta_t\mathbf{Y}_t + \tfrac{1}{2}\phi''(U_t)(\Delta_t\mathbf{Y}_t)^2 + \tfrac{1}{6}\phi'''(A')(\Delta_t\mathbf{Y}_t)^3,\end{aligned}$$

for some random variables A and A'. As $\mathbf{X}_t$ and $\mathbf{Y}_t$ have mean 0 and variance 1 we have $0 = \mathbb{E}[\phi'(U_t)\Delta_t\mathbf{Y}_t] = \mathbb{E}[\phi'(U_t)\Delta_t\mathbf{X}_t]$ and $\mathbb{E}[\phi''(U_t)(\Delta_t)^2] = \mathbb{E}[\phi''(U_t)(\Delta_t\mathbf{Y}_t)^2] = \mathbb{E}[\phi''(U_t)(\Delta_t\mathbf{X}_t)^2]$, so

$$|\mathbb{E}[\phi(f(\mathbf{Z}^{:t-1}))] - \mathbb{E}[\phi(f(\mathbf{Z}^{:t}))]| \le \tfrac{1}{6}\|\phi'''\|_\infty(\mathbb{E}[|\Delta_t\mathbf{X}_t|^3] + \mathbb{E}[|\Delta_t\mathbf{Y}_t|^3]) \le \tfrac{1}{3}\|\phi'''\|_\infty\sigma_t^{-1}\|\Delta_t\|_3^3.$$

In the last inequality, we have viewed the expectation $\mathbb{E}[|\Delta_t\mathbf{X}_t|^3]$ (and similarly the expectation $\mathbb{E}[|\Delta_t\mathbf{Y}_t|^3]$) as being over $\mathbf{X}_t$ and over all of the coordinates in $\mathbf{Z}^{:t-1}$ except for its tth coordinate, noting $\mathbf{X}_t$ depends only on the former random variable whereas Δ_t depends only on the latter random variables. The function Δ_t is the function $\mathrm{D}_t[f]$ applied on random variables satisfying the hypothesis of Lemma 7.9 for $q = 3$. Moreover, $\mathrm{I}_S[\mathrm{D}_t[f]]$ is either 0 when $t \in S$, or $\sigma_t^2\mathrm{I}_{S\cup\{t\}}[f]$ when $t \notin S$, in which case $\mathrm{I}_S[D_t f] \le \sigma_t^2\epsilon$. Hence, by Lemma 7.9 (with $q = 3$), we obtain

$$\|\Delta_t\|_3^3 \le 6^{4.5d}\sigma_t\sqrt{\epsilon}\|\Delta_t\|_2^2 = 6^{4.5d}\sigma_t\sqrt{\epsilon}\cdot\sum_{S\ni t}\hat{f}(S)^2.$$

Hence,

$$\sum_{t=0}^{n}\tfrac{1}{3}\|\phi'''\|_\infty\sigma_t^{-1}\|\Delta_t\|_3^3 \le 6^{4.5d}\sqrt{\epsilon}\tfrac{1}{3}\|\phi'''\|_\infty\sum_S|S|\hat{f}(S)^2 \le 6^{4.5d}\sqrt{\epsilon}\tfrac{d}{3}\|\phi'''\|_\infty W_\emptyset(f).$$

This completes the proof of the theorem since $6^{4.5d}\frac{d}{3} \le 2^{12d}$. □

8.1 Applications of the Invariance Principle

As mentioned in the Introduction, one consequence of our Invariance Principle is a variant of the 'Majority is Stablest' Theorem of Mossel, O'Donnell and Oleszkiewicz [63] (see also [61]). We omit the proof of Corollary 1.10), as it goes along the same lines of [61] (see also [64, Chapter 11]).

As an additional application, one can obtain the following sharp threshold result for *almost* monotone Boolean functions. This statement asserts that any such function which is global has a sharp threshold. Let us remark that we have already established such a result in the sparse regime (see Section 6). On the other hand, the version below applies in the *dense* regime.

With notation as in Section 6, we say that f is (δ, p, q)-almost monotone if $p < q \in (0, 1)$ and choosing $\mathbf{x}, \mathbf{y} \sim D(p, q)$ gives $\Pr[f(\mathbf{y}) = 0, f(\mathbf{x}) = 1] < \delta$. We say that f has an ϵ-coarse threshold in an interval $[p, q]$ if $\mu_p(f) > \epsilon$ and $\mu_q(f) < 1 - \epsilon$.

Corollary 8.3. *For each $\epsilon > 0$, there exists $\delta > 0$, such that the following holds. Let $p < q < \frac{1}{2}$, and suppose that $q > (1 + \epsilon)p$. Let f be a (δ, p, q)-almost monotone Boolean function having an ϵ-coarse threshold in an interval $[p, q]$. Then there exists a set S of size at most $\frac{1}{\delta}$, such that $\mathrm{I}_S[f] \geq \delta$ either with respect to the p-biased measure or with respect to the q-biased measure.*

The proof is similar to the one given by Lifshitz [56], so we only sketch it.

Proof sketch. First we observe that Corollary 1.10 extends to the one sided noise operator. Let $f_1 = f$ be the function viewed as a function on the p-biased cube, and let $f_2 = f$ be the function viewed as a function on the q-biased cube. So assuming for contradiction that $\mathrm{I}_S[f] \leq \delta$ for each S, we obtain an upper bound on $\langle \mathrm{T}^{p \to q} f_1, f_2 \rangle_{\mu_q}$ of the form $\langle \mathrm{T}^{p \to q} H_{\mu_p(f)}, H_{\mu_q(f)} \rangle_{\mu_q}$

However, the (δ, p, q)-almost monotonicity of f implies the lower bound $\langle \mathrm{T}^{p \to q} f_1, f_2 \rangle_{\mu_q} \rangle \geq \mu_p(f) - \delta$.

Standard estimates on $\langle \mathrm{T}^{p \to q} H_{\mu_p(f)}, H_{\mu_q(f)} \rangle_{\mu_q}$ show that the lower bound and the upper bound cannot coexist provided that δ is sufficiently small (see [56]). □

9 Concluding remarks

We are optimistic that our sharp threshold result in the sparse regime will have many applications in the same vein as the applications of the classical sharp threshold results, e.g. to Percolation [12], Complexity Theory [29], Coding Theory [55], and Ramsey Theory [30].

In particular, despite the recent solution of the Kahn–Kalai Threshold Conjecture, there remain challenging open problems on thresholds that are potentially amenable to our sharp threshold theorem (Theorem 1.6).

Our variant of the Kahn–Kalai Isoperimetric Conjecture is only effective in the p-biased setting for small p, whereas the corresponding known results [52, 50] for the uniform measure are substantial weaker. This leaves our current state of knowledge in a rather peculiar state, as in many related problems the small p case seems harder than the uniform case! A natural open problem is give a unified approach extending both results for all p.

Our final open problem is to obtain a generalisation of Hatami's Theorem to the sparse regime, i.e. to obtain a density increase from $\mu_p(f) = o(1)$ to $\mu_q(f) \geq 1 - \varepsilon$ under some pseudorandomness condition on f; we expect that a such result would have profound consequences in Extremal Combinatorics.

Acknowledgment

We would like to thank Yuval Filmus, Ehud Friedgut, Gil Kalai, Nathan Keller, Guy Kindler, Muli Safra and the anonymous referees for various helpful comments and suggestions.

References

[1] Amirali Abdullah and Suresh Venkatasubramanian. A directed isoperimetric inequality with application to bregman near neighbor lower bounds. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 509–518. ACM, 2015.

[2] Dimitris Achlioptas and Ehud Friedgut. A sharp threshold for k-colorability. *Random Structures & Algorithms*, 14(1):63–70, 1999.

[3] Daniel Ahlberg, Erik Broman, Simon Griffiths, and Robert Morris. Noise sensitivity in continuum percolation. *Israel Journal of Mathematics*, 201(2):847–899, 2014.

[4] Rudolf Ahlswede and Levon H Khachatrian. The diametric theorem in hamming spaces—optimal anticodes. *Advances in Applied mathematics*, 20(4):429–449, 1998.

[5] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795–815, 2021.

[6] László Babai and Vera T Sós. Sidon sets in groups and induced subgraphs of cayley graphs. *European Journal of Combinatorics*, 6(2):101–114, 1985.

[7] Mitali Bafna, Boaz Barak, Pravesh K. Kothari, Tselil Schramm, and David Steurer. Playing unique games on certified small-set expanders. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1629–1642, 2021.

[8] Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. High dimensional expanders: Eigenstripping, pseudorandomness, and unique games. In *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, pages 1069–1128, 2022.

[9] Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. Hypercontractivity on high dimensional expanders. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 185–194, 2022.

[10] William Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, pages 159–182, 1975.

[11] Michael Ben-Or and Nathan Linial. Collective coin flipping. *randomness and computation*, 5:91–115, 1990.

[12] Itai Benjamini, Stéphane Boucheron, Gábor Lugosi, and Raphaël Rossignol. Sharp threshold for percolation on expanders. *The Annals of Probability*, 40(1):130–145, 2012.

[13] Itai Benjamini and Jérémie Brieussel. Noise sensitivity of random walks on groups. *arXiv preprint arXiv:1901.03617*, 2019.

[14] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Etudes Sci. Publ. Math.*, 90:5–43, 1999.

[15] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 488–497, 2010.

[16] Béla Bollobás and Andrew G Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[17] Aline Bonami. Étude des coefficients de Fourier des fonctions de $l^p(g)$. In *Annales de l'institut Fourier*, volume 20(2), pages 335–402, 1970.

[18] Christer Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probability Theory and Related Fields*, 70(1):1–13, 1985.

[19] Jean Bourgain and Gil Kalai. Influences of variables and threshold intervals under group symmetries. *Geometric and Functional Analysis*, 7(3):438–461, 1997.

[20] Irit Dinur and Ehud Friedgut. Intersecting families are essentially contained in juntas. *Combinatorics, Probability & Computing*, 18(1-2):107–122, 2009.

[21] Irit Dinur, Ehud Friedgut, and Oded Regev. Independent sets in graph powers are almost contained in juntas. *Geometric and Functional Analysis*, 18(1):77–97, 2008.

[22] Sean Eberhard. Product mixing in the alternating group. *arXiv preprint arXiv:1512.03517*, 2015.

[23] Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596, 1981.

[24] David Ellis, Guy Kindler, and Noam Lifshitz. An analogue of bonami's lemma for functions on spaces of linear maps, and 2-2 games. *arXiv preprint arXiv:2209.04243*, 2022.

[25] Yuval Filmus, Guy Kindler, Noam Lifshitz, and Dor Minzer. Hypercontractivity on the symmetric group. *arXiv preprint arXiv:2009.05503*, 2020.

[26] Peter Frankl and Norihide Tokushige. The Erdős–Ko–Rado theorem for integer sequences. *Combinatorica*, 19(1):55–63, 1999.

[27] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 194(2):475–495, 2021.

[28] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[29] Ehud Friedgut. Sharp thresholds of graph properties, and the k-sat problem (with an appendix by Jean Bourgain). *Journal of the American Mathematical Society*, 12(4):1017–1054, 1999.

[30] Ehud Friedgut, Hiêp Hàn, Yury Person, and Mathias Schacht. A sharp threshold for Van der Waerden's theorem in random subsets. *Discrete Analysis*, 7:19, 2016.

[31] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American mathematical Society*, 124(10):2993–3002, 1996.

[32] Zoltán Füredi, Tao Jiang, and Robert Seiver. Exact solution of the hypergraph Turán problem for k-uniform linear paths. *Combinatorica*, 34(3):299–322, 2014.

[33] N. Fusco, F. Maggi, and A. Pratelli. The sharp quantitative isoperimetric inequality. *Annals of Math.*, 168(3):941–980, 2008.

[34] William T Gowers. Quasirandom groups. *Combinatorics, Probability and Computing*, 17(3):363–387, 2008.

[35] L. Gross. Logarithmic Sobolev inequalities. *American J. Math.*, 97:1061–1083, 1975.

[36] Tom Gur, Noam Lifshitz, and Siqi Liu. Hypercontractivity on high dimensional expanders. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 176–184, 2022.

[37] Elad Haramaty, Amir Shpilka, and Madhu Sudan. Optimal testing of multivariate polynomials over small prime fields. *SIAM J. Comput.*, 42(2):536–562, 2013.

[38] Hamed Hatami. A structure theorem for Boolean functions with small total influences. *Annals of Mathematics*, 176(1):509–533, 2012.

[39] Hao Huang, Po-Shen Loh, and Benny Sudakov. The size of a hypergraph and its matching number. *Combinatorics, Probability and Computing*, 21(03):442–450, 2012.

[40] Vishesh Jain and Huy Tuan Pham. Optimal thresholds for latin squares, steiner triple systems, and edge colorings. *arXiv preprint arXiv:2212.06109*, 2022.

[41] Anders Johansson, Jeff Kahn, and Van Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[42] Jeff Kahn and Gil Kalai. Thresholds and expectation thresholds. *Combinatorics, Probability and Computing*, 16(03):495–502, 2007.

[43] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *Foundations of Computer Science, 1988., 29th Annual Symposium on*, pages 68–80. IEEE, 1988.

[44] Tali Kaufman and Dor Minzer. Improved optimal testing results from global hypercontractivity. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 98–109, 2022.

[45] Peter Keevash. The optimal edge-colouring threshold. *arXiv preprint arXiv:2212.04397*, 2022.

[46] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Sharp thresholds and expanded hypergraphs. 2019.

[47] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Forbidden intersections for codes. *arXiv preprint arXiv:2103.05050*, 2021.

[48] Peter Keevash, Noam Lifshitz, and Dor Minzer. On the largest product-free subsets of the alternating groups. *arXiv preprint arXiv:2205.15191*, 2022.

[49] Peter Keevash and Eoin Long. Stability for vertex isoperimetry in the cube. *arXiv:1807.09618*, 2018.

[50] Peter Keevash and Eoin Long. A stability result for the cube edge isoperimetric inequality. *J. Combin. Theory Ser. A*, 155:360–375, 2018.

[51] Nathan Keller and Noam Lifshitz. The junta method for hypergraphs and Chvátal's simplex conjecture. *arXiv preprint arXiv:1707.02643*, 2017.

[52] Nathan Keller and Noam Lifshitz. Approximation of biased Boolean functions of small total influence by DNF's. *Bulletin of the London Mathematical Society*, 50(4):667–679, 2018.

[53] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Small set expansion in the Johnson graph. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2018.

[54] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 592–601. IEEE, 2018.

[55] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, Eren Sasoglu, and Rüdiger L. Urbanke. Reed–Muller codes achieve capacity on erasure channels. *IEEE Transactions on Information Theory*, 63(7):4298–4316, 2017.

[56] Noam Lifshitz. Hypergraph removal lemmas via robust sharp threshold theorems. *arXiv preprint arXiv:1804.00328*, 2018.

[57] Eyal Lubetzky and Jeffrey Steif. Strong noise sensitivity and random graphs. *The Annals of Probability*, 43(6):3239–3278, 2015.

[58] G. Margulis. Probabilistic characteristic of graphs with large connectivity. In *Problems Info. Transmission*. Plenum Press, 1977.

[59] Madan Lal Mehta. *Random matrices*. Elsevier, 2004.

[60] Richard Montgomery. Spanning trees in random graphs. *arXiv:1810.03299*, 2018.

[61] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.

[62] Elchanan Mossel and Joe Neeman. Robust optimality of Gaussian noise stability. *J. Europ. Math. Soc.*, 17(2):433–482, 2015.

[63] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, pages 295–341, 2010.

[64] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

[65] Jinyoung Park and Huy Tuan Pham. A proof of the kahn-kalai conjecture. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 636–639. IEEE, 2022.

[66] Michał Przykucki and Alexander Roberts. Vertex-isoperimetric stability in the hypercube. *arXiv:1808.02572*, 2018.

[67] Lucio Russo. An approximate zero-one law. *Probability Theory and Related Fields*, 61(1):129–139, 1982.

[68] Oded Schramm and Jeffrey E Steif. Quantitative noise sensitivity and exceptional times for percolation. *Annals of mathematics*, 171(2):619–672, 2010.

[69] Stanislav Smirnov. Critical percolation and conformal invariance. *Proc. ICM*, 2006.

[70] M Talagrand. Approximate 0-1 law. *Ann. Prob*, 22:1576–1587, 1994.