AND Testing and Robust Judgement Aggregation

Yuval Filmus*
filmus.yuval@gmail.com
Department of Computer Science, Technion
Israel

Noam Lifshitz
noamlifshitz@gmail.com
Einstein Institute of Mathematics
Hebrew University of Jerusalem.
Israel

Dor Minzer†
minzer.dor@gmail.com
Institute for Advanced Study
Princeton, NJ, USA

Elchanan Mossel‡
elmos@mit.edu
Department of Mathematics and IDSS, MIT
Cambridge, MA, USA

ABSTRACT

A function $f: \{0,1\}^n \to \{0,1\}$ is called an approximate AND-homomorphism if choosing $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ uniformly at random, we have that $f(\mathbf{x} \wedge \mathbf{y}) = f(\mathbf{x}) \wedge f(\mathbf{y})$ with probability at least $1 - \varepsilon$, where $x \wedge y = (x_1 \wedge y_1, \ldots, x_n \wedge y_n)$. We prove that if $f: \{0,1\}^n \to \{0,1\}$ is an approximate AND-homomorphism, then f is δ-close to either a constant function or an AND function, where $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. This improves on a result of Nehama, who proved a similar statement in which δ depends on n.

Our theorem implies a strong result on judgement aggregation in computational social choice. In the language of social choice, our result shows that if f is ε-close to satisfying judgement aggregation, then it is $\delta(\varepsilon)$-close to an oligarchy (the name for the AND function in social choice theory). This improves on Nehama's result, in which δ decays polynomially with n.

Our result follows from a more general one, in which we characterize approximate solutions to the eigenvalue equation $\mathrm{T}f = \lambda g$, where T is the downwards noise operator $\mathrm{T}f(x) = \mathbb{E}_{\mathbf{y}}[f(x \wedge \mathbf{y})]$, f is $[0,1]$-valued, and g is $\{0,1\}$-valued. We identify all exact solutions to this equation, and show that any approximate solution in which $\mathrm{T}f$ and λg are close is close to an exact solution.

CCS CONCEPTS

• **Theory of computation** → *Streaming, sublinear and near linear time algorithms.*

KEYWORDS

Property Testing, Analysis of Boolean Functions

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 802020-ERC-HARMONIC.
†Supported partially by NSF grant CCF-1412958 and Rothschild Fellowship.
‡Supported by NSF Award DMS-1737944, by AROMURI W911NF1910217 and by a Simons Investigator Award in Mathematics (622132).

STOC '20, June 22–26, 2020, Chicago, IL, USA

ACM ISBN 978-1-4503-6979-4/20/06...$15.00
https://doi.org/10.1145/3357713.3384254

ACM Reference Format:
Yuval Filmus, Noam Lifshitz, Dor Minzer, and Elchanan Mossel. 2020. AND Testing and Robust Judgement Aggregation. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC '20), June 22–26, 2020, Chicago, IL, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3357713.3384254

1 INTRODUCTION

Which functions $f: \{0,1\}^n \to \{0,1\}$ satisfy

$$f(\mathbf{x} \wedge \mathbf{y}) = f(\mathbf{x}) \wedge f(\mathbf{y}) \text{ w.p. } 1 - \varepsilon,$$

where $\mathbf{x}, \mathbf{y}$ are chosen uniformly at random?

If $\varepsilon = 0$, it is not hard to check that f is either constant or an AND of a subset of the coordinates. Nehama [40] showed that when $\varepsilon > 0$, f must be $O((n\varepsilon)^{1/3})$-close to a constant function or to an AND (in other words, $\Pr[f \neq g] = O((n\varepsilon)^{1/3})$, where g is constant or an AND). The main result in this paper implies, as a corollary, a similar statement, in which the distance between f, g vanishes with ε, without any dependence on n.

Theorem 1.1. *For each $\delta > 0$ there is $\varepsilon > 0$ such that if a function $f: \{0,1\}^n \to \{0,1\}$ satisfies*

$$\Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{x} \wedge \mathbf{y}) = f(\mathbf{x}) \wedge f(\mathbf{y})] \geqslant 1 - \varepsilon,$$

then f is δ-close to a constant or an AND.

Our technique is in fact more general, and allows us to study the multi-function version of this problem, in which we are interested in triples of functions $f, g, h: \{0,1\}^n \to \{0,1\}$ that satisfy $f(\mathbf{x} \wedge \mathbf{y}) = g(\mathbf{x}) \wedge h(\mathbf{y})$ with probability at least $1 - \varepsilon$. Quantitatively, we get a quasi-polynomial relationship between ε and δ, and more precisely $\varepsilon = \exp(-\Theta(\log^2(1/\delta)))$, but we expect the correct dependency to be a polynomial.

Theorem 1.1 gives shows that the natural test above is a one-sided error tester for the property of being an AND function. This question was first considered in [43], wherein the authors propose this test; being unable to analyze it the authors consider a more complicated tester and analyze it, but it comes at the cost of being less natural and having more queries as well as two sided error.

If we replace $\wedge$ with $\oplus$ in the above problem, then the result corresponding to Theorem 1.1 is the well-known soundness of the Blum–Luby–Rubinfeld *linearity test* [7], that plays an important role in the construction of PCPs [1, 2, 24]. By now, many proofs for the soundness of this test are known: self-correction [7], Fourier analysis [6, 24], induction [12]. Unfortunately, all of these proofs rely on $\oplus$ being a group operation (either directly or via Fourier analysis), and hence do not extend to our setting.

Our approach recasts the problem as determining the approximate eigenfunctions of a one-sided noise operator. Define the operator T acting on functions $f\colon \{0,1\}^n \to \{0,1\}$ in the following way:

$$(\mathrm{T}f)(x) = \mathop{\mathbb{E}}_{\mathbf{y}}[f(x \wedge \mathbf{y})].$$

Using this operator, the premise of Theorem 1.1 implies that f is an approximate eigenfunction of this operator, i.e. $\mathrm{T}f \approx \lambda f$, where $\lambda = \mathbb{E}[f]$ is the average of f. Here, by approximate solution we mean that the L_1 distance between the two functions is small: $\mathbb{E}_{\mathbf{x}}[|\mathrm{T}f(\mathbf{x}) - \lambda f(\mathbf{x})|] \leqslant \varepsilon$.

If we replace T with the usual two-sided noise operator T_ρ, then a short spectral argument shows that if f is an approximate eigenfunction then it must be close to an exact eigenfunction of T_ρ. Unfortunately, the spectral argument relies on orthogonality of eigenspaces of T_ρ, a property which T doesn't satisfy (its eigenspaces are spanned by ANDs, which aren't orthogonal). Indeed, T has approximate eigenfunctions beyond ANDs, and below we give two examples f_1 and f_2.

$$f_1(x) = \begin{cases} x_1 \vee x_2 & \text{if } |x| \geqslant n/3, \\ x_1 \oplus x_2 & \text{if } |x| < n/3 \end{cases} \qquad f_2(x) = \begin{cases} 1 & \text{if } |x| \geqslant n/3, \\ \mathrm{Ber}(\lambda) & \text{if } |x| < n/3 \end{cases}$$

In f_2, we stress that the function is defined probabilistically: for each x such that $|x| < n/3$ independently, we take $f_2(x) = 1$ with probability $\lambda < 1$. It is easy to verify by case analysis that we have that $\mathrm{T}f_1 \approx \frac{1}{2} f_1$, $\mathrm{T}f_2 \approx \lambda f_2$. Note however, that these functions are not counter-examples to the AND-test above, since both of them pass the test with some constant probability bounded away from 1.

Note that each one of the functions f_1, f_2 is essentially composed of two, completely different "sub-functions": one defined on high Hamming-weight inputs, and another defined on low Hamming-weight inputs. This suggests decoupling the two functions, and considering the generalized eigenvalue problem

$$\mathrm{T}f = \lambda g, \text{ where } f\colon \{0,1\}^n \to [0,1] \text{ and } g\colon \{0,1\}^n \to \{0,1\}.$$

Here f represents the low-weight part, and g represents the high-weight part. We represent the probabilistic aspect of the low-weight part by allowing f to take on values in the interval $[0,1]$.

The two examples above corresponds to *exact* solutions of this generalized problem: $\mathrm{T}(x_1 \oplus x_2) = \frac{1}{2}(x_1 \vee x_2)$ and $\mathrm{T}\lambda = \lambda \cdot 1$. Therefore, as a prerequisite to characterizing approximate eigenfunctions of T, we must first study exact solutions to the more general two-function version. We show:

Theorem 1.2. *If $f\colon \{0,1\}^n \to [0,1]$ and $g\colon \{0,1\}^n \to \{0,1\}$ satisfy $\mathrm{T}f = \lambda g$ then either $f = g = 0$ or there exist disjoint subsets $S_1, \dots, S_m \subseteq [n]$, where $m \leqslant \log_2(1/\lambda)$, such that*

$$f(x) = \bigwedge_{i=1}^{m} \bigoplus_{j \in S_i} x_i, \qquad g(x) = \bigwedge_{i=1}^{m} \bigvee_{j \in S_i} x_i.$$

Moreover, if f is monotone then g is an AND and $f = g$.

Thus if $\mathrm{T}f = \lambda g$ then g is an "AND-OR" and f is the corresponding "AND-XOR" (or $f = g = 0$). When f is monotone, g must be an AND, and so $f = g$. Using Theorem 1.2, we can then actually solve the more general problem of characterizing approximate solutions to the equation $\mathrm{T}f = \lambda g$.

Theorem 1.3. *If $f\colon \{0,1\}^n \to [0,1]$ and $g\colon \{0,1\}^n \to \{0,1\}$ satisfy $\mathrm{T}f \approx \lambda g$ then either $f \approx g \approx 0$ or g is close to an AND-OR and f is close to an AND-XOR.*

Moreover, if f is monotone then f, g are both close to a constant or to an AND.

We also show how to deduce Theorem 1.1 from Theorem 1.3. We remark that Theorem 1.3 is stated in a somewhat informal way: the closeness of the function f to an AND-XOR function has to be stated in a more subtle way (since otherwise it is false), and we defer this point to the formal statement of the theorems in Section 2.

1.1 Other Variants

Other noise rates. Nehama [40] also considers the more general equation

$$f(\mathbf{x_1} \wedge \cdots \wedge \mathbf{x_m}) \approx f(\mathbf{x_1}) \wedge \cdots \wedge f(\mathbf{x_m}),$$

where each one of $\mathbf{x_1}, \dots, \mathbf{x_m}$ is sampled uniformly and independently from $\{0,1\}^n$. We can reduce this problem, in a similar manner, to an eigenfunction of an appropriate operator $\mathrm{T}^{(m)}$, defined by

$$\mathrm{T}^{(m)} f(x) = \mathop{\mathbb{E}}_{\mathbf{y_1}, \dots, \mathbf{y_{m-1}}} f(x \wedge \mathbf{y_1} \wedge \dots \wedge \mathbf{y_{m-1}}).$$

Our techniques also apply to such operators (and in fact to a slightly richer family of noise operators), and we prove variants of Theorem 1.2 and Theorem 1.3 in this case as well:

Theorem 1.4. *Let $m > 2$. If $f\colon \{0,1\}^n \to [0,1]$ and $g\colon \{0,1\}^n \to \{0,1\}$ satisfy $\mathrm{T}^{(m)} f = \lambda g$ then either $f = g = 0$ or $f = g$ is an AND.*

Furthermore, if $\mathrm{T}^{(m)} f \approx \lambda g$ then either $f \approx g \approx 0$ or f, g are close to an AND.

One-sided error version. Finally, we consider the one-sided error version of the equation $\mathrm{T}f = \lambda g$. That is, suppose we have a bounded function $f\colon \{0,1\}^n \to [0,1]$, and a Boolean function g, such that with probability $1 - \varepsilon$ over $\mathbf{x}$: (a) if $g(\mathbf{x}) = 1$, then $f(\mathbf{x} \wedge \mathbf{y}) \geqslant \lambda$ with constant probability over $\mathbf{y}$, and (b) if $g(\mathbf{x}) = 0$ then $f(\mathbf{x} \wedge \mathbf{y}) \leqslant \varepsilon$ with probability $1 - \varepsilon$ over $\mathbf{y}$.

We note that this condition is a relaxation of the approximate eigenvalue condition. In this case, we prove a weaker structural result than in Theorem 1.3, namely that g is close to a monotone junta.

Theorem 1.5. *Suppose that the functions $f\colon \{0,1\}^n \to [0,1]$ and $g\colon \{0,1\}^n \to \{0,1\}$ satisfy the following condition: when $g = 0$, $\mathrm{T}f$ is typically small; and when $g = 1$, $\mathrm{T}f$ is typically at least λ.*[1] *Then g is close to a monotone Boolean junta.*[2]

We remark that while the structural result in this case is weaker, it is for a good reason: for any monotone junta f, choosing $g = f$ yields an approximate, one-sided error solution.

[1] In contrast, in Theorem 1.3 we ask that $\mathrm{T}f$ be typically *close* to λ.
[2] A junta is a function depending on a constant number of coordinates.

1.2 Social Choice Interpretations of Approximate Eigenfunctions

The seminal work of Kornhauser and Sager [32] discusses a situation where three cases A, B, C are considered in court, and by law, one should rule against C if and only if there is a ruling against both A and B. When several judges are involved, their opinions should be aggregated using a function f that preserves this law, that is, satisfies $f(x \wedge y) = f(x) \wedge f(y)$; we say that f is an *AND-homomorphism.* List and Pettit [34, 35] showed that the only non-constant aggregation functions that are AND-homomorphisms are the AND functions, known in the social choice literature as *oligarchies.*

Let the individual opinions of the judges be $x_1, \dots, x_n$ on A, $y_1, \dots, y_n$ on B, and $x_1 \wedge y_1, x_2 \wedge y_2, \dots, x_n \wedge y_n$ on C. The characterization of robust judgement aggregation that we prove in this paper (Theorem 1.1) states that if typically $f(x \wedge y) = f(x) \wedge f(y)$, then f is close to an oligarchy.

The characterization in terms of approximate eigenfunctions (Theorem 1.3) actually shows more. Suppose that opinions are aggregated according to a monotone function f which satisfies the following two conditions:

- There is rarely a ruling against C unless there is a ruling against A and a ruling against B.
- Suppose that there is a ruling against A. If there is also a ruling against B, then with probability roughly q, there is a ruling against C.
  (Formally, for typical x such that $f(x) = 1$, we have $\Pr[f(x \wedge y) = 1 \mid f(y) = 1] \approx q$.)

Then f must be close to an oligarchy or to a constant function, and $q \approx 1$.

In fact, the second condition can be weakened significantly:

- Suppose that there is a ruling against A. Then with probability roughly λ, there is a ruling against C.

Theorem 1.3 implies that f must be close to an oligarchy or to a constant function, and $\lambda \approx \mathbb{E}[f]$.

Similarly, Theorem 1.5 shows that f has to be close to a monotone junta if the second condition above is replaced with either of the following two conditions:

- Suppose that there is a ruling against A. If there is also a ruling against B, then with probability *at least* q, there is a ruling against C.
- Suppose that there is a ruling against A. Then with probability *at least* λ, there is a ruling against C.

Thus our results do not only strengthen robust judgement aggregation in a quantitative way, but also in a qualitative way.

1.3 Our Techniques

Our main result, Theorem 1.1, easily follows from Theorem 1.3, which is our main technical result. Below we sketch the proof idea of Theorem 1.3 (the proofs of Theorem 1.4 and Theorem 1.5 follow similar lines).

Suppose f, g are functions as in Theorem 1.3 that satisfy $\mathrm{T}f \approx \lambda g$. The first step of the proof is to show that the function g is close to a junta h, i.e. to a function depending only on constantly many variables. To get some intuition for that, note that if T was the standard noise operator, then the function $\mathrm{T}f$ has exponentially decreasing tail and hence it is very concentrated on its low Fourier levels. When the operator T is the one-sided noise operator, one can actually use similar reasoning to claim that g again has an exponentially decaying tail (as observed by Lifshitz [33]). Since g is Boolean and $g \approx \frac{1}{\lambda}\mathrm{T}f$, this observation would then allow us to use structure theorems on Boolean functions (more specifically, a result of Bourgain [8] or of Kindler and Safra [31]) to conclude that g is (close to) a junta.

Thus, ignoring some (important) technical details, one can think of g as a function of a constant number of variables, and since the proximity parameter between $\mathrm{T}f$ and g can be taken to be very small (even in comparison to the number of variables g depends on), one may as well think of it as being 0. In other words, the problem essentially boils down to studying exact solutions to the equation $\mathrm{T}f = g$ when n is constant, which is where Theorem 1.2 enters the picture. Using it, we prove the structural result on g; getting the structural result on f then amounts to averaging f over coordinates that g does not depend on (since those could be thought of as a "source of randomness" as in the example of $f_2(x)$ above), and then inverting the operator T acting on functions of a constant number of variables.

This ends the informal description of our techniques. We remark that actually composing the two components, namely the approximation by junta and the solution to the exact equation, is more subtle and requires some care. We also remark that in the case of one-sided error (Theorem 1.5), the Fourier-analytic argument alluded to above, which implies that g is close to a junta, does not seem to be applicable. We thus present an alternative, more combinatorial argument that captures this case as well.

1.4 Related Work

1.4.1 Quantitative social choice theory. Social choice theory studies how to aggregate the opinion of a number of agents. Already in the 18th century, Condorcet [11] noted that natural aggregation rules often result in paradoxes. A large body of work has been developed in economics since the middle of the 20th century, in which it was shown that natural aggregation tasks have *no* good aggregation functions. The two most famous results in this area are Arrow's impossibility theorem [3, 4] and the Gibbard–Satterthwaite (GS) manipulation theorem [21, 45]. The questions of aggregation re-emerged in the context of multi-agent systems in computer science, where the hope was that either the probability of paradoxical outcome is small, or there is computational difficulty in arriving at a paradoxical outcome, see e.g. [5] and the survey [18]. A sequence of results showed that this is not the case by proving strong and general quantitative versions of both Arrow's Theorem [28, 29, 37, 38] and the GS Theorem [19, 20, 27, 39], as well as results interpolating the two theorems [17].

The main motivation for the problem discussed in this paper is Judgement Aggregation. This problem is considered in a fascinating paper in the Yale Law Review by Kornhauser and Sager [32]. In particular, toward the end of the paper, the authors considered legal cases, where the judgement aggregation function f should satisfy $f(x \wedge y) = f(x) \wedge f(y)$. They observe that this does not hold when f is the majority vote on three opinions.

The failure of Majority, which mirrors the failure of Majority in ranking that was observed by Condorcet, led to work by List and Pettit [34, 35], who characterized exactly the functions f that are AND-homomorphisms, i.e., oligarchies. The question of judgement aggregation has attracted much attention in philosophy, social epistemology, and artificial intelligence [44]. In the context of multi-agent systems, when the number of agents is large, it makes sense to ask if it is possible to achieve approximate judgment aggregation. Our results show that this can only be achieved in the obvious way, i.e., by almost-oligarchies.

We note that the study of judgement aggregation extends well beyond AND-homomorphisms, to other types of homomorphisms, and indeed such a theory of polymorphisms is well-developed [13–16, 41, 46]. We leave if for future work to investigate robust versions of these results.

1.4.2 Property testing. The work of Blum, Luby and Rubinfeld [7] has been extended to more general settings by various authors. For example, Moore and Russell [36] and Gowers and Hatami [23] considered approximate representations of finite groups. Other authors had considered infinite groups, see for example the survey of Hyers and Rassias [26]. Theorem 1.1 generalizes Blum–Luby–Rubinfeld in a different direction, to approximate polymorphisms, where there is no group structure.

We remark that Theorem 1.1 implies that the soundness of a property testing algorithm of Parnas, Ron and Samorodnitsky [43], whose goal is to test whether the input function is a dictatorship or more generally an AND function. I.e., the tester should accept with high probability (preferably 1) if the function if an AND function, and reject with probability at least $\frac{2}{3}$ if the function is ε-far from all AND functions. The authors proposed the following natural tester, which they were unable to analyze: test that f has expectation $1/2$ and satisfies $f(x \wedge y) = f(x) \wedge f(y)$. Instead, they proposed a somewhat less natural tester. Our results imply that their original tester also works.

We remark that can design a tester that only queries f on points distributed according to $\mu_{1/2}$ at the cost of introducing an additional query. The tester samples $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w} \sim \mu_{1/2}$ conditioned on $\mathbf{x} \wedge \mathbf{y} = \mathbf{z} \wedge \mathbf{w}$, and checks that $f(\mathbf{x}) \wedge f(\mathbf{y}) = f(\mathbf{z}) \wedge f(\mathbf{w})$. The soundness of this tester follows immediately from the soundness of the current tester by redefining for $a \sim \mu_{1/4}$ the value of $f(a)$ to be the prominent value of $f(\mathbf{x}) \wedge f(\mathbf{y})$ where $\mathbf{x}, \mathbf{y} \sim \mu_{1/2}$ are sampled so that $\mathbf{x} \wedge \mathbf{y} = a$.

It is interesting to explore if there is a relationship between our results and different notions of approximate polymorphisms that appear in the literature [9, 10], which were used to prove hardness of approximation results.

Organization. We formally state our results in Section 2. After some preliminaries in Section 3, we prove the various results in Sections 4, deferring the proof of the rest of the results to the full version of the paper. We close the paper by stating some open questions in Section 5.

2 MAIN RESULTS

Let μ_p denote the p-biased measure on $\{0,1\}^n$. Let $L_2(\{0,1\}^n, \mu_p)$ be the space of real-valued functions on $\{0,1\}^n$ equipped with the inner product $\langle f, g \rangle = \mathbb{E}_{\mathbf{x} \sim \mu_p}[f(\mathbf{x})g(\mathbf{x})]$.

Definition 2.1. *For $q \leqslant p$, the distribution $(\mathbf{y}, \mathbf{x}) \sim \mathbb{D}(q,p)$ over $\{0,1\}^n \times \{0,1\}^n$ is the distribution in which for each $i \in [n]$ independently, we have* $\Pr[\mathbf{y}_i = 1] = q$ *and* $\Pr[\mathbf{x}_i = 1] = p$, *and always* $\mathbf{y}_i \leqslant \mathbf{x}_i$.

One way to generate inputs $(\mathbf{y}, \mathbf{x}) \sim \mathbb{D}(q,p)$ that will be useful for us is as follows. Sample $\mathbf{x} \sim \mu_p$ and $\mathbf{z} \sim \mu_{q/p}$ independently, and output $(\mathbf{x} \wedge \mathbf{z}, \mathbf{x})$. (Here $\wedge$ refers to the coordinatewise AND operation, i.e. for each $i \in [n]$ we have $(\mathbf{x} \wedge \mathbf{z})_i = \mathbf{x}_i \wedge \mathbf{z}_i$.)

For $\rho \in (0,1)$, define the one-sided operator $\mathrm{T}^{\downarrow}_{p,\rho p}$ as follows. For any function $f : (\{0,1\}^n, \mu_{\rho p}) \to \{0,1\}$, we define the function $\mathrm{T}^{\downarrow}_{p,\rho p} f : (\{0,1\}^n, \mu_p) \to \{0,1\}$ by

$$\mathrm{T}^{\downarrow}_{p,\rho p} f(x) = \mathbb{E}_{(\mathbf{y},\mathbf{v}) \sim \mathbb{D}(\rho p, p)}[f(\mathbf{y}) \mid \mathbf{v} = x].$$

Equivalently, we have $\mathrm{T}^{\downarrow}_{p,\rho p} f(x) = \mathbb{E}_{\mathbf{z} \sim \mu_\rho}[f(x \wedge \mathbf{z})]$.

Next, we shall discuss the spectrum (eigenvectors and eigenvalues) of the operator $\mathrm{T}^{\downarrow}_{p,\rho p}$. We remark that throughout this section, the parameters p and ρ should be thought of as constants bounded away from 0, 1.

For each $S \subseteq [n]$, the function $\mathrm{AND}_S : \{0,1\}^n \to \{0,1\}$ defined by $\mathrm{AND}_S(x) = \bigwedge_{i \in S} x_i$ is an eigenvector of $\mathrm{T}^{\downarrow}_{p,\rho p}$ with eigenvalue $\rho^{|S|}$ (we omit the easy proof). Moreover, these are the only eigenvectors of $\mathrm{T}^{\downarrow}_{p,\rho p}$ that are Boolean valued.[3] Our goal in this paper is to find a robust version of this characterization of the Boolean eigenvectors of $\mathrm{T}^{\downarrow}_{p,\rho p}$.

We say that a function f is an η-approximate eigenvector with eigenvalue λ, if $\|\mathrm{T}^{\downarrow}_{p,\rho p} f - \lambda f\|_1 \leqslant \eta$ (here and throughout the paper we will consider the ℓ_1 norm with respect to the μ_p measure). What can be said about the structure of Boolean, approximate eigenvectors of $\mathrm{T}^{\downarrow}_{p,\rho p}$? A natural conjecture would be that any such function has to be close to an exact eigenvector, which by Booleanity would have to be an AND-function over $\approx \log_\rho(\lambda)$ variables. However, this conjecture turns out to be false, as the following example demonstrates.

Set $p = \rho = \frac{1}{2}$, and consider the function f defined by $f(x) = x_1 \vee x_2$ for inputs whose hamming weight is $n/2 \pm \sqrt{n \log n}$, and by $x_1 \oplus x_2$ for the rest of the inputs. It is easy to see that f is far from any AND function on the $\mu_{1/2}$ measure, and we argue that $\|\mathrm{T}^{\downarrow}_{1/2,1/4} f - \frac{1}{2} f\| = o(1)$. By definition, for each x, $\mathrm{T}^{\downarrow}_{1/2,1/4} f(x)$ is the probability that picking $\mathbf{z} \sim \mu_{1/2}$, we have $f(x \wedge \mathbf{z}) = 1$. Except with probability $o(1)$, the hamming weight of $x, x \wedge \mathbf{z}$ is roughly $n/2, n/4$ respectively, and we focus only on this event. We now consider two cases depending on the value $f(x)$, and analyze each one of them separately. In case $f(x) = 0$, we get by definition that $x_1 = x_2 = 0$ so that $(x \wedge z)_1 = (x \wedge z)_2 = 0$ for any choice of z and in particular $f(x \wedge \mathbf{z}) = 0$. Otherwise, if $f(x) = 1$ then $(x_1, x_2) \neq (0,0)$, and by definition $f(x \wedge \mathbf{z}) = x_1 \mathbf{z}_1 \oplus x_2 \mathbf{z}_2$, so that for each fixed x such that $(x_1, x_2) \neq (0,0)$, this is a uniform unbiased bit, and in particular $\mathbb{E}_{\mathbf{z}}[f(x \wedge \mathbf{z})] = \frac{1}{2}$.

[3] To see that, note that any function f can be written as a linear combination of AND functions, and if f is an eigenvector then all of these ANDs are of the same size, say with coefficients $\alpha_1, \ldots, \alpha_m$. Considering the value of f on the minterms of these ANDs, one concludes that all of the α's must be 1, and considering the value of f on the all-1 string, one concludes that $m = 1$.

Remark 2.2. *It is worth noting that for any constant $\lambda > 0$, there are approximate eigenvectors of $\mathrm{T}^{\downarrow}_{1/2,1/4}$ with eigenvalue λ (not only for $\lambda = 2^{-k}$). Indeed, the function f that is constantly 1 on inputs with Hamming weight $n/2 \pm \sqrt{n \log n}$, and on each other point x independently, we take $f(x) = 1$ with probability λ, is (with probability $1 - o(1)$) an approximate eigenvector with eigenvalue λ.*

2.1 The Basic Two-Function Version

Since the previous example is essentially composed of two different functions (one around the middle slice and the other around the $n/4$-slice), it makes sense to consider the two-function version of the approximate eigenvector problem. Namely, let $f: (\{0,1\}^n, \mu_{p\rho}) \to \{0,1\}$, $g: (\{0,1\}^n, \mu_p) \to \{0,1\}$, and $\lambda \in (0,1)$ be such that $\|\mathrm{T}^{\downarrow}_{p,\rho p} f - g\|_1 \leqslant \eta$. What can we say about f and g? We note that in this case, even the exact version of the problem, i.e. determining which functions can satisfy $\mathrm{T}^{\downarrow}_{p,\rho p} f = g$, is already unclear (and in fact, as it turns out, understanding solutions to the exact problem is a key step in solving the approximate problem).

The version of the problem we will consider is actually more general and allows the function f to take values in $[0,1]$. It turns out that the structure of the solutions heavily depends on ρ, and we consider three different regimes: $0 < \rho < \frac{1}{2}$, $\rho = \frac{1}{2}$ and $\frac{1}{2} < \rho < 1$. We remark that all of the results apply in particular for the original approximate eigenvector problem, i.e. the case $f = g$.

The first range, $0 < \rho < \frac{1}{2}$, is the simplest, and we have the following result.

Theorem 2.3. *For any $\zeta > 0$ there is $J \in \mathbb{N}$ such that for any $\varepsilon > 0$ there is $\eta > 0$ such that the following holds. Let $p \in [\zeta, 1-\zeta]$, $\rho \in [\zeta, \frac{1}{2} - \zeta]$ and $\lambda \in [\zeta, 1]$, and let $f: \{0,1\}^n \to [0,1]$ and $g: \{0,1\}^n \to \{0,1\}$ satisfy $\|\mathrm{T}^{\downarrow}_{p,\rho p} f - \lambda g\|_1 \leqslant \eta$. Then either f, g are ε-close to the zero function, or there is a set $T \subseteq [n]$ of size at most J such that:*

- *g is ε-close to AND_T.*
- *After averaging outside T, f is ε-close to $\rho^{-|T|}\lambda \cdot \mathsf{AND}_T$. More precisely, the function $\tilde{f}: \{0,1\}^T \to [0,1]$ given by $\tilde{f}(x) = \mathbb{E}_{\mathbf{y} \sim \mu_{\rho p}^{[n]\setminus T}} [f(\mathbf{y}, x)]$ is ε-close to $\rho^{-|T|}\lambda \cdot \mathsf{AND}_T$.*

(This range corresponds to the operators $\mathrm{T}^{(m)}$ mentioned in Theorem 1.4.)

In the second range, $\rho = 1/2$, the structure of f and g may be more complicated (we have already seen an example in this range where $g = \mathsf{OR}_T$ and $f = \mathsf{XOR}_T$ for T of size 2).

Definition 2.4. *A function $g: \{0,1\}^n \to \{0,1\}$ is called an AND-OR function of width m if there are disjoint sets $A_1, \ldots, A_m$ such that $g(x) = \bigwedge_{i \in [m]} \bigvee_{j \in A_i} x_j$.*

Definition 2.5. *A function $g: \{0,1\}^n \to \{0,1\}$ is called an AND-XOR function of width m if there are disjoint sets $A_1, \ldots, A_m$ such that $g(x) = \bigwedge_{i \in [m]} \bigoplus_{j \in A_i} x_j$.*

Theorem 2.6. *For any $\zeta > 0$ there is $m \in \mathbb{N}$ such that for any $\varepsilon > 0$ there are $\eta > 0$, $J \in \mathbb{N}$ such that the following holds for all $p \in [\zeta, 1-\zeta]$ and $\lambda \in [\zeta, 1]$. If $f: \{0,1\}^n \to [0,1]$ and $g: \{0,1\}^n \to \{0,1\}$ satisfy $\|\mathrm{T}^{\downarrow}_{p,p/2} f - \lambda g\|_1 \leqslant \eta$. Then there is a set $T \subseteq [n]$ of size at most J, and a partition $T = A_1 \cup \cdots \cup A_r$ for $r \leqslant m$ such that either f, g are ε-close to the zero function, or:*

- *g is ε-close to $\bigwedge_{i \in [r]} \bigvee_{j \in A_i} x_j$ (i.e. to an AND-OR function of width at most m).*
- *After averaging, f is close to a multiple of $\bigwedge_{i \in [r]} \bigoplus_{j \in A_i} x_j$. More precisely, the function $\tilde{f}: \{0,1\}^T \to [0,1]$ given by $\tilde{f}(x) = \mathbb{E}_{\mathbf{y} \sim \mu_{p/2}^{[n]\setminus T}} [f(\mathbf{y}, x)]$ is ε-close to $2^r \lambda \cdot \bigwedge_{i \in [r]} \bigoplus_{j \in A_i} x_j$.*

Both Theorem 2.3 and Theorem 2.6 can be shown to be qualitatively tight. For Theorem 2.6, for example, any pair of functions f, g where g is an AND-OR function and f is the corresponding AND-XOR function is an exact solution. To see that some averaging is needed to get a structure for f, note that given a pair of approximate solutions f, g, one may sub-sample f, i.e. change the value on each x such that $f(x) = 1$ with probability $1/2$, to get a new approximate solution with $\lambda/2$, and f has no apparent structure (other than being a multiple of AND-XOR after averaging).

Quantitatively, the dependence of η on ε in Theorem 2.3 is quasi-polynomial $\eta = \exp(-\Theta_\zeta(\log^2(1/\varepsilon)))$. The dependence in Theorem 2.6 is exponentially worse, i.e. $\eta = \exp(-\exp(\Theta_\zeta(\log^2(1/\varepsilon))))$. The source of this difference is that in the case of Theorem 2.3 (and also in Theorem 2.7 and Theorem 2.8) we are able to prove stronger approximation by junta results than in Theorem 2.6. Namely, we show that there is $J(\zeta)$ (independent of the proximity to junta parameter ε), such that if η is a sufficiently small function of ε, then g is ε-close to a J-junta. In the case of Theorem 2.6, we are forced to allow the size of the junta J to also depend on ε. As far as we know, in both cases the dependence of η on ε could be much better, perhaps even polynomial.

In the third range of parameters, $\frac{1}{2} < \rho < 1$, the solutions to the problem have a richer structure. It can be shown, for example, that there are $\rho \in (\frac{1}{2}, 1)$, $\lambda \in (0,1)$ and a function $f: \{0,1\}^n \to [0,1]$ such that f and $g(x) = \mathsf{Maj}(x_1, x_2, x_3)$ are an exact solution to $\mathrm{T}^{\downarrow}_{1/2,\rho/2} f = \lambda g$. In this case we only show a relatively weak structure, namely that g is close to a monotone junta (see Theorem 2.11). We remark that in order to get a stronger structure, one would only need to classify all exact solutions to the equation $\mathrm{T}^{\downarrow}_{p,\rho p} f = \lambda g$ for $\rho > 1/2$.

2.2 Special Cases

We next present our result for a few special cases of interest, in which we are able to prove a stronger structure. The first result is concerned with the case when the approximate eigenvalue is large:

Theorem 2.7. *For every $\zeta, \varepsilon > 0$ there is $\eta > 0$ such that the following holds for any $\rho, p \in [\zeta, 1-\zeta]$ and $\lambda \geqslant \rho + \zeta$. If $f: (\{0,1\}^n, \mu_{\rho p}) \to [0,1]$ and $g: (\{0,1\}^n, \mu_p) \to \{0,1\}$ satisfy $\|\mathrm{T}^{\downarrow}_{p,\rho p} f - \lambda g\|_1 \leqslant \eta$, then g is ε-close to a constant function $\Gamma \in \{0,1\}$, and the average of f according to $\mu_{\rho p}$ is close to $\lambda\Gamma$, i.e. $\left|\mathbb{E}_{\mathbf{x} \sim \mu_{\rho p}} [f(\mathbf{x})] - \lambda\Gamma\right| \leqslant \varepsilon$.*

Next, we consider the case in which f is a monotone function. In this case (and actually for a more relaxed case in which f is "almost monotone"), we show that g must be an AND function and f must be a multiple of that AND function after averaging. We also get quantitatively stronger relation between ε and η.

Theorem 2.8. *For every $\zeta > 0$, $\varepsilon > 0$ there exists $\eta > 0$ such that the following holds for all $p, \rho \in [\zeta, 1-\zeta]$ and $\lambda \in [\zeta, 1]$. If $f : (\{0,1\}^n, \mu_{\rho p}) \to [0,1]$ is monotone and $g : (\{0,1\}^n, \mu_p) \to \{0,1\}$ satisfies $\|\mathrm{T}^{\downarrow}_{p,\rho p} f - \lambda g\|_1 \leqslant \eta$, then:*

- *There exists $T \subseteq [n]$ of size at most $\lceil \log(2/\lambda) \rceil$ and a function h that is either constant (in which case $T = \emptyset$) or AND_T, such that $\|g - h\|_1 \leqslant \varepsilon$.*
- *$\tilde{f} : \{0,1\}^T \to [0,1]$ given by $\tilde{f}(x) = \mathbb{E}_{\mathbf{y} \sim \mu_{\rho p}}[f(x, \mathbf{y})]$ is ε-close in L_∞-norm to $\rho^{-|T|}\lambda \cdot h$.*

The monotonicity condition in Theorem 2.8 can be relaxed to "almost monotonicity", in the sense that flipping any coordinate from 0 to 1 cannot decrease the value of the function too much. To define this relaxation more precisely we need the notion of negative influences:

Definition 2.9. *Let $f : (\{0,1\}^n, \mu_p) \to [0,1]$ and let $i \in [n]$. The negative influence of a variable i on f, denoted by $I^-_i[f]$, is defined to be $\mathbb{E}_{\mathbf{x} \sim \mu_p}[\max(0, f(\mathbf{x}_{-i}, x_i = 0) - f(\mathbf{x}_{-i}, x_i = 1)]$, where $(x_{-i}, x_i = b)$ denoted the points y that agrees with x on all coordinates $j \neq i$, and has value b on coordinate j.*

(Note that whereas $I_i[f]$ is the average of *squared differences*, $I^-_i[f]$ is an average of *differences*.)

With this definition, Theorem 2.8 also holds when we relax the condition of monotonicity of f to the condition that all of its individual negative influences are small, i.e. $I^-_i[f] \leqslant \eta$ for all $i \in [n]$ (the proof of Theorem 2.8 in the full version achieves this stronger statement). One benefit of this relaxation is that it is able to capture the case of "judgement aggregation" as an immediate consequence.

Theorem 2.10. *For all $\zeta, \varepsilon > 0$ there is $\eta > 0$ such that the following holds for all $p, \rho \in [\zeta, 1-\zeta]$. If $f : (\{0,1\}^n, \mu_{\rho p}) \to \{0,1\}$, $g : (\{0,1\}^n, \mu_p) \to \{0,1\}$ and $h : (\{0,1\}^n, \mu_\rho) \to \{0,1\}$ satisfy $\Pr_{\mathbf{x} \sim \mu_p, \mathbf{y} \sim \mu_\rho}[f(\mathbf{x} \wedge \mathbf{y}) = g(\mathbf{x}) \wedge h(\mathbf{y})] \geqslant 1 - \eta$, then one of the following cases must happen.*

1. *f and at least one of the functions g or h are ε-close to the constant 0 function.*
2. *There is a set $T \subseteq [n]$ such that f, g, h are all ε-close to AND_T (each with respect to their input distribution).*

2.3 One-Sided Error

Finally, we consider a more relaxed version of approximate solutions to $\mathrm{T}^{\downarrow}_{p,\rho p} f = g$. We say functions $f : \{0,1\}^n \to [0,1]$ and $g : \{0,1\}^n \to \{0,1\}$ are one-sided error solutions with $\lambda > 0$ and error η if the following two conditions occur:

(1) $\mathrm{T}^{\downarrow}_{p,\rho p} f$ is very small on typical inputs x such that $g(x) = 0$:

$$\mathop{\mathbb{E}}_{\mathbf{x} \sim \mu_p}\left[(1 - g(\mathbf{x})) \cdot \mathrm{T}^{\downarrow}_{p,\rho p} f(\mathbf{x})\right] \leqslant \eta.$$

(2) $\mathrm{T}^{\downarrow}_{p,\rho p} f$ is bounded away from 0 on typical inputs x such that $g(x) = 1$:

$$\Pr_{\mathbf{x} \sim \mu_p}\left[g(\mathbf{x}) = 1, \mathrm{T}^{\downarrow}_{p,\rho p} f(\mathbf{x}) \leqslant \lambda\right] \leqslant \eta.$$

Theorem 2.11. *For any $\varepsilon, \zeta > 0$ there are $\eta > 0$ and $J \in \mathbb{N}$ such that the following holds for any $p, \rho \in [\zeta, 1-\zeta]$ and $\lambda \in [\zeta, 1]$. If $f : \{0,1\}^n \to [0,1]$ and $g : \{0,1\}^n \to \{0,1\}$ are one-sided error solutions with λ and error η, then g is ε-close to a monotone, Boolean J-junta.*

We remark that any monotone junta g is a one-sided error approximate solution (by taking $f = g$), so Theorem 2.11 is tight with respect to the structure of g.

Organization. The proof of Theorem 2.6 is given in Section 4. The proofs of the rest of the results are deferred to the full version of the paper.

3 PRELIMINARIES

For any $p \in (0,1)$, we consider functions $f : (\{0,1\}^n, \mu_p) \to \mathbb{R}$ equipped with the inner product $\langle f, g \rangle = \mathbb{E}_{\mathbf{x} \sim \mu_p}[f(\mathbf{x})g(\mathbf{x})]$. We will use the Fourier–Walsh orthonormal basis $\left\{\chi^p_S\right\}_{S \subseteq [n]}$, where for each $S \subseteq [n]$ we define $\chi^p_S : \{0,1\}^n \to \mathbb{R}$ by

$$\chi^p_S(x) = \prod_{i \in S}\left[(x_i - p)/\sqrt{p(1-p)}\right].$$

Thus, we may write the Fourier expansion of a function $f : \{0,1\}^n \to \mathbb{R}$ by

$$f(x) = \sum_{S \subseteq [n]} \widehat{f_p}(S)\chi^p_S(x), \qquad \text{where } \widehat{f_p}(S) = \langle f, \chi^p_S \rangle.$$

Since $\left\{\chi^p_S\right\}_{S \subseteq [n]}$ is an orthonormal basis, we have Parseval's identity $\|f\|_2^2 = \sum_{S \subseteq [n]} \widehat{f_p}(S)^2$. We will need a few more notions and results from Fourier analysis, such as the Junta Theorems of [8, 31] and the Sensitivity Conjecture proved recently by [25], which we present below.

3.1 Influences

For a function $f : (\{0,1\}^n, \mu_p) \to \mathbb{R}$ and a coordinate $i \in [n]$, we define the p-biased influence of variable i to be

$$I^p_i[f] = \mathop{\mathbb{E}}_{\mathbf{x} \sim \mu_p}\left[(f(\mathbf{x}) - f(\mathbf{x} \oplus e_i))^2\right].$$

When the bias parameter is clear from context, we often write $I_i[f]$.

We will also use the notion of negative influences as given in Definition 2.9. We have the following simple fact, stating that averaging may only decrease negative influences. The proof is deferred to the full version of the paper.

Fact 3.1. *Let $f : (\{0,1\}^n, \mu_p) \to \mathbb{R}$ be a function, and let $i \in [n]$. Consider the function $g : (\{0,1\}^{n-1}, \mu_p) \to \mathbb{R}$ defined by $g(z) = \mathbb{E}_{\mathbf{x} \sim \mu_p}\left[f(\mathbf{x}) \mid \mathbf{x}_{[n] \setminus \{i\}} = z\right]$ (i.e. averaging f over the coordinate i). Then $I^-_j[g] \leqslant I^-_j[f]$ for any $j \in [n] \setminus \{i\}$.*

We also need the following fact that relates negative influences and distance from monotonicity.

Fact 3.2. *For all $p \in (0,1)$, $n \in \mathbb{N}$ and $\tau > 0$, if the function $f : (\{0,1\}^n, \mu_p) \to \mathbb{R}$ satisfies $I^-_i[f] \leqslant \tau$ for all $i \in [n]$, then there is a monotone function $h : (\{0,1\}^n, \mu_p) \to \mathbb{R}$ such that $\|f - h\|_1 \leqslant ((1-p)p)^{-n} n\tau$.*

We remark that the above fact is inspired by [22], wherein a similar statement was proved for Boolean functions for $p = 1/2$, with a better bound ($n\tau$). The proof is deferred to the full version of the paper.

3.2 Junta Theorems

We will use Bourgain's Theorem [8]; the sharp version below is proved in [30]. For $k \in \mathbb{N}$, the Fourier tail $W_{\geqslant k}[f]$ is defined to be $\sum_{|S|\geqslant k} \widehat{f}_p(S)^2$.

Theorem 3.3. *For any $\zeta > 0$ there is a constant $C(\zeta) > 0$ such that for any $k \in \mathbb{N}$, $\varepsilon > 0$ there are $\tau = (k/\varepsilon)^{-C\cdot k}$, $J = (k/\varepsilon)^{C\cdot k}$ such that the following holds for all $p \in [\zeta, 1-\zeta]$. If $f : (\{0,1\}^n, \mu_p) \to \{0,1\}$ satisfies $W_{\geqslant k}[f] \leqslant \frac{\varepsilon}{C\sqrt{k}\log^{1.5}(k)}$, then f is ε-close to a J-junta h.*

Furthermore, h only depends on variables i such that $I_i[g] \geqslant \tau$.

We also need the following result of Kindler and Safra. While being quantitatively weaker in some regards, it has an important feature that will be important for us and is missing from Theorem 3.3. The size of the junta in the result of Kindler and Safra only depends on the parameter k, as opposed to also on the closeness parameter ε as in Theorem 3.3.

Theorem 3.4 ([31]). *For any $\zeta > 0$, $m \in \mathbb{N}$ there exists $J(m,\zeta) \in \mathbb{N}$, $C(m,\zeta) > 0$ such that the following holds for all $p \in [\zeta, 1-\zeta]$. For any $\varepsilon > 0$ there exists $\delta = C(m,\zeta)\cdot\varepsilon$ such that if $f : (\{0,1\}^n, \mu_p) \to \{0,1\}$ is a function such that $W_{\geqslant m}[f] \leqslant \delta$, then f is ε-close to a junta of size $J(m,\zeta)$.*

3.3 Degree and Sensitivity

For any $f : \{0,1\}^n \to \{0,1\}$ and $x \in \{0,1\}^n$, the sensitivity of f at x is equal to the number of coordinates $i \in [n]$ such that $f(x) \neq f(x \oplus e_i)$. The max-sensitivity of a function f is $s(f) = \max_x s(f,x)$. The degree of a function $\deg(f)$ is the maximal size of S such that $\widehat{f}_p(S) \neq 0$ (we remark that this is easily seen to be independent of p).

We will use the following recent result of Huang [25] (formerly known as the sensitivity conjecture [42]) in our proof. We remark that quantitatively weaker results that were known earlier (such as the bound $s(f) \geqslant \Omega(\log(\deg(f)))$) would have been enough for us, but yield to a loss in several parameters.

Theorem 3.5 ([25]). *For any $f : \{0,1\}^n \to \{0,1\}$ we have that $s(f) \geqslant \sqrt{\deg(f)}$.*

4 PROOF OF THEOREM 2.6

In this section, we prove Theorem 2.6. Since we will always consider the downwards noise operator $\mathrm{T}^{\downarrow}_{p,p/2}$, we denote it succinctly by T.

4.1 Main Lemma

Lemma 4.1. *For any $\zeta > 0$ and $n \in \mathbb{N}$ there exists $\eta_0 > 0$ such that the following holds for all $p \in [\zeta, 1-\zeta]$, $\lambda \in [\zeta, 1]$ and $\eta \in [0, \eta_0]$. If $\|\mathrm{T}^{\downarrow}_{p,p/2} f - \lambda g\|_\infty \leqslant \eta$ then:*

- *g is an AND-OR function of width r, where $r \leqslant \lceil \log(2/\zeta) \rceil$.*
- *Let ϕ be the corresponding AND-XOR function. Then $\|f - 2^r\lambda\cdot\phi\|_\infty \leqslant 3^n\eta$.*

This section is devoted to the proof of this lemma, and the proof is divided into several claims. It will be convenient for us to identify vectors in $\{0,1\}^n$ with subsets of $[n]$ by identifying a vector with its support, and consequently think of the inputs of functions as subsets of $[n]$. The definition of the operator T to these language is immediate: $\mathrm{T}f(B) = \mathbb{E}_{\mathbf{C}\subseteq B}[f(\mathbf{C})] = 2^{-|B|}\sum_{C\subseteq B} f(C)$.

Fix ζ, n, and choose $\eta = \zeta 4^{-n^2-4n-4}$. Let f, g be functions as in the statement of the lemma.

Claim 4.2. *g is monotone.*

Proof. Suppose g is not monotone. Then there is an edge (A, B) of the hypercube where $A \subsetneq B$ such that $g(A) = 1$, $g(B) = 0$. We have that $\mathrm{T}f(B) \leqslant \lambda g(B) + \eta = \eta$, which by definition of T implies that $\mathbb{E}_{\mathbf{C}\subseteq B}[f(\mathbf{C})] \leqslant \eta$. Denote $\{i\} = B \setminus A$, and note that with probability $1/2$ we have $i \notin \mathbf{C}$, in which case $\mathbf{C} \subseteq A$. Thus, the non-negativity of f implies that $\mathbb{E}_{\mathbf{C}\subseteq A}[f(\mathbf{C})] \leqslant 2\mathbb{E}_{\mathbf{C}\subseteq B}[f(\mathbf{C})] \leqslant 2\eta$, i.e. $\mathrm{T}f(A) \leqslant 2\eta$. This is in contradiction to $\mathrm{T}f(A) \geqslant \lambda g(A) - \eta = \lambda - \eta$ (by the choice of η). □

Since g is monotone, one can discuss its minterms, i.e. sets $M \subseteq [n]$ such that $g(M) = 1$ but for all $A \subsetneq M$, $g(A) = 0$. The following lemma asserts that the value of f on any minterm of g is determined (up to a small error).

Claim 4.3. *If M is a minterm of g, then $\left|f(M) - \lambda 2^{|M|}\right| \leqslant 4^{|M|}\eta$.*

Proof. Since $g(M) = 1$, we have that $|\mathrm{T}f(M) - \lambda| \leqslant \eta$, and by definition of T we have $\mathrm{T}f(M) = 2^{-|M|}\sum_{A\subseteq M} f(A)$, so by the triangle inequality it follows that $\left|f(M) - 2^{|M|}\lambda\right| \leqslant 2^{|M|}\eta + \sum_{A\subsetneq M} f(A)$, and to finish the proof, we upper bound the last sum. Note that for every $A \subsetneq M$, choosing $\mathbf{B} \subsetneq M$ randomly of size $|M|-1$, we have that $A \subseteq \mathbf{B}$ with probability at least $1/|M|$, hence by the non-negativity of f there is B of size $|M|-1$ such that $\sum_{A\subsetneq M} f(A) \leqslant |M|\sum_{A\subseteq B} f(A)$, and we fix such B.[4] Since $B \subsetneq M$ and M is a minterm of g, we have that $g(B) = 0$, and therefore $\mathrm{T}f(B) \leqslant \eta$ or equivalently $\sum_{A\subseteq B} f(A) \leqslant 2^{|B|}\eta$. Plugging that in we get that $\sum_{A\subsetneq M} f(A) \leqslant |M|\, 2^{|M|-1}\eta$ and the claim follows. □

We next wish to argue all minterms of g are of the same size, and towards this end (and also in other places in the argument) the following proposition will be useful.

Proposition 4.4. *Let $B, Z \subseteq [n]$ be disjoint such that $g(B) = 1$. Then*

$$\left|\sum_{A\subseteq B} f(A\cup Z) - 2^{|B|}\lambda\right| \leqslant 2^{|B|}3^{|Z|}\eta.$$

Proof. Note that $|\mathrm{T}f(B\cup W) - \lambda g(B\cup W)| \leqslant \eta$ for any $W \subseteq Z$. Since g is monotone and $g(B) = 1$, we must have $g(B\cup Y) = 1$ and we get that

$$\left|\sum_{A\subseteq B, Y\subseteq W} f(A\cup Y) - 2^{|B\cup W|}\lambda\right| \leqslant 2^{|B\cup W|}\eta.$$

[4] Alternatively, note that $\sum_{A\subsetneq M} f(A) \leqslant \sum_{|B|=m-1}\sum_{A\subseteq B} f(A)$, and take B maximizing $\sum_{A\subseteq B} f(A)$.

Note that for any $Y \subseteq Z$, $\sum_{W\colon Y\subseteq W\subseteq Z}(-1)^{|Z\setminus W|} = 0$ unless $Y = Z$, in which case the sum is 1, and so we get that

$$\sum_{A\subseteq B} f(A\cup Z) = \sum_{\substack{A\subseteq B, Y\subseteq Z\\ W\colon Y\subseteq W\subseteq Z}} (-1)^{|Z\setminus W|} f(A\cup Y) = \sum_{W\subseteq Z}(-1)^{|Z\setminus W|}\sum_{A\subseteq B, Y\subseteq W} f(A\cup Y).$$

Therefore the triangle inequality implies that

$$\left|\sum_{A\subseteq B} f(A\cup Z) - \lambda\sum_{W\subseteq Z}(-1)^{|Z\setminus W|}2^{|B\cup W|}\right| \leqslant \sum_{W\subseteq Z}\left|\sum_{A\subseteq B, Y\subseteq W} f(A\cup Y) - 2^{|B\cup W|}\lambda\right| \leqslant \sum_{W\subseteq Z} 2^{|B\cup W|}\eta,$$

which is equal to $2^{|B|}3^{|Z|}\eta$. To complete the proof, we observe that by the binomial formula

$$\sum_{W\subseteq Z}(-1)^{|Z\setminus W|}2^{|B\cup W|} = 2^{|B|}\sum_{W\subseteq Z}2^{|W|}(-1)^{|Z|-|W|} = 2^{|B|}(2-1)^{|Z|} = 2^{|B|}.$$

□

We now show two consequences of the above proposition. First, we show that all minterms of g have the same size.

Claim 4.5. *For any two minterms M, M' of g we have $|M| = |M'|$.*

Proof. Let $Z = M' \setminus M$. By Proposition 4.4 we have

$$\sum_{A\subseteq M} f(A\cup Z) \leqslant 2^{|M|}\lambda + 2^{|M|}3^{|Z|}\eta.$$

By the non-negativity of f the left-hand side is at least the value of f for $A = M\cap M'$, i.e. on $A\cup Z = M'$; furthermore, by Claim 4.3 we have $f(M') \geqslant \lambda 2^{|M'|} - 4^{|M'|}\eta$, so combining we get

$$\lambda 2^{|M'|} \leqslant 2^{|M|}\lambda + \eta(2^{|M|}3^{|Z|} + 4^{|M'|}) \leqslant 2^{|M|}\lambda + \lambda/2,$$

where the last inequality is by the choice of η. This implies that $|M'| \leqslant |M|$. The second inequality is proved analogously. □

Denote the size of a minterm of g by m, and note that $m \leqslant \lceil\log(2/\lambda)\rceil$. Indeed, letting M be any minterm of g, by Claim 4.3 we get that $\lambda 2^m \leqslant f(M) + 4^m\eta \leqslant 2$.

We next show that the value of f in a point B must be either close to 0 or close to $2^m\lambda$.

Claim 4.6. *For any $B \in \{0,1\}^n$, it holds that either $f(B) \leqslant 4^{|B|^2}\eta$ or $|f(B) - 2^m\lambda| \leqslant 4^{|B|}\eta$.*

Proof. If $g(B) = 0$, the claim is immediate since $2^{-|B|}f(B) \leqslant \mathrm{T}f(B) \leqslant \lambda g(B) + \eta = \eta$, so we assume that $g(B) = 1$. We prove the claim by induction on $|B|$. If $|B| = m$, then B is a minterm and the claim follows from Claim 4.3. Assume the claim holds for all $|B| \leqslant i$, and let B be of size $i+1$. Since $g(B) = 1$ we get that there must be a minterm $M \subseteq B$ of g. Let $Z = B\setminus M$, then by Proposition 4.4,

$$\left|\sum_{A\subseteq M} f(A\cup Z) - 2^m\lambda\right| \leqslant 2^m 3^{|Z|}\eta.$$

For each $A \subsetneq M$, by the induction hypothesis $f(A\cup Z)$ is either close to 0 (i.e. at most $4^{|A\cup Z|^2}\eta$) or close to $2^m\lambda$ (more precisely, up to $\pm 4^{|A\cup Z|}\eta$).

- If there is $A^\star \subsetneq M$ that falls into the second case, then by non-negativity of f we get that $f(B)$ is equal to
$$\begin{aligned} f(M\cup Z) &\leqslant \sum_{A\subseteq M} f(A\cup Z) - f(A^\star\cup Z)\\ &\leqslant \left|\sum_{A\subseteq M} f(A\cup Z) - 2^m\lambda\right| + \left|f(A^\star\cup Z) - 2^m\lambda\right|\\ &\leqslant 2^m3^{|Z|}\eta + 4^{|A^\star\cup Z|}\eta \leqslant 4^{|B|}\eta,\end{aligned}$$
where in the last inequality we used $|A^\star| \leqslant m-1$, and the claim is proved for B.
- Otherwise, by the triangle inequality
$$\begin{aligned} |f(M\cup Z) - 2^m\lambda| &\leqslant \left|\sum_{A\subsetneq M} f(A\cup Z)\right| + \left|\sum_{A\subseteq M} f(A\cup Z) - 2^m\lambda\right|\\ &\leqslant \sum_{A\subsetneq M} 4^{|A\cup Z|^2} + 2^m3^{|Z|}\eta\\ &\leqslant \left(4^{(m+|Z|-1)^2+m} + 3^{m+|Z|}\right)\eta,\end{aligned}$$
which is at most $4^{(m+|Z|)^2}\eta = 4^{|B|^2}\eta$. Hence the claim is proved for B (as $B = M\cup Z$). □

We are now able to restate Proposition 4.4 in a more convenient form. For each pair of disjoint sets $B, Z \subseteq [n]$ such that $g(B) = 1$, denote $X(B, Z) = \{A\cup Z \mid A\subseteq B\}$.

Corollary 4.7. *Suppose $B, Z \subseteq [n]$ are disjoint and $g(B) = 1$. Then there is a unique $A^\star \subseteq B$ such that:*

- $|f(A^\star\cup Z) - 2^m\lambda| \leqslant 4^n\eta$.
- *For any other $A \subset B$ we have that $f(A\cup Z) \leqslant 4^{n^2}\eta$.*
- *$g(A^\star\cup Z) = 1$ and for any $A \subsetneq A^\star$ we have $g(A\cup Z) = 0$.*

Proof. For the first item, if for all $A\subseteq B$ it holds that $f(A\cup Z) \leqslant 4^{n^2}\eta$, then by Proposition 4.4 we have $2^m\lambda \leqslant \sum_{A\subseteq B} f(A\cup Z) + 6^n\eta \leqslant 4^{n^2+3n}\eta$, which contradicts the choice of η. Therefore, by Claim 4.6 there is $A^\star\subseteq B$ such that $|f(A^\star\cup y) - 2^m\lambda| \leqslant 4^n\eta$.

For the second item, assume towards contradiction there are two such A_1, A_2. By Proposition 4.4 we have

$$2\cdot(2^m\lambda - 4^n\eta) \leqslant f(A_1\cup Z) + f(A_2\cup Z) \leqslant \sum_{A\subseteq B} f(A\cup Z) \leqslant 2^m\lambda + 6^n\eta,$$

and therefore $2^m\lambda \leqslant 6^{n+1}\eta$, which is a contradiction to the choice of η.

For the third item, note that

$$g(A^\star\cup Z) \geqslant \mathrm{T}f(A^\star\cup Z) - \eta \geqslant 2^{-|A^\star\cup Z|}f(A^\star\cup Z) - \eta \geqslant 2^{-n}\lambda - (4^n+1)\eta > 0,$$

and since g is Boolean-valued it follows that $g(A^\star\cup Z) = 1$. Also, for any $A \subsetneq A^\star$ we have

$$g(A\cup Z) \leqslant \mathrm{T}f(A\cup Z) + \eta \leqslant 4^{n^2}\eta + \eta < 1,$$

where in the second inequality we used the definition of T and the second item. Since g is Boolean we get that $g(A \cup Z) = 0$. ☐

To simplify notation, for the rest of the section we often say "the value of $f(S)$ is close to $2^m\lambda$" to express that $|f(S) - 2^m\lambda| \leqslant 4^n\eta$ and "the value of $f(S)$ is close to 0" to express that $f(S) \leqslant 4^{n^2}\eta$.

Consider the m-uniform hypergraph $H = ([n], E)$ whose edges are the minterms of g. In the remainder of this section we show that H is a complete m-partite hypergraph, which is easily seen to be equivalent to g being an ANR-OR function of width m. Towards this end, we will define a coloring $\chi\colon [n] \to [m]$ and show that (a) each edge $e \in E$ is rainbow colored (i.e. no two vertices in it are colored in the same color), and (b) any rainbow colored set $A \subseteq [n]$ is an edge.

Fix a minterm $B \subseteq [n]$, and write $B = \{b_1, \ldots, b_m\}$, where $b_1, \ldots, b_m \in [n]$. We define $\chi(b_i) = i$. We now define $\chi(v)$ for any $v \in [n] \setminus B$. Fix $v \in [n] \setminus B$, and consider the set $X(B, \{v\})$; by Corollary 4.7 there exists a unique $A \subseteq B$ such that $f(A \cup \{v\})$ is close to $2^m\lambda$, and its g-value is 1. Since $g(A \cup \{v\}) = 1$, we must have $|A \cup \{v\}| \geqslant m$, and there are two options:

- If $|A \cup \{v\}| = m + 1$, i.e. $A = B$, define $\chi(v) = \perp$.
- Otherwise, there is $i \in [m]$ such that $A = B \setminus \{b_i\}$, and we define $\chi(v) = \chi(b_i) = i$.

We first show that each minterm of g is colored using only elements from $[m]$ (as opposed to $\perp$).

Claim 4.8. *Let $M \in E$ be a minterm of g. Then for each $v \in M$ we have $\chi(v) \neq \perp$.*

Proof. Assume towards contradiction that this is not the case, and let $v \in M$ be such that $\chi(v) = \perp$. Then by definition of χ this means that $f(B \cup \{v\})$ is close to $2^m\lambda$, and since B is a minterm of g we also know, by Claim 4.3, that $f(B)$ is close to $2^m\lambda$. This gives us two points in $X(M, B \setminus M)$ whose f-value is close to $2^m\lambda$, in contradiction to Corollary 4.7. ☐

Next, we show that each minterm of g is rainbow colored by χ.

Claim 4.9. *Let $M \in E$ be a minterm of g. Then M is rainbow colored.*

Proof. Write $M = \{v_1, \ldots, v_m\}$, and assume towards contradiction the statement is false. Then there are v_i, v_j that are assigned the same color by χ, and without loss of generality we may assume $\chi(v_1) = \chi(v_2) = m$. By definition of χ it follows that $f(\{v_1\} \cup (B \setminus \{b_m\}))$ and $f(\{v_2\} \cup (B \setminus \{b_m\}))$ are both close to $2^m\lambda$. However, note that these are two distinct points in $X(M, B \setminus M)$, and thus we get a contradiction to the second item in Corollary 4.7. ☐

Note that the definition of the coloring χ may depend on the minterm B chosen initially to define it. The following claim shows that this is actually not the case — and more precisely that if we use a different minterm B' to define a coloring χ', then there is a permutation π on $[m]$ such that $\chi' = \pi \circ \chi$.

Claim 4.10. *Let $B' = \{b'_1, \ldots, b'_m\}$ be any minterm of g, and let χ' be a coloring defined as above using B' in place of B. Then there exists $\pi \in S_m$ such that $\chi = \pi \circ \chi'$.*

Proof. Since B is a minterm of g, it follows by Claim 4.9 that it is rainbow colored by both χ and χ', so we define $\pi \in S_m$ by $\pi(\chi'(b_i)) = \chi(b_i)$. We define $\tilde{\chi} = \pi \circ \chi'$ and show that $\chi = \tilde{\chi}$.

Let $v \notin B$, and assume without loss of generality $\chi(v) = m$. Then by definition of χ we must have that $\{b_1, \ldots, b_{m-1}, v\}$ is a minterm of g, and hence by Claim 4.9 it must be rainbow colored by $\tilde{\chi}$. Since $\tilde{\chi}$ agrees with χ on $b_1, \ldots, b_{m-1}$, we must have $\tilde{\chi}(v) = m$, and we are done. ☐

Lastly, we show that each rainbow colored set of size m is a minterm of g.

Claim 4.11. *For any minterm B of g and a coloring χ defined by it, the following holds. If $M \subseteq [n]$ of size m is rainbow colored by χ, then $g(M) = 1$. Consequently, M is a minterm of g.*

Proof. We prove the statement for all B, χ, M by induction on $|B \cap M|$.

Write $M = \{v_1, \ldots, v_m\}$, $B = \{b_1, \ldots, b_m\}$, and assume without loss of generality that $\chi(v_i) = \chi(b_i)$. The base case is $|B \cap M| = m$, in which case $M = B$ and the claim is obvious.

Let $k \leqslant m-1$, assume the statement is true whenever $|B \cap M| \geqslant k+1$, and let M be such that $|B \cap M| = k$. Without loss of generality we may assume that $v_i = b_i$ for all $1 \leqslant i \leqslant k$. Since $\chi(v_{k+1}) = \chi(b_{k+1})$ we get that $B' = (B \setminus \{b_{k+1}\}) \cup \{v_{k+1}\}$ is a minterm of g. Let χ' be the coloring defined by B'. By Claim 4.9 we get that $\chi' = \pi \circ \chi$ for some $\pi \in S_m$, and in particular as M is rainbow colored by χ it is also rainbow colored by χ'. Since $|B' \cap M| = k+1$ we may apply the induction hypotehsis on B' with the coloring χ' to conclude that M is a minterm of g, as required. ☐

It follows that the function g is the function $\bigwedge_{i \in [m]} \bigvee_{j \in A_i} x_j$ where $A_i = \chi^{-1}(i)$. To complete the proof of Lemma 4.1 we must establish the structural result for f, which we do by inverting f.

Claim 4.12. *The operator T has an inverse T^{-1} given by $\mathrm{T}^{-1}h(B) = \sum_{A \subseteq B} (-1)^{|B \setminus A|} 2^{|A|} h(A)$.*

Proof. Let h be in the image of T, i.e. $h(B) = \mathrm{T}\tilde{h}(B)$ for some $\tilde{h}$. We prove by induction on B that $\tilde{h}(B) = \sum_{A \subseteq B} (-1)^{|B \setminus A|} 2^{|A|} h(A)$.

The base case $|B| = 0$ is clear. Assume the statement holds for all B such that $|B| \leqslant k$, and let B be of size $k+1$. By definition of h we have that $2^{|B|}h(B) = \sum_{A \subsetneq B} \tilde{h}(A) + \tilde{h}(B)$. Using the induction hypothesis on each $A \subsetneq B$ we get that

$$\begin{aligned}\sum_{A \subsetneq B} \tilde{h}(A) &= \sum_{A \subsetneq B} \sum_{C \subseteq A} (-1)^{|A \setminus C|} 2^{|C|} \tilde{h}(C) \\ &= \sum_{C} 2^{|C|} \tilde{h}(C) \sum_{\substack{A: \\ C \subseteq A \subsetneq B}} (-1)^{|A \setminus C|} \\ &= \sum_{C} 2^{|C|} \tilde{h}(C) (-1)^{|B \setminus C|+1},\end{aligned}$$

where in the last equality we used the fact that adding the summand corresponding to $A = B$, the sum would be 0. Plugging that into the previous equality and rearranging finishes the inductive step. ☐

Define $\psi = \lambda \sum_{A \subseteq B} (-1)^{|B \setminus A|} 2^{|A|} g(B)$. By Claim 4.12 we have that $\lambda g = \mathrm{T}\psi$.

Claim 4.13. $\|f-\psi\|_\infty \leqslant 3^n\eta$.

Proof. Let $h = \mathrm{T}f$. Using the formula for h from Claim 4.12 and the definition of ψ we have that for all $B \subseteq [n]$,

$$|f(B)-\psi(B)| \leqslant \sum_{A\subseteq B} 2^{|A|}\,|h(A)-\lambda g(A)| \leqslant \sum_{A\subseteq B} 2^{|A|}\eta = 3^{|B|}\eta. \quad \square$$

Let $\phi = \bigwedge_{i=1}^m \left(\bigoplus_{j\in A_i} x_j\right)$. We show that $\mathrm{T}\phi = 2^{-m}g$. Since T is invertible by Claim 4.12 and $\mathrm{T}\psi = \lambda g$, we get $\psi = 2^m\lambda\phi$, and hence Claim 4.13 implies that f is $3^n\eta$-close to $2^m\lambda\phi$ in L_∞, as required.

Claim 4.14. *Let $A_1,\dots,A_m$ be disjoint, non-empty sets, and let $\phi = \bigwedge_{i=1}^m \bigoplus_{j\in A_i} x_j$, $g = \bigwedge_{i=1}^m \bigvee_{j\in A_i} x_j$. Then $\mathrm{T}\phi = 2^{-m}g$.*

Proof. Fix $B \subseteq [n]$, and let $B_i = B\cap A_i$ for each i.

If $g(B) = 0$, then $B_i = \emptyset$ for some i, without loss of generality $i = 1$. Thus, for any $C \subseteq B$ we have that $C\cap A_1 = \emptyset$, and hence $\phi(C) = 0$, so $\mathrm{T}\phi(B) = 0$.

If $g(B) = 1$, then $B_i \neq \emptyset$ for all i. Let $\mathbf{C} \subseteq B$ be chosen uniformly at random, and denote $\mathbf{C}_i = \mathbf{C}\cap A_i$. Note that the distribution of $\mathbf{C}_1,\dots,\mathbf{C}_m$ is of independent uniform subsets of $B_1,\dots,B_m$, and as such the parity of the size of each $\mathbf{C}_i$ is a uniform and independent bit. Thus,

$$\begin{aligned}\mathrm{T}\phi(B) &= \Pr_{\mathbf{C}\subseteq B}[\phi(\mathbf{C}) = 1]\\ &= \Pr_{\mathbf{C}\subseteq B}[|\mathbf{C}_i| \equiv 1 \pmod 2 \text{ for all } i\in[m]] = 2^{-m}. \quad \square\end{aligned}$$

This completes the proof of Lemma 4.1. ◻

4.2 Deducing Theorem 2.6

In this section we use Lemma 4.1 to deduce Theorem 2.6, and we first sketch the argument. Given an approximate solution f, g, we first observe that the function g is noise insensitive — that is, has a small Fourier tail — and hence deduce from Theorem 3.3 that it is close to a junta. We then show that for almost all restrictions β outside the junta variables, we can associate a bounded function $\tilde{f}_\beta$ such that $\tilde{f}_\beta, g_\beta$ are a solution to the equation in L_∞, and we may deduce some structure for g_β and $\tilde{f}_\beta$. Using the fact that the restricted variables barely affect the function g (since it is junta) one can thus deduce the necessary AND-OR structure from g. To get the structural result for the function f, slightly more work is needed. We show that eliminating ORs that are too wide from the g_β's, almost all of them become the same function, and we show that after averaging over the removed variables, f is close to a multiple of the corresponding AND-XOR function.

We first give several statements that will be useful for us in the proof. The following lemma from [33] shows the effect of the operator $\mathrm{T}^{\downarrow}_{p,\rho p}$ on the Fourier expansion of a function.

Lemma 4.15. *If $f = \sum_S \hat{f}(S)\chi_S^{(\rho p)}$ then*

$$\mathrm{T}^{\downarrow}_{p,\rho p} f = \sum_S \left(\frac{(1-p)\rho}{1-\rho p}\right)^{|S|/2} \hat{f}(S)\chi_S^{(p)}.$$

Using the previous lemma we may show that if f, g are approximate solutions, then g has an exponentially small tail.

Lemma 4.16. *Let $f\colon(\{0,1\}^n,\mu_{\rho p}) \to [0,1]$, $g\colon(\{0,1\}^n,\mu_p) \to \{0,1\}$ be functions such that $\|\mathrm{T}^{\downarrow}_{p,\rho p} f - \lambda g\| \leqslant \eta$. Then for any $k\in\mathbb{N}$ we have that $W_{\geqslant k}[g] \leqslant 2\lambda^{-2}(\eta+\rho^k)$.*

Proof. Since f, g are bounded between 0 and 1, $\mathrm{T}f - \lambda g$ is bounded between $-1, 1$ at each point and therefore we get that $\|\mathrm{T}f-\lambda g\|_2^2 \leqslant \|\mathrm{T}f-\lambda g\|_1 \leqslant \eta$. Using Parseval's inequality (and Lemma 4.15 to get the Fourier coefficients of $\mathrm{T}f$ on μ_p), we get that, denoting $\rho_2 = \frac{(1-p)\rho}{1-\rho p} \leqslant \rho$, we have

$$\sum_S (\rho_2^{|S|/2}\widehat{f}_{p\rho}(S) - \lambda\widehat{g}_p(S))^2 \leqslant \eta.$$

Therefore, using $(a+b)^2 \leqslant 2(a^2+b^2)$ we get that

$$\begin{aligned}\lambda^2 \sum_{|S|\geqslant k} \widehat{g}_p(S)^2 &\leqslant 2\eta + 2\sum_{|S|\geqslant k} (\rho_2^{|S|/2}\widehat{f}_{p\rho}(S))^2\\ &\leqslant 2\eta + 2\cdot\rho_2^k \sum_{|S|\geqslant k} \widehat{f}_{p\rho}(S)^2\\ &\leqslant 2\eta + 2\cdot\rho^k,\end{aligned}$$

where in the last inequality we used Parseval to bound the sum of Fourier coefficients of f by 1 and $\rho_2 \leqslant \rho$. ◻

Second, the following will be useful for us in the pruning process of the wide ORs.

Lemma 4.17. *Suppose that g_1, g_2 are AND-OR functions of width at most d that are ε-close with respect to μ_p, and let $\gamma = 2p^{-d}\varepsilon$. Let ψ_1, ψ_2 be the truncations of g_1, g_2 respectively resulting by removing all ORs containing more than $\log_{1/(1-p)}(1/\gamma)$ variables.*

Then $\psi_1 = \psi_2$, and furthermore this function is $d\gamma$-close to g_1.

Proof. We say an OR of g_1 is *small* if the number of variables in it is at most $\log_{1/(1-p)}(1/\gamma)$, and let A_1 be a small OR of g_1. We claim that there is a small OR of g_2, which will be denoted by A_2, that contains A_1. Assume towards contradiction that this is not the case. Thus, restricting A_1-variables to 0, the restricted function $(g_2)_{A_1\to 0}$ does not become identically 0 and it is still an AND-OR function of width at most d, and therefore it gets the value 1 with probability at least p^d. Since the probability that all of the variables in A_1 get the value 0 is at least $(1-p)^{\log_{1/(1-p)}(1/\gamma)} = \gamma$, we get that $\Pr_{\mathbf{x}}\left[\mathrm{OR}_{A_1}(\mathbf{x}) = 0, g_2(\mathbf{x}) = 1\right] \geqslant \gamma\cdot p^d$. However, note that on any such x we have $g_1(x) = 0$ and $g_2(x) = 1$, and by assumption the probability mass on such x's is at most ε, so we get that $\varepsilon \geqslant \gamma p^d$ and contradiction.

Therefore, for each small OR of g_1 there is a small OR in g_2 containing it and vice versa. As the ORs in each function are disjoint in variables, it follows that each small OR of g_1 appears in g_2 and vice versa, so in other words $\psi_1 = \psi_2$.

Finally, since g_1 and ψ_1 may differ only when there is an OR of size at least $\log_{1/(1-p)}(1/\gamma)$ in g_1 that evaluates to 0, and there are at most d such clauses, it follows from the union bound that $\Pr_x[g_1(x) \neq \psi_1(x)] \leqslant d\gamma$. ◻

We are now ready to prove Theorem 2.6.

Proof of Theorem 2.6. Fix $\zeta, \varepsilon > 0$ from Theorem 2.6 (we assume $\varepsilon > 0$ is small enough) and choose $m = \lceil\log(2/\zeta)\rceil$. Let

$C = C(\zeta)$ be from Theorem 3.3, choose $\eta_1 = \zeta^2\varepsilon^2/(4C\log(1/\varepsilon))$, and pick τ, J from Theorem 3.3 for ε and $k = \lceil\log(1/\eta_1)\rceil$. Later in the proof we will also define η_2 and subsequently take $\eta = \min(\eta_1, \eta_2)$.

Let f, g be functions as in the statement of Theorem 2.6, and set $k = \lceil\log(1/\eta_1)\rceil$. From Lemma 4.16 we have that $W_{\geqslant k}[g] \leqslant \lambda^{-2}(\eta + 2^{-k}) \leqslant \varepsilon/(Ck)$, and Theorem 3.3 implies that there is $T \subseteq [n]$ of size J such that g is ε^2-close to a T-junta.

Take $\eta_{4.1}$ from Lemma 4.1 for ζ and $n = J$, and set $\eta_2 = 6^{-J}\eta_{4.1}\varepsilon^2$. We write points $x \in \{0,1\}^n$ as (α, β) where $\alpha \in \{0,1\}^T$ and $\beta \in \{0,1\}^{[n]\setminus T}$. For each $\beta \in \{0,1\}^{[n]\setminus T}$, define $\tilde{f}_\beta \colon \{0,1\}^T \to [0,1]$ and $g_\beta \colon \{0,1\}^T \to \{0,1\}$ by

$$\tilde{f}_\beta(\alpha) = \mathop{\mathbb{E}}_{\boldsymbol{\beta'} \leqslant \beta}\left[f(\alpha, \boldsymbol{\beta'})\right], \quad g_\beta(\alpha) = g(\alpha, \beta),$$

and let $B = \left\{\beta \in \{0,1\}^{[n]\setminus T} \;\middle|\; \|\tilde{f}_\beta - \lambda g_\beta\|_\infty \leqslant 3^{-J}\varepsilon\eta_{4.1}\right\}$. Since for any α, β we have $\mathrm{T}\tilde{f}_\beta(\alpha) = \mathrm{T}f(\alpha, \beta)$, it follows that

$$\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\left[\|\mathrm{T}\tilde{f}_{\boldsymbol{\beta}} - \lambda g_{\boldsymbol{\beta}}\|_1\right] = \|\mathrm{T}f - \lambda g\|_1 \leqslant \eta_2.$$

Therefore Markov's inequality implies that with probability at least $1 - \varepsilon$ over $\boldsymbol{\beta} \sim \mu_p$ we have $\|\mathrm{T}\tilde{f}_{\boldsymbol{\beta}} - \lambda g_{\boldsymbol{\beta}}\|_1 \leqslant 6^{-J}\varepsilon\eta_{4.1}$, in which case $\boldsymbol{\beta} \in B$. In particular, we conclude that

$$\Pr_{\boldsymbol{\beta} \sim \mu_p^{[n]\setminus T}}[\boldsymbol{\beta} \in B] \geqslant 1 - \varepsilon.$$

For each $\beta \in B$, Lemma 4.1 implies that g_β is an AND-OR function of width $r(\beta)$ which is at most $m(\zeta) = O(\log(1/\zeta))$, and that $\tilde{f}_\beta$ is ε-close to $2^r\lambda \cdot$ AND-XOR in L_∞ for the corresponding AND-XOR function; we will use that only later when we establish the structure for f. Since g is ε^2-close to a T-junta, if we choose $\boldsymbol{\beta}, \boldsymbol{\beta'} \in B$ independently (according to μ_p) then on average $g_{\boldsymbol{\beta}}$ and $g_{\boldsymbol{\beta'}}$ are δ-close, where $\delta \leqslant 2\varepsilon^2/\Pr[B] \leqslant 2\varepsilon^2$. Thus there is $\beta^\star \in B$ such that $\mathbb{E}_{\boldsymbol{\beta} \in B}\left[\|g_{\boldsymbol{\beta}} - g_{\beta^\star}\|_1\right] \leqslant 2\varepsilon^2$, so by Markov's inequality, defining $B' \subseteq B$ by $B' = \left\{\beta \in B \;\middle|\; \|g_\beta - g_{\beta^\star}\|_1 \leqslant \varepsilon\right\}$ we have that $\Pr_{\boldsymbol{\beta} \sim \mu_p^{[n]\setminus T}}[\boldsymbol{\beta} \in B'] \geqslant 1 - \frac{2\varepsilon}{\Pr[B]} \geqslant 1 - 4\varepsilon$. We may already argue that g is close to the AND-OR function $g_{\beta^\star}$, however that will not be strong enough to establish the structural result for f and hence we prove a stronger statement. Namely, we show that if we truncate g_β by removing the wide ORs, then almost all of them will produce the same AND-OR function ψ.

Proving the structural result for g. For each $\beta \in B'$, let ψ_β be the AND-OR function g_β where we remove from it all ORs whose width exceeds $\log_{1/(1-p)}(1/(2p^{-m}\varepsilon))$. From Lemma 4.17 we get that there is an AND-OR function ψ (namely, $\psi_{\beta^\star}$) such that $\psi_\beta = \psi$ and $\|g_\beta - \psi\|_1 \leqslant 2mp^{-m}\varepsilon = O_\zeta(\varepsilon)$ for each $\beta \in B'$. Therefore, it follows that

$$\begin{aligned}
&\Pr_{\boldsymbol{\alpha},\boldsymbol{\beta}}[g(\boldsymbol{\alpha}, \boldsymbol{\beta}) \neq \psi(\boldsymbol{\alpha})] \\
&\leqslant \Pr_{\boldsymbol{\beta}}\left[\boldsymbol{\beta} \notin B'\right] + \Pr_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left[g(\boldsymbol{\alpha}, \boldsymbol{\beta}) \neq \psi(\boldsymbol{\alpha}) \;\middle|\; \boldsymbol{\beta} \in B'\right] \\
&= O_\zeta(\varepsilon),
\end{aligned}$$

i.e. g is close to the AND-OR function ψ. Let $T' \subseteq T$ be the set of variables that appear in ψ, and write $\alpha \in \{0,1\}^T$ as $\alpha = (\alpha_1, \alpha_2)$, where $\alpha_1 \in \{0,1\}^{T'}$ and $\alpha_2 \in \{0,1\}^{T\setminus T'}$.

Proving the structural result for f. We show that averaging outside T' makes f close to a multiple of ϕ, where ϕ is the AND-XOR function corresponding to ψ. For each $\beta \in B'$ we denote by ϕ_β the AND-XOR function corresponding to g_β.

For each $\beta \in B'$, define $A_\beta(\alpha_1) = \mathbb{E}_{\boldsymbol{\alpha_2'} \sim \mu_{p\rho}^{T\setminus T'}}\left[\tilde{f}_\beta(\alpha_1, \boldsymbol{\alpha_2'})\right]$, and $A \colon \{0,1\}^T \to [0,1]$ by $A(\alpha_1) = \mathbb{E}_{\boldsymbol{\beta} \sim \mu_p^{[n]\setminus T}}\left[A_{\boldsymbol{\beta}}(\alpha_1)\right]$. For each $\beta \in B'$, Lemma 4.1 implies that $\tilde{f}_\beta$ is ε-close to $2^{r(\beta)}\lambda \cdot \phi_\beta$ in L_∞, hence by averaging over α_2 we conclude that $A_\beta(\alpha_1)$ is ε-close to $2^{r(\beta)}\lambda \cdot \mathbb{E}_{\boldsymbol{\alpha_2} \sim \mu_{p\rho}^{T\setminus T'}}\left[\phi_\beta(\alpha_1, \boldsymbol{\alpha_2})\right] = c(\beta) \cdot \phi(\alpha_1)$, where $c(\beta) \leqslant 2^m$.

Set $K = \mathbb{E}_{\boldsymbol{\beta}}\left[c(\boldsymbol{\beta})1_{\boldsymbol{\beta} \in B'}\right]$. Then

$$\begin{aligned}
\|A - K\phi\|_1 &\leqslant 2^m \Pr_{\boldsymbol{\beta}}\left[\boldsymbol{\beta} \notin B'\right] + \mathop{\mathbb{E}}_{\boldsymbol{\beta}}\left[\|A_{\boldsymbol{\beta}} - c(\boldsymbol{\beta}) \cdot \phi\|_1 1_{\boldsymbol{\beta} \in B'}\right] \\
&\leqslant 2^m \Pr_{\boldsymbol{\beta}}\left[\boldsymbol{\beta} \notin B'\right] + 3^J\eta \\
&\leqslant 2^m O(\varepsilon) + \varepsilon = O_\zeta(\varepsilon).
\end{aligned}$$

Since $A(\alpha_1) = \mathbb{E}_{\boldsymbol{\alpha_2}, \boldsymbol{\beta'} \sim \mu_{p\rho}}[f(\alpha_1, \boldsymbol{\alpha_2}, \boldsymbol{\beta'})]$, this shows that after averaging outside T', the function f is $O_\zeta(\varepsilon)$-close to $K\phi$, and we next show that one may replace K by $\lambda 2^r$, where r is the width of ϕ, and retain this closeness.

If $\psi = 0$ then $\phi = 0$ so the value of K does not matter, so we assume henceforth that $\psi \neq 0$, in which case we clearly have $\|\psi\|_1 \geqslant p^m$. Since T is a contraction, $\|\mathrm{T}A - K \cdot \mathrm{T}\phi\|_1 \leqslant \|A - K\phi\|_1 = O_\zeta(\varepsilon)$. Since by Claim 4.14 we have $\mathrm{T}\phi = 2^{-r}\psi$, it follows that $\|\lambda g - K \cdot 2^{-r}\psi\|_1 = O_\zeta(\varepsilon)$. Since g is $O_\zeta(\varepsilon)$-close to ψ, this gives $|\lambda - K2^{-r}|\,\|\psi\|_1 = O_\zeta(\varepsilon)$, and so $|\lambda - K2^{-r}| = O(p^{-m}\varepsilon) = O_\zeta(\varepsilon)$, implying that $|K - 2^r\lambda| = O_\zeta(2^r\varepsilon) = O_\zeta(\varepsilon)$. Thus, $K \cdot \phi$ is $O_\zeta(\varepsilon)$-close to $2^r\lambda \cdot \phi$, and by the triangle inequality we conclude that $\|A - 2^r\lambda \cdot \phi\|_1 \leqslant O_\zeta(\varepsilon)$, finishing the proof. □

5 OPEN QUESTIONS

Our work raises many open questions. Perhaps the most obvious is the quantitative aspect of our results:

Open Question 1. *What is the optimal dependence between ε and δ in Theorem 1.1?*

We can ask a similar question about the various results listed in Section 2.

Nehama [40] showed that if we allow ε to depend on n, then we can choose $\varepsilon = \Theta(\delta^3/n)$. Theorem 1.1 eliminates the dependence on n in return for an exponential dependence on δ. We conjecture that Theorem 1.5 holds for $\varepsilon = \delta^{\Theta(1)}$.

Nehama situates Theorem 1.1 in the larger context of approximate judgement aggregation, or equivalently, approximate polymorphisms. He considers not only functions satisfying

$$f(\mathbf{x_1} \wedge \cdots \wedge \mathbf{x_m}) \approx f(\mathbf{x_1}) \wedge \cdots \wedge f(\mathbf{x_m}),$$

but also functions satisfying

$$f(\mathbf{x_1} \oplus \cdots \oplus \mathbf{x_m}) \approx f(\mathbf{x_1}) \oplus \cdots \oplus f(\mathbf{x_m}),$$

showing (using linearity testing) that the latter must be close to XORs. More generally, we can replace $\wedge, \oplus$ with an arbitrary Boolean function (or even a function on a larger domain):

Open Question 2. *Fix* $\phi\colon \{0,1\}^m \to \{0,1\}$. *Suppose* $f\colon \{0,1\}^n \to \{0,1\}$ *satisfies*

$$f(\phi(\mathbf{x_1},\ldots,\mathbf{x_m})) \approx \phi(f(\mathbf{x_1}),\ldots,f(\mathbf{x_m}))$$

for random $\mathbf{x_1},\ldots,\mathbf{x_m} \in \{0,1\}^n$, *where* $\phi(x_1,\ldots,x_m)$ *signifies elementwise application.*
What can we say about f*?*

Dokow and Holzman [13] showed (essentially) that when ϕ is a non-trivial function which is not an AND or an XOR, then the only exact solutions are dictatorships. We conjecture that when ϕ is such that the only exact solutions are dictatorships, then approximate solutions are approximate dictatorships.

Finally, let us mention the following tantalizing question:

Open Question 3. *What can be said about functions* $f\colon \{0,1\}^n \to \{0,1\}$ *satisfying*

$$\Pr[f(\mathbf{x} \wedge \mathbf{y}) = f(\mathbf{x}) \wedge f(\mathbf{y})] \geqslant \frac{3}{4} + \varepsilon?$$

We remark that the $\frac{3}{4}$ bound on the right hand side is natural in light of semi-random functions $f\colon \{0,1\}^n \to \{0,1\}$, chosen by taking $f(x)$ to be a uniform bit when $|x| \approx \frac{1}{2}n$, and $f(x) = 0$ otherwise.

REFERENCES

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. 1998. Proof Verification and the Hardness of Approximation Problems. *J. ACM* 45, 3 (1998), 501–555. https://doi.org/10.1145/278298.278306

[2] Sanjeev Arora and Shmuel Safra. 1998. Probabilistic Checking of Proofs: A New Characterization of NP. *J. ACM* 45, 1 (1998), 70–122. https://doi.org/10.1145/273865.273901

[3] K. Arrow. 1950. A difficulty in the theory of social welfare. *J. of Political Economy* 58 (1950), 328–346.

[4] K. Arrow. 1963. *Social choice and individual values.* John Wiley and Sons.

[5] John J Bartholdi, Craig A Tovey, and Michael A Trick. 1989. The computational difficulty of manipulating an election. *Social Choice and Welfare* 6, 3 (1989), 227–241.

[6] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan. 1996. Linearity testing in characteristic two. *IEEE Transactions on Information Theory* 42, 6 (1996), 1781–1795.

[7] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. 1993. Self-testing/correcting with applications to numerical problems. *J. Computer and System Sciences* 47, 3 (1993), 549–595.

[8] Jean Bourgain. 2002. On the distribution of the Fourier spectrum of boolean functions. *Israel Journal of Mathematics* 131, 1 (2002), 269–276.

[9] Joshua Brakensiek and Venkatesan Guruswami. 2020. Symmetric Polymorphisms and Efficient Decidability of Promise CSPs. In *SODA 2020.*

[10] Jonah Brown-Cohen and Prasad Raghavendra. 2015. Combinatorial Optimization Algorithms via Polymorphisms. arXiv:1501.01598 [cs.CC]

[11] Jean-Antoine-Nicolas Condorcet. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* De l'Imprimerie Royale.

[12] Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. 2017. Direct sum testing. *SIAM J. Comput.* 46, 4 (2017), 1336–1369.

[13] Elad Dokow and Ron Holzman. 2009. Aggregation of binary evaluations for truth-functional agendas. *Soc. Choice Welf.* 32 (2009), 221–241.

[14] Elad Dokow and Ron Holzman. 2010. Aggregation of binary evaluations. *J. Econ. Theory* 145 (2010), 495–511.

[15] Elad Dokow and Ron Holzman. 2010. Aggregation of binary evaluations with abstentions. *J. Econ. Theory* 145, 2 (2010), 544–561.

[16] Elad Dokow and Ron Holzman. 2010. Aggregation of non-binary evaluations. *Adv. Appl. Math.* 45 (2010), 487–504.

[17] Dvir Falik and Ehud Friedgut. 2014. Between arrow and Gibbard-Satterthwaite; A representation theoretic approach. *Israel Journal of Mathematics* 201, 1 (01 Jan 2014), 247–297. https://doi.org/10.1007/s11856-014-1064-5

[18] Piotr Faliszewski and Ariel D. Procaccia. 2010. AI's War on Manipulation: Are We Winning? *AI Magazine* 31, 4 (Sep. 2010), 53–64. https://doi.org/10.1609/aimag.v31i4.2314

[19] Ehud Friedgut, Gil Kalai, Nathan Keller, and Noam Nisan. 2011. A quantitative version of the Gibbard–Satterthwaite theorem for three alternatives. *SIAM J. Comput.* 40, 3 (2011), 934–952.

[20] E. Friedgut, G. Kalai, and N. Nisan. 2009. Elections Can be Manipulated Often. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS).* 243–249.

[21] A. Gibbard. 1973. Manipulation of voting schemes: a general result. *Econometrica* 41, 4 (1973), 587–601.

[22] Oded Goldreich, Shafi Goldwasser, Eric Lehman, and Dana Ron. 1998. Testing Monotonicity. In *39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA.* 426–435. https://doi.org/10.1109/SFCS.1998.743493

[23] William Timothy Gowers and Omid Hatami. 2017. Inverse and stability theorems for approximate representations of finite groups. *Sbornik: Mathematics* 208, 12 (dec 2017), 1784–1817. https://doi.org/10.1070/sm8872

[24] Johan Håstad. 2001. Some Optimal Inapproximability Results. *J. ACM* 48, 4 (July 2001), 798–859. https://doi.org/10.1145/502090.502098

[25] Hao Huang. 2019. Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture. *arXiv e-prints,* Article arXiv:1907.00847 (Jul 2019), arXiv:1907.00847 pages. arXiv:1907.00847 [math.CO]

[26] Donald H. Hyers and Themistocles M. Rassias. 1992. Approximate homomorphisms. *Aequationes Mathematicae* 44, 2 (01 Oct 1992), 125–153. https://doi.org/10.1007/BF01830975

[27] M. Isaksson, G. Kindler, and E. Mossel. 2012. The Geometry of Manipulation — a Quantitative Proof of the Gibbard Satterthwaite Theorem. *Combinatorica* 32, 2 (2012), 221–250.

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

[29] Nathan Keller. 2012. A tight quantitative version of Arrow's impossibility theorem. (2012), 1331–1355 pages.

[30] Guy Kindler, Naomi Kirshner, and Ryan O'Donnell. 2018. Gaussian noise sensitivity and Fourier tails. *Israel Journal of Mathematics* 225, 1 (01 Apr 2018), 71–109. https://doi.org/10.1007/s11856-018-1646-8

[31] Guy Kindler and Shmuel Safra. [n.d.]. Noise-resistant boolean functions are juntas. ([n. d.]).

[32] Lewis A Kornhauser and Lawrence G Sager. 1986. Unpacking the court. *Yale LJ* 96 (1986), 82.

[33] Noam Lifshitz. 2018. Hypergraph removal lemmas via robust sharp threshold theorems. *arXiv preprint arXiv:1804.00328* (2018).

[34] Christian List and Philip Pettit. 2002. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy* 18, 1 (2002), 89–110.

[35] Christian List and Philip Pettit. 2004. Aggregating sets of judgments: Two impossibility results compared. *Synthese* 140, 1–2 (2004), 207–235.

[36] Cristopher Moore and Alexander Russell. 2015. Approximate Representations, Approximate Homomorphisms, and Low-Dimensional Embeddings of Groups. *SIAM Journal on Discrete Mathematics* 29, 1 (2015), 182–197. https://doi.org/10.1137/140958578

[37] E. Mossel. 2012. A Quantitative Arrow Theorem. *Probability Theory and Related Fields* 154, 1 (2012), 49–88.

[38] E. Mossel, K. Oleszkiewicz, and A. Sen. 2013. On reverse hypercontractivity. *Geometric and Functional Analysis* 23, 3 (2013), 1062–1097.

[39] E. Mossel and M. Z. Racz. 2015. A quantitative Gibbard–Satterthwaite theorem without neutrality. *Combinatorica* 35, 3 (2015), 317–387.

[40] Ilan Nehama. 2013. Approximately classic judgement aggregation. *Ann. Math. Artif. Intell.* 68 (2013), 91–134.

[41] Klaus Nehring and Clemens Puppe. 2008. Consistent judgement aggregation: the truth-functional case. *Soc. Choice Welf.* 31 (2008), 41–57.

[42] Noam Nisan and Mario Szegedy. 1992. On the Degree of Boolean Functions as Real Polynomials. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada.* 462–467. https://doi.org/10.1145/129712.129757

[43] Michal Parnas, Dana Ron, and Alex Samorodnitsky. 2002. Testing basic Boolean formulae. *SIAM J. Disc. Math.* 16, 1 (2002), 20–46.

[44] Gabriella Pigozzi. 2016. Belief Merging and Judgment Aggregation. In *The Stanford Encyclopedia of Philosophy* (winter 2016 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[45] M. A. Satterthwaite. 1975. Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *J. of Economic Theory* 10 (1975), 187–217.

[46] Mario Szegedy and Yixin Xu. 2018. Impossibility theorems and the universal algebraic toolkit. (2018).