Analyzing Boolean functions on the biased hypercube via higher-dimensional agreement tests*

Irit Dinur† Yuval Filmus‡ Prahladh Harsha§

August 12, 2018

Abstract

We propose a new paradigm for studying the structure of Boolean functions on the *biased* Boolean hypercube, i.e. when the measure is μ_p and p is potentially very small, e.g. as small as $O(1/n)$. Our paradigm is based on the following simple fact: the p-biased hypercube is expressible as a convex combination of many small-dimensional copies of the uniform hypercube. To uncover structure for μ_p, we invoke known structure theorems for $\mu_{1/2}$, obtaining a structured approximation for each copy separately. We then sew these approximations together using a novel "agreement theorem". This strategy allows us to lift structure theorems from $\mu_{1/2}$ to μ_p.

We provide two applications of this paradigm:

- Our main application is a structure theorem for functions that are nearly low degree in the Fourier sense. The structure we uncover in the biased hypercube is *not at all* the same as for the uniform hypercube, despite using the structure theorem for the uniform hypercube as a black box. Rather, new phenomena emerge: whereas nearly low degree functions on the uniform hypercube are close to juntas, when p becomes small, non-juntas arise as well. For example, the function $\max(y_1, \cdots, y_{\varepsilon/p})$ (where $y_i \in \{0,1\}$) is nearly degree 1 despite not being close to any junta.
- A second (technically simpler) application is a test for being low degree in the $GF(2)$ sense, in the setting of the biased hypercube.

In both cases, we use as a black box the corresponding result for $p = 1/2$. In the first case, it is the junta theorem of Kindler and Safra, and in the second case, the low degree testing theorem of Alon *et al.* [*IEEE Trans. Inform. Theory*, 2005] and Bhattacharyya *et al.* [*Proc. 51st FOCS*, 2010].

A key component of our proof is a new local-to-global agreement theorem for higher dimensions, which extends the work of Dinur and Steurer [*Proc. 29th CCC*, 2014]. Whereas their result sews together vectors, our agreement theorem sews together labeled graphs and hypergraphs.

The proof of our agreement theorem uses a novel pruning lemma for hypergraphs, which may be of independent interest. The pruning lemma trims a given hypergraph so that the number of hyperedges in a random induced subhypergraph has roughly a Poisson distribution, while maintaining the expected number of hyperedges.

*This paper combines the results that appeared in two manuscripts [DFH17a, DFH17b] by the authors.

†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

1 Introduction

The p-biased hypercube is the set $\{0,1\}^n$ with the μ_p measure for a parameter $p \in (0,1)$, in which the probability of a string $y = (y_1, \ldots, y_n) \in \{0,1\}^n$ is $\mu_p(y_1, \ldots, y_n) = p^{y_1+\cdots+y_n}(1-p)^{(1-y_1)+\cdots+(1-y_n)}$. There is a great deal known about the structure of Boolean functions in the uniform ($p = 1/2$) case, but less so for the biased setting.

We describe a method for lifting known structure theorems in the $p = 1/2$ case to the general μ_p case. This holds for all values of p, and in particular even when p is potentially very small as a function of n, e.g. $p = O(1/n)$. This stands in contrast to structure theorems proved using hypercontractivity, whose generalization to the μ_p setting typically deteriorates in power as p gets smaller.

The key idea in this method is as follows: to study a function f over the p-biased hypercube, we consider restrictions of the function f to subcubes $\{0,1\}^S$ obtained by fixing all coordinates not in S to 0. The crucial observation is that if we choose S according to the measure μ_{2p} (i.e., $i \in S$ with probability $2p$) and then choose a point x in the subcube $\{0,1\}^S$ uniformly at random, then the point x is distributed according to μ_p.

We study the structure of the function f on μ_p by looking at its restrictions to these small uniform hypercubes $\{0,1\}^S$. One can apply as a black box known structure theorems for the uniform case, and obtain for each hypercube *separately* an approximate structure. To be able to say something coherent about the global structure of our function, we must then be able to "sew" these approximations together. To this end, we design a new "agreement theorem" that stitches together an ensemble of *local* functions that satisfy some local consistency into a single *global* function.

In an agreement theorem, the input is a collection of local functions (e.g. one function per local restriction). In addition, it is also known that the local functions satisfy with high probability some local consistency, i.e., most local functions agree with each other whenever their domains overlap. From this, the agreement theorem concludes the existence of a global function that agrees with most of the initial data of local functions. Agreement theorems originally come from the PCP literature, where they generalize low degree tests and direct product tests. We prove a new "higher-dimensional" agreement theorem, and use this theorem to prove two new results about Boolean functions on the biased hypercube.

Our first and main application of this method is a structure theorem for functions that are nearly low degree in the Fourier sense. A second (technically simpler) application regards testing $GF(2)$ low degreeness. In both cases we use as a black box the corresponding result for $p = 1/2$.

1.1 A new higher-dimensional "agreement theorem"

We now turn to describe the new agreement theorem. In order to motivate the setup, let us fix on the first application: analyzing the structure of a low degree nearly-Boolean[1] function on the biased hypercube.

Let $f : \{0,1\}^n \to \mathbb{R}$ have degree d and suppose it is ε-close to Boolean. The idea is to consider restrictions of the function f to subcubes $\{0,1\}^S$ obtained by fixing all coordinates not in S to 0. As pointed out earlier, if we choose S according to the measure μ_{2p} (i.e., $i \in S$ with probability $2p$) and then choose a point $x \in \{0,1\}^S$ uniformly at random, then the point x is distributed according to μ_p.

Let us denote by $f|_S$ the restriction of f to $\{0,1\}^S$. Note that S is chosen according to μ_{2p}, and the distribution on $\{0,1\}^S$ conditioned on S is the standard uniform measure (i.e., $\mu_{1/2}$). An averaging argument implies that $f|_S$ itself is close to being Boolean, where closeness is now according to the $\mu_{1/2}$ measure. We can then use a known structure theorem for $\mu_{1/2}$ to obtain information about $f|_S$ locally on each subcube $\{0,1\}^S$. For example, from the theorem of Kindler and Safra we get a function g_S that is a junta on S, and approximates $f|_S$ well on $\{0,1\}^S$.

The next step is to obtain global information about f on the p-biased hypercube, by "patching" the local pieces g_S to a global function g on the entire hypercube which agrees on most of the local pieces.

[1] It is equivalent to analyzing functions that are Boolean and nearly-low degree.

If the pieces g_S were completely arbitrary, then it would be impossible to patch them to a global function. However, since the g_S's were obtained from local restrictions of the same function f, we are typically able to show that if we choose $S_1, S_2 \sim \mu_{2p}$ in a coupled way which guarantees that S_1, S_2 have significant overlap, then the local functions g_{S_1}, g_{S_2} completely agree on the intersection of their domains with probability $1 - O(\varepsilon)$.

Let us recall the agreement theorem of Dinur and Steurer [DS14]. An equivalent rephrasing[2] of their result concerns an ensemble of local functions $v_S \colon S \to \Sigma$ for every $S \subseteq [n]$, where Σ is some finite alphabet. Suppose that we choose a pair of sets S_1, S_2 according to the distribution $\mu_{p,\alpha}$, in which $i \in S_1 \cap S_2$ with probability $p\alpha$, $i \in S_1 \setminus S_2$ with probability $p(1-\alpha)$, and $i \in S_2 \setminus S_1$ with probability $p(1-\alpha)$. The result of Dinur and Steurer states that if $\Pr[v_{S_1} \neq v_{S_2}] = \varepsilon$ then there exists a global function $v \colon [n] \to \Sigma$ such that $\Pr_{S \sim \mu_p}[v_S \neq v|_S] = O(\varepsilon)$.

This theorem goes in the correct local-to-global spirit but as is it is not useful for us, since the local data we have per S cannot be described by a vector $v_S : S \to \Sigma$. This motivates a different but analogous agreement theorem that is "higher-dimensional" .

More precisely, we can identify each local function g_S with a multi-dimensional function $f_S \colon \binom{S}{\leq d} \to \Sigma$. Our main technical result is that the agreement theorem of Dinur and Steurer can be extended to this high-dimensional setting. More precisely:

Theorem 1.1 (High-dimensional agreement theorem via majority decoding)**.** *For every positive integer d and finite alphabet Σ, there exists a constant $p_0 \in (0, 1/2)$ such that for all $p \in (0, p_0)$, all $\alpha \in (0, 1)$, and all 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*

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

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 $\mu_p([n])$) satisfies

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

We remark that the above theorem shows that the global function G can be obtained from the local functions f_S by the natural *majority decoding* (more accurately, *plurality decoding*) procedure. For instance, in the one-dimensional setting ($d = 1$) of Dinur and Steurer, we have that the value of $v(i)$ is the μ_p-most common value of $v_S(i)$ among all sets S containing i. The agreement theorem of Dinur and Steurer doesn't specify how v is constructed from the v_S, whereas our theorem guarantees that v is formed using majority decoding. Our new result therefore improves on the Dinur–Steurer result even in the one-dimensional case. We note that this strengthening of the agreement theorem (even for the one-dimensional case) is needed for technical reasons in one of our applications.

We now turn to describe the two applications of the agreement theorem.

1.2 The structure of Boolean functions with low real degree

We study the structure of "simple" Boolean functions in the p-biased hypercube. A well-accepted measure of simplicity is the approximate Fourier degree of the function. Nisan and Szegedy [NS94] showed that a Boolean function on the hypercube that is exactly of degree $\leq d$ must be a junta (i.e., a function that depends only on a constant number of variables). Kindler and Safra [KS02, Kin03] extended this to degree d functions which are merely *close* to being Boolean, showing that such functions are *close* to juntas. The earlier work of Friedgut, Kalai and Naor [FKN02] proved a similar theorem for the case $d = 1$.

The closeness in the above theorems is with respect to the uniform measure on $\{0,1\}^n$. In many applications, one is interested in studying the hypercube with respect to biased measures. It is easy to see that both

[2] Dinur and Steurer state their main result in a different language.

the Friedgut–Kalai–Naor theorem and the Kindler–Safra theorem extend for any fixed $p \in (0,1)$, but when p tends to 0, new behavior emerges. For example, the function $y_1 + \cdots + y_{\sqrt{\varepsilon}/p}$ is a degree 1 function which is $O(\varepsilon)$-close to Boolean but not $O(\varepsilon)$-close to any junta.[3] It was shown by the second-named author [Fil16] that such functions are essentially the only degree 1 functions which are close to Boolean. We call this result *the biased FKN theorem*.

We show a similar result for larger degree polynomials, in particular a common generalization of the Kindler–Safra theorem and the biased FKN theorem. As demonstrated by the example $y_1 + \cdots + y_{\sqrt{\varepsilon}/p}$, the class of juntas does not suffice to characterize all degree d functions that are ε-close to Boolean functions for small p. Thus, we must first *uncover* the "correct" class of simple functions, which we refer to as *sparse juntas*.

We note that the function $y_1 + \cdots + y_{\sqrt{\varepsilon}/p}$ satisfies two properties: (a) the non-zero coefficients in the "polynomial expansion" of the function come from a finite set (independent of n); and (b) a random input (distributed according to the μ_p measure) zeroes out all but $O(1)$ monomials in the polynomial expansion with probability $1 - O(\varepsilon)$. *Our main result shows that any low-degree function close to a Boolean function is close to a function satisfying these two properties.*

The first step towards defining *sparse juntas*, is to define the notion of "polynomial expansion" we employ.

Definition 1.2 (y-expansion)**.** *The y-expansion of a function $f\colon \{0,1\}^n \to \mathbb{R}$ is the unique multilinear expansion $f(y) = \sum_S \tilde{f}(S) y_S(x)$, where $\{y_S\}_S$ is the basis of functions given by $y_S = \prod_{i \in S} y_i$.*

We use the terminology y-expansion to stress that this is *not* the standard Fourier expansion of f (under $\mu_{1/2}$), which is its expansion as a multilinear polynomial in ± 1 input variables. Even more importantly, the basis of the y-expansion is independent of p and is not the set of p-biased Fourier characters, which form the standard μ_p-orthonormal basis while working with functions on $\{0,1\}^n$ under the μ_p measure.

The biased FKN theorem 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 degree 1 function whose non-zero y-expansion coefficients are all in the set $\{\pm 1\}$. This motivates the following definition of *quantized polynomials*.

Definition 1.3 (quantized polynomial)**.** *Given a finite set $A \subset \mathbb{R}$, a function $f\colon \{0,1\}^n \to \mathbb{R}$ is said to be an A-quantized polynomial of degree d if all the non-zero coefficients of the y-expansion of f belong to A.*

The class of sparse juntas consists of 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 $r \geq 0$, 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 the main theorem of this section:

Theorem 1.4 (biased Kindler–Safra theorem)**.** *If $f\colon \{0,1\}^n \to \mathbb{R}$ is a degree d function which is ε-close to Boolean with respect to the μ_p measure for some $p \leq 1/2$ then f is $O(\varepsilon)$-close to a "sparse junta" degree d polynomial g in the sense that:*

1. *$\|f - g\|^2 = O(\varepsilon)$.*
2. ***(g is quantized)** All non-zero coefficients of the y-expansion of g belong to a finite set $Q(d)$ which is independent of p, ε, and n. (When $d = 1$, $Q(d) = \{\pm 1\}$.)*

[3] Throughout the paper we say that f is ε-close to g if $\|f - g\|^2_{\mu_p} := \mathbb{E}_{x \sim \mu_p}[(f(x) - g(x))^2] \leq \varepsilon$. Similarly, f is ε-close to Boolean if f is ε-close to some Boolean function.

3. *(g **has bounded branching factor**) For each $e \le d$, the function g has $O((1/p)^e)$ monomials of degree e. Moreover, at most $O((1/p)^{e-t})$ monomials of degree e are multiples of $y_{i_1} \cdots y_{i_t}$ for any $i_1, \ldots, i_t$.*

4. *(g **is nearly Boolean**) The function g is Boolean on $1 - O(\varepsilon)$ of its inputs.*

5. *(g **is sparse**) A random input (distributed μ_p) zeroes out all but $O(1)$ monomials of g with probability $1 - O(\varepsilon)$.*

(See Theorem 10.1 for a formal statement.) We also show that the above theorem actually provides a characterization of all degree d functions which are ε-close to Boolean, in the sense that every function which satisfies the properties listed above is $O(\varepsilon)$-close to Boolean (see Lemma 11.1). In this sense, Theorem 1.4 is similar to Hatami's celebrated result [Hat12], which characterizes functions on the p-biased hypercube with low total influence.

When $d = 1$, all sparse juntas have the same structure: either $\sum_{i=1}^m y_i$ or $1 - \sum_{i=1}^m y_i$, where $m = O(\sqrt{\varepsilon}/p)$. The situation gets considerably more complex for higher d. Here are some of the possibilities for $d = 2$:

1. Disjoint pairs: $\sum_{i=1}^m x_i y_i$ for $m = O(\sqrt{\varepsilon}/p^2)$.

2. Non-disjoint pairs: $\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} x_i y_{i,j}$ for $m_1 m_2 = O(\sqrt{\varepsilon}/p^2)$.

3. Intertwined XOR: $\sum_{i=1}^m y_i - 2\sum_{1 \le i < j \le m} y_i y_j$ for $m = O(\sqrt[3]{\varepsilon}/p)$.

4. Intertwined OR: $\sum_{i=1}^m y_i - \sum_{1 \le i < j \le m} y_i y_j$ for $m = O(\sqrt[4]{\varepsilon}/p)$.

For $d = 2$, we have a complete list of all Boolean degree 2 functions,[4] and so in principle we can describe all sparse juntas of degree 2. For general d there is a combinatorial explosion of possibilities (indeed, even the largest number of coordinates that such a function depends on is unknown), and so all we can hope for is a characterization along the lines provided by our main theorem.

To illustrate the usefulness of the structure uncovered by our main theorem, we give two corollaries. The first is a large deviation bound:

Lemma 1.5 (Large deviation bound)**.** *If $f\colon \{0,1\}^n \to \mathbb{R}$ is a degree d function which is ε-close to Boolean with respect to the μ_p measure for some $p \le 1/2$, then for large t,*

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

Our second corollary shows that every degree d function which is ε-close to Boolean must be quite biased:

Lemma 1.6 (Sparse juntas are biased)**.** *If $f\colon \{0,1\}^n \to \mathbb{R}$ is a degree d function which is ε-close to Boolean with respect to the μ_p measure for some $p \le 1/2$, then f is $O(\varepsilon^{C_d} + p)$-close to a constant function, where $C_d < 1$ depends only on d.*

This shows that if we are willing to settle with an $O(\varepsilon^C)$-approximation for some fixed $C < 1$, then we can replace the sparse junta in Theorem 1.4 with a constant function.

Extension to quantized functions All the results stated above hold in greater generality. Instead of requiring the functions to be close to Boolean, it suffices to assume that they are close to being A-valued, where A is an arbitrary finite set; the parameters appearing in the various results now depend not only on d, but also on A. The advantage of this point of view is that it allows us to formulate the following corollary of Theorem 1.4:

[4] Up to permutation and negation of inputs and output, every Boolean degree 2 function is one of the following: 0, x, xy, $x(1-y) + (1-x)y$, $xy + (1-x)z$, $[x = y = z]$, $[x \le y \le z \le w \vee x \ge y \ge z \ge w]$.

If a degree d function over $\{0,1\}^n$ is close to being A-valued, then the coefficients of its polynomial expansion are close to being B-valued, where B is a finite set depending only on d and A.

This point of view inspired us to give a new proof of the Kindler–Safra theorem, very different from the original one, which proceeds by induction on the degree. This proof can be found in Section 13.

1.3 $GF(2)$-low degree testing in the biased hypercube

As a second illustration of our method, we lift the low degree test [AKK+05, BKS+10] to the biased setting using a straightforward application of the agreement theorem. This is similar to the way in which the analysis of the "uniform BLR" test was lifted from the middle slice to an arbitrary slice by David *et al.* [DDG+17].

Alon *et al.* [AKK+05] studied a 2^{d+1}-query test T_d to test low-degreeness. Bhattacharyya *et al.* [BKS+10] gave an optimal analysis of this test to show that $\delta_d(f) = O_d(\mathrm{rej}_d(f))$, where $\delta_d(f)$ refers to the distance of f to the closest degree d function under the $\mu_{1/2}$ measure (i.e., $\delta_d(f) = \min_{\mathrm{bdeg}(g)\le d} \Pr_{\mu_{1/2}}[f \neq g]$), and $\mathrm{rej}_d(f)$ is the rejection probability of the test T_d on input function f. We would like to extend the test T_d to the p-biased setting, wherein we measure closeness of f to Boolean degree function with respect to the μ_p measure instead of $\mu_{1/2}$ measure. More precisely, $\delta_d^{(p)}(f) := \min_{\mathrm{bdeg}(g)\le d} \Pr_{\mu_p}[f \neq g]$. To this end, we study the following test $T_{p,d}$.

- Test $T_{p,d}$: Input $f\colon \{0,1\}^n \to \{0,1\}$
  - Pick $S \subseteq [n]$ according to the distribution μ_{2p}.
  - Let $f|_S\colon \{0,1\}^S \to \{0,1\}$ denote the restriction of f to $\{0,1\}^S$ by zeroing out all the coordinates outside S.
  - Pick $x, a_1, \ldots, a_{d+1} \in \{0,1\}^S$ independently from the distribution $\mu_{1/2}^{\otimes S}$, subject to the constraint that $a_1, \ldots, a_{d+1}$ are linearly independent.
    (If $|S| \le d$, skip this and the following step, and immediately accept.)
  - Accept iff
$$\sum_{I\subseteq[d+1]} f|_S\left(x + \sum_{i\in I} a_i\right) = 0 \pmod 2 .$$

We use the agreement theorem to show that this natural extension is a valid low-degree test for the p-biased setting.

Theorem 1.7. *For every d, there exists a $p_0 = p_0(d)$ such that for all $p \in (0, p_0)$ the 2^{d+1}-query test $T_{p,d}$ (described above) satisfies the following properties.*

- *Completeness: if f has $GF(2)$-degree at most d then $\mathrm{rej}_{T_{p,d}}(f) = 0$.*
- *Soundness: $\delta_d^{(p)}(f) = O_d(\mathrm{rej}_{T_{p,d}}(f))$, where the hidden constant is independent of p.*

We remark that we actually prove a stronger theorem which works for all $p \in (0,1)$, not just $p \in (0, p_0(d))$. However, the test for other ranges of p is not $T_{p,d}$ but a slight variant of it (see Theorem 8.7 for exact details).

1.4 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. Starting from the result of Nisan and Szegedy [NS94] (which was recently improved by Chiarelli, Hatami and Saks [CHS18]), 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 by 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 of Benjamini and Kalai [BK18, 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 $f(x) = \max_{S:|S|\leq d} \tilde{f}(S) y_S(x)$. 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$^+$15], the real code [KM13], and the Grassmann code [KMS17]. In all of these, understanding the structure of Boolean functions with nearly low degree seems important. A recent line of work [KMS17, DKK$^+$18b, DKK$^+$18a, KMS18] proved the 2-to-1 conjecture by analyzing 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 $O(1/n)$. Along this vein a recent work [KMMS18] analyzed certain small set expansion of the Johnson scheme which is the "fixed slice" version of the biased hypercube.

- 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.

- Relation of agreement theorem to property testing: Agreement testing is similar to property testing in that we study the relation between a global object and its local views. In property testing we have access to a single global object, and we restrict ourselves to look only at random local views of it. In agreement tests, we don't get access to a global object but rather to an *ensemble of local functions* that are not a priori guaranteed to come from a single global object. Another difference is that unlike in property testing, in an agreement test the local views are pre-specified and are a part of the problem description, rather than being part of the algorithmic solution. Consider the following special case of the agreement theorem for $d = 2$ and $\Sigma = \{0,1\}$, which gives an interesting statement about combining small pieces of a graph into a global one.

  Corollary 1.8 (Agreement test for graphs)**.** *There exist a constant $C > 1$ such that for all $\alpha, \beta \in (0,1)$ satisfying $\alpha + \beta \leq 1$ and all for all positive integers $n \geq k \geq t \geq 4$ satisfying $n \geq Ck$, $t \geq \alpha k$ and $k - t \geq \max\{\beta k, 2\}$ the following holds:*

  Let $\{G_S\}$ be an ensemble of graphs, where S is a k element subset of $[n]$ and G_S is a graph on vertex set S. Suppose that

  $$\Pr_{\substack{S_1,S_2 \in \binom{[n]}{k} \\ |S_1 \cap S_2| = t}} \left[G_{S_1}|_{S_1 \cap S_2} = G_{S_2}|_{S_1 \cap S_2} \right] \geq 1 - \varepsilon.$$

  Then there exists a single global graph $G = ([n], E)$ satisfying $\Pr_{S \in \binom{[n]}{k}}[G_S = G|_S] = 1 - O(\varepsilon)$.

  There is an interesting interplay between Corollary 1.8, which talks about combining an ensemble of local graphs into one global graph, and graph property testing. Suppose we focus on some testable graph property, and suppose further that the test proceeds by choosing a random set of vertices and reading all of the edges in the induced subgraph, and checking that the property is satisfied there (many graph properties are testable this way, for example bipartiteness [GGR98]). Suppose we only allow ensembles $\{G_S\}$ where for each subset S, the local graph G_S satisfies the property (e.g. it is bipartite). This fits into our formalism by specifying the space of allowed functions $\mathcal{F}_S$ to consist only of accepting local views. This is analogous to requiring, in the low degree test, that the local function on each line has low degree as a univariate polynomial. By Corollary 1.8, we know that if these local graphs agree with each other with probability $1 - \varepsilon$, there is a global graph G that agrees with $1 - O(\varepsilon)$ of them. In particular, this graph *passes the property test*, so must itself be close to having the property! At this point it is absolutely crucial that the agreement theorem provides the stronger guarantee that $G|_S = G_S$ (and not $G|_S \approx G_S$) for $1 - O(\varepsilon)$ of the S's. We can thus conclude that not only is there a global graph G, but actually that this global G is close to having the property.

  This should be compared to the low degree agreement test, where we only allow local functions with low degree, and the conclusion is that there is a global function that itself has low degree.

Organization

The rest of the paper is organized as follows. We begin with a few preliminaries in Section 2. In Section 3, we define the branching factor and discuss some of its properties. The rest of the paper is divided into two parts; in Part I we prove the agreement theorem and in Part II we prove the two applications, Theorem 1.4 and Theorem 1.7.

Part I: We begin this part in Section 4 by (re-)proving dimension one case of the agreement theorem (namely the result of Dinur and Steurer [DS14]), in a manner that generalizes to higher dimension. We then generalize the proof of the $d = 1$ theorem to higher dimensions (Theorem 5.1) in Section 5. This almost proves the agreement theorem, but for the majority decoding part. In Section 6, we prove the hypergraph pruning lemma, a crucial ingredient in the generalization to higher dimensions. Finally, in Section 7, we use the hypergraph pruning lemma (again) to prove the majority decoding of Theorem 7.2, thus completing the proof of Theorem 1.1.

Part II: The application to low degree testing and the proof of Theorem 1.7 appears in Section 8. We generalize the classical Kindler–Safra theorem to A-valued functions in Section 9. We then prove the main result regarding structure of Boolean functions with nearly low degree (Theorem 1.4) in Section 10. In Section 11, we prove the converse to Theorem 1.4. We discuss some applications in Section 12 and give an alternate proof of the classical Kindler–Safra theorem in Section 13.

Summary of results For the benefit of the reader, we summarize below the list of results proved in the paper:

1. Higher-dimensional agreement theorem, Theorem 1.1, proved in Section 7.
2. Hyperergraph pruning lemma, Lemma 6.1.
3. Versions of items 1 and 2 for the uniform setting, in which $(\{0,1\}^n, \mu_p)$ is replaced with the slice $\binom{[n]}{np}$: Theorem 7.2 (agreement theorem) and Lemma 3.5 (hypergraph pruning lemma).
4. Biased low degree test, Theorem 8.7.
5. Biased Kindler–Safra theorem, Theorem 10.1, and a converse, Lemma 11.1.
6. Two corollaries: a large deviation bound, Corollary 12.5, and a bound on the deviation from being constant, Corollary 12.7.
7. A new proof of the unbiased Kindler–Safra theorem, Theorem 13.7 (see also Theorem 9.1, in which the A-valued version of the Kindler–Safra theorem is derived from its Boolean version).

2 Preliminaries

We will need the following definitions:

- We define $\operatorname{dist}(x, A) = \min_{y \in A} |x - y|$.
- We define $\operatorname{round}(x, A)$ as an element in A whose distance from x is $\operatorname{dist}(x, A)$.
- 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 (defined on page 3) naturally corresponds to a hypergraph $H_f \subset 2^{[n]}$, which we sometimes refer to as the *support* of f.
- 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

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

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

We will need the following theorems.

Theorem 2.1 (Nisan–Szegedy [NS94], Chiarelli–Hatami–Saks [CHS18])**.** *If $f\colon \{0,1\}^n \to \{0,1\}$ is a degree k function, then f is an $O(2^k)$-junta.*

Theorem 2.2 ($(2,p)$ hypercontractivity (see [O'D14, Chapter 9]))**.** *Let $p \geq 2$, then for any function $f\colon \{0,1\}^n \to \mathbb{R}$ of degree at most k, we have $\|f\|_p \leq (p-1)^{k/2} \cdot \|f\|_2$.*

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 $\deg g_1, \deg g_2 \leq d$. 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. □

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 12, 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 $A \not\subseteq B$ 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 $A \not\subseteq B$ 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 $A \not\subseteq B$ is at least

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

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

Part I
Agreement testing

Agreement tests are a type of PCP tests that capture fundamental local-to-global phenomena. A key example is the line vs. line [GLR+91, RS96] low degree test in the original proof of the PCP theorem. The simplest agreement theorem is the classic direct product test. In the 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 $\mu_{p,\alpha}$ 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 $p\alpha$, to $S_1 \setminus S_2$ with probability $p(1-\alpha)$, to $S_2 \setminus S_1$ with probability $p(1-\alpha)$, and to neither set otherwise. Here, we assume $p \le 1/2$ and $q \in (0,1)$. The direct product testing results [DG08, IKW12, DS14] state that if the local functions agree most of the time, i.e.,

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

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

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

It will be convenient for us to reformulate the direct test as follows: the global function G can be viewed as specifying the coefficients of a linear form $\sum_{i=1}^n G(i)x_i$ over variables $x_1, \ldots, x_n$. For each S, the local function f_S specifies the partial linear form only over the variables in S. This f_S is supposed to be equal to G on the part of the domain where $x_i = 0$ for all $i \notin S$. Given an ensemble $\{f_S\}$ whose elements are promised to agree with each other on average, the agreement theorem allows us to conclude the existence of a global linear function that agrees with most of the local pieces.

The agreement theorem required to prove Theorem 1.4 is a high-degree analogue of the above direct product test. Here, the global function G is a degree d polynomial with coefficients in Σ, namely $G(x) = \sum_T G(T)x_T$, where we sum over subsets $T \subset [n]$, $|T| \le d$. The local functions f_S will be polynomials of degree $\le d$, supposedly obtained by zeroing out all variables outside S. Two local functions f_{S_1}, f_{S_2} are said to agree, denoted $f_{S_1} \sim f_{S_2}$, if every monomial that is induced by $S_1 \cap S_2$ has the same coefficient in both polynomials. Our new agreement theorem states that in this setting as well, local agreement implies global agreement.

Theorem 1.1 (Restated; Agreement theorem via majority decoding) *For every positive integer d, finite alphabet Σ, and positive $\eta > 0$, the following holds for all $p \in (0, 1-\eta)$, $\alpha \in (0,1)$, and all n. Let $\{f_S\colon \binom{S}{\le d} \to \Sigma \mid S \in \{0,1\}^n\}$ be an ensemble of functions satisfying*

$$\Pr_{S_1,S_2\sim\mu_{p,\alpha}}[f_{S_1}|_{S_1\cap S_2} \ne f_{S_2}|_{S_1\cap S_2}] \le \varepsilon.$$

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

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

For $d = 1$, this theorem is precisely the direct product theorem of Dinur and Steurer [DS14] but for the fact that the Dinur-Steurer theorem only proved that that a global function exists and did not show that the global function obtained by plurality decoding works. This strengthens our theorem by naming the popular vote function as a candidate global function that explains most of the local functions even for the dimension one case.

Proof sketch of the agreement theorem

Our proof of Theorem 1.1 proceeds by induction on the dimension d. For $d = 1$, this is the direct product test theorem of Dinur and Steurer [DS14], which we reprove in a way that more readily generalizes to higher dimensions. Given an ensemble $\{f_S\}$, it is easy to define the global function G, by popular vote ("majority decoding"). The main difficulty is to prove that for a typical set S, f_S agrees with $G|_S$ on all elements $i \in S$ (and later on all d-sets).

Our proof doesn't proceed by defining G as majority vote right away. Instead, like in many previous proofs [DG08, IKW12, DS14], we condition on a certain event (focusing say on all subsets that contain a certain set T, and such that $f_S|_T = \alpha$ for a certain value of α), and define a "restricted global" function, for each T, by taking majority just among the sets in the conditioned event. This boosts the probability of agreement inside this event. After this boost, we can afford to take a union bound and safely get agreement with the restricted global function G_T. The proof then needs to perform another agreement step which stitches the restricted global functions $\{G_T\}_T$ into a completely global function. The resulting global function does not necessarily equal the majority vote function G, and a separate argument is then carried out to show that the conclusion is correct also for G.

In higher dimensions $d > 1$, these two steps of agreement (first to restricted global and then to global) become a longer sequence of steps, where at each step we are looking at restricted functions that are defined over larger and larger parts of the domain.

The technical main difficulty is that a single event $f_S = F|_S$ consists of $\binom{k}{d}$ little events, namely $f_S(A) = F(A)$ for all $A \in \binom{S}{d}$, that each have some probability of failure. We thus need to boost the failure probability from ε to ε/k^d so that we can afford to take a union bound on the $\binom{k}{d}$ different sub-events. How do we get this large boost? Our strategy is to proceed by induction, where at each stage, we condition on the global function from the previous stage, boosting the probability of success further.

Hypergraph pruning lemma An important technical component that yields this boosting is the following hypergraph pruning lemma (Lemma 3.5). This lemma allows approximating a given hypergraph H by a subhypergraph $H' \subset H$ that has a *bounded branching factor*.

Lemma 3.5 (hypergraph pruning lemma)**.** *Fix constants $\varepsilon > 0$ and $d \geq 1$. There exists $p_0 > 0$ (depending on d, ε) such that for every $n \geq k \geq 2d$ satisfying $k/n \leq p_0$ and every d-uniform hypergraph H on $[n]$ there exists a subhypergraph H' obtained by removing hyperedges such that*

1. $\Pr_{S\sim\nu_{n,k}}[H'|_S \neq \emptyset] = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset])$.
2. *For every $e \in H'$,* $\Pr_{S\sim\nu_{n,k}}[H'|_S = \{e\} \mid S \supset e] \geq 1 - \varepsilon$.

Here $H'|_S$ is the hypergraph induced on the vertices of S.

The lemma can be interpreted by viewing a hypergraph as specifying the minterms of a monotone DNF of width at most d. The lemma allows to prune the DNF so that the new sub-DNF still has similar density (the fraction of inputs on which it is 1), but also has a structural property which we call *bounded branching factor* and which implies that for typical inputs, only a single minterm is responsible for the function evaluating to 1.

Our proof of the hypergraph pruning lemma produces a sub-hypergraph with branching factor $\rho = O(n/k)$. The branching factor is responsible for the second item in the lemma, which guarantees that usually if a set S contains a hyperedge from H, it contains a *unique* hyperedge from H'.

The importance of this is roughly for "inverting union bound arguments". It essentially allows us to estimate the probability of an event of the form "S contains some hyperedge of H'" as the sum, over all hyperedges, of the probability that S contains a specific hyperedge.

The proof of the lemma is subtle and proceeds by induction on the dimension d. It essentially describes an algorithm for obtaining H' from H and the proof of correctness uses the FKG inequality. We illustrate how Lemma 3.5 is used by its application to majority decoding.

Majority decoding The most natural choice for the global function F in the conclusion of Theorem 7.2 is the majority decoding, where $F(A)$ is the most common value of $f_S(A)$ over all S containing A. This is the content of the "furthermore" clause in the statement of the theorem. Neither the proof strategy of Dinur and Steurer [DS14] nor our generalization promises that the produced global function F is the majority decoding. Our inductive strategy produces a global function which agrees with most local functions, but we cannot guarantee immediately that this global function corresponds to majority decoding. What we are able to show is that *if* there is a global function agreeing with most of the local functions *then* the function obtained via majority decoding also agrees with most of the local functions. We outline the argument below. Suppose that $\{f_S\}$ is an ensemble of local functions that mostly agree with each other, and suppose that they also mostly agree with some global function F. Let G be the function obtained by majority decoding: $G(A)$ is the most common value of $f_S(A)$ over all S containing A. Our goal is to show that G also mostly agrees with the local functions, and we do this by showing that F and G mostly agree.

Suppose that $F(A) \neq G(A)$. We consider two cases. If the distribution of $f_S(A)$ is very skewed toward $G(A)$, then $f_S(A) \neq F(A)$ will happen very often. If the distribution of $f_S(A)$ is very spread out, then $f_{S_1}(A) \neq f_{S_2}(A)$ will happen very often. Since both events $f_S(A) \neq F(A)$ and $f_{S_1}(A) \neq f_{S_2}(A)$ are known to be rare, we would like to conclude that $F(A) \neq G(A)$ happens for very few A's.

Here we face a problem: the bad events (either $f_S(A) \neq F(A)$ or $f_{S_1}(A) \neq f_{S_2}(A)$) corresponding to different A's are not necessarily disjoint. A priori, there might be many different A's such that $F(A) \neq G(A)$, but the bad events implied by them could all coincide.

The hypergraph pruning lemma enables us to overcome this difficulty. Let $H = \{A : F(A) \neq G(A)\}$, and apply the hypergraph pruning lemma to obtain a subhypergraph H'. The lemma states that with constant probability, a random set S sees at most one disagreement between F and G. This implies that the bad events considered above can be associated, with constant probability, with a *unique* A. In this way, we are able to obtain an upper bound on the probability that F, G disagree on an input from H'. The hypergraph pruning lemma then guarantees that the probability that F, G disagree (on *any* input) is also bounded.

4 One-dimensional agreement theorem

In this section, we prove the following direct product agreement testing theorem for dimension one in the uniform setting. This theorem is a special case of the more general theorem (Theorem 5.1) proved in the next section and also follows from the work of Dinur and Steurer [DS14]. However, we give the proof for the dimension one case as it serves as a warmup to the general dimension case.

Theorem 4.1 (Agreement theorem, dimension 1)**.** *There exists constants $C > 1$ such that for all $\alpha, \beta \in (0,1)$ satisfying $\alpha + \beta \leq 1$, all positive integers n, k, t satisfying $n \geq Ck$ and $t \geq \alpha k$ and $k - t \geq \beta k$, and all finite alphabets Σ, the following holds: Let $f = \{f_S : S \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying $\mathrm{agree}_{\nu_{n,k,t}}(f) \geq 1 - \varepsilon$, that is,*

$$\Pr_{S_1,S_2\sim\nu_{n,k,t}}[f_{S_1}|_{S_1\cap S_2} = f_{S_2}|_{S_1\cap S_2}] \geq 1-\varepsilon,$$

where $\nu_{n,k,t}$ is the uniform distribution over pairs of k-sized subsets of $[n]$ of intersection exactly t. Then there exists a global function $F : [n] \to \Sigma$ satisfying $\Pr_{S\in\binom{[n]}{k}}[f_S = F|_S] = 1 - O_{\alpha,\beta}(\varepsilon)$.

The distribution $\nu_{n,k.t}$ is the distribution induced on the pair of sets $(S_1, S_2) \in \binom{[n]}{k}^2$ by first choosing uniformly at random a set $U \subset [n]$ of size t and then two sets S_1 and S_2 of size k of $[n]$ uniformly at random conditioned on $S_1 \cap S_2 = U$. We can think of picking these two sets as first choosing uniformly at random a set T of size $t-1$, then a random element $i \in [n] \setminus T$, setting $U = T + i$ and then choosing two sets S_1 and S_2 such that $S_1 \cap S_2 = T + i$. Clearly, the probability that the functions f_{S_1} and f_{S_2} disagree is the sum of the probabilities of the following two events: (A) $f_{S_1}|_T \neq f_{S_2}|_T$, (B) $f_{S_1}|_T = f_{S_2}|_T$ but $f_{S_1}(i) \neq f_{S_2}(i)$. This motivates the following definitions for any $T \in \binom{[n]}{t-1}$ and $i \in [n] \setminus T$.

$$\varepsilon_T(\emptyset) = \Pr_{\substack{S_1,S_2 \sim \nu(k,t) \\ S_1 \cap S_2 \supseteq T}} [f_{S_1}|_T \neq f_{S_2}|_T],$$
$$\varepsilon_T(i) = \Pr_{\substack{S_1,S_2 \sim \nu(k,t) \\ S_1 \cap S_2 = T+i}} [f_{S_1}|_T = f_{S_2}|_T \text{ and } f_{S_1}(i) \neq f_{S_2}(i)].$$

It is easy to see that for a typical T, both $\varepsilon_T(\emptyset)$ and $\mathbb{E}_{i \notin T}[\varepsilon_T(i)]$ is $O(\varepsilon)$. This suggests the following strategy to prove Theorem 4.1. For each typical T, construct a "global" function $g_T \colon [n] \to \Sigma$ based on the most popular value of f_S among the f_S's that agree on T (see Section 4.2 for details) and show that most g_T's agree with each other. More precisely, we prove the theorem in 3 steps as follows: In the first step (Section 4.1), we bound $\varepsilon_T(\emptyset)$ and $\varepsilon_T(i)$ for typical T's and i. In the second step (Section 4.2), we construct for a typical T, a "global" function g_T that explains most "local" $\{f_S\}_{S \supset T}$. In the final step (Section 4.3), we show that the global functions corresponding to most pairs of typical T's agree with each other, thus demonstrating the existence of a single global function F (in particular a random global function g_T) that explains most of the "local" functions f_S even corresponding to S's which do not contain T.

4.1 Step 1: Bounding $\varepsilon_T(\emptyset)$ and $\varepsilon_T(i)$

We begin by showing that for a typical T of size $t-1$, we can upper bound $\varepsilon_T(\emptyset)$ and $\mathbb{E}_{i \notin T}[\varepsilon_T(i)]$.

Lemma 4.2. *We have* $\mathbb{E}_T[\varepsilon_T(\emptyset)] \leq \varepsilon$ *and* $\mathbb{E}_{T, i \notin T}[\varepsilon_T(i)] \leq \frac{\varepsilon}{t}$.

Proof. For a non-negative integer j, let ε_j be the probability that the functions f_{S_1} and f_{S_2} corresponding to a pair of sets (S_1, S_2) picked according to the distribution $\nu_n(k,t)$ disagree on exactly j elements in $S_1 \cap S_2$. By assumption of Theorem 4.1, we have $\sum_{j=1}^{t} \varepsilon_j \leq \varepsilon$. Furthermore, it is easy to see that $\mathbb{E}_T[\varepsilon_T(\emptyset)] = \left(1 - \frac{1}{t}\right) \varepsilon_1 + \sum_{j>1} \varepsilon_j$ and $\mathbb{E}_{T,i}[\varepsilon_T(i)] = \varepsilon_1 / t$. The lemma follows from these observations. $\square$

We will need the following auxiliary lemma in our analysis.

Lemma 4.3. *Let $c \in (0,1)$ and $n \geq 4k/c$. Consider the bipartite inclusion graph between $[n]$ and $\binom{[n]}{k}$ (ie., (i,S) is an edge if $i \in S$). Let $B \subset [n]$ and $T \subset \binom{[n]}{k}$ be such that for each $i \in B$, the set of neighbours of i in T (denoted by $T_i := \{S \in T \mid S \ni i\}$) is of size at least $c\binom{n-1}{k-1}$. Then either*

$$\Pr_{S \sim \nu_{n,k}} [S \in T] \geq \max\left\{ \frac{ck}{2} \cdot \Pr_i[i \in B], \frac{c^2}{16} \right\}.$$

Proof. Let S be a random set of size k. To begin with, we can assume that $|B| \leq n/2$ since otherwise $\Pr_S[S \in T] \geq c/2 \geq c^2/16$ and we are done. Let i be any element in B. The probability that $S \cap B = \{i\}$ conditioned on the event that S contains i is given as follows:

$$\Pr[S \cap B = \{i\} \mid i \in S] = \prod_{i=1}^{|B|-1} \left(1 - \frac{k-1}{n-i}\right) \geq \left(1 - \frac{k-1}{n-|B|}\right)^{|B|} \geq 1 - \frac{k}{n/2}|B|.$$

Hence, for any $i \in B$, $\Pr[S \in T_i \text{ and } S \cap B = \{i\} \mid i \in S] \geq c - \frac{2k}{n} \cdot |B|$. It follows that

$$\Pr[S \in T] \geq \sum_{i \in B} \Pr[S \in T_i \text{ and } S \cap B = \{i\}] \geq \frac{k}{n} \sum_{i \in B} \Pr[S \in T_i \text{ and } S \cap B = \{i\} \mid i \in S] \geq \frac{k}{n}|B| \left(c - \frac{2k}{n}|B|\right).$$

If the above is true for B, it is also true for any $B' \subset B$. Now, if $|B| \geq cn/4k$, then consider $B' \subset B$ of size $\lfloor cn/4k \rfloor \geq cn/8k$. Then applying the above inequality for B', we have $\Pr[S \in T] \geq \frac{c}{8} \cdot \frac{c}{2} = \frac{c^2}{16}$. Other wise $|B| < cn/4k$, now again appealing to the above inequality, we have $\Pr[S \in T] \geq \frac{ck}{2} \cdot \Pr[i \in B]$. $\square$

4.2 Step 2: Constructing global functions for typical T's

We prove the following lemma in this section.

Lemma 4.4. *For all $\alpha \in (0,1)$ and positive integers n, k, t satisfying $n \geq 8k$ and $t \geq \alpha k$ and alphabet Σ the following holds: Let $\{f_S : S \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying*

$$\Pr_{\substack{S_1, S_2 \in \binom{[n]}{k} \\ |S_1 \cap S_2| = t}} [f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \varepsilon,$$

then there exists an ensemble $\{g_T : [n] \to \Sigma \mid T \in \binom{[n]}{t-1}\}$ of global functions such when a random $T \in \binom{[n]}{t-1}$ and $S \in \binom{[n]}{k}$ are chosen such that $S \supset T$, then $\Pr[g_T|_S \neq f_S] = O_\alpha(\varepsilon)$.

By Lemma 4.2, we know that a typical T of size $t-1$ satisfies $\varepsilon_T(\emptyset) = O(1)$. We prove the above lemma, by constructing for each such typical T a global function g_T that explains most local functions f_S for $S \supset T$. For the rest of this section fix such a T.

Given $X = \binom{[n]}{k}$, let $X_T := \{S \in X \mid S \supset T\}$. Let $n' = n - (t-1)$ and $k' = k - (t-1)$. For $i \notin T$, let $X_{T,i} := X_{T+i} = \{S \in X_T \mid i \in S\}$.

We now define the "global" function $g_T : [n] \to \Sigma$ as follows. We first define the value of g_T (we will drop the subscript T when T is clear from context) for $i \in T$ and then for each $i \notin T$. Define $g|_T : T \to \Sigma$ to be the most popular restriction of the functions $f_S|_T$ for $S \in X_T$. In other words, $g|_T$ is the function that maximizes $\Pr_{S \in X_T}[g|_T = f_S|_T]$. Let $X^{(0)} := \{S \in X_T \mid f_S|_T = g|_T\}$ be the set of S's that agree with this most popular value. For each $i \notin T$, let $X^{(0)}_{T,i} := X^{(0)} \cap X_{T,i}$. For each such i, define $g(i)$ to be the most popular value $f_S(i)$ among $S \in X^{(0)}_{T,i}$. This completes the definition of the function g.

We now show that if $\varepsilon_T(\emptyset)$ is small, then the function g_T agrees with most functions $f_S, S \in X_T$.

$$\begin{aligned}
\Pr_{S \in X_T}[f_S \neq g|_S] &\leq \Pr_{S \in X_T}[f_S|_T \neq g|_T] + \sum_{i \notin T} \Pr_{S \in X_T}[i \in S \text{ and } f_S|_T = g|_T \text{ and } f_S(i) \neq g(i)] \\
&= \Pr_{S \in X_T}[f_S|_T \neq g|_T] + \frac{k'}{n'} \sum_{i \notin T} \Pr_{S \in X_{T,i}}[f_S|_T = g|_T \text{ and } f_S(i) \neq g(i)] \\
&= \Pr_{S \in X_T}[f_S|_T \neq g|_T] + \frac{k'}{n'} \sum_{i \notin T} \Pr_{S \in X_{T,i}}[S \in X^{(0)}_{T,i}] \cdot \Pr_{S \in X^{(0)}_{T,i}}[f_S(i) \neq g(i)] \qquad (1)
\end{aligned}$$

This motivates the definition of the following quantities which we need to bound.

$$\gamma(\emptyset) := \Pr_{S \in X_T}[f_S|_T \neq g|_T]; \qquad \gamma(i) := \Pr_{S \in X^{(0)}_{T,i}}[f_S(i) \neq g(i)]; \qquad \rho(i) := \Pr_{S \in X_{T,i}}[S \in X^{(0)}_{T,i}].$$

We now bound $\gamma(\emptyset)$ and $\gamma(i)$ in terms $\varepsilon_T(\emptyset)$ and $\mathbb{E}_{i \notin T}[\varepsilon_T(i)]$ via the following (disagreement) probabilities.

$$\kappa(\emptyset) := \Pr_{S_1, S_2 \in X_T}[f_{S_1}|_T \neq f_{S_2}|_T]; \qquad \kappa(i) := \Pr_{S_1, S_2 \in X^{(0)}_{T,i}}[f_{S_1}(i) \neq f_{S_2}(i)].$$

Claim 4.5 (Bounding $\gamma(\emptyset)$). $\gamma(\emptyset) \leq \kappa(\emptyset) \leq 2\varepsilon_T(\emptyset)$.

Proof. By definition, we have $\kappa(\emptyset) = \mathbb{E}_{S_1 \in X_T}\left[\Pr_{S_2 \in X_T}[f_{S_T}|_T \neq f_{S_2}|_T]\right] \geq \gamma(\emptyset)$ since $g|_T$ is the most popular value among $f_S|_T$ for $S \in X_T$. The only difference between $\kappa(\emptyset)$ and $\varepsilon_T(\emptyset)$ is the distribution from which the pairs (S_1, S_2) are drawn; for $\kappa(\emptyset)$, (S_1, S_2) is drawn uniformly from all pairs $X_T \times X_T$ while for $\varepsilon_T(\emptyset)$, (S_1, S_2) is drawn from $\nu_n(k,t)$. To complete the argument, we choose $S_1, S_2, S \in X_T$ in the following coupled fashion such that $(S_1, S_2) \sim X_T^2$ while $(S_1, S), (S_2, S) \sim \nu_n(k,t)$. First choose $S_1, S_2 \in X_T$ at random, then choose $i_1 \in S_1 \setminus T$ and $i_2 \in S_2 \setminus T$ at random, and choose $S \in X_T$ at random such that $S_1 \cap S = T + i_1$ and $S_2 \cap S = T + i_2$. We now have $(S_1, S), (S_2, S) \sim \nu_n(k,t)$. Clearly, if $f_{S_1}|_T \neq f_{S_2}|_T$, then either $f_{S_1}|_T \neq f_S|_T$ or $f_{S_2}|_T \neq f_S|_T$. Hence, $\kappa(\emptyset) \leq 2\varepsilon_T(\emptyset)$. ☐

Claim 4.6 (Bounding $\gamma(i)$)**.** *If $3k - 2t \le n$, then $\gamma(i) \le \kappa(i) \le 2\varepsilon_T(i)/\rho(i)^3$.*

Proof. The proof of this claim proceeds similar to the proof of the previous claim. By definition, we have $\kappa(i) = \mathbb{E}_{S_1 \in X_{T,i}^{(0)}}\left[\Pr_{S_2 \in X_{T,i}^{(0)}}[f_{S_T}(i) \neq f_{S_2}(i)]\right] \geq \gamma(i)$ since $g(i)$ is the most popular value among $f_S(i)$ for $S \in X_{T,i}^{(0)}$. We then observe that

$$\kappa(i) = \Pr_{S_1,S_2 \in X_{T,i}}[f_{S_1}(i) \neq f_{S_2}(i) \mid S_1, S_2 \in X_{T,i}^{(0)}] = \frac{\Pr_{S_1,S_2 \in X_{T,i}}[S_1, S_2 \in X_{T,i}^{(0)} \text{ and } f_{S_1}(i) \neq f_{S_2}(i)]}{\rho(i)^2}$$

We now choose S_1, S_2, S in a coupled fashion as follows. Let $\mathbf{B}$ be the distribution of $|S_1 \cap S_2|$ when S_1, S_2 are chosen at random from $X_{T,i}$. First choose $S \in X_{T,i}^{(0)}$ at random. Then choose $B \sim \mathbf{B}$, so $B \geq t$. Choose disjoint sets I, I_1, I_2 disjoint from S of sizes $B - t, k - B, k - B$ respectively, and let $S_j = I_j \cup I \cup T \cup \{i\}$ for $j \in \{1, 2\}$. Here, we have used the fact that $3k - B - t \leq n$. The joint distribution (S_1, S_2, S) satisfy that $(S_1, S_2) \sim X_{T,i} \times X_{T,i}$ and $(S_j, S) \sim \nu_n(k, t)$ conditioned on $S_j \in X_{T,i}$ and $S \in X_{T,i}^{(0)}$. Furthermore, if $S_1, S_2 \in X_{T,i}^{(0)}$ (i.e., $f_{S_1}|_T = f_{S_2}|_T = g|_T$) and $f_{S_1}(i) \neq f_{S_2}(i)$ then one of the following must hold:

1. $f_{S_1}|_T = f_S|_T$ and $f_{S_1}(i) \neq f_S(i)$, or
2. $f_{S_2}|_T = f_S|_T$ and $f_{S_2}(i) \neq f_S(i)$.

(The first parts always hold, and the second parts cannot both not hold.) This shows that $\kappa(i)$ is bounded above by

$$\begin{aligned}\kappa(i) &\leq \frac{2}{\rho(i)^2} \cdot \Pr_{\substack{S_1 \in X_{T,i} \\ S \in X_{T,i}^{(0)} \\ S_1 \cap S = T \cup \{i\}}}[f_{S_1}|_T = f_S|_T \text{ and } f_{S_1}(i) \neq f_S(i)] \\ &\leq \frac{2}{\rho(i)^3} \cdot \Pr_{\substack{S_1, S \in X_{T,i} \\ S_1 \cap S = T \cup \{i\}}}[f_{S_1}|_T = f_S|_T \text{ and } f_{S_1}(i) \neq f_S(i)] = \frac{2\varepsilon_T(i)}{\rho(i)^3}.\end{aligned}$$
□

Claim 4.7. *If $8k \leq n$ and $\varepsilon_T(\emptyset) \leq \frac{1}{128}$, then $\Pr_{i \notin T}\left[\rho(i) \leq \frac{1}{2}\right] \leq O(\varepsilon_T(\emptyset)/k')$.*

Proof. This follows from an application of Lemma 4.3 by setting $c = \frac{1}{2}$ and $B := \{i \notin T \mid \rho(i) \leq \frac{1}{2}\}$. Then, either $\gamma(\emptyset) \geq 1/64$ or $\Pr[i \in B] \leq 4\gamma(\emptyset)/k' \leq 8\varepsilon_T(\emptyset)/k'$. □

We now return to bounding $\Pr[f_S \neq g|_{S \cup T}]$ from (1) as follows:

Claim 4.8. *If $n \geq 8k$ and $\varepsilon_T(\emptyset) \leq \frac{1}{128}$, then $\Pr_{S,T:\, S \supset T}[f_S \neq g_T|_S] = O(\varepsilon_T(\emptyset) + k' \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)])$.*

Proof.

$$\begin{aligned}\Pr[f_S \neq g_T|_S] &\leq \Pr_{S \in X_T}[f_S|_T \neq g|_T] + \frac{k'}{n'} \cdot \sum_{i \notin T} \Pr_{S \in X_{T,i}}[S \in X_{T,i}^{(0)}] \cdot \Pr_{S \in X_{T,i}^{(0)}}[f_S(i) \neq g(i)] \\ &= \gamma(\emptyset) + \left(\frac{k'}{n'} \cdot \sum_{i \notin T, \rho(i) \leq 1/2} 1\right) + \left(\frac{k'}{n'} \cdot \sum_{i \notin T, \rho(i) > 1/2} \rho(i) \cdot \gamma(i)\right) \\ &\leq 2\varepsilon_T(\emptyset) + 8\varepsilon_T(\emptyset) + \left(\frac{k'}{n'} \cdot \sum_{i \notin T, \rho(i) > 1/2} \frac{2\varepsilon_T(i)}{\rho(i)^2}\right) = O\left(\varepsilon_T(\emptyset) + k' \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)]\right).\end{aligned}$$
□

We now complete the proof of the main lemma of this section.

Proof of Lemma 4.4. By Lemma 4.2, we have $\mathbb{E}_T[\varepsilon_T(\varnothing)] \le \varepsilon$. Hence, $\Pr_T[\varepsilon_T(\varnothing) \le \frac{1}{128}] = 1 - O(\varepsilon)$. We call such a T typical. For non-typical T, we define g_T arbitrarily (this happens with probability at most $O(\varepsilon)$). For every typical T, we have from the global function g_T satisfies

$$\Pr_{S \in X_T}[f_S \neq g_T|_S] = O\left(\varepsilon_T(\varnothing) + (k-(t-1)) \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)]\right).$$

If $t \ge k\alpha$, the right hand side of the above inequality can be further bounded (using Lemma 4.2) as $O\left(\varepsilon_T(\varnothing) + (k-(t-1)) \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)]\right) = O(\varepsilon + k \cdot \varepsilon/t) = O_\alpha(\varepsilon)$. This completes the proof of Lemma 4.4. □

4.3 Step 3: Obtaining a single global function

In the final step, we show that the global function g_T corresponding to a random typical T explains most local functions f_S corresponding to S's not necessarily containing T. We will first prove this under the assumption that $k - 2(t-1) = \Omega(k)$. For concreteness, let us assume $t \le k/3$. We will then show how to extend it to any t satisfying $k - t \ge \beta k$.

Suppose we choose two $(t-1)$-sets T_1, T_2 at random, and a k-set S containing $T_1 \cup T_2$ at random (here we use $2(t-1) \le k$). Then,

$$\Pr[g_{T_1}|_S \neq g_{T_2}|_S] = O(\varepsilon).$$

This prompts defining

$$\delta_{T_1,T_2} := \Pr_{S \supseteq T_1 \cup T_2}[g_{T_1}|_S \neq g_{T_2}|_S],$$

so that $\mathbb{E}[\delta_{T_1,T_2}] = O(\varepsilon)$.

If g_{T_1}, g_{T_2} disagree on $T_1 \cup T_2$ then $\delta_{T_1,T_2} = 1$, which happens with probability at most $O(\varepsilon)$. Assume this is not the case. Denote by B the set of points of $\overline{T_1 \cup T_2}$ on which g_{T_1}, g_{T_2} disagree, and let $n' = n - |T_1 \cup T_2| = \Theta(n)$, $k' = k - |T_1 \cup T_2| = \Theta(k)$. Applying Lemma 4.3 (with $c = 1$) shows that unless $\delta_{T_1,T_2} > 1/8$ (which happens with probability at most $O(\varepsilon)$), we have $|B|/n' = O(\delta_{T_1,T_2}/k')$, and so $|B|/n = O(\delta_{T_1,T_2}/k)$. This shows that if $\delta_{T_1,T_2} \le 1/8$ then

$$\Pr_{i \in [n]}[g_{T_1}(i) \neq g_{T_2}(i) \mid \delta_{T_1,T_2} \le 1/8] \le O(\delta_{T_1,T_2}/k).$$

Choose a random $S \in \binom{[n]}{k}$ containing a random T_2 (but not necessarily T_1). Then

$$\begin{aligned}
\mathbb{E}_{T_1}\left[\Pr_{T_2, S\colon S \supset T_2}[g_{T_1}|_S \neq g_{T_2}|_S]\right] &= \Pr_{T_1,T_2,S\colon S \supset T_2}[g_{T_1}|_S \neq g_{T_2}|_S] \\
&\le \Pr[\delta_{T_1,T_2} > 1/8] + \Pr[\exists i, i \in S \text{ and } g_{T_1}(i) \neq g_{T_2}(i) \mid \delta_{T_1,T_2} \le 1/8] \\
&= O(\varepsilon) + n \cdot \frac{(k-(t-1))}{(n-(t-1))} \cdot \frac{O(\mathbb{E}[\delta_{T_1,T_2} \mid \delta_{T_1,T_2} \le 1/8])}{k} \\
&= O\left(\varepsilon + \frac{\mathbb{E}[\delta_{T_1,T_2}]}{\Pr[\delta_{T_1,T_2} \le 1/8]}\right) = O(\varepsilon).
\end{aligned}$$

Choose a set T_1 such that the above event holds (i.e., $\Pr_{T_2,S\colon S \supset T_2}[g_{T_1}|_S \neq g_{T_2}|_S] = O(\varepsilon)$), and define $F = g_{T_1}$. Then

$$\Pr_S[f_S \neq F|_S] \le \Pr_{S,T_2\colon S \supset T_2}[f_S \neq g_{T_2}|_S] + \Pr_{S,T_2\colon S \supset T_2}[g_{T_1}|_S \neq g_{T_2}|_S] = O(\varepsilon).$$

We have proved the following lemma.

Lemma 4.9. *For all $\alpha \in (0, 1/3)$ if $n \ge 4k$ and $\alpha k \le t \le k/3$, there exists a function $F\colon [n] \to \Sigma$ such that $\Pr[f_S \neq F|_S] = O_\alpha(\varepsilon)$.*

Proof of Theorem 4.1. Consider the following coupling argument. Let $S_1, S_2 \sim \nu_n(k, t')$. Let S be a random set of size k containing $S_1 \cap S_2$ as well as $t - t'$ random elements from S_1, S_2 each and the rest of the elements chosen from $\overline{S_1 \cup S_2}$. This can be done as long as $k \geq 2(t - t') + t' = 2t - t'$. Clearly, $(S, S_j) \sim \nu_n(k, t)$ for $j = 1, 2$. Furthermore,

$$\Pr[f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \Pr[f_{S_1}|_{S_1 \cap S} \neq f_S|_{S_1 \cap S}] + \Pr[f_{S_2}|_{S_2 \cap S} \neq f_S|_{S_2 \cap S}] \leq 2\varepsilon.$$

This demonstrates that if the hypothesis for the agreement theorem is true for a particular choice of n, k, t, then the hypothesis is also true for n, k, t' by increasing ε to 2ε provided $k - t \geq (k - t')/2$. Thus, given the hypothesis is true for some t satisfying $k - t \geq \beta k$, we can perform the above coupling argument a constant number of times to to reduce t to less than $k/3$ and then conclude using Lemma 4.9. $\square$

5 Agreement theorem for high dimensions

Theorem 5.1 (Agreement theorem). *For all positive integers d there exists a constant $C > 1$ such that for all $\alpha, \beta \in (0, 1)$ satisfying $\alpha + \beta \leq 1$, for all positive integers n, k, t satisfying $n \geq Ck$, $t \geq \max\{\alpha k, d\}$ and $k - t \geq \max\{\beta k, d\}$, and for all alphabets Σ, the following holds: Let $\{f_S \colon \binom{S}{\leq d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of functions satisfying*

$$\Pr_{\substack{S_1, S_2 \in \binom{[n]}{k} \\ |S_1 \cap S_2| = t}} [f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \varepsilon,$$

then there exists a function $F \colon \binom{[n]}{\leq d} \to \Sigma$ satisfying $\Pr_{S \in \binom{[n]}{k}}[f_S \neq F|_S] = O_{\alpha,\beta,d}(\varepsilon)$. Here, $F|_S$ refers to the restriction $F|_{\binom{S}{\leq d}}$.

As before, we let $\nu_n(k, t)$ denote the distribution induced on the pair of sets $(S_1, S_2) \in \binom{[n]}{k}^2$ by first choosing uniformly at random a set $U \subset [n]$ of size t and then two sets S_1 and S_2 of size k of $[n]$ uniformly at random conditioned on $S_1 \cap S_2 = U$. The proof of this theorem proceeds similar to the dimension one setting in three steps. In the first step (Section 5.1), we prove some preliminary lemmata which help in bounding the error of a "typical" subset T of $[n]$ of size $t - d$. In the second step (Section 5.2), we define for each $T \subset [n]$ of size $t - d$, a "global" function $g_T \colon \binom{[n]}{\leq d} \to \Sigma$ such that when we pick a random pair $T \subset S$ where $|T| = t - d$ and $|S| = k$, then $\Pr_{T,S\colon T \subset S}[g_T|_S = f_S] = O(\varepsilon)$. In other words, for a random $T \subset S$, the global function explains the local function. Finally, in step (Section 5.3), we argue that a random "global" function g_T explains most "local" functions f_S corresponding to S (not necessarily ones that contain T).

First for some notation. Let $n' := n - (t - d)$ and $k' := k - (t - d)$. For any set $T \subset [n]$ of size $t - d$, we let $\overline{T} := [n] \setminus T$. Let $X_T := \{S \in \binom{[n]}{k} \mid S \supset T\}$. For $A \subset \overline{T}$, $|A| = i \leq d$, we define $X_{T,A} := X_{T \cup A} = \{S \in \binom{[n]}{k} \mid S \supset T \cup A\}$.

For $i = -1, 0, \ldots, d$, Define $T^{(i)} := \{U \in \binom{[n]}{\leq d} \mid |U \setminus T| \leq i\}$. Clearly, $\emptyset = T^{(-1)} \subset \binom{T}{\leq d} = T^{(0)} \subset T^{(1)} \subset \ldots \subset T^{(d-1)} \subset T^{(d)} = \binom{[n]}{\leq d}$. For $A \subset \overline{T}$ and $|A| = i$, define $T^{(A)} := \{U \in \binom{[n]}{\leq d} \mid U \setminus T \subset A\} = \binom{T \cup A}{\leq d}$. Clearly, $T^{(i)} = \bigcup_{A \in \binom{\overline{T}}{i}} T^{(A)}$. For $S \in X_{(A)}$, let $f_S|_{T,A}$ denote the restriction $f_S|_{T^{(A)} \cap \binom{S}{\leq d}}$. Similarly, $f_S|_{T,i} := f_S|_{T^{(i)} \cap \binom{S}{\leq d}}$. Note that $f_S|_{T,i}$ refers to the restriction of f_S to the set of all subsets of size at most d which have at most i elements outside T. Given two local functions f_{S_1} and f_{S_2}, we say that they agree (denoted by $f_{S_1} \sim f_{S_2}$) if they agree on the intersection of their domains (ie., $f_{S_1}(a) = f_{S_2}(a)$ for all $a \in \binom{S_1 \cap S_2}{\leq d}$). Similarly, we say that two restrictions $f_{S_1}|_{T,i}$ and $f_{S_2}|_{T,i}$ agree (denoted by $f_{S_1}|_{T,i} \sim f_{S_2}|_{T,i}$) if $f_{S_1}(a) = f_{S_2}(a)$ for all $a \in \binom{S_1 \cap S_2}{\leq d} \cap T^{(i)}$.

5.1 Step 1: some preliminary lemmata

Lemma 5.2. *For all $0 \le i \le d$,*

$$\Pr_{\substack{S_1,S_2\sim\nu_n(k,t)\\ T\subseteq S_1\cap S_2,|T|=t-d}}\left[f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1} \not\sim f_{S_2}\right] = O_{d,\alpha}(k^{-i}\varepsilon).$$

Proof. We can rewrite the above probability as

$$\Pr_{S_1,S_2\sim\nu_n(k,t)}[f_{S_1}\not\sim f_{S_2}]\cdot \mathbb{E}_{\substack{S_1,S_2\sim\nu_n(k,t)\\ f_{S_1}\not\sim f_{S_2}}}\left[\Pr_{T\subseteq S_1\cap S_2,|T|=t-d}[f_{S_1}|_{T,i-1}\sim f_{S_2}|_{T,i-1}]\right].$$

The first factor is clearly at most ε. Now consider any S_1, S_2 of size k intersecting at a set of size t such that $f_{S_1} \not\sim f_{S_2}$, say $f_{S_1}(A) \neq f_{S_2}(A)$ for some $A \subseteq S_1 \cap S_2$. Hence, if f_{S_1} and f_{S_2} agree on all sets in $T^{(i-1)} \cap \binom{S_1\cap S_2}{\le d}$, it must be the case that $|A \setminus T| \ge i$. Hence,

$$\Pr_{T\subseteq S_1\cap S_2,|T|=t-d}[f_{S_1}|_{T,i-1}\sim f_{S_2}|_{T,i-1}] \le \Pr_{T\subseteq S_1\cap S_2,|T|=t-d}[|A\setminus T|\ge i].$$

Let $U = S_1 \cap S_2$. We can estimate the probability on the right by

$$\Pr_{T\subseteq U,|T|=t-d}[|A\setminus T|\ge i] \le \sum_{B\subseteq A,|B|=i}\Pr_{T\subseteq U,|T|=t-d}[U\setminus T\supseteq B] = \binom{d}{i}\frac{d(d-1)\cdots(d-i+1)}{t(t-1)\cdots(t-i+1)} = O_d(t^{-i}) = O_{d,\alpha}(k^{-i}),$$

where in the last step we have used the fact $t \ge \alpha k$. □

We deduce the following corollaries.

Corollary 5.3. *Let $|T| = t-d$ and $|A| = i \le d$ be disjoint sets. Define*

$$\varepsilon_{T,A} := \Pr_{\substack{S_1,S_2\sim\nu(k,t)\\ S_1\cap S_2\supseteq T\cup A}}[f_{S_1}|_{T,i-1}\sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1}|_{T,A}\not\sim f_{S_2}|_{T,A}].$$

Then $\mathbb{E}_{T,A}[\varepsilon_{T,A}] = O(k^{-i}\varepsilon)$ where the expectation it taken over T and A such that $|T| = t-d, |A| = i$ and $T\cap A = \emptyset$.

Proof. This follows from the simple observation that

$$\begin{aligned}
\mathbb{E}_{T,A}[\varepsilon_{T,A}] &= \mathbb{E}_{T,A}\left[\Pr_{\substack{S_1,S_2\sim\nu(k,t)\\ S_1\cap S_2\supseteq T\cup A}}[f_{S_1}|_{T,i-1}\sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1}|_{T,A}\not\sim f_{S_2}|_{T,A}]\right]\\
&\le \mathbb{E}_{T,A}\left[\Pr_{\substack{S_1,S_2\sim\nu(k,t)\\ S_1\cap S_2\supseteq T\cup A}}[f_{S_1}|_{T,i-1}\sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1}\not\sim f_{S_2}]\right]\\
&= \Pr_{\substack{S_1,S_2\sim\nu_n(k,t)\\ T\subseteq S_1\cap S_2,|T|=t-d}}[f_{S_1}|_{T,i-1}\sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1}\not\sim f_{S_2}]\\
&= O(k^{-i}\varepsilon).
\end{aligned}$$

□

Corollary 5.4. *Let $|T| = k-d$ and let $0 \le i \le d$. Define $\varepsilon_{T,i} := \mathbb{E}_{A\subset\overline{T},|A|=i}[\varepsilon_{T,A}]$. Then $\mathbb{E}_T[\varepsilon_{T,i}] = O(k^{-i}\varepsilon)$.*

We also need the following lemma (which in some sense is the generalization of Lemma 4.3 to general d). However the proof of this lemma is far more elaborate and requires the hypergraph pruning lemma (Lemma 3.5 proved in Section 6).

Lemma 5.5. *Fix $d \geq 1$ and $c > 0$. There exists $p_0 > 0$ (depending on c, d) such that the following holds for every $n \geq k \geq 2d$ satisfying $k/n \leq p_0$.*

Let F be a d-uniform hypergraph, and for each $A \in F$, let $Y_A \subseteq X_A = \{S \in \binom{[n]}{k} \mid S \supseteq A\}$ have density at least c in X_A. Then

$$\Pr_{S\colon |S|=k}\left[S \in \bigcup_{A \in F} X_A\right] = O_{c,d}\left(\Pr_{S\colon |S|=k}\left[S \in \bigcup_{A \in F} Y_A\right]\right).$$

Proof. Let $\varepsilon = c/2$, and apply the uniform hypergraph pruning lemma (Lemma 3.5) setting $H := F$ to get a subhypergraph F' of F. For every $A \in F'$,

$$\Pr_{S\colon |S|=k}[S \in Y_A \text{ and } F'|_S = \{A\} \mid S \in X_A] \geq c - \Pr_{S\colon |S|=k}[F'|_S \neq \{A\} \mid S \in X_A] \geq c - \varepsilon = c/2.$$

Summing over all $A \in F'$, we get

$$\Pr_{S\colon |S|=k}\left[S \in \bigcup_{A \in F} Y_A\right] \geq \sum_{A \in F'} \Pr_{S\colon |S|=k}[S \in Y_A \text{ and } F'|_S = \{A\}] \geq$$
$$\frac{c}{2} \sum_{A \in F'} \Pr_{S\colon |S|=k}[S \in X_A] \geq \frac{c}{2} \Pr_{S\colon |S|=k}[F'|_S \neq \emptyset] = \Omega_{c,d}\left(\Pr_{S\colon |S|=k}[F|_S \neq \emptyset]\right).$$

This completes the proof since the right-hand side is exactly the left-hand side of the statement of the lemma. ☐

5.2 Step 2: Constructing a global function for a typical T

We prove the following lemma in this section.

Lemma 5.6. *For all $\alpha, \beta \in (0,1)$ and positive integers d, there exists a constant $p \in (0,1)$ such that for all positive integers n, k, t satisfying $k \leq pn$, $t \geq \max\{\alpha k, d\}$, $k - t \geq \max\{\beta k, d\}$ and alphabet Σ the following holds: Let $\{f_S\colon \binom{S}{\leq d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying*

$$\Pr_{\substack{S_1, S_2 \in \binom{[n]}{k} \\ |S_1 \cap S_2| = t}} [f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \varepsilon,$$

then there exists an ensemble $\{g_T\colon \binom{[n]}{\leq d} \to \Sigma \mid T \in \binom{[n]}{t-d}\}$ of global functions such when a random $T \in \binom{[n]}{t-d}$ and $S \in \binom{[n]}{k}$ are chosen such that $S \supset T$, then $\Pr[g_T|_S \neq f_S] = O_{\alpha,\beta,d}(\varepsilon)$.

We now define the "global" function $g_T\colon \binom{[n]}{\leq d} \to \Sigma$. We will drop the subscript T for ease of notation. We will define g incrementally by first defining $g|_{T^{(-1)}}$ (the empty function) and then inductively extending the definition of g from the domain $T^{(i-1)}$ to $T^{(i)}$ (recall that $T^{(-1)} \subset T^{(0)} \subset \cdots \subset T^{(d)} = \binom{[n]}{\leq d}$). To begin with set $X^{(-1)} := X_T$ and $\delta_{-1} := 1 - \frac{|X^{(-1)}|}{|X_T|} = 0$. Let $g\colon T^{(-1)} \to \Sigma$ be the empty function. For $i := 0 \ldots d$ do, we inductively extend the definition of g from $T^{(i-1)}$ to $T^{(i)}$ as follows. If $\delta_{i-1} > \frac{1}{2}$, set $g := \perp$ and exit. For each $A \in \overline{T}, |A| = i$, let

$$X_{(A)}^{(i-1)} := \{S \in X^{(i-1)} \mid S \supset A\},$$

and g_A be the most popular $f_S|_{T,A}$ among $S \in X_{(A)}^{(i-1)}$ (breaking ties arbitrarily). Let $\gamma(A)$ denote the probability that a random value in $X_{(A)}^{(i-1)}$ is not the popular value, more precisely

$$\gamma(A) := \Pr_{S \in X_{(A)}^{(i-1)}} [f_S|_{T,A} \neq g_A],$$

and $\rho(A) := \frac{|X^{(i-1)}_{(A)}|}{|X_{(A)}|}$. Note that $g_A \colon \binom{T \cup A}{\leq d} \to \Sigma$ and g_A agrees with g on the domain $T^{(i-1)}$ (ie., the domain where it has been defined so far). We now extend g from $T^{(i-1)}$ to $T^{(i)}$ as follows: for each $B \in T^{(i)} \setminus T^{(i-1)}$, let A be the unique subset in $\binom{\overline{T}}{i}$ such that $B = B' \cup A$ for some $B' \in T$. Set $g(B) := g_A(B)$. Set

$$X^{(i)} := \left\{ S \in X^{(i-1)} \mid \forall A \subset S \setminus T, |A| = i, f_S|_{T,A} = g|_{T^{(A)} \cap \binom{S}{\leq d}} \right\},$$

and $\delta_i := 1 - \frac{|X^{(i)}|}{|X_T|}$ before proceeding to the next i. Thus, $X^{(i)}$ refers to the set of S's where the global function $g \colon T^{(i)} \to \Sigma$ agrees with local functions f_S and δ_i is the density of those S's that disagree with the global function.

We would like to bound the probability that the global function g defined above agrees with local functions, namely $\Pr_{S \colon S \supset T}[g_T|_S \neq f_S]$. Note that this probability is upper bounded by the probability δ_d. We now inductively bound $\delta_i, i = 0, \ldots, d$. First we need the following claims on $\gamma(A)$ and $\rho(A)$.

Claim 5.7 (Estimating $\gamma(A)$)**.** *If $t + d \leq k$ and $3k \leq n$, then $\gamma(A) \leq 2\varepsilon_{T,A}/\rho(A)^3$.*

Proof. By definition, we have $\gamma(A) = \min_\alpha \Pr_{S \in X^{(i-1)}_{(A)}}[f_S|_{T,A} \neq \alpha]$. Hence, we have

$$\gamma(A) \leq \Pr_{S_1, S_2 \in X^{(i-1)}_{(A)}}\left[f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A}\right] \leq \frac{1}{\rho(A)^2} \cdot \Pr_{S_1, S_2 \in X_{(A)}}\left[S_1, S_2 \in X^{(i-1)}_{(A)} \text{ and } f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A}\right].$$

Let $\mathbf{M}$ be the distribution of $|S_1 \cap S_2|$ when S_1, S_2 are chosen at random from $X_{(A)}$. Choose $S \in X^{(i-1)}_{(A)}$ at random, and draw $m \sim \mathbf{M}$ (so $m \geq t - d + i$). Choose two disjoint subsets R_1, R_2 of $S \setminus (T \cup A)$ of size $d - i$, two disjoint subsets I_1, I_2 of $\overline{S}$ of size $k - m - d + i$, and a subset I disjoint from I_1, I_2, S of size $m - i - t + d$; this is possible since $t + d \leq k$ and $3k \leq n$. Let $S_j = A \cup R_j \cup I_j \cup I \cup T$ (which have size $i + (d - i) + (k - m - d + i) + (m - i - t + d) + (t - d) = k$, so that $S_1 \cap S_2 = A \cup I \cup T$ has size $i + (m - i - t + d) + (t - d) = m$ and $S_j \cap S = A \cup R_j \cup T$ have size $i + (d - i) + (t - d) = t$. The joint distribution (S_1, S_2, S) satisfy that $(S_1, S_2) \sim X_{(A)} \times X_{(A)}$ and $(S_j, S) \sim \nu_n(k, t)$ conditioned on $S_j \in X_{(A)}$ and $S \in X^{(i-1)}_{(A)}$. Furthermore, if $f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A}$ and $S_1, S_2 \in X^{(i-1)}_{(A)}$ (i.e., for all $A_1 \in S_1 \setminus T$ of size i, $f_{S_1}|_{T,A_1} = g|_{T^{(A_1)} \cap \binom{S}{\leq d}}$ and for all $A_2 \in S_2 \setminus T$ of size i, $f_{S_2}|_{T,A} = g|_{T^{(A)} \cap \binom{S}{\leq d}}$), then one of the following must hold:

1. $f_{S_1}|_{T,i} \sim f_S|_{T,i}$ and $f_{S_1}|_{T,A} \not\sim f_S|_{T,A}$, or
2. $f_{S_2}|_{T,i} \sim f_S|_{T,i}$ and $f_{S_2}|_{T,A} \not\sim f_S|_{T,A}$.

Hence,

$$\gamma(A) \leq \frac{2}{\rho(A)^2} \cdot \Pr_{\substack{S_1 \in X_{(A)} \\ S \in X^{(i-1)}_{(A)} \\ |S_1 \cap S| = t}}[f_{S_1}|_{T,i} \sim f_S|_{T,i} \text{ and } f_{S_1}|_{T,A} \not\sim f_S|_{T,A}] \leq$$

$$\frac{2}{\rho(A)^3} \cdot \Pr_{\substack{S_1, S \in X_{(A)} \\ |S_1 \cap S| = t}}[f_{S_1}|_{T,i} \sim f_S|_{T,i} \text{ and } f_{S_1}|_{T,A} \not\sim f_S|_{T,A}] \leq \frac{2\varepsilon_{T,A}}{\rho(A)^3}. \quad \square$$

Claim 5.8 (Estimating $\rho(A)$)**.** *If $k \geq t + d$ and $k \leq p_0 n$, then $\Pr_{S \in X_T}\left[\exists A \subset \overline{T}, |A| = i, S \supset A, \rho(A) < \frac{1}{2}\right] = O(\delta_{i-1})$.*

Proof. Let $F = \{|A| = i \mid \rho(A) \le 1/2\}$. Define $Y_{(A)} = \{S \in X_{(A)} \mid S \notin X_{(A)}^{(i-1)}\}$. If $A \in F$ then $|Y_{(A)}|/|X_{(A)}| = 1 - \rho(A) \ge 1/2$. Then applying Lemma 5.5 (setting $d = d, c = 1/2, n = n - (t-d), k = k - (t-d)$, we have

$$\Pr_{S \in X_T}[S \supseteq A \text{ for some } A \in F] = O(\Pr_{S \in X_T}[S \in Y_{(A)} \text{ for some } A \in F]).$$

The conditions for Lemma 5.5 require $k - (t-d) \ge 2d$ and $k - (t-d) \le p_0(n - (t-d))$ which are satisfied if $k \ge t + d$ and $k \le p_o n$. If $S \in Y_{(A)}$ for *any* A then $S \notin X^{(i-1)}$, and so the probability on the right is at most $\Pr_{S \in X_T}[S \notin X^{(i-1)}] = \delta_{i-1}$. Therefore

$$\Pr_{S \in X_T}\left[\rho(A) < 1/2 \text{ for some } A \in \binom{S \setminus T}{i}\right] = O(\delta_{i-1}).$$

□

Claim 5.9. *If $k - t \ge \beta k$ and $\delta_{i-1} \le \frac{1}{2}$, then $\delta_i = O_\beta(\delta_{i-1} + k^i \varepsilon_{T,i})$.*

Proof.

$$\begin{aligned}
\delta_i &= \Pr_{S \in X_T}\left[S \notin X^{(i)}\right] = \Pr_{S \in X_T}\left[S \notin X^{(i-1)}\right] + \Pr_{S \in X_T}\left[S \in X^{(i-1)} \text{ and } S \notin X^{(i)}\right] \\
&= \delta_{i-1} + \Pr_{S \in X_T}\left[\exists A \in \overline{T}, |A| = i, S \supset A \text{ and } S \in X^{(i-1)} \text{ and } f_S|_{T,A} \ne g|_{T^{(A)} \cap \binom{S}{\le d}}\right] \\
&= \delta_{i-1} + \Pr_{S \in X_T}\left[\exists A \in \overline{T}, |A| = i, S \supset A, \rho(A) < \frac{1}{2}\right] \\
&+ \Pr_{S \in X_T}\left[\exists A \in \overline{T}, |A| = i, S \supset A, \rho(A) \ge \frac{1}{2}, S \in X^{(i-1)} \text{ and } f_S|_{T,A} \ne g|_{T^{(A)} \cap \binom{S}{\le d}}\right] \\
&= O\left(\delta_{i-1}\right) + \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \Pr_{S \in X_T}\left[S \supset A, S \in X^{(i-1)} \text{ and } f_S|_{T,A} \ne g|_{T^{(A)} \cap \binom{S}{\le d}}\right] \quad \text{[By Claim 5.8]} \\
&= O\left(\delta_{i-1}\right) + \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \Pr_{S \in X_T}\left[S \in X_{(A)}\right] \cdot \Pr_{S \in X_{(A)}}\left[S \in X_{(A)}^{(i-1)}\right] \cdot \Pr_{S \in X_{(A)}^{(i-1)}}\left[f_S|_{T,A} \ne g|_{T^{(A)} \cap \binom{S}{\le d}}\right] \\
&\le O\left(\delta_{i-1}\right) + \frac{\binom{n'-i}{k'-i}}{\binom{n'}{k'}} \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \rho(A) \cdot \gamma(A) \\
&\le O\left(\delta_{i-1}\right) + \left(\frac{k'}{n'}\right)^i \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \frac{2\varepsilon_{T,A}}{\rho(A)^2} \quad \text{[By Claim 5.7]} \\
&\le O\left(\delta_{i-1}\right) + 8\left(\frac{k'}{n'}\right)^i \sum_{A\colon A \in \binom{\overline{T}}{i}} \varepsilon_{T,A} \\
&= O_\beta\left(\delta_{i-1} + k^i \varepsilon_{T,i}\right) \quad \text{[Since } k' = k - (t-d) = \Theta(k)\text{]}
\end{aligned}$$

□

We are now ready to complete the proof of Lemma 5.6

Proof of Lemma 5.6. Given T, we have shown above how to construct a function g_T, given that $\delta_i \le c_\delta$ for all i. If the latter condition fails, define g_T arbitrarily.

We have defined above a sequence $\delta_{-1} = 0, \delta_0, \ldots, \delta_d$. We have defined δ_i only given $\delta_{i-1} \le \frac{1}{2}$. If $\delta_{i-1} > \frac{1}{2}$, we define $\delta_i = 1$. Note that $\Pr[f_S \ne g_T|_S] \le \delta_d$.

We have shown above that if $\delta_{i-1} \le \frac{1}{2}$ then $\delta_i = O(\delta_{i-1} + k^i \varepsilon_{T,i})$. It is always the case that $\delta_i = O(\delta_{i-1} + k^i \varepsilon_{T,i}) + 1_{\delta_{i-1} > \frac{1}{2}}$. We now prove by induction on i that $\mathbb{E}_T[\delta_i] = O_i(\varepsilon)$. This clearly holds when $i = -1$. Assume that it holds for $i-1$, i.e., $\mathbb{E}_T[\delta_{i-1}] = O_{i-1}(\varepsilon)$. Then, $\Pr[\delta_{i-1} > \frac{1}{2}] = O_{i-1}(\varepsilon)$. Also, by Corollary 5.4, we have $\mathbb{E}_T[\varepsilon_{T,i}] = O(k^{-i}\varepsilon)$. We now have for i,

$$\mathbb{E}[\delta_i] = O(\mathbb{E}[\delta_{i-1}] + k^i \mathbb{E}[\varepsilon_{T,i}]) + \Pr[\delta_{i-1} > \frac{1}{2}] = O_i(\varepsilon).$$

We conclude that $\Pr_{T,S}[g_T|S \neq f_S] \le \mathbb{E}[\delta_d] = O_d(\varepsilon)$. □

5.3 Step 3: Obtaining a single global function

Given the set of local functions $\{f_S\}_{S \in \binom{[n]}{k}}$, we constructed a set of global functions $\{g_T\}_{T \in \binom{[n]}{t-d}}$ such that for most pairs $S \supset T$, the global function g_T agrees with the local function f_S (Lemma 5.6). In this step, we conclude that a random global function g_T agrees with most local functions f_S (not necessarily S's that contain T).

We will first prove this under the assumption that $k - 2(t-1) = \Omega(k)$. For concreteness, let us assume $t \le k/3$. We will then show to extend it to any t satisfying $k - t \ge \beta k$. To begin with, we observe that Lemma 5.6 immediately implies the following claim.

Claim 5.10. *For T_1, T_2 of size $t-d$, define $\delta_{T_1,T_2} := \Pr_{S \supseteq T_1 \cup T_2}[g_{T_1}|_S \neq g_{T_2}|_S]$. Then $\mathbb{E}_{T_1,T_2}[\delta_{T_1,T_2}] = O(\varepsilon)$.*

We now move to more general S in the following sense: S contains T_2 but not necessarily T_1.

Claim 5.11. *For all T_1, T_2, $\Pr_{|S|=k, S \supseteq T_2}[g_{T_1}|_S \neq g_{T_2}|_S] = O(\delta_{T_1,T_2})$.*

Proof. We will prove this be choosing $L = O_d(1)$ collection of k-sets $(S, S_1, \ldots, S_L)$ in a coupled fashion such that each S is a random k-set containing T_1 and for each $j \ge 1$, S_j is a random k-set containing $T_1 \cup T2$ with the additional property that $\binom{S}{\le d} \subseteq \bigcup_{j \ge 1} \binom{S}{\le d}$. Given such a distribution, the lemma follows by a union bound.

The coupled distribution is obtained in the following fashion. Let $k - |T_1 \cup T_2| \ge k/3$. We proceed to find a collection of $O(1)$ subsets $R_i \subseteq [k]$ of size at most $k/3$ such that $\binom{[k]}{d} = \bigcup_i \binom{R_i}{d}$. The idea is to split $[k]$ into $O(d)$ parts of size at most $k/(3d)$, and to take as R_i the union of any d of these. Given a random k-set $S \in \binom{[n]}{d}$ containing T_2, choose a random permutation mapping $[k]$ to S, apply it to the R_i, remove from the resulting sets any elements of $T_1 \cup T_2$, and complete them to sets $\tilde{R}_i$ of size $k - |T_1 \cup T_2|$ randomly and set $S_j = \tilde{R}_j \cup T_1 \cup T_2$. Clearly, if S is a random k-set containing T_2, the sets S_j are individually random sets of size k containing $T_1 \cup T_2$. □

We can now complete the proof of Theorem 5.1

Proof of Theorem 5.1. As in the dimension one setting, we first prove Theorem 5.1 if $\alpha k \le t \le k/3$ and then extend it to any t satisfying $k - t \ge \beta k$. From Claim 5.10 and Claim 5.11, we have that

$$\mathbb{E}_{T_1}\left[\Pr_{T_2, S \colon S \supset T_2}[g_{T_1}|_S \neq g_{T_2}|_S]\right] = O(\delta_{T_1,T_2}) = O(\varepsilon).$$

Choose a T_1 such that the inner probability is $O(\varepsilon)$ and set $F = g_{T_1}$. We now have,

$$\begin{aligned}\Pr_S[f_S \neq F|_S] &= \Pr_{S, T_2 \colon S \supset T_2}[f_S \neq F|_S] \\ &\le \mathbb{E}_{T_2}\left[\Pr_{S \colon S \supset T_2}[F|_S \neq g_{T_2}|_S]\right] + \mathbb{E}_{T_2}\left[\Pr_{S \colon S \supset T_2}[f_S \neq g_{T_2}|_S]\right] = O(\varepsilon).\end{aligned}$$

This completes the proof for $t \le k/3$ (in particular to any t satisfying $k - 2t = \Omega(k)$.

To extend the proof to all t satisfying $k - t = \Omega(k)$, we employ the following coupling argument as in the dimension one setting. Let $S_1, S_2 \sim \nu_n(k, t')$. Let S be a random set of size k containing $S_1 \cap S_2$ as well as $t - t'$ random elements from S_1, S_2 each and the rest of the elements chosen from $\overline{S_1 \cup S_2}$. This can be done as long as $k \geq 2(t - t') + t' = 2t - t'$. Clearly, $(S, S_j) \sim \nu_n(k, t)$ for $j = 1, 2$. Furthermore,

$$\Pr[f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \Pr[f_{S_1}|_{S_1 \cap S} \neq f_S|_{S_1 \cap S}] + \Pr[f_{S_2}|_{S_2 \cap S} \neq f_S|_{S_2 \cap S}] \leq 2\varepsilon.$$

This demonstrates that if the hypothesis for the agreement theorem is true for a particular choice of n, k, t, then the hypothesis is also true for n, k, t' by increasing ε to 2ε provided $k - t \geq (k - t')/2$. Thus, given the hypothesis is true for some t satisfying $k - t \geq \beta k$, we can perform the above coupling argument a constant number of times to to reduce t to less than $k/3$ and then conclude using the above argument for $t \leq k/3$. □

6 Hypergraph Pruning Lemma

We begin with a a few definitions. The number of hyperedges in a hypergraph H is denoted $|H|$. For a vertex set V, μ_p refers to the biased distribution over subsets S of V defined by choosing each $v \in V$ to be in S independently with probability p while $\nu_{n,k}$ refers to the uniform distribution over subsets S of V of size k. For a hypergraph H and a subset S of the vertices, $H|_S$ is the subhypergraph induced by the vertices in S while $H|_{S=\emptyset}$ is obtained by removing all vertices in S from all hyperedges of H. For a hypergraph H, $\iota_p(H) := \Pr_{S \sim \mu_p}[H|_S \neq \emptyset]$. And finally, we recall the definition of branching factor from the introduction. 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.

The main goal of this section is to prove the following two hypergraph pruning lemmata; one under the biased μ_p distribution and the other under the uniform $\nu_{n,k}$ distribution, which was stated in the introduction. These pruning lemmata show that any hypergraph H has a subgraph H' with bounded branching factor with almost the same $\iota_p(H)$.

Lemma 6.1 (hypergraph pruning lemma (biased setting))**.** *Fix constants $c > 0$ and $d \geq 0$. There exists $p_0 > 0$ (depending on c, d) such that for every $p \in (0, p_0)$ and every d-uniform hypergraph H there exists a subhypergraph H' obtained by removing hyperedges such that*

1. *H' has branching factor c/p.*
2. *$\iota_p(H') = \Omega_{c,d}(\iota_p(H))$.*

Lemma 3.5 (Restated) (hypergraph pruning lemma (uniform setting)) *Fix constants $\varepsilon > 0$ and $d \geq 1$. There exists $p_0 > 0$ (depending on d, ε) such that for every $n \geq k \geq 2d$ satisfying $k/n \leq p_0$ and every d-uniform hypergraph H on $[n]$ there exists a subhypergraph H' obtained by removing hyperedges such that*

1. *$\Pr_{S \sim \nu_{n,k}}[H'|_S \neq \emptyset] = \Omega_{d,\varepsilon}(\Pr_{S \sim \nu_{n,k}}[H|_S \neq \emptyset])$.*
2. *For every $e \in H'$, $\Pr_{S \sim \nu_{n,k}}[H'|_S = \{e\} \mid S \supset e] \geq 1 - \varepsilon$.*

Here $H'|_S$ is the hypergraph induced on the vertices of S.

6.1 Proof in the μ_p biased setting

The hypergraph pruning lemma (Lemma 6.1) is proved by induction on d. The proof is divided into several steps, expressed in the following lemmata. We begin with an easy claim.

The first lemma identifies a "critical depth" for H.

Lemma 6.2. *For every integer $d, c > 0$ and $p \in (0,1)$ the following holds. Let H be a d-uniform hypergraph. Then, either H has a subhypergraph H' with branching factor c/p such that $\iota_p(H') \ge \iota_p(H)/(d+1)$, or for some there $1 \le r \le d$, there exists a $(d-r)$-uniform hypergraph I, and a subhypergraph H' of H such that*

1. *Each hyperedge in I has at least $(c/p)^r$ extensions in H'.*
2. *For every $e \in I$ and every $A \ne \emptyset$ disjoint from e, $e \cup A$ has at most $(c/p)^{r-|A|}$ extensions in H'.*
3. *$\iota_p(I) \ge \iota_p(H)/(d+1)$.*

Proof. We define a sequence of graphs H_r, B_r for $0 \le r \le d$ as follows:

- $H_0 = H$ and B_0 is the empty d-uniform hypergraph.
- B_r contains all sets $|A| = d - r$ which have at least $(c/p)^r$ extensions in H_{r-1}.
- H_r contains all hyperedges in H_{r-1} which are not extensions of a set in B_r.

It's not hard to check that $\iota_p(H_r) \le \iota_p(H_{r+1}) + \iota_p(B_{r+1})$, and so

$$\iota_p(H) \le \iota_p(B_1) + \cdots + \iota_p(B_d) + \iota_p(H_d).$$

Hence one of the values on the right-hand side is at least $\iota_p(H)/(d+1)$.

The construction guarantees that for every r, every set A of size at least $d-r$ has at most $(c/p)^{d-|A|}$ extensions in H_r. In particular, H_d has branching factor c/p. This completes the proof when $\iota_p(H_d) \ge \iota_p(H)/(d+1)$. If $\iota_p(B_r) \ge \iota_p(H)/(d+1)$ for some $r \ge 1$ then we take $I = B_r$ and $H' = H_{r-1}$. The first property in the statement of the lemma follows directly from the construction of B_r, and the second follows from the guarantee stated earlier for H_{r-1} applied to $e \cup A$, which has size $d - r + |A|$ which is at least $d - (r-1)$. □

The strategy now is to apply induction on I to reduce its branching factor, and then to "complete" it to a d-uniform hypergraph. The completion is accomplished in two steps. The first step adds all hyperedges which can be associated with more than one hyperedge of the pruned I.

Lemma 6.3. *For every integer $d, c > 0$ and $p \in (0,1)$ the following holds. Let H be a d-uniform hypergraph and I a $(d-r)$-uniform hypergraph for some $1 \le r \le d$ such that*

1. *For every $e \in I$ and every $A \ne \emptyset$ disjoint from e, $e \cup A$ has at most $(c/p)^{r-|A|}$ extensions in H.*
2. *I has branching factor c/p.*

Then the subhypergraph K of H consisting of all hyperedges of H which extend at least two hyperedges of I has branching factor $O_d(c/p)$.

Proof. Fix a set B of size $d-s$, where $s \ge 1$. We have to bound the number of extensions of B in K. Each of these extensions belongs to one of the following types:

- Type 1: Extends $e_1 \ne e_2 \in I$, where $B \not\subseteq e_1$.
- Type 2: Extends $e_1 \ne e_2 \in I$, where $B \subseteq e_1 \cap e_2$.

We consider each of these types separately.

Type 1. Let $B' = B \cap e_1$. There are at most $2^{|B|} \le 2^d$ choices for B'. Since I has branching factor c/p and $e_1 \supseteq B'$, given $B' \subseteq B$ there are at most $(c/p)^{d-r-|B'|}$ choices for e_1. By assumption, $A := B \setminus e_1$ is non-empty, and moreover $|A| = |B| - |B \cap e_1| = d - s - |B'|$. Hence the first property of I implies that $e_1 \cup B = e_1 \cup A$ has at most $(c/p)^{r-|A|} = (c/p)^{r+s-d+|B'|}$ extensions in H. In total, we have counted at most $2^d \cdot (c/p)^{d-r-|B|'} \cdot (c/p)^{r+s-d+|B'|} = 2^d (c/p)^s$ extensions.

Type 2. Since $e_1 \supseteq B$ and I has branching factor c/p, there are at most $(c/p)^{(d-r)-(d-s)} = (c/p)^{s-r}$ choices for e_1. Let $e_\cap = e_1 \cap e_2$, and note that given e_1, there are at most $2^{|e_1|} \le 2^d$ choices for $e_\cap$. Given $e_\cap$, since I has branching factor c/p, there are at most $(c/p)^{d-r-|e_\cap|}$ choices for e_2. By assumption, $A := e_2 \setminus e_1$ is non-empty, and moreover $|A| = |e_2| - |e_\cap| = d - r - |e_\cap|$. Hence the first property of I implies that $e_1 \cup e_2 = e_1 \cup A$ has at most $(c/p)^{r-|A|} = (c/p)^{2r-d+|e_\cap|}$ extensions in H. In total, we have counted at most $(c/p)^{s-r} \cdot 2^d \cdot (c/p)^{d-r-|e_\cap|} \cdot (c/p)^{2r-d+|e_\cap|} = 2^d (c/p)^s$ extensions.

Summing over both types, there are at most $2^{d+1}(c/p)^s \le (2^{d+1}c/p)^s$ extensions, completing the proof. □

The second completion step guarantees that the completion contains enough hyperedges.

Lemma 6.4. *For every integer d, $c > 0$, there exists $p_0 = p_0(c,d) \in (0,1)$ such that the following holds for all $p \in (0, p_0)$. Let H be a d-uniform hypergraph and I a $(d-r)$-uniform hypergraph for some $1 \le r \le d$ such that*

1. *Each hyperedge in I has at least $(c/p)^r$ extensions in H.*
2. *For every $e \in I$ and every $A \neq \emptyset$ disjoint from e, $e \cup A$ has at most $(c/p)^{r-|A|}$ extensions in H.*
3. *I has branching factor c/p.*

Then there exists a subhypergraph K of H such that

1. *K contains $\Omega_d(|I|(c/p)^r)$ hyperedges.*
2. *K has branching factor $O_d(c/p)$.*

Proof. We choose p_0 so that $\lfloor (c/p)^r \rfloor \ge (c/p)^r/2$.[5]

Let K' be the subhypergraph constructed in Lemma 6.3. Every hyperedge in $H \setminus K'$ extends at most one hyperedge of I. For every hyperedge $e \in I$, let n_e be the number of extensions of e in K', let $m_e = \max(\lfloor (c/p)^r \rfloor - n_e, 0)$, and let H_e be a set of m_e extensions of e in $H \setminus K'$. We let $K = K' \cup \bigcup_{e \in I} H_e$.

By construction, every $e \in I$ has at least $(c/p)^r/2$ extensions in K. A given hyperedge can extend at most 2^d many hyperedges of I, so K contains at least $|I|(c/p)^r/2^{d+1}$ hyperedges.

It remains to bound the branching factor of K. Fix a set B of size $d-s$, where $s \ge 1$. We will bound the number of extensions of B in $K \setminus K'$.

Let $B' = B \cap e$. There are at most $2^{|B|} \le 2^d$ choices for B'. Since I has branching factor c/p, given B' there are at most $(c/p)^{d-r-|B'|}$ choices for e. Let $A := B \setminus e$, so that $|A| = |B| - |B \cap e| = d - s - |B'|$. If $A \neq \emptyset$ then the second property of I implies that $e \cup B = e \cup A$ has at most $(c/p)^{r-|A|} = (c/p)^{r+s-d+|B'|}$ extensions in H and so in $K \setminus K'$. If $A = \emptyset$ then we get the same conclusion by construction since $e \cup B = e$. In total, we have counted at most $2^d \cdot (c/p)^{d-r-|B'|} \cdot (c/p)^{r+s-d+|B'|} = 2^d (c/p)^s \le (2^d c/p)^s$ extensions, completing the proof. □

We will argue about the completion using the following fundamental lemma, which is also important for applications.

Lemma 6.5. *For every integer d, $c > 0$ and $\varepsilon \in (0,1)$, there exists $f(c,d,\varepsilon) \in (0,1)$ satisfying $\lim_{c \to 0} f(c,d,\varepsilon) = 1$ for every d, ε such that the following holds. Let H be a d-uniform hypergraph, and let $p \in (0, 1-\varepsilon)$. If H has branching factor c/p then for every hyperedge $e \in H$, $\Pr_{S \sim \mu_p}[H|_S = \{e\}] \ge f(c,d,\varepsilon) p^d$.*

Before proceeding to the proof of the lemma, we first recall the statement of FKG inequality.

Lemma 6.6 (FKG inequality)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be two monotonically increasing (or decreasing) family of subsets. Then*

$$\mu_p(\mathcal{A} \cap \mathcal{B}) \ge \mu_p(\mathcal{A}) \cdot \mu_p(\mathcal{B}).$$

[5] Another possibility, which slightly affects the proof, is to choose p_0 so that $\lceil (c/p)^r \rceil \le 2(c/p)^r$.

Proof of Lemma 6.5. Let $K := H|_{e=\emptyset} \setminus \emptyset = (H-e)|_{e=\emptyset}$. Note that $\Pr_{S\sim\mu_p}[H|_S = \{e\}] = p^d \Pr_{S\sim\mu_p}[K|_S = \emptyset]$. Lemma 3.3 shows that $H|_{e=\emptyset}$ has branching factor $O_d(c/p)$. In particular, for every s it has at most $O_d(c/p)^s$ hyperedges of cardinality s. For every hyperedge $e' \in K$, let $E_{e'}$ denote the event $e' \notin K|_S$ (i.e., $S \not\supseteq e'$), where $S \sim \mu_p$. Note that

$$\Pr[E_{e'}] = 1 - p^s = \exp\left(\frac{\log(1-p^s)}{p^s} p^s\right).$$

Now $\log(1-x)/x = -1 - x/2 - \cdots$ is decreasing (its derivative is $-1/2 - 2x/3 - \cdots$), and so $p^s \le p \le 1-\varepsilon$ implies that $\log(1-p^s)/p^s \ge \log \varepsilon/(1-\varepsilon)$. In other words, $\Pr[E_{e'}] \ge e^{-O_\varepsilon(p^s)}$.

Since the events $E_{e'}$ are monotone decreasing, the FKG lemma shows that they positively correlate, hence

$$\Pr_{S\sim\mu_p}[K|_S = \emptyset] \ge \prod_{s=1}^{d} (1-p^s)^{O_d(c/p)^s} \ge \prod_{s=1}^{d} e^{-O_{d,\varepsilon}(c^s)} =: f(c,d,\varepsilon).$$

The lemma follows since clearly $\lim_{c\to 0} f(c,d,\varepsilon) = 1$. □

We can now complete the inductive proof of Lemma 6.1.

Proof of Lemma 6.1. The proof is by induction on d. When $d = 0$ we can take $H' = H$, so we can assume that $d \ge 1$. Let $\gamma = c/M_d$, where $M_d \ge 1$ will be chosen later. We apply Lemma 6.2 to H with $c := \gamma$. If H has a subhypergraph H' with branching factor γ/p such that $\iota_p(H') \ge \iota_p(H)/(d+1)$ then we are done, so suppose that there exists some $d-r$ uniform hypergraph I and a subhypergraph H' satisfying the properties of the lemma. Apply the induction hypothesis to construct a subhypergraph I' of I that has branching factor γ/p and satisfies $\iota_p(I') = \Omega_{\gamma,d}(\iota_p(I)) = \Omega_{\gamma,d}(\iota_p(H))$ (this requires $p \le p_0'(\gamma,d)$). Next, apply Lemma 6.4 with $c := \gamma$, $H := H'$, and $I := I'$ (this requires $p \le p_0''(\gamma,d)$) to obtain a subhypergraph K of H' (and so of H) satisfying

- K contains $\Omega_d(|I'|(\gamma/p)^r)$ hyperedges.
- K has branching factor $O_d(\gamma/p)$.

We choose M_d so that K has branching factor c/p, and let $p_0 = \min(p_0'(\gamma,d), p_0''(\gamma,d))$, which depends only on c,d.

We will take $H' := K$, so it remains to show that $\iota_p(K) = \Omega_{c,d}(\iota_p(H))$. Since $p \le p_0$, Lemma 6.5 shows that for every hyperedge $e \in K$, $\Pr_{S\sim\mu_p}[K|_S = \{e\}] = \Omega_{c,d}(p^d)$. For different hyperedges these events are disjoint, hence $\iota_p(K) = \Omega_{c,d}(|K|p^d) = \Omega_{c,d}(|I'|p^{d-r})$. On the other hand, the union bound shows that $\iota_p(I') \le |I'|p^{d-r}$, and so $\iota_p(K) = \Omega_{c,d}(\iota_p(I')) = \Omega_{c,d}(\iota_p(H))$, completing the proof. □

As a corollary, we obtain the following useful result.

Corollary 6.7. *Fix constants $\varepsilon > 0$ and $d \ge 0$. There exists $p_0 > 0$ (depending on d,ε) such that for every $p \in (0,p_0)$ and every d-uniform hypergraph H there exists a subhypergraph H' obtained by removing hyperedges such that*

1. *$\iota_p(H') = \Omega_{d,\varepsilon}(\iota_p(H))$.*
2. *For every $e \in H'$, $\Pr_{S\sim\mu_p}[H'|_S = \{e\}] \ge (1-\varepsilon)p^d$.*

Proof. Let $c > 0$ be a constant to be chosen later, and define $p_0 \le 1/2$ so that the theorem applies. The theorem gives us a subhypergraph satisfying the first property. Moreover, for every $e \in H'$, Lemma 6.5 (applied with $\varepsilon := 1/2$) shows that $\Pr_{S\sim\mu_p}[H|_S = \{e\}] \ge f(c,d)p^d$, where $\lim_{c\to 0} f(c,d) = 1$. Take c so that $f(c,d) > 1-\varepsilon$ to complete the proof. □

6.2 Proof in the uniform setting

We now use Corollary 6.7 to transfer the hypergraph pruning lemma to the uniform setting (Lemma 3.5). Recall that distribution $\nu_{n,k}$ refers to the uniform distribution over $\binom{[n]}{k}$.

Proof of Lemma 3.5. Let $p = k/n$. Notice that

$$\Pr_{S\sim\mu_p}[H|_S \neq \emptyset] \geq \sum_{\ell=k}^{n} \Pr[\mathrm{Bin}(n,p)=\ell] \Pr_{S\sim\nu_{n,\ell}}[H|_S \neq \emptyset] \geq \Pr[\mathrm{Bin}(n,p)\geq k] \Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset].$$

It is well-known that the median[6] of $\mathrm{Bin}(n,p)$ is one of $\lfloor np \rfloor, \lceil np \rceil$. Since $np = k$, we deduce that the median is k and $\Pr[\mathrm{Bin}(n,p) \geq k] \geq 1/2$. Therefore $\iota_p(H) \geq \Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]/2$. Applying Corollary 6.7 with $\varepsilon := \min(\varepsilon/2, 1/2)$, we thus get a subhypergraph H' such that

$$\iota_p(H') = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]),$$

which implies that

$$|H'| = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]/p^d).$$

Let now $e \in H'$ be an arbitrary hyperedge. We are given that $\Pr_{S\sim\mu_p}[H'|_S = \{e\} \mid e \in S] \geq 1 - \varepsilon/2$. For $K = H'|_{e=\emptyset} \setminus \{\emptyset\}$, the left-hand side is $\Pr_{S\sim\mu_p}[K|_S = \emptyset]$. As before, we have

$$\Pr_{S\sim\nu_{n,k}}[K|_S \neq \emptyset] \leq 2 \Pr_{S\sim\mu_p}[K|_S \neq \emptyset] \leq \varepsilon,$$

and so we get the second property. For the first property, we have

$$\Pr_{S\sim\nu_{n,k}}[H'|_S \neq \emptyset] \geq \sum_{e\in H'} \Pr_{S\sim\nu_{n,k}}[H'|_S = \{e\}] \geq (1-\varepsilon)|H'|\frac{k^{\underline{d}}}{n^{\underline{d}}}.$$

By assumption $k^{\underline{d}}/n^{\underline{d}} \geq (p/2)^d$, and so

$$\Pr_{S\sim\nu_{n,k}}[H'|_S \neq \emptyset] \geq (1-\varepsilon)\cdot\Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]/p^d)\cdot(p/2)^d = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]). \qquad \square$$

7 Agreement theorem via majority decoding

A nice application of the hypergraph pruning lemma is to show that majority decoding always works for agreement testing. In particular, if the agreement theorem (Theorem 5.1) holds, then one might without loss of generality assume that the global function is the one obtained by majority/plurality decoding.

Lemma 7.1. *For every positive integer d and alphabet Σ, there exists a $p \in (0,1)$ such that for $\alpha \in (0,1)$ and all positive integers n, k, t satisfying $n \geq k \geq t \geq \max\{2d, \alpha k\}$ and $k \leq pn$ the following holds.*

Suppose an ensemble of local functions $\{f_S : \binom{S}{d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ and a global function $F : \binom{[n]}{d} \to \Sigma$ satisfy

$$\Pr_{S_1,S_2\sim\nu_{n,k,t}}[f_{S_1}|_{S_1\cap S_2} \neq f_{S_2}|_{S_1\cap S_2}] = \varepsilon, \qquad \Pr_{S\sim\nu_{n,k}}[f_S \neq F|_S] = \delta.$$

Then, the global function $G : \binom{[n]}{d} \to \Sigma$ defined by plurality decoding (ie., $G(T)$ is the most popular value of $f_S(T)$ over all S containing T, breaking ties arbitrarily) satisfies

$$\Pr_{S\sim\nu_{n,k}}[f_S \neq G|_S] = O_{d,\alpha}(\varepsilon+\delta).$$

[6] The median of a distribution X on the integers is the integer m such that $\Pr[X \geq m], \Pr[X \leq m] \geq 1/2$.

Proof. All probabilities below, unless specified otherwise, are over $S \sim \nu_{n,k}$.

Since $\Pr[f_S \neq G|_S] \leq \Pr[f_S \neq F|_S] + \Pr[F|_S \neq G|_S] = \delta + \Pr[F|_S \neq G|_S]$, it suffices to bound $\Pr[F|_S \neq G|_S]$. Let $H := \{T : G(T) \neq F(T)\}$, so that $\Pr[F|_S \neq G|_S] = \Pr[H|_S \neq \emptyset]$. Note that F and G are functions, while H is a hypergraph. Apply Lemma 3.5 on the hypergraph H, for a constant $\varepsilon = \eta := 1/(2|\Sigma|) > 0$, to get a subhypergraph H' ($p = p_0(d, \varepsilon)$ is chosen such that $k \leq pn$).

For any edge $T \in H'$ and $\sigma \in \Sigma$, define the following quantities

$$p(T, \sigma) := \Pr[H'|_S = \{T\} \text{ and } f_S(T) = \sigma \mid S \supseteq T], \qquad p(T) := \max_\sigma p(T, \sigma)$$

$$q(T, \sigma) := \Pr[f_S(T) = \sigma \mid S \supseteq T], \qquad q(T) := \max_\sigma q(T, \sigma)$$

Note that $G(T)$ by definition satisfies $q(T) = q(T, G(T))$. Since by the hypergraph pruning lemma, we have $\Pr[H'|_S = \{T\} | S \supset T] \geq 1 - \eta$, we have $q(T, \sigma) \geq (1 - \eta) \cdot p(T, \sigma)$ for all σ. Hence, $q(T, G(T)) = q(T) \geq (1 - \eta) \cdot p(T)$. On the other hand for any σ, $p(T, \sigma) \geq q(T, \sigma) - \eta$. In particular, $p(T, G(T)) \geq q(T, G(T)) - \eta \geq q(T, G(T))/2$ (since $q(T, G(T)) \geq 1/|\Sigma|$ and $\eta \leq 1/(2|\Sigma|)$). Combining these, we have that for all $T \in H'$,

$$p(T, G(T)) \geq (1 - \eta) \cdot p(T)/2. \tag{2}$$

We now relate the probabilities $p(T)$ and $p(T, G(T))$ to δ and ε in the lemma statement.

By the hypergraph pruning lemma, we have $\Pr[H'|_S = \{T\} | S \supset T] \geq 1 - \eta$ or equivalently $\sum_\sigma p(T, \sigma) \geq 1 - \eta$. For each hyperedge $T \in H'$, we have

$$\Pr_{S_1, S_2 \sim \nu_{n,k}}[f_{S_1}(T) \neq f_{S_2}(T) \text{ and } H'|_{S_1} = H'|_{S_2} = \{T\} \mid S_1 \cap S_2 \supseteq T] = \sum_{\sigma_1 \neq \sigma_2} p(T, \sigma_1) p(T, \sigma_2)$$
$$\geq \sum_{\sigma_1} p(T, \sigma_1)(1 - \eta - p(T, \sigma_1)) \geq \sum_{\sigma_1} p(T, \sigma_1)(1 - \eta - p(T)) \geq (1 - \eta)(1 - \eta - p(T)).$$

Consider now the following coupling. Choose $S_1, S_2 \sim \nu_{n,k}$ containing T, and choose a set S intersecting each of S_1, S_2 in exactly t elements including T (this is possible since k/n is small enough). If $f_{S_1}(T) \neq f_{S_2}(T)$ then either $f_{S_1}(T) \neq f_S(T)$ or $f_{S_2}(T) \neq f_S(T)$, and so

$$(1 - \eta)(1 - \eta - p(T)) \leq 2 \Pr_{S_1, S \sim \nu_{n,k,t}}[f_{S_1}(T) \neq f_S(T) \text{ and } H'|_{S_1} = \{T\} \mid S_1 \cap S \supseteq T].$$

Summing over all edges in H', we deduce that

$$\varepsilon \geq \sum_{T \in H'} \frac{(1 - \eta)(1 - \eta - p(T))}{2} \Pr_{S_1, S_2 \sim \nu_{n,k,t}}[S_1 \cap S_2 \supseteq T] = \sum_{T \in H'} \frac{(1 - \eta)(1 - \eta - p(T))}{2} \Omega_\alpha(\Pr[S \supseteq T]), \tag{3}$$

since $t \geq \alpha k$.

We now relate δ to $p(T, H(T))$. We clearly have

$$\Pr_{S \sim \nu_{n,k}}[f_S(T) \neq F(T) \text{ and } H'|_S = \{T\} \mid S \supseteq T] \geq \Pr_{S \sim \nu_{n,k}}[f_S(T) = G(T) \text{ and } H'|_S = \{T\} \mid S \supseteq T] = p(T, G(T)).$$

Summing over all edges in H', we deduce that

$$\delta \geq \sum_{T \in F'} p(T, G(T)) \cdot \Pr[S \supseteq T]. \tag{4}$$

Either $p(T) \leq 1/3$ in which case $(1 - \eta)(1 - \eta - p(T))/2 = \Omega(1)$ or $p(T) \geq 1/3$ and hence $p(T, G(T)) \geq 1/6 = \Omega(1)$ (from (2)). Thus, in either case, adding (4) and (3), we have

$$\varepsilon + \delta \geq \sum_{T \in H'} \Omega_\alpha(\Pr[S \supseteq T]) = \Omega_\alpha(\Pr[H'|_S \neq \emptyset]) = \Omega_{d,\alpha}(\Pr[H|_S \neq \emptyset]).$$

We conclude that $\Pr[H|_S \neq \emptyset] = O_{d,\alpha}(\varepsilon + \delta)$, completing the proof. □

We can now combine the above lemma with the agreement theorem (Theorem 5.1) proved earlier to obtain the following strenghtened agreement theorem, which is the "uniform version" of the agreement theorem stated in the introduction.

Theorem 7.2 (Main)**.** *For every positive integer d and alphabet Σ, there exists a constant $C > 1$ such that for all $\alpha, \beta \in (0,1)$ satisfying $\alpha + \beta \leq 1$ and all positive integers $n \geq k \geq t$ satisfying $n \geq Ck$ and $t \geq \max\{\alpha k, 2d\}$ and $k - t \geq \max\{\beta k, d\}$, the following holds: Let $f = \{f_S \colon \binom{S}{\leq d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying $\text{agree}_{\nu_{n,k,t}}(f) \geq 1 - \varepsilon$, that is,*

$$\Pr_{S_1, S_2 \sim \nu_{n,k,t}} [f_{S_1}|_{S_1 \cap S_2} = f_{S_2}|_{S_1 \cap S_2}] \geq 1 - \varepsilon,$$

where $\nu_{n,k,t}$ is the uniform distribution over pairs of k-sized subsets of $[n]$ of intersection exactly t. Then there exists a global function $G \colon \binom{[n]}{\leq d} \to \Sigma$ satisfying $\Pr_{S \in \binom{[n]}{k}}[f_S = G|_S] = 1 - O_{d,\alpha,\beta}(\varepsilon)$.

Here, $F|_S$ refers to the restriction $F|_{\binom{S}{\leq d}}$.

Furthermore, we may assume that the global function G is the one given by "popular vote", namely for each $A \in \binom{[n]}{\leq d}$ set $G(A)$ to be the most frequently occurring value among $\{f_S(A) \mid S \supset A\}$ (breaking ties arbitrarily).

Proof of Theorem 7.2. By Theorem 5.1, we have a global function $F \colon \binom{[n]}{\leq d} \to \Sigma$ (not necessarily G) satisfying

$$\Pr_{S \in \binom{[n]}{k}} [f_S \neq F|_S] = O(\varepsilon).$$

For each $i \in \{0, 1, \ldots, d\}$, let $f^{(i)}|_S := f_S|_{\binom{S}{i}}$, $F^{(i)} := F|_{\binom{[n]}{i}}$ and $G^{(i)} := G|_{\binom{[n]}{i}}$. Clearly, we have for each i,

$$\Pr_{S_1, S_2 \sim \nu_{n,k,t}} [f^{(i)}_{S_1}|_{S_1 \cap S_2} \neq f^{(i)}_{S_2}|_{S_1 \cap S_2}] = \varepsilon, \qquad \Pr_{S \sim \nu_{n,k}} [f^{(i)}_S \neq F^{(i)}|_S] = O(\varepsilon).$$

Hence, by Lemma 7.1, we have

$$\Pr_{S \sim \nu_{n,k}} [f^{(i)}_S \neq G^{(i)}|_S] = O(\varepsilon).$$

This implies $\Pr_{S \sim \nu_{n,k}}[f_S \neq G|_S] = d \cdot O(\varepsilon) = O_d(\varepsilon)$. ☐

The entire discussion in this part so far has been with respect to the distribution $\nu_{n,k}$, the uniform distribution over k-sized subsets of $[n]$. We can extend these results to the biased setting μ_p using a trick, thus proving the agreement theorem (Theorem 1.1) stated in the introduction. In this setting, the distribution $\nu_{n,k,t}$ is replaced by the distribution $\mu_{p,\alpha}$, which is a distribution over pairs S_1, S_2 of subsets of $[n]$ defined as follows. For each element x independently, we put x only in S_1 or only in S_2 with probability $p(1 - \alpha)$ (each), and we put x in both with probability $p\alpha$. This is possible if $p(2 - \alpha) \leq 1$ (we assume below $p \leq 1/2$ and hence $p(2 - \alpha) \leq 1$). Note that if sets S_1, S_2 are picked according to the distribution $\mu_{p,\alpha}$ then the marginal distribution of each of S_1 and S_2 is μ_p.

Proof of Theorem 1.1. Let N be a large integer and define $K = \lfloor pN \rfloor$, $T = \lfloor p\alpha N \rfloor$. For every $S \in \binom{[N]}{K}$, define $\tilde{f}_S = f_{S \cap [n]}$. In other words, for all $A \subset S \in \binom{[N]}{K}, |A| \leq d$. let $\tilde{f}_S(A) = f_{S \cap [n]}(A \cap [n])$. If $S_1, S_2 \sim \nu_{N,K,T}$ then the distribution of $S_1 \cap [n], S_2 \cap [n]$ is close to $\mu_{p,\alpha}$, and so for large enough N we have

$$\Pr_{S_1, S_2 \sim \nu_{N,K,T}} [\tilde{f}_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \varepsilon/2.$$

Hence, the ensemble of functions $\{\tilde{f}_S\}_{S \in \binom{[N]}{K}}$ satisfies the hypothesis of the ageement theorem (Theorem 7.2) with ε replaced by $3\varepsilon/2$. Hence, by Theorem 7.2, if we define $\tilde{G} \colon \binom{[N]}{\leq d} \to \Sigma$ by plurality decoding then

$\Pr_{S\sim\nu_{N,K}}[\tilde{f}_S \neq \tilde{G}|_S] = O_d(\varepsilon)$. Since $\tilde{f}_S$ depends only on $S \cap [n]$, there exists a function $\hat{G}\colon \binom{[n]}{d} \to \Sigma$ such that $\tilde{G}(T) = \hat{G}(T \cap [n])$. Moreover, for large enough N the distribution of $S \cap [n]$ approaches μ_p, and so $\hat{G} = G$. (There's a fine point here: there could be several most common values. Fortunately, this doesn't invalidate the proof — just choose the correct G.) This completes the proof. □

7.1 More parameter settings

Our main agreement theorem, Theorem 1.1, only holds for $p \leq p_0$ for some constant $p_o := p_0(d)$. For the application to testing Reed–Muller codes, we need an agreement theorem that holds for all $p \in (0,1)$. The following theorem shows that it is easy to prove a counterpart of Theorem 1.1 if one is allowed a multiplicative decay of p^{-d}. Note that for $p \geq p_0$, this loss is not an issue for our applications (as we think of d as a constant and $p \geq p_0$ is a constant). This is no longer true when $p = o(1)$, a regime in which we need the stronger result proved in Theorem 1.1, which is independent of p (but still dependent on d).

Theorem 7.3. *For every positive integer d and alphabet Σ, the following holds for all p. If $\{f_S\colon \binom{S}{\leq d} \to \Sigma \mid S \in \{0,1\}^n\}$ is an ensemble of functions satisfying*

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

then the global function $G\colon \binom{[n]}{\leq d} \to \Sigma$ defined by plurality decoding satisfies

$$\Pr_{S\sim\mu_p}[f_S \neq G|_S] \leq 2p^{-d}\varepsilon.$$

Proof. The main observation behind the proof is this: if we choose $S_1, S_2 \sim \mu_p$ independently, then $(S_1, S_2) \sim \mu_{p,p}$.

Consider now an arbitrary f_{S_1}. Suppose that $f_{S_1} \neq G|_{S_1}$, and choose an entry T such that $f_{S_1}(T) \neq G(T)$. If we choose $S_2 \sim \mu_p$, then $T \subseteq S_2$ with probability $p^{|T|} \geq p^d$. Moreover, since $f_{S_1}(T) \neq G(T)$, the probability that $f_{S_2}(T) = f_{S_1}(T)$ is at most $1/2$. Therefore the probability that f_{S_1} and f_{S_2} disagree on their intersection is at least $p^d/2$. This shows that

$$\varepsilon = \Pr_{S_1,S_2\sim\mu_{p,p}}[f_{S_1}|_{S_1\cap S_2} \neq f_{S_2}|_{S_1\cap S_2}] \geq \frac{p^d}{2} \Pr_{S_1\sim\mu_p}[f_{S_1} \neq G|_{S_1}].$$ □

Part II
Structure Theorems

8 Testing Reed–Muller codes

Every Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ can be written in a unique way as $P \bmod 2$, where P is a *Boolean polynomial*, that is, a sum of distinct multilinear monomials. The *Boolean degree* of f, denoted $\mathrm{bdeg}(f)$, is the degree of this polynomial.

The well-known BLR test [BLR93, BCH+96] checks whether a given Boolean function has Boolean degree 1. Alon *et al.* [AKK+05] developed the following 2^{d+1}-query test T_d, which is a generalization of the BLR test to large Boolean degrees.

- Test T_d: Input $f\colon \{0,1\}^n \to \{0,1\}$
  - Pick $x, a_1, \ldots, a_{d+1} \in \{0,1\}^n$ independently from the distribution $\mu_{1/2}^{\otimes n}$, subject to the constraint that $a_1, \ldots, a_{d+1}$ are linearly independent.
  - Accept iff
$$\sum_{I \subseteq [d+1]} f\left(x + \sum_{i \in I} a_i\right) = 0 \pmod 2 .$$

This test is closely related to the Gowers norms. An optimal analysis of the test was provided by Bhattacharyya *et al.* [BKS+10]. We need a few definitions to state their result.

Definition 8.1. *Let $f\colon \{0,1\}^n \to \{0,1\}$ and let $d \geq 0$. The* distance to degree d of f *is defined as follows:*
$$\delta_d(f) := \min_{\mathrm{bdeg}(g) \leq d} \Pr_{\mu_{1/2}}[f \neq g].$$

Definition 8.2. *Let $f\colon \{0,1\}^n \to \{0,1\}$ and let $d \geq 0$. The* failure probability of test T_d *is*
$$\mathrm{rej}_d(f) := \Pr_{\substack{x, a_1, \ldots, a_{d+1} \sim \mu_{1/2}^{\otimes n} \\ a_1, \ldots, a_{d+1} \text{ linearly independent}}} \left[\sum_{I \subseteq [d+1]} f\left(x + \sum_{i \in I} a_i\right) \neq 0 \pmod 2 \right].$$

Remark 8.3. *Another variant of T_d, which is closer to the definition of the Gowers norms, samples $a_1, \ldots, a_{d+1}$ without requiring them to be linearly independent. When $a_1, \ldots, a_{d+1}$ are linearly dependent, the test always succeeds. On the other hand, when $n \geq d+1$, the probability that $a_1, \ldots, a_{d+1}$ are linearly dependent is lower-bounded by a positive constant. Therefore when $n \geq d+1$, removing the constraint of linear independence only affects the rejection probability by a constant factor. Finally, when $n \leq d$ the test is pointless, since every function has Boolean degree at most d.*

Theorem 8.4 ([BKS+10])**.** *For every integer $d \geq 1$ there exists a constant $\varepsilon_d > 0$ such that for all Boolean functions $f\colon \{0,1\}^n \to \{0,1\}$,*
$$\mathrm{rej}_d(f) \geq \min(2^d \delta_d(f), \varepsilon_d).$$

Corollary 8.5. *For every integer $d \geq 1$ and all Boolean functions $f\colon \{0,1\}^n \to \{0,1\}$,*
$$\delta_d(f) = O_d(\mathrm{rej}_d(f)).$$

Proof. If $2^d \delta_d(f) \leq \varepsilon_d$ then $\delta_d(f) \leq 2^{-d} \mathrm{rej}_d(f)$. Otherwise, $\mathrm{rej}_d(f) \geq \varepsilon_d$, and so $\delta_d(f) \leq 1 \leq \varepsilon_d^{-1} \mathrm{rej}_d(f)$. $\square$

Our goal in this section is to extend the analysis of Bhattacharyya *et al.* to the μ_p setting (wherein we measure closeness of f to degree d with respect to the μ_p measure instead of the $\mu_{1/2}$ measure). More precisely:

Definition 8.6. *Let $f\colon \{0,1\}^n \to \{0,1\}$ and let $d \geq 0$. The* distance to degree d *of f is defined as follows:*

$$\delta_d^{(p)}(f) := \min_{\mathrm{bdeg}(g) \leq d} \Pr_{\mu_p}[f \neq g].$$

To this end, we consider the following natural extension $T_{p,d}$ of the AKKLR test T_d to the μ_p measure.

- Test $T_{p,d}$: Input $f\colon \{0,1\}^n \to \{0,1\}$
  - Pick $S \subseteq [n]$ according to the distribution μ_{2p}.
  - Let $f|_S\colon \{0,1\}^S \to \{0,1\}$ denote the restriction of f to $\{0,1\}^S$ by zeroing out all the coordinates outside S.
  - Pick $x, a_1, \ldots, a_{d+1} \in \{0,1\}^S$ independently from the distribution $\mu_{1/2}^{\otimes S}$, subject to the constraint that $a_1, \ldots, a_{d+1}$ are linearly independent.
    (If $|S| \leq d$, skip this and the following step, and immediately accept.)
  - Accept iff

$$\sum_{I \subseteq [d+1]} f|_S\left(x + \sum_{i \in I} a_i\right) = 0 \pmod 2 .$$

Observe that the points $(x + \sum_{i\in I} a_i)$, when viewed as points in $\{0,1\}^n$ (by filling the coordinates outside S with 0's), are distributed individually according to $\mu_p^{\otimes n}$.

Let $\mathrm{rej}_T(f)$ denote the rejection probability of a test T on input function f. Let us say that a test T is *valid for p* if $\mathrm{rej}_T(f) = 0$ whenever $\mathrm{bdeg}(f) \leq d$ (completeness), and there exists a universal constant C such that $\delta_d^{(p)}(f) \leq C\,\mathrm{rej}_T(f)$ (soundness). Corollary 8.5 states that T_d (modified so that it always accepts when the dimension is at most d) is valid for $1/2$. In the rest of this section, we prove the following theorem.

Theorem 8.7 (p-biased version of the BKSSZ Theorem)**.** *For every d and $p \in (0,1)$ there exists a 2^{d+1}-query test T that satisfies the following properties:*

- *Completeness: if $\mathrm{bdeg}(f) \leq d$ then $\mathrm{rej}_{T_{d,p}}(f) = 0$.*
- *Soundness: $\delta_d^{(p)}(f) = O_d(\mathrm{rej}_{T_{d,p}}(f))$, where the hidden constant is independent of p.*

Remark 8.8. *The most natural candidate for the test T (mentioned in the theorem above) is the test $T_{p,d}$ defined above. In fact, we prove below that for small p, this is indeed the case. In other words, for small p, the test $T_{p,d}$ is valid for p. For other p, we prove that slight variants of this natural test work, though we believe that the natural test $T_{p,d}$ works for all $p \in (0,1/2)$. The variants of the test $T_{p,d}$ are obtained using the following simple observation. Given a test T which is valid for p, we can obtain a test $\overline{T}$ which is valid for $1-p$ by running T on the function $\overline{f} = (x_1, \ldots, x_n) \mapsto f(1-x_1, \ldots, 1-x_n)$. The end result is a test which sets some coordinates to zero, other coordinates to one, and only then invokes T_d.*

Let us say that *agreement holds for* (p, α) if a statement of the form of Theorem 1.1 holds for the given values of p, α, with $\Pr[f_S \neq g|_S] \leq C\varepsilon$ for an *absolute* constant C. Thus Theorem 1.1 shows that for any fixed α and $p < p_0(d)$, agreement holds for (p, α), and Theorem 7.3 shows that for fixed $p_1 > 0$, agreement holds for (p, p) for all $p \geq p_1$.

Lemma 8.9. *Suppose that $\mathcal{T}$ is valid for r, that agreement holds for $(r^{-1}p, \alpha)$, where $p \leq r \leq \alpha$, and that there exists a constant $c > 0$ such that $\min(r/\alpha, 1-r/\alpha) \geq c$. Then the test $\mathcal{U} := \mathcal{T}^{(r^{-1}p)}$ which runs $\mathcal{T}$ on $f|_S$ (the restriction of f to $\{0,1\}^S$ obtained by zeroing out all other coordinates) for $S \sim \mu_{r^{-1}p}$ is valid for p.*

Proof. Let us start by noticing that completeness is clear, since $\mathrm{bdeg}(f|_S) \le \mathrm{bdeg}(f)$. It remains to prove soundness.

By construction, $\mathrm{rej}_{\mathcal{U}}(f) = \mathbb{E}_{S \sim \mu_{r^{-1}p}}[\mathrm{rej}_{\mathcal{T}}(f|_S)]$. The assumption that $\mathcal{T}$ is valid for r guarantees the existence of a Boolean polynomial P_S over $\{x_i : i \in S\}$ of degree at most d satisfying $\delta_S := \Pr_{\mu_r}[f|_S \neq P_S \bmod 2] = O(\mathrm{rej}_{\mathcal{T}}(f|_S))$. Our goal now is to use agreement in order to sew the various polynomials P_S. We will show that the polynomials P_S agree with each other using the Schwartz–Zippel lemma, in the following biased form.

Claim 8.10. *Suppose that P is a non-zero polynomial with* $\mathrm{bdeg}(f) \le d$. *Then for all $\theta \in (0,1)$,*

$$\Pr_{\mu_\theta}[P \bmod 2 = 1] \ge \min(\theta^d, 1-\theta) \ge \min(\theta, 1-\theta)^d.$$

Proof. Let I be an inclusion-maximal set such that P contains the monomial $\prod_{i \in I} x_i$. For every setting of the variables not in I, there is at least one setting of the variables in I for which $P \bmod 2 = 1$, and this setting has probability at least $\min(\theta^d, 1-\theta)$ in μ_θ. □

Let $\theta = r/\alpha$. For two sets $S \supseteq T$, we define

$$\delta_{S,T} = \Pr_{\mu_\theta}[f|_T \neq P_S|_T \bmod 2].$$

Let us say that (S,T) is *good* if $\delta_{S,T} < \min(\theta, 1-\theta)^d/2$. If both (S_1,T) and (S_2,T) are good then

$$\Pr_{\mu_\theta}[P_{S_1}|_T \neq P_{S_2}|_T \bmod 2] < \min(\theta, 1-\theta)^d,$$

and so Claim 8.10 shows that $P_{S_1}|_T = P_{S_2}|_T$.

Notice that

$$\mathop{\mathbb{E}}_{T \sim \mu_\alpha(S)}[\delta_{S,T}] = \delta_S = O(\mathrm{rej}_{\mathcal{T}}(f|_S)).$$

This implies that for fixed S, if we choose $T \sim \mu_\alpha(S)$ then (S,T) is good with probability $1 - O(\mathrm{rej}_{\mathcal{T}}(f|_S))$. If we sample $(S_1,S_2) \sim \mu_{r^{-1}p,\alpha}$ then $S_1 \cap S_2 \sim \mu_\alpha(S_1)$. Therefore

$$\Pr_{(S_1,S_2) \sim \mu_{r^{-1}p,\alpha}}[(S_1,S_2) \text{ good}] = 1 - \mathop{\mathbb{E}}_{S_1 \sim \mu_{r^{-1}p}}[O(\mathrm{rej}_{\mathcal{T}}(f|_{S_1}))] = 1 - O(\mathrm{rej}_{\mathcal{U}}(f)).$$

The same holds with the roles of S_1, S_2 reversed, and so

$$\Pr_{(S_1,S_2) \sim \mu_{r^{-1}p,\alpha}}[P_{S_1}|_T \neq P_{S_2}|_T] = O(\mathrm{rej}_{\mathcal{U}}(f)).$$

We can now apply the agreement assumption to deduce that there exists a degree d polynomial P over $x_1, \ldots, x_n$ such that $\Pr_{S \sim \mu_{r^{-1}p}}[P|_S \neq P_S] = O(\mathrm{rej}_{\mathcal{U}}(f))$. It follows that

$$\Pr_{\mu_p}[f \neq P \bmod 2] = \mathop{\mathbb{E}}_{S \sim \mu_{r^{-1}p}}[\Pr_{\mu_r}[f|_S \neq P|_S]] \le O(\mathrm{rej}_{\mathcal{U}}(f)) + \mathop{\mathbb{E}}_{S \sim \mu_{r^{-1}p}}[\Pr_{\mu_r}[f|_S \neq P_S]] =$$
$$O(\mathrm{rej}_{\mathcal{U}}(f)) + \mathop{\mathbb{E}}_{S \sim \mu_{r^{-1}p}}[O(\mathrm{rej}_{\mathcal{T}}(f|_S))] = O(\mathrm{rej}_{\mathcal{U}}(f)). \quad \square$$

As mentioned above, T_d is valid for $r = 1/2$, and there exists a constant $p_0 > 0$ depending on d such that agreement holds for $(2p, 2/3)$ whenever $p \le p_0/2$. Lemma 8.9 shows that $U_p := T_d^{(2p)}$ (which is the natural test $T_{p,d}$) is valid for all $p \le p_0/2$, and so $V_r := \overline{U_{1-r}}$ is valid for all $r \ge 1 - p_0/2$.

We now make use of the following corollary of Lemma 8.9.

Corollary 8.11. *Suppose that $\mathcal{T}$ is valid for $r \geq 1/2$, and let $C > 1$. For all $Cr^2 \leq p \leq r$, the test $\mathcal{U} := \mathcal{T}^{(r^{-1}p)}$ is valid for p.*

Proof. Let $\alpha = r^{-1}p$. Theorem 7.3 shows that agreement holds for $(r^{-1}p, r^{-1}p)$, since $r^{-1}p > Cr \geq C/2$. Note that $r/\alpha = r^2/p \geq r \geq 1/2$ and $r/\alpha = r^2/p \leq 1/C$. Lemma 8.9 therefore applies, implying the corollary. □

Let $C = 1 + p_0/2$, so that $C(1 - p_0/2) < 1$. Define $r_0 := 1 - p_0/2$ and $r_{t+1} := Cr_t^2$ for $t \geq 0$. Induction shows that $r_{t+1} < r_0$ (since $Cr_0 < 1$), and so $r_t \leq (Cr_0)^t r_0$. Therefore $r_t \leq 1/2$ for some finite t. Applying Corollary 8.11 (t times), starting with the test V_r described above, we obtain tests for all $p \geq 1/2$, and so also for all $p \leq 1/2$.

9 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 9.1. *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 9.2. *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 9.3. *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 9.4. *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 9.2 and thus prove Theorem 9.1

Proof of Theorem 9.1. 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 $\varepsilon \le \varepsilon_0$.

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
$$\operatorname{dist}(f_a, \{0,1\}) \le |f_a| = \frac{|\delta|}{|y - b|} \prod_{b \neq a,y} \frac{|y - b + \delta|}{|a - b|}.$$
If $|\delta| \le \delta_0$ then $\operatorname{dist}(f_a, \{0,1\}) = O(|\delta|)$, and otherwise $\operatorname{dist}(f_a, \{0,1\}) = O(|\delta|^m)$. If $y = a$ then
$$\operatorname{dist}(f_a, \{0,1\}) \le |f_a - 1| = \left| \prod_{b \neq a} \left| 1 + \frac{\delta}{a - b} \right| - 1 \right|.$$
Once again, if $|\delta| \le \delta_0$ 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 $|\delta| \le \delta_0$ and $|\delta| > \delta_0$. The first case is simple:
$$\mathbb{E}[\operatorname{dist}(f_a, \{0,1\})^2 1_{|\delta| \le \delta_0}] \le O(\mathbb{E}[\delta^2]) = O(\varepsilon).$$
For the second case, we use Cauchy–Schwartz and the bound $\Pr[\delta^2 \ge \delta_0^2] = O(\varepsilon)$ (recall δ_0 is a constant):
$$\mathbb{E}[\operatorname{dist}(f_a, \{0,1\})^2 1_{|\delta| \ge \delta_0}] \le \sqrt{\mathbb{E}[\delta^{2m}]} O(\sqrt{\varepsilon}).$$
Let $C := 2 \max_{a \in A} |a|$. If $|f| \ge \max_{a \in A} |a|$ then clearly $|\delta| \le |f|$, and otherwise $|\delta| \le |f| + \max_{a \in A} |A| \le C$. Therefore it always holds that $|\delta| \le \max(C, |f|)$. This shows that
$$\mathbb{E}[\delta^{2m}] \le C^{2m} + \mathbb{E}[f^{2m}] = O(1) + \|f\|_{2m}^{2m}.$$
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 $\|f_a - g_a\|^2 \ge \|g_a^{>d}\|^2 = \Omega(1)$ (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 $\deg g_a \le d$.

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 9.4 (with $\alpha = 1/2$). □

10 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 10.1 (Restatement of Theorem 1.4)**.** *For every $p \le 1/2$ 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 10.2. *Let $d \ge 0$ be any positive integer and $A \subseteq \mathbb{R}$ any finite set. For every $p \le 1/2$ 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 1.1. 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 preliminary claims till the proof of Theorem 10.1, we further assume that $p \le p_0$. 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 $\mathbb{E}_{\mu_p}[\operatorname{dist}(f, A)^2] = \varepsilon$

The main result of this section extends the generalized Kindler–Safra theorem Theorem 9.1, which holds only for constant p, to all values of p via the agreement theorem Theorem 1.1. 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 $f|_S \colon \{0,1\}^S \to \mathbb{R}$ 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 1.1. 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 10.3. *For every set $S \subseteq [n]$, let*

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

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 9.1. 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 1.1, 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 10.4. *For every set $S \subseteq [n]$, let the y-expansion of the junta g_S given in Claim 10.3 be as follows:*

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

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

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

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,

$$\Pr_{S_1,S_2\sim\mu_{4p,1/2}}[g_{S_1}|_{S_1\cap S_2} \neq g_{S_2}|_{S_1\cap S_2}] \le 2 \Pr_{\substack{S\sim\mu_{4p}\\ T\sim\mu_{1/2}(S)}}[g_S|_T \neq h_T] \tag{5}$$

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,

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

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

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

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 9.2). 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 10.3

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

We now return to (5), 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 1.1). 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 10.1 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 10.1.

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 10.5. *The function g from Claim 10.4 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 10.3. Recall that $g_S = \sum_B d_{S,B} y_B$. Since g_S is a junta (by Claim 9.2), $\sum_B d_{S,B}^2 = O(1)$. 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. $\square$

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

Lemma 10.6. *Let g be the function from Claim 10.4, 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 10.5 via Lemma 3.2, and the second from Claim 10.4.

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 9.2). 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 10.7. *Let g be the function from Claim 10.4. Then $\|f - g\|^2 = \mathbb{E}_{\mu_p}[(f-g)^2] = O(\varepsilon)$.*

Proof. Let $F = \mathrm{round}(f, A)$, and let g_S, g, G be the functions defined in Claim 10.3, Claim 10.4, and Lemma 10.6. 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 10.3 allows us to estimate ε_1, using the L_2^2 triangle inequality:

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

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 $|x| \geq M$ then $x^2 \leq C\prod_{a\in A}(x-a)^2$. Let $g = g_{\leq M} + g_{>M}$, where $g_{\leq M} = g 1_{|g| \leq M}$. The L_2^2 triangle inequality shows that

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

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

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

using Claim 10.4. The defining property of M shows that

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

using the fourth property of Lemma 10.6. 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 $p \leq p_0$.

Proof of Theorem 10.1. Suppose that $p \leq p_0$, and let g be the function constructed in Claim 10.4. The first property follows from Lemma 10.7. The second property follows from Claim 10.4. The third property follows from the definition of g. The fourth property follows from Claim 10.5. Finally, Claim 10.4 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 9.1, guaranteeing the first property (we use the fact that the big O constant varies continuously with p). Claim 9.2 shows that g is an A-valued junta, implying all the other properties. □

Corollary 10.2 is proved along similar lines.

Proof of Corollary 10.2. Apply the theorem to $f := F^{\leq d}$, which satisfies $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$. The L_2^2 triangle inequality shows that $\|F - g\|^2 \leq 2\|F - f\|^2 + 2\|f - g\|^2 = O(\varepsilon)$. For the second property,

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

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

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

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

11 A converse to the main result

Given a degree d function f such that $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$, Theorem 10.1 gives a function g such that $\|f - g\|^2 = O(\varepsilon)$ and:

- $\deg g \leq d$
- 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 $\mathbb{E}[\mathrm{dist}(g, A)^2] = \varepsilon$, and in this sense Theorem 10.1 is a complete characterization of degree d functions close (in L_2) to A.

Lemma 11.1. *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 10.6. 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

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

12 Corollaries of the structure theorem

Our main theorem, Theorem 10.1, 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.

12.1 Large deviation bounds

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

Lemma 12.1. *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 12.2. *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 12.1 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 12.3. *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 12.1 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 12.4. *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 12.3 just as Lemma 12.2 follows from Lemma 12.1. □

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

Corollary 12.5 (Restatement of Lemma 1.5). *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 $\mathbb{E}[\text{dist}(f, A)^2] = \varepsilon$ 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 10.1 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. □

12.2 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 9.2 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 12.6. *Fix a constant $d \geq 0$ and a finite set A. There exist constants $C, \varepsilon_0 > 0$ such that for all $p \leq 1/4$*[7] *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 9.2. We will show that there exist constants $L, K > 0$ such that either $\delta = O(p)$ or

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

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

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

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. □

Applying Corollary 10.2, we obtain a similar result for bounded degree almost quantized functions.

Corollary 12.7 (Restatement of Lemma 1.6)**.** *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 $\mathbb{E}[\text{dist}(f, A)^2] = \varepsilon$. Then there exists $a \in A$ such that $\Pr[\text{round}(f, A) \neq a] = O(\varepsilon^C + p)$.

Proof. Let $F = \text{round}(f, A)$. Corollary 10.2 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 $\Pr[g \neq a] = O(\varepsilon^C + p)$ for some $a \in A$, and the corollary follows. □

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

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

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

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.

13 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 $\{\pm1\}^n$, and we denote their inputs by $x_1,\ldots,x_n \in \{\pm1\}$.

When we write $x \sim \{\pm1\}^n$, we always mean that x is chosen according to the uniform distribution over $\{\pm1\}^n$.

13.1 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 13.1. *Fix a finite set A. Let $f\colon \{\pm1\}^n \to \mathbb{R}$ 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 $J_0 = \{i : \hat{f}(i)^2 \geq m\varepsilon\}$. 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 $\mathbb{E}[\mathrm{dist}(f|_\alpha, A)^2] \leq \varepsilon$. This implies that for some c,

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

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 $0 \leq i \leq |A|$, define

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

For every $0 \leq i \leq |A|$, let $a_i = \mathrm{round}(v_i, A)$. Since $v_i - v_{i-1} = 2\hat{f}(i) > 0$, we can assume that $a_i \geq a_{i-1}$. By assumption, $|v_i - a_i|^2 \leq 2^{|J|}\varepsilon = 2^{|A|}\varepsilon$ for all i. If $a_i = a_{i-1}$, then this implies that $(v_i - v_{i-1})^2 \leq 2^{|A|+2}\varepsilon$ (using the L_2^2 triangle inequality), which contradicts the upper bound, $(v_i - v_{i-1})^2 = 4\hat{f}(i)^2 \geq 4m\varepsilon = 2^{|A|+3}\varepsilon$. 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 13.2. *Fix a finite set A, and let m be the constant from Lemma 13.1. There exists a constant $\varepsilon_0 > 0$ (depending on A) such the following holds for all $\varepsilon \leq \varepsilon_0$.*

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

Proof. Markov's inequality shows that each of the events $(f - \mathbb{E}[f])^2 \leq 3(2+m)\varepsilon$ and $\mathrm{dist}(f, A)^2 \leq 3\varepsilon$ 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 $a \in A$,

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

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

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

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

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

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

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

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

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

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

We now carry out the induction.

Lemma 13.3. *Fix a finite set A, and let m, ε_0 be the constants from Lemma 13.2. The following holds for all $\varepsilon \le \varepsilon_0$.*

Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree 1 function satisfying $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$, 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 $J = \{1, \dots, N\}$ 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 $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$, there must exist an assignment α to $x_1, \dots, x_i$ such that $\mathbb{E}[\mathrm{dist}(f|_\alpha, A)^2] \le \varepsilon$. Then $g = f|_\alpha$ satisfies $\mathbb{E}[\mathrm{dist}(g, A)^2] \le \varepsilon$ and $\mathbb{V}[g] \le (2 + m)\varepsilon$. Lemma 13.2 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 13.4. *For every finite set A and every x, y we have $(x - \mathrm{round}(y, A))^2 = O((x - y)^2 + \mathrm{dist}(x, A)^2)$.*

Proof. Let $a = \mathrm{round}(x, A)$ and $b = \mathrm{round}(y, A)$. If $a = b$ then $(x - b)^2 = (x - a)^2 = \mathrm{dist}(x, A)^2$. 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 13.5. *Fix a finite set A, and let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree 1 function satisfying $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$. 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 13.3. Suppose first that $\varepsilon \le \varepsilon_0$. The lemma defines a set J of size at most $|A| - 1$ (according to Lemma 13.1) such that $h := \hat{f}(\emptyset) + \sum_{i \in J} \hat{f}(i) x_i$ satisfies $\|f - h\|^2 \le 2\varepsilon$. Let $g = \mathrm{round}(h, A)$, which also depends only on the coordinates in J. Lemma 13.4 shows that $\|f - g\|^2 = O(\|f - h\|^2 + \mathbb{E}[\mathrm{dist}(f, A)^2]) = O(\varepsilon)$.

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 $\mathbb{E}[f^2] \le 2\,\mathbb{E}[\mathrm{round}(f, A)^2] + 2\,\mathbb{E}[\mathrm{dist}(f, A)^2] = O_A(1 + \varepsilon)$. 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 13.6. *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 $\mathbb{E}[\mathrm{dist}(f, A)^2] \le \mathbb{E}[(f - F)^2] = \varepsilon$. 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 $\Pr[F \ne g] = \mathbb{E}[1_{F \ne g}] = O_A(\mathbb{E}[(F - g)^2]) = O_A(\varepsilon)$. □

13.2 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 13.7. *Fix a finite set A and a degree d. Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be a degree d function satisfying $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$. 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 13.6.

Corollary 13.8. *Fix a finite set A and a degree d. Let $F\colon \{\pm 1\}^n \to A$ be a degree d function satisfying $\mathbb{E}[\operatorname{dist}(f,A)^2] = \varepsilon$. 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 $\Pr[F \neq g] = O_{A,d}(\varepsilon)$.*

The theorem clearly holds when $d = 0$ (take $g = \operatorname{round}(f,A)$), and it holds for $d = 1$ due to Theorem 13.5. Consider now $d > 1$. Assuming Theorem 13.7 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 $\mathbb{E}[\operatorname{dist}(f,A)^2] = \varepsilon$. As in the proof of Theorem 13.5, 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 13.9. *We have $\|f\|^2 = O_A(1)$.*

Proof. The L_2^2 triangle inequality shows that

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

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

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

Claim 13.10. *For all S,*

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

Proof. We have

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

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 13.11. *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

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

We complete the proof using Claim 13.9. $\square$

For each S, y, we apply Theorem 13.7 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 13.12. *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 13.10. □

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

Claim 13.13. *There exists a finite set C (depending only on A, d) such that for all $S \subseteq [n]$ and $T \in \binom{S}{d-1}$,*

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

Proof. Theorem 13.5, 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 $\|h_{S,T} - u_{S,T}\|^2 = O_{A,d}(\mathbb{E}[\mathrm{dist}(h_{S,T}, B)^2])$. 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 13.14. *We have*

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

Proof. Summing over T in Claim 13.13 and using Claim 13.12, we get that for all $S \subseteq [n]$,

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

Taking expectation with respect to $S \sim \mu_\delta$, where $\delta = \varepsilon^{1/d}$, Claim 13.11 shows that

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

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

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

This claim prompts defining

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

Claim 13.15. *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 13.14 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 13.9, 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. $\square$

The next step is an application of Theorem 13.7 for degree $d-2$.

Claim 13.16. *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 13.14, and so the L_2^2 triangle inequality shows that $\mathbb{E}[\mathrm{dist}(\tilde{f}, A)^2] \le 2\,\mathbb{E}[\mathrm{dist}(f, A)^2] + 2\|f - \tilde{f}\|^2 = O_{A,d}(\varepsilon + \varepsilon^{1/d}) = O_{A,d}(\varepsilon^{1/d})$ (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 $\mathbb{E}[\mathrm{dist}(f^{<d-1}, E)^2] = O_{A,d}(\varepsilon^{1/d})$.

Applying Theorem 13.7 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 $\|f - (g+h)\|^2 = O_{A,d}(\varepsilon^{1/d})$. $\square$

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

Claim 13.17. *We have $\|f - (g+h)\|^2 = O_{A,d}(\varepsilon)$.*

Proof. Let $s := f - (g+h)$. Since $\mathbb{E}[\mathrm{dist}(f, A)^2] = \varepsilon$ and $g + h$ is $(D+E)$-valued (where $D+E$ is the Minkowski sum), we see that $\mathbb{E}[\mathrm{dist}(s, V)^2] \le \varepsilon$, 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 $\mathrm{round}(s, V) = 0$ or $\mathrm{round}(s, V) = \Omega_A(1)$. Hence

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

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 13.16, 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 13.7. Let $r = \mathrm{round}(g + h, A)$, and note that r depends on $O_{A,d}(1)$ coordinates. Lemma 13.4 shows that $\|f - r\|^2 = O(\|f - (g+h)\|^2 + \mathbb{E}[\mathrm{dist}(f, A)^2]) = O_{A,d}(\varepsilon)$. 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 13.5, in this case $\|f - a\|^2 = O_{A,d}(\varepsilon)$ for any $a \in A$. $\square$

References

[AKK+05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. *Testing Reed-Muller codes*. IEEE Trans. Inform. Theory, 51(11):4032–4039, 2005. (Preliminary version in *7th RANDOM*, 2003). 5, 32

[BCH+96] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan. *Linearity testing in characteristic two*. IEEE Trans. Inform. Theory, 42(6):1781–1795, 1996. (Preliminary version in *36th FOCS*, 1995). 32

[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. 6

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

[BK18] Itai Benjamini and Gil Kalai. *Around two theorems and a lemma by Lucio Russo*. Mathematics and Mechanics of Complex Systems, 6(2), 2018. 6

[BKS+10] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. *Optimal testing of Reed-Muller codes*. In *Proc. 51st IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 488–497. 2010. arXiv:0910.0641, eccc:2009/TR09-086. 5, 32

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. *Self-testing/correcting with applications to numerical problems*. J. Comput. Syst. Sci., 47(3):549–595, December 1993. (Preliminary version in *22nd STOC*, 1990). 32

[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*). 6

[CHS18] John Chiarelli, Pooya Hatami, and Michael Saks. *Tight bound on the number of relevant variables in a bounded degree Boolean function*, 2018. (manuscript). arXiv:1801.08564. 6, 9

[DDG+17] Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. *Direct sum testing*. SIAM J. Comput., 46(4):1336–1369, 2017. (Preliminary version in *6th ITCS*, 2015). eccc:2014/TR14-002. 5

[DFH17a] Irit Dinur, Yuval Filmus, and Prahladh Harsha. *Agreement tests on graphs and hypergraphs*, 2017. (manuscript). arXiv:1711.09426, eccc:2017/TR17-181. 1

[DFH17b] ———. *Low degree almost Boolean functions are sparse juntas*, 2017. (manuscript). arXiv:1711.09428, eccc:2017/TR17-180. 1

[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. 11, 12

[DKK+18a] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. *On non-optimally expanding sets in Grassmann graphs*. In *Proc. 50th ACM Symp. on Theory of Computing (STOC)*, pages 940–951. 2018. eccc:2017/TR17-094. 6

[DKK+18b] ———. *Towards a proof of the 2-to-1 games conjecture?* In *Proc. 50th ACM Symp. on Theory of Computing (STOC)*, pages 376–389. 2018. eccc:2016/TR16-198. 6

[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. 2, 8, 11, 12, 13

[Fil16] Yuval Filmus. *Friedgut-Kalai-Naor theorem for slices of the Boolean cube*. Chic. J. Theoret. Comput. Sci., 2016(14), 2016. arXiv:1410.7834. 3, 45

[FK96] Ehud Friedgut and Gil Kalai. *Every monotone graph property has a sharp threshold*. Proc. Amer. Math. Soc., 124(10):2993–3002, 1996. 6

[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. 2, 6

[Fri98] Ehud Friedgut. *Boolean functions with low average sensitivity depend on few coordinates*. Combinatorica, 18(1):27–35, 1998. 6

[Fri99] ———. *Sharp thresholds of graph properties, and the k-SAT problem*. J. Amer. Math. Soc., 12(4):1017–1054, 1999. 6

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. *Property testing and its connection to learning and approximation*. J. ACM, 45(4):653–750, July 1998. (Preliminary version in *37th FOCS*, 1996). eccc:1996/TR96-057. 7

[GLR+91] Peter Gemmell, Richard J. Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. *Self-testing/correcting for polynomials and for approximate functions*. In *Proc. 23rd ACM Symp. on Theory of Computing (STOC)*, pages 32–42. 1991. 11

[Hat12] Hamed Hatami. *A structure theorem for Boolean functions with small total influences*. Ann. of Math., 176(1):509–533, 2012. 4, 6

[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. 11, 12

[Kin03] Guy Kindler. *Property Testing, PCP and Juntas*. Ph.D. thesis, Tel Aviv University, 2003. eccc:2016/TR16-007. 2, 45

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. *The influence of variables on Boolean functions (extended abstract)*. In *Proc. 29th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 68–80. 1988. 6

[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. 7

[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. 6

[KMMS18] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. *Small set expansion in the Johnson graph*, 2018. (manuscript). eccc:2018/TR18-078. 7

[KMS17] Subhash Khot, Dor Minzer, and Muli Safra. *On independent sets, 2-to-2 games, and Grassmann graphs*. In *Proc. 49th ACM Symp. on Theory of Computing (STOC)*, pages 576–589. 2017. eccc:2016/TR16-124. 6

[KMS18] ———. *Pseudorandom sets in Grassmann graph have near-perfect expansion*. In *Proc. 59th IEEE Symp. on Foundations of Comp. Science (FOCS)*. 2018. (To appear). eccc:2018/TR18-006. 6

[KS02] Guy Kindler and Shmuel Safra. *Noise-resistant Boolean functions are juntas*, 2002. (manuscript). 2

[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). 2, 6, 9

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi:10.1017/CBO9781139814782. 6, 9

[RS96] Ronitt Rubinfeld and Madhu Sudan. *Robust characterizations of polynomials with applications to program testing*. SIAM J. Comput., 25(2):252–271, April 1996. (Preliminary version in *23rd STOC*, 1991 and *3rd SODA*, 1992). 11