Electronic Colloquium on Computational Complexity, Report No. 180 (2017)

Low degree almost Boolean functions are sparse juntas

Irit Dinur* Yuval Filmus† Prahladh Harsha‡

November 19, 2017

Abstract

Nisan and Szegedy showed that low degree Boolean functions are juntas. Kindler and Safra showed that low degree functions which are *almost* Boolean are close to juntas. Their result holds with respect to μ_p for every *constant* p. When p is allowed to be very small, new phenomena emerge. For example, the function $y_1 + \cdots + y_{\varepsilon/p}$ (where $y_i \in \{0, 1\}$) is close to Boolean but not close to a junta.

We show that low degree functions which are almost Boolean are close to a new class of functions which we call *sparse juntas*. Roughly speaking, these are functions which on a random input look like juntas, in the sense that only a finite number of their monomials are non-zero. This extends a result of the second author for the degree 1 case.

As applications of our result, we show that low degree almost Boolean functions must be very biased, and satisfy a large deviation bound.

An interesting aspect of our proof is that it relies on a local-to-global agreement theorem. We cover the p-biased hypercube by many smaller dimensional copies of the uniform hypercube, and approximate our function locally via the Kindler–Safra theorem for constant p. We then stitch the local approximations together into one global function that is a sparse junta.

*Weizmann Institute of Science, ISRAEL. email: irit.dinur@weizmann.ac.il.
†Technion Israel Institute of Technology, ISRAEL. email: yuvalfi@cs.technion.ac.il
‡Tata Institute of Fundamental Research, INDIA. email: prahladh@tifr.res.in

ISSN 1433-8092

1 Introduction

We study the structure of "simple" Boolean functions in the p-biased hypercube, for all values of p, and in particular when $p(n) \to 0$. We introduce a new class of functions that we call *sparse juntas* which generalize the standard juntas. Our main result is that every Boolean function that has at most ε of its ℓ_2^2 mass above degree d, is $O(\varepsilon)$ close to a degree d sparse junta. Throughout the paper we say that f is ε-close to g if $\|f-g\|_{\mu_p}^2 := \mathbb{E}_{x\sim\mu_p}[(f(x)-g(x))^2] \le \varepsilon$.

Nisan and Szegedy showed that Boolean functions that are exactly low degree must be juntas [NS94], namely functions that depend on a constant number of coordinates. Classical theorems in the analysis of Boolean functions describe the structure of Boolean functions that are close to being "simple" functions, where closeness is measured with respect to the uniform measure. Notions of "simple" include functions that are noise-stable, or nearly low degree, or have low total-influence [Fri98, FKN02, Bou02, KS02]. These results invariably prove that the function depends on a few coordinates (a dictator or a junta). For example, Friedgut, Kalai and Naor [FKN02] prove that a function whose ℓ_2 mass is almost all on Fourier levels $0, 1$ must be a function that depends on at most one variable (dictator, anti-dictator or constant). Bourgain [Bou02] and Kindler and Safra [KS02] studied Boolean functions with small mass on the Fourier levels above d. Kindler and Safra proved that such functions are close to juntas.

Theorem 1.1 (Kindler–Safra [KS02, Kin03])**.** *Fix $d \ge 0$. For every $0 < p < 1$ there exists $\varepsilon_0 = \varepsilon_0(p,d)$ such that for every $\varepsilon < \varepsilon_0$, if $f\colon \{\pm1\}^n \to \{\pm1\}$ satisfies $\|f^{>d}\|_{\mu_p}^2 = \varepsilon$ then there exists a degree d function $g\colon \{\pm1\}^n \to \{\pm1\}$ (which necessarily depends on $O_d(1)$ coordinates) satisfying $\|f-g\|_{\mu_p}^2 \le \varepsilon(1+O(\varepsilon^{1/4}(1/p)^d))$. In particular, when $\varepsilon = o(p^{4d})$ and $\varepsilon \le \varepsilon_0$, $\|f-g\|^2 = \varepsilon + o(\varepsilon)$.*

The term junta was actually coined in an earlier paper of Friedgut who proved that any Boolean function with small total influence is close to a junta [Fri98].

Theorem 1.1 is a generalization to degree d of the earlier theorem of Friedgut, Kalai and Naor [FKN02] mentioned above, which states that functions which are close to degree 1 are close to dictators or constants. An alternative way of saying this is that given a function f with only ε fraction of its ℓ_2^2-mass outside levels $0, 1$, if $f(x) \in \{\pm1\}$ for all $x \in \{\pm1\}^n$, then it must be $O(\varepsilon)$-close to a function g such that $\hat{g}(S) \in \{0, -1, 1\}$ for all $S \subseteq [n]$ (it is easy to verify that such Boolean functions are exactly the dictators, anti-dictators or the constant functions).

It is natural to wonder if the condition that the range is Boolean, namely $f(x) \in \{\pm1\}$ for all x, can be replaced by $f(x) \in A$ for all x, for any arbitrary finite set $A \subset \mathbb{R}$. What can be said about such a function that has ε of its mass outside levels $0, 1$? The answer becomes more complicated as the size of A grows, and the function need not depend on just one variable, as can be seen by the function $f(x) = x_1 + x_2$ that takes only three distinct values but depends on more than one variable. Nevertheless, we show that a similarly flavored statement is true: if the function f takes values in a finite set $A \subset \mathbb{R}$ and has only ε mass outside levels $0, \dots, d$, then there is a finite set $A' \subset \mathbb{R}$ such that f is close to a function whose Fourier coefficients belong to the set A'.

Theorem 1.2 (A-valued functions with low degree)**.** *Let $A \subset \mathbb{R}$ be a finite set, let $d \in \mathbb{N}$, and let $f\colon \{\pm1\}^n \to A$ be a function that has at most ε fraction of its ℓ_2^2-mass outside levels $0, \dots, d$, that is, $\|f^{>d}\|_{\mu_{1/2}}^2 < \varepsilon$. Then f is $O_{A,d}(\varepsilon)$-close to a function g of degree d with Fourier coefficients in a finite set $A' \subset \mathbb{R}$.*

This theorem is not difficult to prove given Theorem 1.1, but it turns out to be quite useful. In fact, generalizing from Boolean to A-valued allows us to give an new proof of Theorem 1.1 that proceeds by induction on the degree d (see Section 8).

Having warmed up, we turn to the main focus of this paper, which is understanding the structure of Boolean (or A-valued) functions that are nearly degree d in the p-biased hypercube. The p-biased hypercube is the set $\{\pm1\}^n$ equipped with the μ_p measure (given by $\mu_p(x) = p^{(n+\sum_i x_i)/2}(1-p)^{(n-\sum_i x_i)/2}$). We think of p as being possibly very small, for example $p = 1/\sqrt{n}$.

The theorem of Kindler and Safra [KS02] continues to hold under the μ_p measure, but the quality of the approximation deteriorates with p. Indeed, the class of junta functions does not seem to be the correct class of functions for approximating low degree functions that are μ_p-almost Boolean. This is demonstrated by the following simple example: Let $f(x) = \sum_{i\in S} \frac{1-x_i}{2}$ be a degree 1 function, and let g be the Boolean function closest to f. If $|S| = O(\sqrt{\varepsilon}/p)$ then g is ε-close to f, and yet it depends on many coordinates. It turns out that this example is canonical: in previous work [Fil16], the second named author has proved that all functions that are nearly degree one, in μ_p, essentially look like this one.

The function f considered above is very biased: with probability roughly $1-\sqrt{\varepsilon}$, it is equal to zero. More generally, the result of [Fil16] implies that if φ is a degree 1 function which is ε-close to Boolean then φ is $O(\sqrt{\varepsilon})$-close to a constant function. Still, one can hope for an even better approximation, with error of $O(\varepsilon)$, and indeed this is possible: φ is $O(\varepsilon)$-close to a linear function similar to the function f considered above (or its negation).

Generalizing this theorem to higher degrees requires coming up with a new syntactic class of simple functions that are the good approximators for low degree Boolean functions. As before, constants give an $O(\varepsilon^C)$-approximation for some $C < 1$, a fact which comes as a consequence of our main theorem (see Lemma 1.7), but our goal here will be to find an even better approximation, on the order of $O(\varepsilon)$. The first step is to move away from the Fourier basis whose basis functions depend on p and are thus non-canonical. Instead, we will rely on the *y-expansion* of f:

Definition 1.3 (y-expansion)**.** *The* y-expansion *of a function* $f\colon \{\pm 1\}^n \to \mathbb{R}$ *is the unique expansion* $f(x) = \sum_S \tilde{f}(S) y_S(x)$ *where* $\{y_S\}_S$ *is a basis of functions given by* $y_i = \frac{1-x_i}{2}$ *and for* $S \subseteq [n]$*, we define* $y_S = \prod_{i\in S} y_i$*.*

The y-expansion is the standard expansion of f as a multilinear polynomial in $\{0,1\}$ variables instead of $\{\pm 1\}$ variables. We stress that this is *not* the Fourier expansion of f (under $\mu_{1/2}$), which is its expansion as a multilinear polynomial in $\{\pm 1\}$ input variables. The y-expansion is better suited for working with μ_p for small p. The result mentioned above [Fil16] states that any degree 1 function that is close to being Boolean in the p-biased hypercube can be approximated by a function whose y-expansion coefficients are all in $\{-1, 0, 1\}$.

This motivates the following generalization:

Definition 1.4 (quantized polynomial)**.** *Given a finite set* $A \subset \mathbb{R}$*, a function* $f\colon \{\pm 1\}^n \to \mathbb{R}$ *is said to be an* A-quantized *polynomial of degree* d *if all coefficients of the y-expansion of* f *belong to* A*.*

As part of our main result, stated below as Theorem 1.5, we show that for all $p \le 1/2$, a low degree function that is ε-close under μ_p to being A-valued, is close to an A'-quantized polynomial for some finite set $A' = A'(d, A)$. This can be nicely rephrased as follows: For all $d \ge 0$ and sets A, there exists $A' = A'(d, A)$ such that for all $p \le 1/2$:

> *If a function has degree $\le d$ and is ε-close under μ_p to an A-valued function, then its y-expansion is $O(\varepsilon)$-close to being A'-quantized.*

Observe that the y-expansion is important for making such a statement. It could not be made for the Fourier expansion since the coefficients would have to depend on p.

This generalizes Theorem 1.2 above since in the uniform $\mu_{\frac{1}{2}}$ setting a quantized polynomial that has bounded norm *must* be a junta. Indeed, substituting $y_i = \frac{1-x_i}{2}$ shows that if $|S| = d$ then $\hat{f}(S) = (-1)^d 2^{-d} \tilde{f}(S)$ (where $\tilde{f}(S)$ is the coefficient of y_S in the y-expansion of f), and so Parseval's identity shows that there is a constant number of non-zero $\tilde{f}(S)$ with $|S| = d$. Removing them can only increase the ℓ_2 norm by a constant, and so applying the same reasoning inductively shows that f is a junta.

Our main theorem gives a somewhat stronger syntactic characterization, showing that A-valued functions with nearly low degree are close to being *sparse juntas*. These are quantized polynomials that have an additional structural property which we call bounded *branching factor*. The branching factor of a quantized polynomial g is best explained by considering the hypergraph whose edges correspond to all non-zero coefficients in the y-expansion of g. This hypergraph has branching factor $\rho = O(1/p)$ if for all subsets $A \subseteq [n]$ and integers $r \ge 0$, there are at most ρ^r hyperedges in H of cardinality $|A| + r$ containing A.

While this is the syntactic definition, the meaning of having small branching factor is that the function is "empirically" a junta, because a typical input only leaves a constant number of monomials non-zero. This is why we call these functions sparse juntas. Finally, we can state our main theorem:

Theorem 1.5 (Main)**.** *For every positive integer d and finite set $A \subset \mathbb{R}$, there exists a finite set $A' = A'(d, A)$ such that the following holds. For every $p \leq 1/2$ and $f : \{\pm 1\}^n \to \mathbb{R}$ of degree d there exists a function $g : \{\pm 1\}^n \to \mathbb{R}$ of degree d that satisfies the following properties for $\varepsilon := \mathbb{E}_{\mu_p}[\mathrm{dist}(f, A)^2]$:*

1. $\|f - g\|^2 = O(\varepsilon)$.
2. $\Pr[g \notin A] = O(\varepsilon)$
3. *g is an A'-**sparse junta**, that is, it is an A'-quantized polynomial of degree d with branching factor $O(1/p)$.*
4. *If $x \sim \mu_p$ then $g(x)$ is the sum of $O(1)$ coefficients of g with probability $1 - O(\varepsilon)$.*

We also show a converse to the above theorem (see Lemma 6.1) in the sense that the second and third properties are a complete characterization of degree d functions that are $O(\varepsilon)$-close to A (i.e., $\mathbb{E}_{\mu_p}[\mathrm{dist}(f, A)^2] = O(\varepsilon)$).

As applications of our theorem we show a large deviation bound for degree d functions close to a finite set A:

Lemma 1.6 (Large deviation bound)**.** *Fix an integer d and a finite set A. Suppose that $f : \{0, 1\}^n \to \mathbb{R}$ is a degree d function satisfying $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$ with respect to μ_p for some $p \leq 1/2$. For large t,*

$$\Pr[|f| \geq t] \leq \exp\left(-\Omega(t^{1/d}) + O(\varepsilon/t^2)\right).$$

We also prove that such functions must by very biased:

Lemma 1.7 (Sparse juntas are very biased)**.** *Fix a constant $d \geq 0$ and a finite set A. There exist constants $C, \varepsilon_0 > 0$ such that for all $p \leq 1/2$ and $\varepsilon \leq \varepsilon_0$, the following holds.*

Suppose that $g : \{0, 1\}^n \to \mathbb{R}$ is a degree d function with branching factor $O(1/p)$ such that $\Pr[g \notin A] = \varepsilon$. Then there exists $a \in A$ such that $\Pr[g \neq a] = O(\varepsilon^C + p)$.

Combining this with our main theorem implies that if an A-valued function is ε close to degree d, it must be very biased.

A local-to-global aspect of the proof

Let us highlight an interesting aspect of the proof of our main theorem. Previous works analyzing the structure of Boolean functions rely on hypercontractivity. When $p \to 0$ the hypercontractive behavior breaks down, and this is responsible for the deterioration of the approximation in Theorem 1.1. Our proof doesn't go down this path, and instead proceeds by breaking up the p-biased hypercube into many small sub-cubes that are obtained by setting many variables to 0 (using the $\{0, 1\}$ convention for the inputs). The measure on these sub-cubes becomes the uniform measure, and so we are able to approximate f locally on them using the classical Kindler-Safra theorem, Theorem 1.2. This gives us a separate junta function f_S on each sub-cube S. Moving from local to global, we rely on a recent so-called agreement theorem proven by the authors [DFH17] that gives us a single global function g that agrees with most of the local approximations (after ensuring that the local pieces typically agree with each other).

To complete the proof of our main theorem, we use a crucial feature of the agreement theorem proven in [DFH17], namely that *agreement is reached by consensus*. This means that each coefficient of the y-expansion of g is chosen by picking the most "popular" value appearing in all relevant f_S. In turns out that this feature guarantees that g has branching factor $O(1/p)$.

A new proof of the Kindler–Safra theorem

Our new proof of Theorem 1.1 demonstrates the power of our view of the theorem as stating that if a low degree function is close to being quantized, then its Fourier expansion is close to being quantized. Our inductive proof also makes essential use of the generalization to A-valued, rather than just Boolean, functions: even when starting with a Boolean function, A-valued functions arise in the proof.

Given a function f of degree d which is close to a finite set A, we use the theorem for degree $d - 1$ (assumed to hold by induction) together with the A-valued FKN theorem to show that the degree $d - 1$

and degree d coefficients are almost quantized (this is the heart of the proof). This allows us to replace the two highest levels of f with a quantized polynomial, which must be a junta. Removing these two levels altogether, we get an A'-valued function for some A' depending on A, d. Applying the theorem for degree $d-2$ completes the proof.

Related work

Understanding the structure of Boolean functions that are simple according to some measure such as being nearly low degree is a basic complexity goal. Similar structure theorems such as the KKL theorem [KKL88], Friedgut's junta theorem [Fri98], and the FKN theorem [FKN02], have found numerous applications. The analogous questions for the p-biased hypercube are understood only to some extent, yet the questions are natural and play an important role in several areas in combinatorics and the theory of computation.

- A major motivation for studying Boolean functions under the μ_p measure comes from trying to understand the sharp threshold behavior of graph properties, and of satisfiability of random k-CNF formulae.

  A large area of combinatorics is concerned with understanding properties of graphs selected from the random graph model of Erdős and Rényi, $G(n, p)$. A graph property is described via a Boolean function f whose $N = \binom{n}{2}$ input variables describe the edges of a graph and the function is 1 iff the property is satisfied. Selecting a graph at random from the $G(n, p)$ distribution is equivalent to selecting a random input to f with distribution μ_p. The density of this function is the probability that the property holds, and so its fine behavior as p increases from 0 to 1 is the business of sharp threshold theorems. For many of the most interesting graph properties, such as connectivity and appearance of a triangle, a phase transition occurs for very small values of p (corresponding to $p \approx 1/\sqrt{N}$). Friedgut and Kalai [FK96] used the theorem of Kahn, Kalai and Linial [KKL88] to prove that *every* monotone graph property has a narrow threshold.

  A famous theorem of Friedgut [Fri99] characterizes which graph and hypergraph properties have sharp threshold. As an application, Friedgut establishes the existence of a sharp threshold for the satisfiability of random k-CNF formulae. This is done through analyzing the structure of p-biased Boolean functions with low total influence, which corresponds to *not* having a sharp threshold. The same question was also studied by Bourgain [Bou99] and subsequently by Hatami [Hat12], who proved that such functions must be "pseudo-juntas" (see [O'D14, Chapter 10] for a discussion of these results). We recommend the nice recent survey [BK17, Section 3] for a description of some related questions and conjectures.

  Our condition of having nearly degree d is a strictly stronger condition than having low total influence, and indeed our sparse juntas are in particular pseudo-juntas. Unlike sparse juntas, the pseudo-junta property is not syntactic (it does not define a class of functions, but rather a property of the given function), and it is interesting to understand the relation between pseudo-juntas and sparse juntas.

  Friedgut conjectured that every monotone function that has a coarse threshold is approximable by a narrow DNF, which is a function that can be written as $f(x) = \max_{S:|S|\le d} \tilde{f}(S) y_S(x)$. This is quite similar to our class of sparse juntas (in fact, they coincide for degree $d = 1$), except that our functions are expressed as a sum of monomials rather than their maximum, and thus we must restrict ourselves to functions with bounded branching factor. The assumption of having a coarse threshold is weaker than having nearly degree d, yet it is interesting whether our techniques can be applied toward resolution of this conjecture.

- Hardness of approximation: The p-biased hypercube has been used as a gadget for proving hardness of approximation of vertex cover, where the relevant regime is some constant $p < 1/2$. Other variants of the hypercube have been used or suggested as gadgets for proving inapproximability, including the short code [BGH+15], the real code [KM13], and the Grassmann code [KMS17]. In all of these, understanding the structure of Boolean functions with nearly low degree seems crucial. In the Grassmann code, one considers subspaces of small dimension inside a large-dimensional vector space. Some conjectures were made in [DKK+16, DKK+17] regarding the structure of Boolean functions whose domain is the set of subspaces and that have non-negligible mass on the space of functions that corresponds to having low degree. Thinking of subspaces as subsets of points, this is analogous to the p-biased case, when p is very small, on the order of $O(1/n)$. Toward understanding that question, it is natural to first pursue such a study on the simpler model of the p-biased hypercube for very small p, where the analysis is potentially easier since the space is a product space.

- Relatively recent work [KKM+17] proves that Reed–Muller codes achieve capacity on the erasure channel, using the Bourgain–Kalai sharp threshold theorem for affine-invariant functions [BK97]. The regime of this result is only for codes with constant rate, and it seems that extending it to lower rates would require understanding the structure of affine-invariant functions under the p-biased measure for small p.

Organization

The rest of the paper is organized as follows. We begin with a few preliminaries in Section 2, which includes the agreement testing results. In Section 3, we define the branching factor and discuss some of its properties. We generalize the classical Kindler-Safra theorem to A-valued functions in Section 4. We then prove the main result of the paper (Theorem 1.5) in Section 5. In Section 6, we prove the converse to our main result. We discuss some applications in Section 7 and give an alternate proof to the classical Kindler-Safra theorem in Section 8.

2 Preliminaries

We will need the following definitions:

- We define $\mathrm{dist}(x, A) = \min_{y \in A} |x - y|$.
- We define $\mathrm{round}(x, A)$ as an element in A whose distance from x is $\mathrm{dist}(x, A)$.
- For a function $f\colon \{0,1\}^n \to \mathbb{R}$ and a set $S \subseteq [n]$, the function $f|_S\colon \{0,1\}^S \to \mathbb{R}$ results from substituting zero to all coordinates outside of S.
- For a function $f\colon \{0,1\}^n \to \mathbb{R}$, the support of its y-expansion naturals corresponds to a hypergraph $H_f \subset 2^{[n]}$ which we sometimes refer to as the support of g.
- For a set S, $\mu_p(S)$ is a distribution over subsets of S in which each element of S is chosen independently with probability p.
- The L_2^2 triangle inequality states that $(a+b)^2 \le 2(a^2+b^2)$. It implies that

$$\mathrm{dist}(x+y, A)^2 = \min_{a \in A}(x+y-a)^2 \le \min_{a \in A}[2(x-a)^2 + 2y^2] \le 2\,\mathrm{dist}(x, A)^2 + 2y^2.$$

- For any $p, q \in (0,1)$ satisfying $2p - pq \le 1$, the distribution $\mu_{p,q}$ is defined to be the distribution on pairs S_1, S_2 in which each element belongs only to S_1 with probability $p(1-q)$, only to S_2 with probability $p(1-q)$, and to both S_1 and S_2 with probability pq.

We will need the following theorems.

Theorem 2.1 (Nisan–Szegedy). *If $f\colon \{0,1\}^n \to \{0,1\}$ is a degree k function, then f is a $k2^{k-1}$-junta.*

Theorem 2.2 ($(2,p)$ hypercontractivity). *Let $p \ge 2$, then for any function $f\colon \{0,1\}^n \to \mathbb{R}$ of degree at most k, we have $\|f\|_p \le (p-1)^{k/2} \cdot \|f\|_2$.*

We also need the following result about quantization.

Lemma 2.3. *For every finite set V and integer d there exists a finite set U such that the following holds. Suppose that $\deg g_1, \deg g_2 \le d$. If all coefficients of the y-expansion of g_1, g_2 belong to V, then all coefficients of the y-expansion of $g_1 g_2$ belong to U.*

Proof. Let $g := g_1 g_2$, and let $|A| \leq 2d$ (otherwise $\tilde{g}(A) = 0$). Since $y_{A_1} y_{A_2} = y_{A_1 \cup A_2}$, we have

$$\tilde{g}(A) = \bigcup_{A_1 \cup A_2 = A} \tilde{g_1}(A_1)\tilde{g_2}(A_2).$$

The lemma follows from the fact that the sum contains at most 3^{2d} terms. □

2.1 Agreement testing

Agreement tests are a type of PCP tests that capture local-to-global phenomena. Our proof of the main result uses an agreement test recently analyzed by the authors [DFH17], which is an extension of the direct product test to higher dimensions. In the standard direct product test, one is given a ground set $[n]$ and an ensemble of local functions $\{f_S\}_{S \subset [n]}$ containing a local function $f_S \colon S \to \{0,1\}$ for each subset $S \subset [n]$. The direct product test is specified by the distribution $\mu_{p,q}$ over pairs of sets (S_1, S_2), in which each element $i \in [n]$ is independently added to $S_1 \cap S_2$ with probability pq, to $S_1 \setminus S_2$ with probability $p(1-q)$, to $S_2 \setminus S_1$ with probability $p(1-q)$, and to neither set with probability $1-(2p-pq)$. Here, we assume $p \leq 1/2$ and $q \in (0,1)$. The direct product testing results [DG08, IKW12, DS14] state that if the local functions agree most of the time, ie.,

$$\Pr_{(S_1,S_2) \sim \mu_{p,q}} [f_{S_1}|_{S_1 \cap S_2} = f_{S_2}|_{S_1 \cap S_2}] = 1 - \varepsilon,$$

then there must exist a global function $G \colon [n] \to \{0,1\}$ that explains most of the local functions:

$$\Pr_{S \sim \mu_p} [f_S = G|_S] = 1 - O(\varepsilon).$$

In recent work [DFH17], the authors extended this direct product to higher dimensions, wherein the local functions are functions not only on the vertices of S but also on hyperedges supported by S, i.e., $f_S \colon \binom{S}{\leq d} \to \{0,1\}$ instead of $f_S \colon S \to \{0,1\}$. Furthermore, they demonstrated that the function obtained by majority decoding serves as a good candidate for the global function. Formally:

Theorem 2.4 (Agreement theorem via majority decoding)**.** *For every positive integer d and alphabet Σ, there exists a constant $p_0 \in (0, 1/2)$ such that for all $p \in (0, p_0)$ and $q \in (0,1)$ and sufficiently large n, the following holds. Let $\{f_S \colon \binom{S}{\leq d} \to \Sigma \mid S \in \{0,1\}^n\}$ be an ensemble of functions satisfying*

$$\Pr_{S_1,S_2 \sim \mu_{p,q}} [f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \varepsilon.$$

Then the global function $G \colon \binom{[n]}{\leq d} \to \Sigma$ defined by plurality decoding (ie., $G(T)$ is the most popular value of $f_S(T)$ over all S containing T, chosen according to the distribution $\mu_p([n])$, i.e., $\Pr_{S \sim \mu_p}[f_S(T) = G(T)] = \max_\sigma \Pr_{S \sim \mu_p}[f_S(T) = \sigma]$) satisfies

$$\Pr_{S \sim \mu_p} [f_S \neq G|_S] = O_{d,q}(\varepsilon).$$

3 Branching factor

The analog of juntas for small p are quantized functions with branching factor $O(1/p)$. Let us start by formally defining this concept,

Definition 3.1 (branching factor)**.** *For any $\rho \geq 1$, a hypergraph H over a vertex set V is said to have* branching factor ρ *if for all subsets $A \subset V$ and integers $k \geq 0$, there are at most ρ^k hyperedges in H of cardinality $|A| + k$ containing A.*

A function $g \colon \{0,1\}^n \to \mathbb{R}$ is said to have branching factor ρ if the corresponding hypergraph H_g (given by the support of the y-expansion of g) has branching factor ρ.

In what sense is a function with branching factor $O(1/p)$ similar to a junta? If f is a junta and $y \sim \mu_{1/2}$, then $f(y)$ is the sum of a bounded number of coefficients of the y-expansion of f. Let us call such a coefficient *live*. In other words, the coefficients left alive by S are all $\tilde{f}(S)$ for which $y_S = 1$.

We want a similar property to hold for a function f with respect to an input $y \sim \mu_p$ for small p. As a first approximation, we need the *expected* number of live coefficients to be bounded. If $\deg f = d$ then the expected number of live coefficients is

$$\sum_{e=0}^{d} p^e N_e, \text{ where } N_e = |\{|S| = e : \tilde{f}(S) \neq 0\}|.$$

This sum is bounded if $N_e = O(1/p^e)$ for all e. A drawback of this definition is that it is not closed under substitution: if the expected number of live coefficients of f is bounded, this doesn't guarantee the same property for $f|_{y_i=1}$. For example, consider the function

$$f = y_0(y_1 + \cdots + y_{1/p^2}).$$

While the expected number of live coefficients is $p^2/p^2 = 1$, if we substitute $y_0 = 1$ then the expected number of live coefficients jumps to $p/p^2 = 1/p$. The recursive nature of the definition of branching factor guarantees that this cannot happen.

Functions with branching factor $O(1/p)$ also have several other desirable properties, such as the large deviation bound proved in Section 7, and Lemma 3.4 below.

In the rest of this section we prove several elementary properties of the branching factor. We start by estimating the branching factor of a sum or product of functions.

Lemma 3.2. *Suppose that φ_1, φ_2 have degree d and branching factor ρ. Then $\varphi_1\varphi_2$ and $\varphi_1 + \varphi_2$ have branching factor $O(\rho)$, where the hidden constant depends on d.*

Proof. The claim about $\varphi_1 + \varphi_2$ is obvious, so let us consider $\varphi = \varphi_1\varphi_2$. Given A, e, we have to show that the number of non-zero coefficients in φ which extend A by e elements is $O(\rho^e)$.

If $\tilde{\varphi}(B) \neq 0$ then $B = B_1 \cup B_2$ for some B_1, B_2 such that $\tilde{\varphi}_i(B_i) \neq 0$. Let $B_1 = A_1 \cup C_1 \cup F$ and $B_2 = A_2 \cup C_2 \cup F$, where $A_1 \cup A_2 = A$, and C_1, C_2, F are disjoint and disjoint from A, so that $|C_1 \cup C_2 \cup F| = e$. Denote the sizes of C_1, C_2, F by c_1, c_2, f.

There are $O(1)$ options for A_1, A_2. Given A_1, there are at most ρ^{c_1+f} non-zero coefficients in φ_1 extending A_1 by $c_1 + f$ elements, and for each such extension, there are $O(1)$ options for F. Given A_2, F, there are at most ρ^{c_2} non-zero coefficients in φ_2 extending $A_2 \cup F$ by c_2 elements. In total, we deduce that for each of the $O(1)$ choices of c_1, c_2, f, the number of non-zero coefficients extending A by e elements is $O(1) \cdot \rho^{c_1+f} \cdot O(1) \cdot \rho^{c_2} = O(\rho^e)$. □

As mentioned above, substitution has a bounded effect on the branching factor.

Lemma 3.3. *If H has branching factor ρ then $H|_{A=\emptyset}$ has branching factor $2^{|A|}\rho$.*

Proof. It's enough to prove the theorem when $A = \{i\}$. Let B, k be given. We will show that the number of hyperedges in $H|_{i=\emptyset}$ extending B by k elements is at most $(2\rho)^k$. If $k = 0$ then this is clear. Otherwise, for each such hyperedge e, either e or $e + i$ belongs in H. The former case includes all hyperedges of H extending B by k elements, and the latter all hyperedges of H extending $B + i$ by k elements. Since H has branching factor ρ, we can upper bound the number of hyperedges by $2\rho^k \leq (2\rho)^k$. □

One of the crucial properties of functions with branching factor $O(1/p)$ is that given that a certain y-coefficient is live, there is constant probability that no other y-coefficient is live.

Lemma 3.4 (Uniqueness)**.** *Suppose that φ has branching factor $O(1/p)$ and degree $d = O(1)$, where $p \leq 1/2$. For every B, the probability that $y_B = 1$ and $y_A = 0$ for all $A \nsubseteq B$ in the support of φ is $\Omega(p^{|B|})$.*

Proof. Let H be the hypergraph formed by the support of φ (that is, C is a hyperedge if $\tilde{\varphi}(C) \neq 0$). Given that $y_B = 1$, the probability that $y_A = 0$ for all $A \nsubseteq B$ is exactly equal to $\Pr_{S \sim \mu_p}[(H|_{B=1} \setminus \{\emptyset\})|_S = \emptyset]$. Lemma 3.3 shows that $H|_{B=1}$ has branching factor $O(1/p)$, and so it has $O(p^{-e})$ hyperedges of size e. The probability that each such edge survives is $1 - p^e$, and so the FKG lemma shows that given that $y_B = 1$, the probability that $y_A = 0$ for all $A \nsubseteq B$ is at least

$$\prod_{e=1}^{d} (1 - p^e)^{O(p^{-e})} = \Omega(1).$$

This completes the proof, since $\Pr[y_B = 1] = p^{|B|}$. □

4 Generalized Kindler–Safra theorem to A-valued functions

In this section, we prove the following generalization of Kindler-Safra to quantized function (i.e, A-valued functions for some finite set A). Everything that follows holds with respect to μ_p for *fixed* $p \in (0,1)$. All hidden constants depend continuously on p.

Theorem 4.1. *For all integers d and finite sets A the following holds. If $f\colon \{0,1\}^n \to \mathbb{R}$ is a degree d and $\varepsilon := \mathbb{E}[\operatorname{dist}(f,A)^2]$ then f is $O(\varepsilon)$-close to a degree d function $g\colon \{0,1\}^n \to A$.*

We start with the following easy claim which is an easy consequence of the Nisan-Szegedy theorem (Theorem 2.1).

Claim 4.2. *For all integers d and finite sets A there exists M such that the following holds. If $f\colon \{0,1\}^n \to A$ has degree d then f depends on at most M coordinates.*

Proof. For all $a \in A$, define

$$f_a = \prod_{b \neq a} \frac{f-b}{a-b}.$$

The function f_a has degree at most $d(|A|-1)$ and is Boolean, and so it depends on at most M_0 coordinates. Since

$$f = \sum_{a \in A} a f_a,$$

we see that f depends on at most $M_0|A|$ coordinates. □

Suppose we are dealing with degree d functions which are close to some finite set A (ie., $\mathbb{E}[\operatorname{dist}(h,A)^2] = O(\varepsilon)$) and we wish to show that $\|h\|^2 = O(\varepsilon)$. The following trick (using hypercontractivity Theorem 2.2) shows that is suffices to show $\|h\|^2 = O(\varepsilon^\alpha)$ for some $\alpha < 1$.

Claim 4.3. *Fix an integer d, a finite set A, and an exponent $\alpha < 1$. If $h\colon \{0,1\}^n \to \mathbb{R}$ is a degree d function satisfying $\mathbb{E}[\operatorname{dist}(h,A)^2] = O(\varepsilon)$ and $\|h\|^2 = O(\varepsilon^\alpha)$ then $\|h\|^2 = O(\varepsilon)$.*

Proof. We can assume that $\varepsilon \leq 1$, since otherwise the theorem is trivial. Similarly, we can assume that $0 \in A$, since adding 0 can only decrease $\mathbb{E}[\operatorname{dist}(h,A)^2]$.

Let $z \in A$ denote the element of A closest to h. Then

$$O(\varepsilon) \geq \mathbb{E}[\operatorname{dist}(h,A)^2] \geq \mathbb{E}[h^2 1_{z=0}] = \mathbb{E}[h^2] - \mathbb{E}[h^2 1_{z\neq 0}].$$

If $z \neq 0$ then $z = \Omega(1)$, and so $h^2 = O(h^k)$ for any integer $k \geq 2$. In particular, for $k = \lceil 2/\alpha \rceil$, this shows that

$$\mathbb{E}[h^2 1_{z\neq 0}] = O(\mathbb{E}[h^k]) = O(\|h\|_k^k) = O(\|h\|_2^k) = O(\varepsilon^{k(\alpha/2)}) = O(\varepsilon),$$

using hypercontractivity and $\varepsilon \leq 1$. It follows that $\mathbb{E}[h^2] = O(\varepsilon)$. □

Corollary 4.4. *Fix an integer d, finite sets A, B, and an exponent $\alpha < 1$. If $f, g\colon \{0,1\}^n \to \mathbb{R}$ are degree d functions satisfying $\mathbb{E}[\operatorname{dist}(f,A)^2] = O(\varepsilon)$, $\mathbb{E}[\operatorname{dist}(g,B)^2] = O(\varepsilon)$, and $\|f-g\|^2 = O(\varepsilon^\alpha)$, then $\|f-g\|^2 = O(\varepsilon)$.*

Proof. Let $h = f - g$. The L_2^2 triangle inequality shows that $\mathbb{E}[\operatorname{dist}(h, A-B)^2] = O(\varepsilon)$. Also, $\|h\|^2 = O(\varepsilon^\alpha)$. The lemma therefore shows that $\|h\|^2 = O(\varepsilon)$. □

We now generalize the Kindler–Safra theorem to the A-valued setting, using the decomposition of Claim 4.2 and thus prove Theorem 4.1

Proof of Theorem 4.1. Pick some arbitrary $a \in A$ and arbitrary constant $\varepsilon_0 > 0$. The L_2^2 triangle inequality shows that $\|f-a\|^2 = O(1+\varepsilon)$. If $\varepsilon > \varepsilon_0$, the conclusion of the theorem is trivially satisfied with $g = a$. Therefore from now on we assume that $\varepsilon \leq \varepsilon_0$.

For $a \in A$, define

$$f_a(x) = \prod_{b \neq a} \frac{f(x)-b}{a-b}.$$

Also, let $y(x) \in A$ be the element in A closest to $f(x)$, and let $\delta(x) := (f(x) - y(x))$. Note $\operatorname{dist}(f(x), A) = |\delta(x)|$. We will usually drop the argument x from all these functions. Finally, define $m = |A| - 1$.

Our first goal is to bound $\operatorname{dist}(f_a, \{0,1\})$ in terms of δ. Let $\delta_0 > 0$ be a small constant. We consider two cases. If $y \neq a$ then

$$\operatorname{dist}(f_a, \{0,1\}) \leq |f_a| = \frac{|\delta|}{|y-b|} \prod_{b \neq a,y} \frac{|y-b+\delta|}{|a-b|}.$$

If $|\delta| \leq \delta_0$ then $\operatorname{dist}(f_a, \{0,1\}) = O(|\delta|)$, and otherwise $\operatorname{dist}(f_a, \{0,1\}) = O(|\delta|^m)$. If $y = a$ then

$$\operatorname{dist}(f_a, \{0,1\}) \leq |f_a - 1| = \left| \prod_{b \neq a} \left| 1 + \frac{\delta}{a-b} \right| - 1 \right|.$$

Once again, if $|\delta| \leq \delta_0$ then $\operatorname{dist}(f_a, \{0,1\}) = O(|\delta|)$, and otherwise $\operatorname{dist}(f_a, \{0,1\}) = O(|\delta|^m)$.

We can now obtain a rough bound on $\mathbb{E}[\operatorname{dist}(f_a, \{0,1\})^2]$ by considering separately the cases $|\delta| \leq \delta_0$ and $|\delta| > \delta_0$. The first case is simple:

$$\mathbb{E}[\operatorname{dist}(f_a, \{0,1\})^2 1_{|\delta| \leq \delta_0}] \leq O(\mathbb{E}[\delta^2]) = O(\varepsilon).$$

For the second case, we use Cauchy–Schwartz and the bound $\Pr[\delta^2 \geq \delta_0^2] = O(\varepsilon)$ (recall δ_0 is a constant):

$$\mathbb{E}[\operatorname{dist}(f_a, \{0,1\})^2 1_{|\delta| \geq \delta_0}] \leq \sqrt{\mathbb{E}[\delta^{2m}]} O(\sqrt{\varepsilon}).$$

Let $C := 2 \max_{a \in A} |a|$. If $|f| \geq \max_{a \in A} |a|$ then clearly $|\delta| \leq |f|$, and otherwise $|\delta| \leq |f| + \max_{a \in A} |A| \leq C$. Therefore it always holds that $|\delta| \leq \max(C, |f|)$. This shows that

$$\mathbb{E}[\delta^{2m}] \leq C^{2m} + \mathbb{E}[f^{2m}] = O(1) + \|f\|_{2m}^{2m}.$$

Since $\deg f = d$, we have $\|f\|_{2m} = O(\|f\|_2)$. The L_2^2 triangle inequality shows that $\|f\|_2^2 = O(\max_{a \in A} |a| + \varepsilon) = O(1)$, and in total this case contributes $O(\sqrt{\varepsilon})$. We conclude that

$$\mathbb{E}[\operatorname{dist}(f_a, \{0,1\})^2] = O(\sqrt{\varepsilon}).$$

The L_2^2 triangle inequality also allows us to bound $\|f_a\|_2^2$ by $O(1)$, by writing it as a polynomial in f and bounding separately all the summands.

The Kindler–Safra theorem shows that f_a is $O(\sqrt{\varepsilon})$-close to a Boolean junta g_a depending on the variables J_a. If $\deg g_a > d$ then $\|f_a - g_a\|^2 \geq \|g_a^{>d}\|^2 = \Omega(1)$ (since there are finitely many options for g_a, up to the choice of J_a), and so $\varepsilon = \Omega(1)$. Choosing ε_0 appropriately, we can assume that $\deg g_a \leq d$.

Define now $g = \sum_{a \in A} a g_a$, and note that this is an A-valued junta of degree at most d. The L_2^2 inequality shows that

$$\|f - g\|^2 = \left\| \sum_{a \in A} a(f_a - g_a) \right\|^2 = O\left(\sum_{a \in A} \|f_a - g_a\|^2 \right) = O(\sqrt{\varepsilon}).$$

The theorem now follows directly from Corollary 4.4 (with $\alpha = 1/2$). □

5 Main result: sparse juntas

In this section, we prove our main result, an analog of the Kindler-Safra theorem for all $p \in (0, 1/2)$.

Theorem 5.1 (Restatement of Theorem 1.5)**.** *For every $p \leq 1/2$ and $f \colon \{0,1\}^n \to \mathbb{R}$ of degree d there exists a function $g \colon \{0,1\}^n \to \mathbb{R}$ of degree d that satisfies the following properties for $\varepsilon := \mathbb{E}[\operatorname{dist}(f, A)^2]$:*

1. $\|f - g\|^2 = O(\varepsilon)$.
2. $\Pr[g \notin A] = O(\varepsilon)$
3. *The coefficients of the y-expansion of g belong to a finite set (depending only on d, A).*

4. *The support of g has branching factor $O(1/p)$.*

5. *If $x \sim \mu_p$ then $g(x)$ is the sum of $O(1)$ coefficients of g with probability $1 - O(\varepsilon)$.*

The following corollary (proved at the end of this section) for A-valued functions which have light Fourier tails follows from the above the theorem.

Corollary 5.2. *Let $d \geq 0$ be any positive integer and $A \subseteq \mathbb{R}$ any finite set. For every $p \leq 1/2$ and $F\colon \{0,1\}^n \to A$ there exists a function $g\colon \{0,1\}^n \to \mathbb{R}$ of degree d that satisfies the following properties for $\varepsilon := \|F^{>d}\|^2$:*

1. $\|F - g\|^2 = O(\varepsilon)$.

2. $\Pr[F \neq g] = O(\varepsilon)$.

3. *All other properties of g (alone) stated in the theorem.*

Given d and alphabet A, let p_0 be the constant given by the agreement theorem Theorem 2.4. For the rest of this section, we fix the constant d, set A and p_0. All hidden constants will depend only on d and A. For all the prelimary claims till the proof of Theorem 5.1, we further assume that $p \leq p_0$. Finally, as in the hypothesis of the theorem, we assume f is a function from $\{0,1\}^n$ to $\mathbb{R}$ of degree d satisfying $\mathbb{E}_{\mu_p}[\text{dist}(f,A)^2] = \varepsilon$

The main result of this section extends the generalized Kindler-Safra theorem Theorem 4.1, which holds only for constant p, to all values of p via the agreement theorem Theorem 2.4. The idea is to consider, for each subset $S \subset [n]$, a "restriction" of f obtained by fixing the inputs outside S to be 0. Namely, we define $f|_S : \{0,1\}^S \to \mathbb{R}$ by $f|_S(x) = f(x \circ 0_{\bar{S}})$ where $x \circ 0_{\bar{S}} \in \{0,1\}^n$ is the input that agrees with x on the coordinates of S and is zero outside of S. We will find an approximate structure for each $f|_S$, and then stitch them together using the agreement theorem Theorem 2.4. We start by applying the generalized Kindler-Safra theorem to $f|_S$ for subsets S selected according to two constant values of p (namely, $p = 1/2$ and $p = 1/4$).

Claim 5.3. *For every set $S \subseteq [n]$, let*

$$\varepsilon_S := \mathop{\mathbb{E}}_{\mu_{1/4}}[\text{dist}(f|_S, A)^2], \qquad\qquad \delta_S := \mathop{\mathbb{E}}_{\mu_{1/2}}[\text{dist}(f|_S, A)^2]$$

Then $\mathbb{E}_{S \sim \mu_{4p}}[\varepsilon_S] = \mathbb{E}_{S \sim \mu_{2p}}[\delta_S] = \varepsilon$, and for every S there exist A-valued degree d juntas $g_S\colon \{0,1\}^S \to A$ and $h_S\colon \{0,1\}^S \to A$ such that $\mathbb{E}_{\mu_{1/4}}[(f|_S - g_S)^2] = O(\varepsilon_S)$ and $\mathbb{E}_{\mu_{1/2}}[(f|_S - h_S)^2] = O(\delta_S)$.

Proof. If $S \sim \mu_{4p}$ and $x \sim \mu_{1/4}(S)$ then $x \sim \mu_p$, and this explains why $\mathbb{E}_{S \sim \mu_{4p}}[\varepsilon_S] = \varepsilon$. The fact that $\mathbb{E}_{\mu_{1/4}}[(f|_S - g_S)^2] = O(\varepsilon_S)$ follows from the generalized Kindler-Safra theorem Theorem 4.1. The proof of $\mathbb{E}_{S \sim \mu_{2p}}[\delta_S] = \varepsilon$ and $\mathbb{E}_{\mu_{1/2}}[(f|_S - h_S)^2] = O(\delta_S)$. □

Towards applying the agreement theorem Theorem 2.4, we need to prove that the collection of local juntas $\{g_S\}_S$ typically agree with each other. We do so by showing that typically g_{S_1} and g_{S_2} agree on the intersection of their domains with $h_{S_1 \cap S_2}$. In the next claim, we show that if the pair of sets (S_1, S_2) are chosen according to the distribution $\mu_{4p,1/2}$, then the two juntas g_{S_1} and g_{S_2} agree with $h_{S_1 \cap S_2}$ with probability $1 - O(\varepsilon)$. We will then apply the agreement theorem using majority decoding to obtain a single degree d function $g\colon \{0,1\}^n \to \mathbb{R}$ that explains most of the juntas g_S.

Claim 5.4. *For every set $S \subseteq [n]$, let the y-expansion of the junta g_S given in Claim 5.3 be as follows:*

$$g_S = \sum_{\substack{T \subseteq S \\ |T| = d}} d_{S,T} y_T.$$

For every $|T| \leq d$, let d_T be the plurality value of $d_{S,T}$ among all $S \supseteq T$ (measured according to μ_{4p}), and define

$$g := \sum_{|T| \leq d} d_T y_T.$$

Then $\Pr_{S \sim \mu_{4p}}[g_S = g|_S] = 1 - O(\varepsilon)$, and so $\Pr_{\mu_p}[g \in A] = 1 - O(\varepsilon)$.

Proof. To apply the agreement theorem we would like to first bound the probability $\Pr_{S_1,S_2\sim\mu_{4p,1/2}}[g_{S_1}|_{S_1\cap S_2} \neq g_{S_2}|_{S_1\cap S_2}]$ when the pair of sets (S_1, S_2) are chosen according to $\mu_{4p,1/2}$. Now for $(S_1, S_2) \sim \mu_{4p,1/2}$, let $T := S_1 \cap S_2$. Notice that $S_1, S_2 \sim \mu_{4p}$, while $T \sim \mu_{1/2}(S_1)$. Consider the three juntas g_{S_1}, g_{S_2} and h_T. Clearly, if $g_{S_1}|_T \neq g_{S_2}|_T$ then one of $g_{S_1}|_T \neq h_T$ or $g_{S_2}|_T \neq h_T$ must hold. Thus,

$$\Pr_{S_1,S_2\sim\mu_{4p,1/2}}[g_{S_1}|_{S_1\cap S_2} \neq g_{S_2}|_{S_1\cap S_2}] \leq 2 \Pr_{\substack{S\sim\mu_{4p}\\ T\sim\mu_{1/2}(S)}}[g_S|_T \neq h_T] \tag{1}$$

Thus, it suffices to bound the probability $\Pr_{S,T}[g_S|_T \neq h_T]$ where $S \sim \mu_{4p}$ and $T \sim \mu_{1/2}(S)$.

For any $T \subseteq S \subseteq [n]$, the L_2^2 triangle inequality shows that,

$$\mathop{\mathbb{E}}_{\mu_{1/2}}[(g_S|_T-h_T)^2] \leq 2 \mathop{\mathbb{E}}_{\mu_{1/2}}[(g_S|_T-f|_T)^2]+2 \mathop{\mathbb{E}}_{\mu_{1/2}}[(f|_T-h_T)^2] = 2 \mathop{\mathbb{E}}_{\mu_{1/2}}[(g_S|_T-f|_T)^2]+O(\mathop{\mathbb{E}}_{\mu_{1/2}}[\mathrm{dist}(f|_T, A)^2]).$$

Taking expectation over $T \sim \mu_{1/2}(S)$, we see that

$$\mathop{\mathbb{E}}_{T\sim\mu_{1/2}(S)} \mathop{\mathbb{E}}_{\mu_{1/2}}[(g_S|_T - h_T)^2] \leq 2 \mathop{\mathbb{E}}_{\mu_{1/4}}[(g_S - f|_S)^2] + O(\mathop{\mathbb{E}}_{\mu_{1/4}}[\mathrm{dist}(f|_S, A)^2]) = O(\mathop{\mathbb{E}}_{\mu_{1/4}}[\mathrm{dist}(f|_S, A)^2]).$$

Here we used the fact that if $T \sim \mu_{1/2}(S)$ and $x \sim \mu_{1/2}(T)$ then $x \sim \mu_{1/4}(S)$.

Both $g_S|_T$ and h_T are A-valued degree d juntas (see Claim 4.2). Hence either they agree, or $\mathbb{E}_{\mu_{1/2}}[(g_S|_T - h_T)^2] = \Omega(1)$. Therefore

$$\Pr_{T\sim\mu_{1/2}(S)}[g_S|_T \neq h_T] = O(\mathop{\mathbb{E}}_{\mu_{1/4}}[\mathrm{dist}(f|_S, A)^2]) = O(\varepsilon_S).$$

Now, taking expectation over $S \sim \mu_{4p}$, we obtain via Claim 5.3

$$\Pr_{\substack{S\sim\mu_{4p}\\ T\sim\mu_{1/2}(S)}}[g_S|_T \neq h_T] = \mathbb{E}_{S\sim\mu_{4p}}[O(\varepsilon_S)] = O(\varepsilon).$$

We now return to (1), to conclude that

$$\Pr_{S_1,S_2\sim\mu_{4p,1/2}}[g_{S_1}|_{S_1\cap S_2} \neq g_{S_2}|_{S_1\cap S_2}] = O(\varepsilon).$$

We have thus satisfied the hypothesis of the agreement theorem (Theorem 2.4). Invoking the agreement theorem, we deduce that $\Pr_{S\sim\mu_{4p}}[g_S = g|_S] = 1 - O(\varepsilon)$. Since g_S is A-valued,

$$\Pr_{\mu_p}[g \in A] \geq \Pr_{\substack{S\sim\mu_{4p}\\ x\sim\mu_{1/4}(S)}}[g(x) = g_S(x)] \geq \Pr_{S\sim\mu_{4p}}[g|_S = g_S] = 1 - O(\varepsilon). \qquad \square$$

We have thus constructed the function g indicated in the Theorem 5.1 and shown that $\Pr_{\mu_p}[g \notin A] = O(\varepsilon)$. In the remaining claims, we show the other properties of g mentioned in Theorem 5.1.

First, we observe that since the g_S are juntas, the coefficients $d_{S,T}$, and so d_T, belong to a finite set depending only on d, A. We can easily deduce an upper bound on the support of g.

Claim 5.5. *The function g from Claim 5.4 has branching factor $O(1/p)$.*

Proof. Let R, e be given. We want to show that the number of $B \supseteq R$ such that $|B| = |R| + e$ and $d_B \neq 0$ is $O(p^{-e})$. Let us denote by $\mathcal{B} = \{B \supseteq R : |B \setminus R| = e\}$ the collection of all such potential B.

Let g_S be the functions from Claim 5.3. Recall that $g_S = \sum_B d_{S,B} y_B$. Since g_S is a junta (by Claim 4.2), $\sum_B d_{S,B}^2 = O(1)$. Therefore

$$\mathop{\mathbb{E}}_{\substack{S\sim\mu_{4p}\\ S\supseteq R}}\left[\sum_{\substack{B\in\mathcal{B}\\ B\subseteq S}} d_{S,B}^2\right] = O(1).$$

Given that S contains R, the probability that it also contains a specific $B \in \mathcal{B}$ is $(4p)^{|B|-|S|} = (4p)^e$, and so

$$\sum_{B\in\mathcal{B}} \mathop{\mathbb{E}}_{\substack{S\sim\mu_{4p}\\ S\supseteq B}}[d_{S,B}^2] = O(p^{-e}).$$

Since there are only finitely many possible values for $d_{S,B}$ (since g_S is an A-valued junta) and we chose d_B as the plurality value, the inner expectation is $\Omega(d_B^2)$, and so

$$\sum_{B \in \mathcal{B}} d_B^2 = O(p^{-e}).$$

Again due to the finitely many possible values for d_B, each non-zero d_B^2 is $\Omega(1)$. We conclude that the number of non-zero d_B for $B \in \mathcal{B}$ is $O(p^{-e})$, as needed. □

Our next step is to consider an auxiliary function derived from g.

Lemma 5.6. *Let g be the function from Claim 5.4, and define*

$$G = \prod_{a \in A} (g - a).$$

Then G satisfies the following properties:

1. *G has branching factor $O(1/p)$.*
2. *$\Pr_{\mu_p}[G = 0] = 1 - O(\varepsilon)$.*
3. *The number of sets B of size e such that $\tilde{G}(B) \neq 0$ is $O(p^{-e}\varepsilon)$.*
4. *$\mathbb{E}_{\mu_p}[G^2] = O(\varepsilon)$.*

Proof. The first property follows from Claim 5.5 via Lemma 3.2, and the second from Claim 5.4.

For the third property, we start by bounding the number N_e of sets B of size e such that $\tilde{G}(B) \neq 0$ but $\tilde{G}(R) = 0$ for all $R \subsetneq B$. For each such B, Lemma 3.4 shows that the probability that $y_B = 1$ and $y_C = 0$ for all other C in the support of G is $\Omega(p^e)$. If this event happens, then $G = \tilde{G}(B) \neq 0$. Since these events are disjoint, we deduce that $\Pr[G \neq 0] = \Omega(p^e N_e)$, which implies that $N_e = O(p^{-e}\varepsilon)$.

We can associate with each B of size e such that $\tilde{G}(B) \neq 0$ a subset $B' \subseteq B$ such that $\tilde{G}(B') \neq 0$ but $\tilde{G}(R) = 0$ for all $R \subsetneq B$. For each $e' \leq e$, there are $N_{e'} = O(p^{-e'}\varepsilon)$ options for the set B'. Since G has branching factor $O(1/p)$, the set B' has $O(p^{-(e-e')})$ extensions of size e in the support of G. In total, for each e' there are $O(p^{-e'}\varepsilon) \cdot O(p^{-(e-e')}) = O(p^{-e}\varepsilon)$ sets B with $|B'| = e'$. Considering the $e+1$ possible values of e', we deduce the third property.

For the fourth property, write

$$G^2 = \sum_B y_B \sum_{B_1 \cup B_2 = B} \tilde{G}(B_1)\tilde{G}(B_2).$$

Lemma 2.3 implies that $|\tilde{G}(B)| = O(1)$ (recalling that the coefficients d_B of g belong to a finite set depending only on d, A, due to Claim 4.2). Denoting by M_e the number of pairs B_1, B_2 such that $\tilde{G}(B_1), \tilde{G}(B_2) \neq 0$ and $|B_1 \cup B_2| = e$, it follows that $\mathbb{E}[G^2] = O(\sum_e p^e M_e)$. Since the sum contains finitely many terms ($\deg G \leq d|A|$), the fourth property will follow if we show that $M_e = O(p^{-e}\varepsilon)$.

Given e, it remains to bound the number of pairs B_1, B_2 such that $\tilde{G}(B_1), \tilde{G}(B_2) \neq 0$ and $|B_1 \cup B_2| = e$. For each $e_1, e_2, e_\cap$, we will count the number of such pairs with $|B_i| = e_i$ and $|B_1 \cap B_2| = e_\cap$. The third property shows that there are $O(p^{-e_1}\varepsilon)$ many choices for B_1. For each such B_1, there are $O(1)$ many choices for $B_1 \cap B_2$, and given $B_1 \cap B_2$, the first property shows that there are $O(p^{-(e_2 - e_\cap)})$ choices for B. In total, there are $O(p^{-e_1}\varepsilon) \cdot O(1) \cdot O(p^{-(e_2-e_\cap)}) = O(p^{-(e_1+e_2-e_\cap)}\varepsilon) = O(p^{-e}\varepsilon)$ choices for B_1, B_2. The fourth property follows since there are $O(1)$ many choices for $e_1, e_2, e_\cap$. □

Using the function G, we can finally compare f and g.

Lemma 5.7. *Let g be the function from Claim 5.4. Then $\|f - g\|^2 = \mathbb{E}_{\mu_p}[(f-g)^2] = O(\varepsilon)$.*

Proof. Let $F = \operatorname{round}(f, A)$, and let g_S, g, G be the functions defined in Claim 5.3, Claim 5.4, and Lemma 5.6. We have

$$\mathop{\mathbb{E}}_{\mu_p}[(F-g)^2] = \mathop{\mathbb{E}}_{S \sim \mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}}[(F|_S - g|_S)^2] = \underbrace{\mathop{\mathbb{E}}_{S \sim \mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}}[(F|_S - g|_S)^2 1_{g|_S = g_S}]}_{\varepsilon_1} + \underbrace{\mathop{\mathbb{E}}_{S \sim \mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}}[(F|_S - g|_S)^2 1_{g|_S \neq g_S}]}_{\varepsilon_2}.$$

Claim 5.3 allows us to estimate ε_1, using the L_2^2 triangle inequality:

$$\varepsilon_1 = \mathop{\mathbb{E}}_{S\sim\mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}} [(F|_S - g_S)^2] \le 2 \mathop{\mathbb{E}}_{S\sim\mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}} [(F|_S - f|_S)^2] + 2 \mathop{\mathbb{E}}_{S\sim\mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}} [(f|_S - g_S)^2] =$$
$$2 \mathop{\mathbb{E}}_{\mu_p}[(F-f)^2] + 2 \mathop{\mathbb{E}}_{S\sim\mu_{4p}}[\varepsilon_S] = O(\varepsilon) + O(\varepsilon) = O(\varepsilon).$$

We estimate ε_2 by truncation. Since $x^2 = O(\prod_{a\in A}(x-a)^2)$ as $x \to \infty$, we can find constants $M, C > 0$ (depending only on A) such that if $|x| \ge M$ then $x^2 \le C\prod_{a\in A}(x-a)^2$. Let $g = g_{\le M} + g_{>M}$, where $g_{\le M} = g 1_{|g|\le M}$. The L_2^2 triangle inequality shows that

$$\varepsilon_2 \le \underbrace{2 \mathop{\mathbb{E}}_{S\sim\mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}} [(F|_S - g_{\le M}|_S)^2 1_{g|_S \ne g_S}]}_{\varepsilon_{2,1}} + \underbrace{2 \mathop{\mathbb{E}}_{S\sim\mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}} [g_{>M}|_S^2 1_{g|_S\ne g_S}]}_{\varepsilon_{2,2}}.$$

Because both F and $g_{\le M}$ are bounded, we can estimate $\varepsilon_{2,1}$ by

$$\varepsilon_{2,1} = O(\Pr_{S\sim\mu_{4p}}[g|_S \ne g_S]) = O(\varepsilon),$$

using Claim 5.4. The defining property of M shows that

$$\varepsilon_{2,2} \le C \mathop{\mathbb{E}}_{S\sim\mu_{4p}} \mathop{\mathbb{E}}_{\mu_{1/4}} [G|_S^2] = O(\mathop{\mathbb{E}}_{\mu_p}[G^2]) = O(\varepsilon),$$

using the fourth property of Lemma 5.6. Altogether, we deduce that $\mathbb{E}_{\mu_p}[(F-g)^2] = O(\varepsilon)$. Since $\mathbb{E}_{\mu_p}[(F-f)^2] = \varepsilon$ by definition, the L_2^2 triangle inequality completes the proof. □

We can now prove our main theorem. Recall that the statement of the theorem does not make any assumptions on p though all the above claims use the fact that $p \le p_0$.

Proof of Theorem 5.1. Suppose that $p \le p_0$, and let g be the function constructed in Claim 5.4. The first property follows from Lemma 5.7. The second property follows from Claim 5.4. The third property follows from the definition of g. The fourth property follows from Claim 5.5. Finally, Claim 5.4 shows that $\Pr_{S\sim\mu_{4p}}[g|_S = g_S] = 1 - O(\varepsilon)$. Hence if we choose $S \sim \mu_{4p}$ and $x \sim \mu_{1/4}(S)$ (so that $x \sim \mu_p$), we get that $g(x) = g_S(x)$ with probability $1 - O(\varepsilon)$, implying the fifth property since g_S is a junta.

When $p \in [p_0, 1/2]$, we choose g using the generalized Kindler–Safra theorem, Theorem 4.1, guaranteeing the first property (we use the fact that the big O constant varies continuously with p). Claim 4.2 shows that g is an A-valued junta, implying all the other properties. □

Corollary 5.2 is proved along similar lines.

Proof of Corollary 5.2. Apply the theorem to $f := F^{\le d}$, which satisfies $\mathbb{E}[\mathrm{dist}(f,A)^2] = \varepsilon$. The L_2^2 triangle inequality shows that $\|F-g\|^2 \le 2\|F-f\|^2 + 2\|f-g\|^2 = O(\varepsilon)$. For the second property,

$$\Pr[F \ne g] \le \Pr[g \notin A] + \Pr[F \ne g \text{ and } g \in A] = \Pr[F \ne g \text{ and } g \in A] + O(\varepsilon).$$

When $g(x) \in A$, if $F(x) \ne g(x)$ then $(F(x) - g(x))^2 = \Omega(1)$. Therefore

$$\Pr[F \ne g \text{ and } g \in A] = \mathop{\mathbb{E}}_{\mu_p}[1_{F\ne g \text{ and } g\in A}] \le \mathop{\mathbb{E}}_{\mu_p}[(F-g)^2] = O(\varepsilon).$$

Altogether we get that $\Pr[F \ne g] = O(\varepsilon)$. All other properties are inherited from the theorem. □

6 A converse to the main result

Given a degree d function f such that $\mathbb{E}[\mathrm{dist}(f,A)^2] = \varepsilon$, Theorem 5.1 gives a function g such that $\|f-g\|^2 = O(\varepsilon)$ and:

- $\deg g \le d$
- g has branching factor $O(1/p)$.

- $\Pr[g \notin A] = O(\varepsilon)$.
- The coefficients of the y-expansion of g belong to some finite set depending only on d, A.

In this short section, we show that a function g satisfying these properties also satisfies $\mathbb{E}[\operatorname{dist}(g, A)^2] = \varepsilon$, and in this sense Theorem 5.1 is a complete characterization of degree d functions close (in L_2) to A.

Lemma 6.1. *Fix $d \geq 0$ and finite sets A, B. Suppose that g satisfies the following properties, for some small enough p:*

- $\deg g \leq d$
- *g has branching factor $O(1/p)$.*
- $\Pr[g \notin A] = \varepsilon$.
- *The coefficients of the y-expansion of g belong to B.*

Then $\mathbb{E}[\operatorname{dist}(g, A)^2] = O(\varepsilon)$.

Proof. The first step is to apply the argument of Lemma 5.6. This lemma defines

$$G = \prod_{a \in A} (g - a),$$

and proves that $\mathbb{E}[G^2] = O(\varepsilon)$, using only the listed properties.

Since $\operatorname{dist}(x, A)^2 = O(\prod_{a \in A}(x - a)^2)$, there exists M such that $\operatorname{dist}(g, A)^2 \leq G^2$ whenever $|g| \geq M$. For an arbitrary $a \in A$ we have

$$\mathbb{E}[\operatorname{dist}(g, A)^2] = \mathbb{E}[\operatorname{dist}(g, A)^2 1_{g \notin A, |g| \leq M}] + \mathbb{E}[\operatorname{dist}(g, A)^2 1_{g \notin A, |g| \geq M}] \leq (M + |a|)^2 \Pr[g \notin A] + \mathbb{E}[G^2] = O(\varepsilon). \quad \square$$

7 Applications

Our main theorem, Theorem 5.1, describes the approximate structure of degree d functions which are close in L_2^2 to a fixed finite set ("almost quantized functions"): all such functions are close to sparse juntas. This allows us to deduce properties of bounded degree almost quantized functions from properties of sparse juntas.

We give two examples of applications of this sort in this section: we prove a large deviation bound, and we show that when p is small, every bounded degree almost quantized function must be very biased.

7.1 Large deviation

Our first application is a large deviation bound, proved via estimating moments. We start by analyzing the simpler case of hypergraphs.

Lemma 7.1. *Let H be a d-uniform hypergraph with branching factor C/p. For $S \sim \mu_p$, let X be the number of hyperedges in $H|_S$. For all integer k,*

$$\mathbb{E}[X^k] \leq (Ckd)^{kd}.$$

Proof. Let $e_1, \ldots, e_k$ be a k-tuple of hyperedges. We can consider the hypergraph whose vertices are $e_1 \cup \cdots \cup e_k$ and whose hyperedges are $e_1, \ldots, e_k$. This is a hypergraph on at most kd vertices which we call a *pattern*. We can crudely upper bound the number of patterns by $(kd)^{kd}$.

Let P be a pattern on $m = m(P)$ vertices. Our goal is to show that the number of k-tuples of hyperedges conforming to this pattern is at most $(C/p)^m$. Suppose that we have already chosen $e_1, \ldots, e_{i-1}$, and suppose that $t_i = |e_i \setminus (e_1 \cup \cdots \cup e_{i-1})|$. Since H has branching factor C/p, there are at most $(C/p)^{t_i}$ choices for e_i. In total, the number of k-tuples is at most $(C/p)^{t_1 + \cdots + t_k} = (C/p)^m$.

We can estimate the kth moment by

$$\mathbb{E}[X^k] = \sum_{e_1, \ldots, e_k} \Pr[e_1 \cup \cdots \cup e_k \subseteq S] = \sum_{e_1, \ldots, e_k} p^{|e_1 \cup \cdots \cup e_k|} \leq \sum_P p^{m(P)} (C/p)^{m(P)} \leq (Ckd)^{kd}. \quad \square$$

This implies a large deviation bound for hypergraphs.

Lemma 7.2. *Let H be a d-uniform hypergraph with branching factor C/p. For $S \sim \mu_p$, let X be the number of edges in $H|_S$. For large enough t,*

$$\Pr[X \geq t] = \exp -\Omega(t^{1/d}/C).$$

Proof. Let $k = t^{1/d}/(eCd)$. We perform the calculation under the assumption that k is an integer; in general k should be taken to be $\lfloor t^{1/d}/(eCd) \rfloor$, but the difference disappears for large t.

Lemma 7.1 shows that $\mathbb{E}[X^k] \leq (t^{1/d}/e)^{kd} = t^k/e^{kd}$, and so Markov's inequality shows that $\Pr[X^k \geq t^k] \leq t^k/\mathbb{E}[X^k] = e^{-kd}$. The lemma follows since $kd = t^{1/d}/(eC)$. □

These two results also apply, with minor changes, to functions with bounded coefficients.

Lemma 7.3. *Let f be a degree d function with branching factor C/p, the coefficients of whose y-expansion are bounded in magnitude by M. For all integer $k \geq 1$,*

$$\mathbb{E}[|f|^k] \leq M^k(2Ckd)^{kd}.$$

Proof. Let H be the support of f. The triangle inequality shows that at a given point S, the value of $|f|^k$ is bounded by M^k times the number of k-tuples $e_1, \ldots, e_k \in H$ such that $e_1, \ldots, e_k \subseteq S$. We can then run the argument of Lemma 7.1 as written, the only difference being that now the hyperedges have *at most* d vertices. This increases the number of patterns to at most (say) $(kd+1)^{kd} \leq (2kd)^{kd}$. □

Lemma 7.4. *Let f be a degree d function with branching factor C/p, the coefficients of whose y-expansion are bounded in magnitude by M. For large enough t,*

$$\Pr[|f| \geq Mt] = \exp -\Omega(t^{1/d}/C).$$

Proof. This lemma follows from Lemma 7.3 just as Lemma 7.2 follows from Lemma 7.1. □

Applying our main theorem, we deduce a large deviation bound for bounded degree almost quantized functions.

Corollary 7.5 (Restatement of Lemma 1.6)**.** *Fix an integer d and a finite set A. Suppose that $f \colon \{0,1\}^n \to \mathbb{R}$ is a degree d function satisfying $\mathbb{E}[\operatorname{dist}(f, A)^2] = \varepsilon$ with respect to μ_p for some $p \leq 1/2$. For large enough t,*

$$\Pr[|f| \geq t] \leq \exp -\Omega(t^{1/d}) + O(\varepsilon/t^2).$$

Proof. Theorem 5.1 shows that there exists a function g satisfying the conditions of the lemma such that $\|f - g\|^2 = O(\varepsilon)$. If $|f| \geq t$ then either $|f - g| \geq t/2$ or $|g| \geq t/2$. The corollary follows from Markov's inequality and the lemma. □

7.2 Distance from being constant

Suppose that f is a bounded degree A-valued function. How does the empirical distribution of f under μ_p look like, for small p? Claim 4.2 shows that f is a junta. All coordinates it depends upon are zero with probability $(1-p)^{O(1)} = 1 - O(p)$, and so for small p the empirical distribution of f is very biased.

What happens when f is just *close* to being A-valued? Consider for example the function $f = y_1 + \cdots + y_{c/p}$, for some small c. The empirical distribution of f is close to Poisson with expectation c, and so $\Pr[f = 0] \approx e^{-c} \approx 1 - c$, $\Pr[f = 1] \approx e^{-c}c \approx c - c^2$, and so $\Pr[f \notin \{0,1\}] \approx c^2$. Taking $c = \sqrt{\varepsilon}$, we see that f is ε-close to $\{0,1\}$, but only $\sqrt{\varepsilon}$-biased (that is, the most probable element in the range is attained with probability roughly $1 - \sqrt{\varepsilon}$). We think of ε as a "small constant" much larger than p, and this shows that almost $\{0,1\}$-valued functions can be much less biased than truly $\{0,1\}$-valued functions.

In this section our goal is to estimate how biased can bounded degree almost quantized functions be. We start by analyzing the situation for sparse juntas.

Lemma 7.6 (Restatement of Lemma 1.7)**.** *Fix a constant $d \geq 0$ and a finite set A. There exist constants $C, \varepsilon_0 > 0$ such that for all $p \leq 1/4$*[1] *and $\varepsilon \leq \varepsilon_0$, the following holds.*

Suppose that $g\colon \{0,1\}^n \to \mathbb{R}$ is a degree d function with branching factor $O(1/p)$ such that $\Pr[g \notin A] = \varepsilon$. Then there exists $a \in A$ such that $\Pr[g \neq a] = O(\varepsilon^C + p)$.

Proof. Lemma 3.4 shows that $\Pr[g = \tilde{g}(\emptyset)] = \Omega(1)$. Choosing ε_0 small enough, we can guarantee that $\tilde{g}(\emptyset) \in A$.

Denote $a := \tilde{g}(\emptyset)$ and $\delta := \Pr[g \neq a]$. Let $S_e = \{|B| = e : \tilde{g}(B) \neq 0\}$. If $g \neq a$ then $y_B \neq 0$ for some B such that $\tilde{g}(B) \neq 0$, and this shows that $\delta \leq \sum_{e=1}^d p^e |S_e|$. Therefore there exists $1 \leq e \leq d$ such that $|S_e| \geq \delta p^{-e}/d$.

Let M be the constant from Claim 4.2. We will show that either there exist constants $L, C > 0$ such that that either $\delta = O(p)$ or

$$\Pr_{S \sim \mu_{2p}}[g|_{y_S=1} \text{ depends on more than } M \text{ and at most } L \text{ coordinates}] = \Omega(\delta^C).$$

If $g|_{y_S=1}$ depends on more than M coordinates then it cannot be A-valued. If it also depends on at most L coordinates, the probability (with respect to $\mu_{1/2}$) that it is not A-valued is $\Omega(1)$. Hence

$$\Pr[g \notin A] = \Pr_{\substack{S \sim \mu_{2p} \\ x \sim \mu_{1/2}(S)}}[g(x) \notin A] \geq \Omega\big(\Pr_{S \sim \mu_{2p}}[g|_{y_S=1} \text{ depends on } > M \text{ and } \leq L \text{ coordinates}]\big) = \Omega(\delta^C),$$

as claimed.

Let M_0 be a constant such that M_0 distinct hyperedges of cardinality at most d span more than M vertices. Note that M_0 such hyperedges also span at most $L := dM_0$ vertices. If $|S_e| < M_0$ then $\delta = O(p^e) = O(p)$, so we can assume that $|S_e| \geq M_0$.

Consider the collection $\mathcal{S}$ of all M_0-tuples of hyperedges from $|S_e|$. Since $|S_e| \geq M_0$, we have $|\mathcal{S}| = \Omega(|S_e|^{M_0}) = \Omega(\delta^{M_0} p^{-eM_0})$. For each M_0-tuple of hyperedges, we can consider the set of vertices contained in these hyperedges. Let $\mathcal{V}$ denote the collection of all such sets of vertices formed from $\mathcal{S}$. Since every set in $\mathcal{S}$ can be obtained from $O(1)$ tuples of $\mathcal{V}$, we have $|\mathcal{V}| = \Omega(\delta^{M_0} p^{-eM_0})$. Every set in $\mathcal{V}$ contains at most eM_0 vertices.

For every $U \in \mathcal{V}$, Lemma 3.3 shows that $g|_{y_U=1}$ has branching factor $O(1/p)$. Hence Lemma 3.4 shows that when $S \sim \mu_{2p}$, with probability $\Omega((2p)^{|U|}) = \Omega(p^{eM_0})$ the vertex support of $g|_{y_S=1}$ contains no vertex outside of U. In fact, since U is the set of vertices contained in an M_0-tuple of hyperedges, the vertex support of $g|_{y_S=1}$ is exactly U, and so $g|_{y_S=1}$ depends on more than M and at most L coordinates. The corresponding events for different U are disjoint, and we conclude that

$$\Pr_{S \sim \mu_{2p}}[g|_{y_S=1} \text{ depends on } > M \text{ and } \leq L \text{ coordinates}] = \Omega(p^{eM_0})|\mathcal{V}| = \Omega(p^{eM_0}) \cdot \Omega(\delta^{M_0} p^{-eM_0}) = \Omega(\delta^{M_0}),$$

completing the proof. $\square$

Applying Corollary 5.2, we obtain a similar result for bounded degree almost quantized functions.

Corollary 7.7. *Fix a constant $d \geq 0$ and a finite set A. There exists constant $C, \varepsilon_0 > 0$ such that for all $p \leq 1/4$ and $\varepsilon \leq \varepsilon_0$, the following holds.*

Suppose that $f\colon \{0,1\}^n \to \mathbb{R}$ is a degree d function satisfying $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$. Then there exists $a \in A$ such that $\Pr[\mathrm{round}(f, A) \neq a] = O(\varepsilon^C)$.

Proof. Let $F = \mathrm{round}(f, A)$. Corollary 5.2 shows that there exists a degree d function $g\colon \{0,1\}^n \to \mathbb{R}$ which has branching factor $O(1/p)$ and satisfies $\Pr[g \notin A] = O(\varepsilon)$ and $\Pr[F \neq g] = O(\varepsilon)$. The lemma shows that $\Pr[g \neq a] = O(\varepsilon^C)$ for some $a \in A$, and the corollary follows. $\square$

Discussion What is the correct exponent of ε? Let us focus on $A = \{0,1\}$. Let $n = \delta/p$, and consider the function

$$f_d = \sum_{i_1} y_{i_1} - \sum_{i_1 < i_2} y_{i_1} y_{i_2} + \cdots \pm \sum_{i_1 < \cdots < i_d} y_{i_1} \cdots y_{i_d}.$$

[1] This constant is arbitrary. Any constant less than 1 can be used.

When exactly m of the coordinates are 1, we have

$$f_d = \sum_{e=1}^{d} (-1)^{e-1} \binom{m}{e} = 1 - \sum_{e=0}^{d} (-1)^e \binom{m}{e}.$$

When $m \le d$, we have

$$f_d = 1 - \sum_{e=0}^{m} (-1)^e \binom{m}{e} = 1 - (1-1)^m = \begin{cases} 0 & \text{if } m = 0, \\ 1 & \text{otherwise.} \end{cases}$$

When $m = d+1$, we have

$$f_d = 1 - \sum_{e=0}^{m} (-1)^e \binom{m}{e} + (-1)^m \binom{m}{m} = 1 - (1-1)^m + (-1)^m = \begin{cases} 0 & \text{if } d \text{ is even,} \\ 2 & \text{if } d \text{ is odd.} \end{cases}$$

For small p, the distribution of m is roughly Poisson with expectation δ, and so for small δ:

- $\Pr[f_d = 0] \ge \Pr[m = 0] \approx e^{-\delta} \approx 1 - \delta$.
- When d is odd, $\Pr[f_d \notin \{0,1\}] \le \Pr[m > d] \approx \Pr[m = d+1] \approx e^{-\delta} \frac{\delta^{d+1}}{(d+1)!} \approx \frac{\delta^{d+1}}{(d+1)!}$.
- When d is even, $\Pr[f_d \notin \{0,1\}] \le \Pr[m > d+1] \approx \Pr[m = d+2] \approx e^{-\delta} \frac{\delta^{d+2}}{(d+2)!} \approx \frac{\delta^{d+2}}{(d+2)!}$.

This shows that a degree d function which is ε-close to A can be $\Omega(\varepsilon^{1/(d+1)})$-far from constant, and even $\Omega(\varepsilon^{1/(d+2)})$-far when d is even. When $d = 1$, the sparse FKN theorem [Fil16] shows that the exponent $1/2$ is tight.

8 New proof of classical Kindler–Safra theorem

In this section we give a self-contained proof of the Kindler–Safra theorem in the $\mu_{1/2}$ setting. The proof can easily be extended to the μ_p setting for any constant p. Our functions are on the domain $\{\pm 1\}^n$, and we denote their inputs by $x_1, \ldots, x_n \in \{\pm 1\}$.

When we write $x \sim \{\pm 1\}^n$, we always mean that x is chosen according to the uniform distribution over $\{\pm 1\}^n$.

8.1 A-valued FKN theorem

As a prerequisite for our proof of the Kindler–Safra theorem, we need to extend the FKN theorem to the A-valued setting. Our proof closely follows the proof in Kindler's thesis [Kin03]. In contrast to the classical FKN theorem, in which the approximating functions are dictators, in the A-valued setting we only get juntas. Indeed, if $A = \{0, 1, \ldots, a\}$ then the function $\sum_{i=1}^{a} \frac{1+x_i}{2}$ is A-valued.

We start by identifying the junta variables.

Lemma 8.1. *Fix a finite set A. Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree 1 function satisfying $\mathbb{E}[\operatorname{dist}(f, A)^2] = \varepsilon$. There exists a constant $m > 0$ (depending on A) such that $\hat{f}(i)^2 \ge m\varepsilon$ for at most $|A| - 1$ many coefficients $\hat{f}(i)$.*

Proof. Let $m = 2^{|A|+1}$, and let $J_0 = \{i : \hat{f}(i)^2 \ge m\varepsilon\}$. Our goal is to show that $|J_0| < |A|$. If not, we can choose a subset $J \subseteq J_0$ of size exactly $|A|$. There is an assignment α to the coordinates outside J such that $\mathbb{E}[\operatorname{dist}(f|_\alpha, A)^2] \le \varepsilon$. This implies that for some c,

$$\mathbb{E}[\operatorname{dist}(\sum_{i \in J} \hat{f}(i) x_i + c, A)^2] \le \varepsilon.$$

We can assume, without loss of generality, that $\hat{f}(i) > 0$ for all $i \in J$ (otherwise, we can define a new function obtained from f by flipping the appropriate inputs). Assume also, for simplicity, that $J = \{1, \ldots, |A|\}$. For $0 \le i \le |A|$, define

$$v_i = c + \sum_{j=0}^{i-1} \hat{f}(j) - \sum_{j=i}^{|A|} \hat{f}(j).$$

For every $0 \le i \le |A|$, let $a_i = \operatorname{round}(v_i, A)$. Since $v_i - v_{i-1} = 2\hat{f}(i) > 0$, we can assume that $a_i \ge a_{i-1}$. By assumption, $|v_i - a_i|^2 \le 2^{|J|}\varepsilon = 2^{|A|}\varepsilon$ for all i. If $a_i = a_{i-1}$, then this implies that $(v_i - v_{i-1})^2 \le 2^{|A|+2}\varepsilon$ (using the L_2^2 triangle inequality), which contradicts the upper bound, $(v_i - v_{i-1})^2 = 4\hat{f}(i)^2 \ge 4m\varepsilon = 2^{|A|+3}\varepsilon$. We conclude that $a_i > a_{i-1}$, and so $a_0 < a_1 < \cdots < a_{|A|}$. However, this is impossible, since A contains only $|A|$ elements. This contradiction shows that $|J_0| < |A|$. □

The idea now is to truncate f to its junta part, and to show that the noisy part has small norm. We do this in an inductive fashion, using the following lemma.

Lemma 8.2. *Fix a finite set A, and let m be the constant from Lemma 8.1. There exists a constant $\varepsilon_0 > 0$ (depending on A) such the following holds for all $\varepsilon \le \varepsilon_0$.*

If $f\colon \{\pm 1\}^n \to \mathbb{R}$ is a degree 1 function satisfying $\mathbb{V}[f] \le (2+m)\varepsilon$ and $\mathbb{E}[\operatorname{dist}(f, A)^2] = \varepsilon$, then in fact $\mathbb{V}[f] \le 2\varepsilon$.

Proof. Markov's inequality shows that each of the events $(f - \mathbb{E}[f])^2 \le 3(2+m)\varepsilon$ and $\operatorname{dist}(f, A)^2 \le 3\varepsilon$ occurs with probability $2/3$, and so there is a point at which both occur simultaneously. The L_2^2 triangle inequality implies that for some $a \in A$,

$$(\mathbb{E}[f] - a)^2 \le 6(2+m)\varepsilon + 6\varepsilon = (18+6m)\varepsilon.$$

Let $\mathcal{E}$ denote the event that $\operatorname{round}(f, A) = a$. Then

$$\varepsilon \ge \mathbb{E}[\operatorname{dist}(f, A)^2 1_{\mathcal{E}}] = \mathbb{E}[(f-a)^2 1_{\mathcal{E}}] = \mathbb{E}[(f-a)^2] - \underbrace{\mathbb{E}[(f-a)^2 1_{\overline{\mathcal{E}}}]}_{\delta}.$$

When $\operatorname{round}(f, A) \ne a$, necessarily $(f-a)^2 = \Omega_A(1)$, and so $(f-a)^2 = O_A((f-a)^4)$. This shows that

$$\delta \le O_A(\mathbb{E}[(f-a)^4]) = O_A(\|f-a\|_4^4) \overset{(*)}{=} O_A(\|f-a\|_2^4) = O_A(\mathbb{E}[(f-a)^2]^2),$$

using hypercontractivity in $(*)$. The L_2^2 triangle inequality shows that

$$\mathbb{E}[(f-a)^2] \le 2\,\mathbb{V}[f] + 2(\mathbb{E}[f]-a)^2 \le 2(2+m)\varepsilon + 2(18+6m)\varepsilon = (40+14m)\varepsilon.$$

Choosing ε_0 small enough (as a function of A), we can guarantee that

$$\varepsilon \ge \mathbb{E}[(f-a)^2](1 - O_A(40+14m)\varepsilon) \ge \frac{1}{2}\,\mathbb{E}[(f-a)^2],$$

and so $\mathbb{E}[(f-a)^2] \le 2\varepsilon$. The lemma follows from the well-known inequality $\mathbb{V}[f] \le \mathbb{E}[(f-a)^2]$. □

We now carry out the induction.

Lemma 8.3. *Fix a finite set A, and let m, ε_0 be the constants from Lemma 8.2. The following holds for all $\varepsilon \le \varepsilon_0$.*

Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree 1 function satisfying $\mathbb{E}[\operatorname{dist}(f, A)^2] = \varepsilon$, let $J = \{i : \hat{f}(i)^2 \ge m\varepsilon\}$, and define $g = \hat{f}(\emptyset) + \sum_{i \in J} \hat{f}(i) x_i$. Then $\|f - g\|^2 \le 2\varepsilon$.

Proof. Assume without loss of generality that $J = \{1, \ldots, N\}$ for some $N < |A|$. We will prove by reverse induction on $i \ge N$ that $\sum_{j > i} \hat{f}(j)^2 \le 2\varepsilon$. The lemma will follow since $\|f - g\|^2 = \sum_{j > N} \hat{f}(j)^2$.

The base case $i = n$ is obvious, so assume that $\sum_{j > i+1} \hat{f}(j)^2 \le 2\varepsilon$ for some $i \ge N$. The definition of J guarantees that $\sum_{j > i} \hat{f}(j)^2 \le (2+m)\varepsilon$. Since $\mathbb{E}[\operatorname{dist}(f, A)^2] = \varepsilon$, there must exist an assignment α to $x_1, \ldots, x_i$ such that $\mathbb{E}[\operatorname{dist}(f|_\alpha, A)^2] \le \varepsilon$. Then $g = f|_\alpha$ satisfies $\mathbb{E}[\operatorname{dist}(g, A)^2] \le \varepsilon$ and $\mathbb{V}[g] \le (2+m)\varepsilon$. Lemma 8.2 shows that $\mathbb{V}[g] \le 2\varepsilon$, and so $\sum_{j > i} \hat{f}(j)^2 \le 2\varepsilon$. □

To complete the proof, we need the following simple lemma.

Lemma 8.4. *For every finite set A and every x, y we have $(x - \operatorname{round}(y, A))^2 = O((x-y)^2 + \operatorname{dist}(x, A)^2)$.*

Proof. Let $a = \text{round}(x, A)$ and $b = \text{round}(y, A)$. If $a = b$ then $(x-b)^2 = (x-a)^2 = \text{dist}(x, A)^2$. Otherwise, without loss of generality $a < b$. Note that $x \le \frac{a+b}{2} \le y$. If $|x-a| \le \frac{b-a}{4}$ then $|x-y| \ge |x - \frac{a+b}{2}| \ge \frac{b-a}{4}$. Therefore $(x-b)^2 \le 2(x-a)^2 + 2(a-b)^2 \le 2(x-a)^2 + 32(x-y)^2$. If $|x-a| \ge \frac{b-a}{4}$ then $(x-b)^2 \le 2(x-a)^2 + 2(a-b)^2 \le 34(x-a)^2$. (In both cases, we used the L_2^2 triangle inequality.) □

The main theorem easily follows.

Theorem 8.5. *Fix a finite set A, and let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree 1 function satisfying $\mathbb{E}[\text{dist}(f, A)^2] = \varepsilon$. There exists a degree 1 function $g\colon \{\pm 1\}^n \to A$, depending on at most $|A| - 1$ coordinates, such that $\|f - g\|^2 = O_A(\varepsilon)$.*

Proof. Let ε_0 be the constant from Lemma 8.3. Suppose first that $\varepsilon \le \varepsilon_0$. The lemma defines a set J of size at most $|A| - 1$ (according to Lemma 8.1) such that $h := \hat{f}(\emptyset) + \sum_{i \in J} \hat{f}(i) x_i$ satisfies $\|f - h\|^2 \le 2\varepsilon$. Let $g = \text{round}(h, A)$, which also depends only on the coordinates in J. Lemma 8.4 shows that $\|f - g\|^2 = O(\|f - h\|^2 + \mathbb{E}[\text{dist}(f, A)^2]) = O(\varepsilon)$.

It remains to show that $\deg g \le 1$. There are finitely many A-valued functions on $|A| - 1$ coordinates. Hence if $g^{>1} \ne 0$ then $g^{>1} = \Omega_A(1)$, and so $\|f - g\|^2 \ge \|(f-g)^{>1}\|^2 = \|g^{>1}\|^2 = \Omega_A(1)$. By possibly reducing ε_0, we can rule out this case, and so $\deg g \le 1$.

If $\varepsilon > \varepsilon_0$ then we take $g = a$ for an arbitrary $a \in A$. The L_2^2 triangle inequality shows that $\mathbb{E}[f^2] \le 2\,\mathbb{E}[\text{round}(f, A)^2] + 2\,\mathbb{E}[\text{dist}(f, A)^2] = O_A(1 + \varepsilon)$. Another application of the triangle inequality shows that $\mathbb{E}[(f-g)^2] \le 2\,\mathbb{E}[f^2] + 2a^2 = O_A(1+\varepsilon)$. Since $\varepsilon \ge \varepsilon_0$, in fact $\mathbb{E}[(f-g)^2] = O_A(1+\varepsilon) = O_A(\varepsilon)$. □

Corollary 8.6. *Fix a finite set A, and let $F\colon \{\pm 1\}^n \to A$ satisfy $\|F^{>1}\|^2 = \varepsilon$. There exists a degree 1 function $g\colon \{\pm 1\}^n \to A$, depending on at most $|A| - 1$ coordinates, such that $\|F - g\|^2 = O_A(\varepsilon)$ and $\Pr[F \ne g] = O_A(\varepsilon)$.*

Proof. Let $f = F^{\le 1}$, which satisfies $\mathbb{E}[\text{dist}(f, A)^2] \le \mathbb{E}[(f - F)^2] = \varepsilon$. The theorem gives an A-valued function g which depends on at most $|A| - 1$ coordinates and satisfies $\|f - g\|^2 = O_A(\varepsilon)$. The L_2^2 triangle inequality shows that $\|F - g\|^2 \le 2\|f - g\|^2 + 2\|f - F\|^2 = O_A(\varepsilon)$. If $F(x) \ne g(x)$ then $(F(x) - g(x))^2 = \Omega_A(1)$, and so $\Pr[F \ne g] = \mathbb{E}[1_{F \ne g}] = O_A(\mathbb{E}[(F-g)^2]) = O_A(\varepsilon)$. □

8.2 A-valued Kindler–Safra theorem

We now prove the A-valued Kindler–Safra theorem by induction on the degree. We start by stating the theorem.

Theorem 8.7. *Fix a finite set A and a degree d. Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree d function satisfying $\mathbb{E}[\text{dist}(f, A)^2] = \varepsilon$. There exists a degree d function $g\colon \{\pm 1\}^n \to A$, depending on $O_{A,d}(1)$ coordinates, such that $\|f - g\|^2 = O_{A,d}(\varepsilon)$.*

We also get a corollary whose omitted proof is the same as that of Corollary 8.6.

Corollary 8.8. *Fix a finite set A and a degree d. Let $F\colon \{\pm 1\}^n \to A$ be a degree d function satisfying $\mathbb{E}[\text{dist}(f, A)^2] = \varepsilon$. There exists a degree d function $g\colon \{\pm 1\}^n \to A$, depending on $O_{A,d}(1)$ coordinates, such that $\|F - g\|^2 = O_{A,d}(\varepsilon)$ and $\Pr[F \ne g] = O_{A,d}(\varepsilon)$.*

The theorem clearly holds when $d = 0$ (take $g = \text{round}(f, A)$), and it holds for $d = 1$ due to Theorem 8.5. Consider now $d > 1$. Assuming Theorem 8.7 for smaller d, we will prove it for the given d.

Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree d function satisfying $\mathbb{E}[\text{dist}(f, A)^2] = \varepsilon$. As in the proof of Theorem 8.5, if $\varepsilon > 2^{-d}$ then $\|f - a\|^2 = O_A(\varepsilon)$ for any $a \in A$, allowing us to take $g = a$, so assume that $\varepsilon \le 2^{-d}$. This has the following implication:

Claim 8.9. *We have $\|f\|^2 = O_A(1)$.*

Proof. The L_2^2 triangle inequality shows that

$$\|f\|^2 \le 2\,\mathbb{E}[\text{round}(f, A)^2] + 2\,\mathbb{E}[\text{dist}(f, A)^2] = O_A(1 + \varepsilon) = O_A(1).$$ □

For a set $S \subseteq [n]$ and an assignment $y \in \{\pm 1\}^{\overline{S}}$, let $f_{S,y} \colon \{\pm 1\}^S \to \mathbb{R}$ be the function obtained by restricting the variables in $\overline{S}$ to the values in y, and define

$$\varepsilon_{S,y} = \mathbb{E}[\mathrm{dist}(f_{S,y}, A)^2].$$

Claim 8.10. *For all S,*

$$\mathop{\mathbb{E}}_{y \sim \{\pm 1\}^{\overline{S}}}[\varepsilon_{S,y}] = \varepsilon.$$

Proof. We have

$$\mathop{\mathbb{E}}_{y \sim \{\pm 1\}^{\overline{S}}}[\varepsilon_{S,y}] = \mathop{\mathbb{E}}_{\substack{y \sim \{\pm 1\}^{\overline{S}} \\ z \sim \{\pm 1\}^S}}[\mathrm{dist}(f(y,z), A)^2] = \mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon. \qquad \square$$

For all S and $y \in \{\pm 1\}^{\overline{S}}$, define

$$\gamma_{S,y} = \|f_{S,y}^{=d}\|^2,$$

and let $\gamma_S = \mathbb{E}_y[\gamma_{S,y}]$.

Claim 8.11. *The value $\gamma_{S,y}$ doesn't depend on y, and*

$$\mathop{\mathbb{E}}_{S \sim \mu_{\varepsilon^{1/d}}([n])}[\gamma_S] = \varepsilon \|f^{=d}\|^2 = O_A(\varepsilon).$$

Proof. Note first that for all y,

$$f_{S,y}^{=d} = \sum_{\substack{|T|=d \\ T \subseteq S}} \hat{f}(T) x_T.$$

Therefore $\gamma_{S,y}$ doesn't depend on y, and

$$\mathop{\mathbb{E}}_{S \sim \mu_{\varepsilon^{1/d}}([n])}[\gamma_S] = \sum_{|T|=d} \mathop{\Pr}_{S \sim \mu_{\varepsilon^{1/d}}([n])}[T \subseteq S] \hat{f}(T)^2 = \sum_{|T|=d} (\varepsilon^{1/d})^d \hat{f}(T)^2 = \varepsilon \|f^{=d}\|^2.$$

We complete the proof using Claim 8.9. $\square$

For each S, y, we apply Theorem 8.7 to the degree $d-1$ function $f_{S,y}^{<d}$ which satisfies

$$\mathbb{E}[\mathrm{dist}(f_{S,y}^{<d}, A)^2] \le 2\,\mathbb{E}[\mathrm{dist}(f_{S,y}, A)^2] + 2\|f_{S,y}^{=d}\|^2 = 2\varepsilon_{S,y} + 2\gamma_S.$$

The theorem gives us an A-valued function $g_{S,y}$ which depends on $O_{A,d}(1)$ coordinates and satisfies

$$\|f_{S,y}^{<d} - g_{S,y}\|^2 = O_{A,d}(\varepsilon_{S,y} + \gamma_S).$$

Since $g_{S,y}$ is an A-valued junta, there exists a finite set B (depending only on A, d) such that all Fourier coefficients of $g_{S,y}$ belong to B.

A simple calculation shows that for all $T \subseteq S$ of size $d-1$,

$$h_{S,T}(y) := \hat{f}_{S,y}(T) = \hat{f}(T) + \sum_{i \notin S} \hat{f}(T+i) y_i.$$

We think of this as a degree 1 function $h_{S,T} \colon \{\pm 1\}^{\overline{S}} \to \mathbb{R}$.

Claim 8.12. *For all $S \subseteq [n]$ we have*

$$\sum_{T \in \binom{S}{d-1}} \mathbb{E}[\mathrm{dist}(h_{S,T}, B)^2] = O_{A,d}(\varepsilon + \gamma_S).$$

Proof. For each $y \in \{\pm 1\}^{\overline{S}}$ we have

$$\sum_{T \in \binom{S}{d-1}} \mathrm{dist}(h_{S,T}(y), B)^2 \le \sum_{T \in \binom{S}{d-1}} (\hat{f}_{S,y}(T) - \hat{g}_{S,y}(T))^2 \le \|f_{S,y}^{<d} - g_{S,y}\|^2 = O_{A,d}(\varepsilon_{S,y} + \gamma_S).$$

Taking expectation over y, we complete the proof using Claim 8.10. $\square$

On the other hand, an application of the generalized FKN theorem gives the following:

Claim 8.13. *There exists a finite set C (depending only on A, d) such that for all $S \subseteq [n]$ and $T \in \binom{S}{d-1}$,*

$$\operatorname{dist}(\hat{f}(T), C)^2 + \sum_{i \notin S} \operatorname{dist}(\hat{f}(T+i), C)^2 = O_{A,d}(\mathbb{E}[\operatorname{dist}(h_{S,T}, B)^2]).$$

Proof. Theorem 8.5, applied to $f := h_{S,T}$ and $A := B$, gives a B-valued function $u_{S,T}$ depending on at most $|B|-1$ coordinates such that $\|h_{S,T} - u_{S,T}\|^2 = O_{A,d}(\mathbb{E}[\operatorname{dist}(h_{S,T}, B)^2])$. All the Fourier coefficients of $u_{S,T}$ belong to some finite set C, and so the claim follows from Parseval's identity since the coefficients of the Fourier expansion of $h_{S,T}$ are $\hat{h}_{S,T}(\emptyset) = \hat{f}(T)$ and $\hat{h}_{S,T}(i) = \hat{f}(T+i)$ for all $i \notin S$. □

Putting both claims together, we deduce:

Claim 8.14. *We have*

$$\sum_{d-1 \le |T| \le d} \operatorname{dist}(\hat{f}(T), C)^2 = O_{A,d}(\varepsilon^{1/d}).$$

Proof. Summing over T in Claim 8.13 and using Claim 8.12, we get that for all $S \subseteq [n]$,

$$\sum_{T \in \binom{S}{d-1}} \left[\operatorname{dist}(\hat{f}(T), C)^2 + \sum_{i \notin S} \operatorname{dist}(\hat{f}(T+i), C)^2 \right] = \sum_{T \in \binom{S}{d-1}} O_{A,d}(\mathbb{E}[\operatorname{dist}(h_{S,T}, B)^2]) = O_{A,d}(\varepsilon + \gamma_S).$$

Taking expectation with respect to $S \sim \mu_\delta$, where $\delta = \varepsilon^{1/d}$, Claim 8.11 shows that

$$\mathop{\mathbb{E}}_{S \sim \mu_\delta} \left[\sum_{T \in \binom{S}{d-1}} \operatorname{dist}(\hat{f}(T), C)^2 + \sum_{i \notin S} \operatorname{dist}(\hat{f}(T+i), C)^2 \right] = O_{A,d}(\varepsilon).$$

A set T of size $d-1$ appears in the sum with probability δ^{d-1}, and a set of size d appears with probability $d\delta^{d-1}(1-\delta)$. Since $\delta \le (2^{-d})^{1/d} = 1/2$ by assumption, we deduce that

$$\sum_{d-1 \le |T| \le d} \operatorname{dist}(\hat{f}(T), C)^2 = O_{A,d}(\varepsilon/\delta^{d-1}) = O_{A,d}(\varepsilon^{1/d}).$$ □

This claim prompts defining

$$h = \sum_{d-1 \le |T| \le d} \operatorname{round}(\hat{f}(T), C) x_T.$$

Claim 8.15. *There exists a finite set D (depending only on A, d) such that h is a D-valued function depending on $O_{A,d}(1)$ coordinates and satisfying $\|h\|^2 = O_{A,d}(1)$.*

Proof. Claim 8.14 shows that $\|h - f^{\ge d-1}\|^2 = O_{A,d}(\varepsilon^{1/d}) = O_{A,d}(1)$. Since $\|f\|^2 = O_{A,d}(1)$ by Claim 8.9, it follows that $\|h\|^2 = O_{A,d}(1)$ and so $\sum_S \hat{h}(S)^2 = O_{A,d}(1)$. As all Fourier coefficients of h belong to C, we deduce that h has $O_{A,d}(1)$ non-zero coefficients. Since all of them involve at most d coordinates, it follows that h depends on $O_{A,d}(1)$ coordinates. Each value of h is a signed sum of $O_{A,d}(1)$ elements of C, and so h is D-valued for some finite set D. □

The next step is an application of Theorem 8.7 for degree $d-2$.

Claim 8.16. *There exists a finite set E (depending only on A, d) and an E-valued degree $d-2$ function g depending on $O_{A,d}(1)$ coordinates such that $\|f - (g+h)\|^2 = O_{A,d}(\varepsilon^{1/d})$.*

Proof. Let $\tilde{f} = f^{<d-1} + h$. Then $\|f - \tilde{f}\|^2 = \|f^{\ge d-1} - h\|^2 = O_{A,d}(\varepsilon^{1/d})$ by Claim 8.14, and so the L_2^2 triangle inequality shows that $\mathbb{E}[\operatorname{dist}(\tilde{f}, A)^2] \le 2\,\mathbb{E}[\operatorname{dist}(f, A)^2] + 2\|f - \tilde{f}\|^2 = O_{A,d}(\varepsilon + \varepsilon^{1/d}) = O_{A,d}(\varepsilon^{1/d})$ (using $\varepsilon \le 2^{-d}$). Setting E to be the Minkowski difference $A - D$ and using the fact that h is D-valued, we deduce that $\mathbb{E}[\operatorname{dist}(f^{<d-1}, E)^2] = O_{A,d}(\varepsilon^{1/d})$.

Applying Theorem 8.7 to the degree $d-2$ function $f^{<d-1}$, we obtain an E-valued degree $d-2$ function g depending on $O_{A,d}(1)$ coordinates such that $\|f^{<d-1} - g\|^2 = O_{A,d}(\varepsilon^{1/d})$. Together with $\|f^{\ge d-1} - h\|^2 = O_{A,d}(\varepsilon^{1/d})$ and the L_2^2 triangle inequality, this shows that $\|f - (g+h)\|^2 = O_{A,d}(\varepsilon^{1/d})$. □

Using the fact that $\mathbb{E}[\text{dist}(f,A)^2] = \varepsilon$, we can improve the bound on $\|f - (g+h)\|^2$.

Claim 8.17. *We have* $\|f - (g+h)\|^2 = O_{A,d}(\varepsilon)$.

Proof. Let $s := f - (g+h)$. Since $\mathbb{E}[\text{dist}(f,A)^2] = \varepsilon$ and $g+h$ is $(D+E)$-valued (where $D+E$ is the Minkowski sum), we see that $\mathbb{E}[\text{dist}(s,V)^2] \le \varepsilon$, where $V = A - (D+E)$ is a finite set. We can assume without loss of generality that $0 \in V$ (this can only decrease the distance). At any point in the domain, either $\text{round}(s,V) = 0$ or $\text{round}(s,V) = \Omega_A(1)$. Hence

$$\varepsilon \ge \mathbb{E}[\text{dist}(s,V)^2 1_{\text{round}(s,V)=0}] = \mathbb{E}[s^2 1_{\text{round}(s,V)=0}] = \mathbb{E}[s^2] - \mathbb{E}[s^2 1_{\text{round}(s,V)\neq 0}] \ge \mathbb{E}[s^2] - O_A(\mathbb{E}[s^{2d}]).$$

Since $\deg(s^{2d}) \le 2d^2$, hypercontractivity shows that $\mathbb{E}[s^{2d}] = \|s\|_{2d}^{2d} = O_d(\|s\|_2^{2d})$, and so Claim 8.16, which states that $\mathbb{E}[s^2] = O_{A,d}(\varepsilon^{1/d})$, implies that

$$\mathbb{E}[s^2] \le \varepsilon + O_{A,d}(\mathbb{E}[s^2]^d) = O_{A,d}(\varepsilon). \qquad \square$$

We can now complete the proof.

Completion of the proof of Theorem 8.7. Let $r = \text{round}(g+h, A)$, and note that r depends on $O_{A,d}(1)$ coordinates. Lemma 8.4 shows that $\|f-r\|^2 = O(\|f-(g+h)\|^2 + \mathbb{E}[\text{dist}(f,A)^2]) = O_{A,d}(\varepsilon)$. If $\deg r > d$ then since r is an A-valued function depending on $O_{A,d}(1)$ coordinates, we have $\|r^{>d}\|^2 = \Omega_{A,d}(1)$, implying that $\|f - r\|^2 = \Omega_{A,d}(1)$ and so $\varepsilon = \Omega_{A,d}(1)$. As in the proof of Theorem 8.5, in this case $\|f - a\|^2 = O_{A,d}(\varepsilon)$ for any $a \in A$. $\square$

References

[BGH+15] Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer. *Making the long code shorter.* SIAM J. Comput., 44(5):1287–1324, 2015. (Preliminary version in *53rd FOCS*, 2012). arXiv:1111.0405, eccc:2011/TR11-142, doi:10.1137/130929394. 4

[BK97] Jean Bourgain and Gil Kalai. *Influences of variables and threshold intervals under group symmetries.* Geom. Funct. Anal., 7(3):438–461, 1997. doi:10.1007/s000390050. 5

[BK17] Itai Benjamini and Gil Kalai. *Around two theorems and a lemma by Lucio Russo*, 2017. (manuscript). 4

[Bou99] Jean Bourgain. *On sharp threshold of monotone properties.* J. Amer. Math. Soc., 12(4):1017–1054, 1999. (Appendix to *E. Friedgut, Sharp Thresholds of Graph Properties, and the k-Sat Problem*). 4

[Bou02] ———. *On the distribution of the Fourier spectrum of Boolean functions.* Israel Journal of Mathematics, 131(1):269–276, 2002. doi:10.1007/BF02785861. 1

[DFH17] Irit Dinur, Yuval Filmus, and Prahladh Harsha. *Agreement tests on graphs and hypergraphs*, 2017. (manuscript). 3, 6

[DG08] Irit Dinur and Elazar Goldenberg. *Locally testing direct product in the low error range.* In *Proc. 49th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 613–622. 2008. doi:10.1109/FOCS.2008.26. 6

[DKK+16] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. *Towards a proof of the 2-to-1 games conjecture?* Technical Report 2016/TR16-198, Elect. Colloq. on Comput. Complexity (ECCC), 2016. eccc:2016/TR16-198. 4

[DKK+17] ———. *On non-optimally expanding sets in Grassmann graphs.* Technical Report 2017/TR17-094, Elect. Colloq. on Comput. Complexity (ECCC), 2017. eccc:2017/TR17-094. 4

[DS14] Irit Dinur and David Steurer. *Direct product testing.* In *Proc. 29th IEEE Conf. on Comput. Complexity*, pages 188–196. 2014. eccc:2013/TR13-179, doi:10.1109/CCC.2014.27. 6

[Fil16] Yuval Filmus. *Friedgut-Kalai-Naor theorem for slices of the boolean cube.* Chicago J. Theor. Comput. Sci., 2016(14), 2016. arXiv:1410.7834, doi:10.4086/cjtcs.2016.014. 2, 17

[FK96] Ehud Friedgut and Gil Kalai. *Every monotone graph property has a sharp threshold.* Proc. Amer. Math. Soc., 124(10):2993–3002, 1996. doi:10.1090/S0002-9939-96-03732-X. 4

[FKN02] Ehud Friedgut, Gil Kalai, and Assaf Naor. *Boolean functions whose Fourier transform is concentrated on the first two levels and neutral social choice.* Adv. Appl. Math., 29(3):427–437, 2002. doi:10.1016/S0196-8858(02)00024-6. 1, 4

[Fri98] Ehud Friedgut. *Boolean functions with low average sensitivity depend on few coordinates.* Combinatorica, 18(1):27–35, 1998. doi:10.1007/PL00009809. 1, 4

[Fri99] ———. *Sharp thresholds of graph properties, and the k-sat problem.* J. Amer. Math. Soc., 12(4):1017–1054, 1999. doi:10.1090/S0894-0347-99-00305-7. 4

[Hat12] Hamed Hatami. *A structure theorem for Boolean functions with small total influences.* Ann. of Math., 176(1):509–533, 2012. doi:10.4007/annals.2012.176.1.9. 4

[IKW12] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. *New direct-product testers and 2-query PCPs.* SIAM J. Comput., 41(6):1722–1768, 2012. (Preliminary version in *41st STOC*, 2009). eccc:2009/TR09-090, doi:10.1137/09077299X. 6

[Kin03] Guy Kindler. *Property Testing, PCP and Juntas.* Ph.D. thesis, Tel Aviv University, 2003. eccc:2016/TR16-007. 1, 17

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. *The influence of variables on Boolean functions (extended abstract).* In *Proc.* 29*th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 68–80. 1988. doi:10.1109/SFCS.1988.21923. 4

[KKM+17] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, Eren Sasoglu, and Rüdiger L. Urbanke. *Reed-Muller codes achieve capacity on erasure channels.* IEEE Trans. Inform. Theory, 63(7):4298–4316, 2017. (Preliminary version in *48th STOC*, 2016). arXiv:1601.04689, doi:10.1109/TIT.2017.2673829. 5

[KM13] Subhash Khot and Dana Moshkovitz. NP-*hardness of approximately solving linear equations over reals.* SIAM J. Comput., 42(3):752–791, 2013. (Preliminary version in *43rd STOC*, 2011). eccc:2010/TR10-053, doi:10.1137/110846415. 4

[KMS17] Subhash Khot, Dor Minzer, and Muli Safra. *On independent sets, 2-to-2 games, and Grassmann graphs.* In *Proc.* 49*th ACM Symp. on Theory of Computing (STOC)*, pages 576–589. 2017. eccc:2016/TR16-124, doi:10.1145/3055399.3055432. 4

[KS02] Guy Kindler and Shmuel Safra. *Noise-resistant Boolean functions are juntas*, 2002. (manuscript). 1

[NS94] Noam Nisan and Mario Szegedy. *On the degree of Boolean functions as real polynomials.* Comput. Complexity, 4:301–313, 1994. (Preliminary version in *24th STOC*, 1992). doi:10.1007/BF01263419. 1

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi:10.1017/CBO9781139814782. 4

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il