

Random Restrictions of Bounded Low Degree Polynomials Are Juntas

Sreejata Kishor Bhattacharya  

School of Technology and Computer Science, Tata Institute of Fundamental Research,
Mumbai, India

Abstract

We study the effects of random restrictions on low degree functions that are *bounded* on every point of the Boolean cube. Our main result shows that, with high probability, the restricted function can be approximated by a junta of arity that is just polynomial in the original degree. More precisely, let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d polynomial ($d \geq 2$) and let ρ denote a random restriction with survival probability $O(\log(d)/d)$. Then, with probability at least $1 - d^{-\Omega(1)}$, there exists a function $g : \{\pm 1\}^n \rightarrow [0, 1]$ depending on at most $d^{O(1)}$ coordinates such that $\|f_\rho - g\|_2^2 \leq d^{-1-\Omega(1)}$.

Our result has the following consequence for the well known, outstanding conjecture of Aaronson and Ambainis. The Aaronson-Ambainis conjecture was formulated to show that the acceptance probability of a quantum query algorithm can be well approximated almost everywhere by a classical query algorithm with a polynomial blow-up: it speculates that a polynomial $f : \{\pm 1\}^n \rightarrow [0, 1]$ with degree d has a coordinate with influence $\geq \text{poly}(1/d, \text{Var}[f])$. Our result shows that this is true for a non-negligible fraction of random restrictions of f assuming $\text{Var}[f]$ is not too low.

Our work combines the ideas of Dinur, Friedgut, Kindler and O’Donnell [7] with an approximation theoretic result, first reported in the recent work of Filmus, Hatami, Keller and Lifshitz [8].

2012 ACM Subject Classification Theory of computation \rightarrow Quantum query complexity

Keywords and phrases Analysis of Boolean Functions, Quantum Query Algorithms

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.17

Related Version *Full Version*: <https://arxiv.org/abs/2402.13952>

Funding *Sreejata Kishor Bhattacharya*: Google Fellowship 2024.

Acknowledgements I want to thank Francisco Escudero Gutierrez for pointing out typos and a small mistake in a previous version of this paper.

1 Introduction

The analysis of Boolean functions is a vibrant field with connections and applications to diverse areas. While several outstanding results have been established for Boolean functions, the closely related class of *bounded* functions, i.e. those which take value in the real interval $[0, 1]$ at every point of the Boolean cube, remains far less understood. Indeed, such bounded functions are of great significance. As mentioned in Dinur, Friedgut, Kindler and O’Donnell [7], “such bounded functions often arise naturally, particularly as weighted averages of boolean functions; e.g., as Fourier transforms of boolean functions, as noise-convolutions of boolean functions, in the context of random walks on the discrete cube, and in hardness-of approximation theory in computational complexity”, and it is often important to understand when such functions depend essentially only on a few number of coordinates.

In this work, we study the effects of random restrictions on such bounded functions of low degree. The study of random restrictions of Boolean functions has been invaluable for establishing circuit lower bounds (see for example, [9, 10, 21]), proof complexity (see for example, [5, 11]), de-randomization (see for example, [12]), and several other areas. Our main result shows a remarkable similarity in the behavior of bounded functions with their Boolean cousins, when hit with random restrictions.



© Sreejata Kishor Bhattacharya;
licensed under Creative Commons License CC-BY 4.0
16th Innovations in Theoretical Computer Science Conference (ITCS 2025).
Editor: Raghu Meka; Article No. 17; pp. 17:1–17:21



Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

17:2 Random Restrictions of Bounded Low Degree Polynomials Are Juntas

By now, it is worth recalling that a classical result of Dinur, Friedgut, Kindler and O’Donnell [7] showed that bounded functions whose Fourier tail above degree k is $\leq \exp(-O(k^2))$ are approximable by juntas with arity $\exp(O(k))$. They also showed that this result is tight in general. We are interested in the case when the function has small degree d . What additional structure follows from this stronger condition? To take cue on what to expect, let’s recall the case of Boolean functions. If $f : \{\pm 1\}^n \rightarrow \{0, 1\}$ is a degree d Boolean function, a seminal result of Beals, Buhrman, Cleve, Mosca, and de Wolf [4] shows that f can be represented by a DNF of width $d^{O(1)}$. By the celebrated Hastad’s switching lemma, this implies that if f is hit with a random restriction with survival probability $1/d^{O(1)}$, with high probability f becomes constant. We show something similar is true for bounded functions as well: if $f : \{\pm 1\}^n \rightarrow [0, 1]$ is a degree d , *bounded* function and f is hit with a random restriction with survival probability $O(\log(d)/d)$, with high probability it can be well-approximated by a junta depending only on $\text{poly}(d)$ coordinates. This is despite the fact that the expected number of variables that remain alive is $O\left(n \cdot \frac{\log d}{d}\right)$.

► **Theorem 1 (Main Theorem).** *For any constants $C_1, C_2 > 0$, there exist constants $C_3, C_4, C_5 > 0$ such that the following holds:*

Let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d polynomial and let ρ be a random restriction with survival probability $\frac{\log(d)}{C_3 d}$. With probability at least $1 - d^{-C_2}$, f_ρ is a $(d^{-C_1} \text{Var}[f], \text{Var}[f]^{-2} d^{C_4})$ junta.

Moreover, if J denotes the set of coordinates on which the junta depends, for each $j \in J$ we have $\text{Inf}_j[f] \geq \text{Var}[f]^2 d^{-C_5}$.

This result has interesting consequences for the Aaronson-Ambainis conjecture, which we explain next.

One of the central open problems in the field of quantum query complexity is determining if there exists a partial function which is defined on a large fraction of the Boolean hypercube (say, constant) but whose quantum query complexity and classical query complexity are super-polynomially separated. The result of Beals et. al. [4] shows that no such separation is possible when the function is defined on the entire hypercube. On the other hand, functions for which we know such a separation (e.g. - Forrelation [1], Bernstein-Vazirani [6]) are defined on an exponentially small fraction of the hypercube. A possible explanation as to why all known functions exhibiting large gaps between quantum and classical query complexity have very small support size would be the following folklore conjecture:

► **Conjecture 2.** *Let Q be a quantum query algorithm with Boolean output on n qubits making q queries. Let $P : \{\pm 1\}^n \rightarrow [0, 1]$ be given by $P(x) = \Pr[Q \text{ outputs } 1 \text{ on } x]$. For any $\epsilon > 0$, there exists a classical query algorithm A such that $\mathbb{E}[(A(x) - Q(x))^2] \leq \epsilon$ and A makes at most $\text{poly}\left(q, \frac{1}{\epsilon}\right)$ queries.*

It is known that if Q makes at most q queries, then P is given by a polynomial of degree at most $2q$. Although P has more structure than any arbitrary low degree bounded polynomial, it is further conjectured that such structure is not necessary. In other words, we forget the fact that P arises from a quantum query algorithm and instead try to construct a classical query algorithm for *any* bounded low-degree polynomial. This led to the following conjecture (also folklore).

► **Conjecture 3.** *Let $P : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d polynomial. For any $\epsilon > 0$, there exists a classical decision tree T of depth at most $\text{poly}(d, 1/\epsilon)$ such that $\mathbb{E}[(P(x) - T(x))^2] \leq \epsilon$.*

Aaronson and Ambainis [2] proposed the following query algorithm to estimate P : suppose the variance of the function is sufficiently small. Then we terminate the query algorithm and output the average over the unqueried coordinates. If not, we query the coordinate with the highest *influence* and restrict the function according to the response received. We keep doing this until we have made too many queries or the variance has become sufficiently low. In order to show that this algorithm gives an accurate estimate, [2] observed that it is sufficient to prove the following conjecture.

► **Conjecture 4** (Aaronson-Ambainis conjecture). *Let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d polynomial. Then, there exists a coordinate j such that $\text{Inf}_j[f] \geq \text{poly}(1/d, \text{Var}[f])$.*

As a side remark, we mention that O’Donnell et al. [18] had shown previously that functions which can be approximated by decision trees have a coordinate with high influence. So conjectures 3 and 4 are equivalent.

The Aaronson-Ambainis Conjecture has received significant attention in the past few years. In 2006, a result of Dinur, Friedgut, Kindler and O’Donnell [7] showed that the conjecture is true if $\text{poly}(d)$ is replaced by $\exp(d)$. In 2012, Montanaro [16] proved the conjecture in the special case of block-multilinear forms where all coefficients have the same magnitude. In 2016, O’Donnell and Zhao [19] showed that it suffices to prove the conjecture for a special class of polynomials known as *one-block decoupled polynomials*. In 2020, Keller and Klein [13] claimed to have found a proof for the conjecture but their paper had a subtle flaw and turned out to be wrong. More recently, Lovett and Zhang [15] initiated a new line of attack using the notions of *fractional block sensitivity* and *fractional certificate complexity*. In 2022, Bansal, Sinha and de Wolf [3] proved that this conjecture is true for *completely bounded block multilinear forms* – a class of polynomials that captures a special kind of quantum query algorithms.

In this work, we show that Aaronson-Ambainis conjecture is true for a large fraction of random restrictions of f assuming $\text{Var}[f]$ is not too low. More precisely, we have:

► **Theorem 5.** *There exist constants $C_1, C_2 > 0$ such that the following holds: let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d polynomial ($d \geq 2$) with $\text{Var}[f] \geq 1/d$. Let ρ denote a random restriction with survival probability $\frac{\log(d)}{C_1 d}$. Then,*

$$\Pr \left[f_\rho \text{ has a coordinate with influence } \geq \frac{\text{Var}[f]^2}{d^{C_2}} \right] \geq \frac{\text{Var}[f] \log(d)}{50C_1 d}.$$

We view this result as evidence that the Aaronson-Ambainis Conjecture is true, and hope it gives new insights into the conjecture.

We also observe that our result implies one of the two main unconditional results proven in a recent paper of Lovett and Zhang [15] about the existence of small sensitive blocks albeit with a slightly different set of parameters.

Let $f : \{\pm 1\}^n \rightarrow [0, 1]$. An input $x \in \{\pm 1\}^n$ is said to be (r, ϵ) sensitive if there exists a y such that $d(x, y) \leq r$ and $|f(x) - f(y)| \geq \epsilon$. [15] proves the following:

► **Theorem 6** (Lovett and Zhang [15]). *If $f : \{\pm 1\}^n \rightarrow [0, 1]$ has degree d , then at least $\Omega(\text{Var}[f])$ fraction of the inputs are (r, ϵ) sensitive where $\epsilon = \text{poly}(\text{Var}[f]/d)$, $r = \text{poly}(d, 1/\epsilon, \log(n))$*

An immediate consequence of our result is that at least $\Omega((\text{Var}[f]/d)^{O(1)})$ fraction of inputs are $(1, \epsilon)$ sensitive where $\epsilon = \text{poly}(\text{Var}[f]/d)$. Thus, while we lose a bit in the fraction of sensitive inputs, we gain by letting our block size be exactly 1 instead of $\text{poly}(d, 1/\epsilon, \log(n))$.

It is interesting to remark that our techniques and those of Lovett and Zhang seem quite different. We further remark that [15] shows that if one can show *every* input has a small sensitive block, Aaronson-Ambainis conjecture will follow.

We believe that Theorem 5 opens up some possible lines of attack to the Aaronson-Ambainis conjecture:

- Assuming a supposed counterexample $f : \{\pm 1\}^n \rightarrow [0, 1]$, modify it appropriately (e.g., by composing it with some appropriate gadget or applying a low noise operator) to get a function $\tilde{f} : \{\pm 1\}^{\tilde{n}} \rightarrow [0, 1]$ such that most of its random restrictions remain a counterexample. Combined with our result, this will prove Aaronson-Ambainis conjecture. This approach is discussed in a bit more detail in the conclusion.
- If we can construct a decision tree that queries y (the input to the fixed part) in a few coordinates and outputs an influential coordinate of f_y (the restricted function) with high probability, Aaronson-Ambainis conjecture will follow from the work of Keller and Klein. Again, this approach is discussed in a bit more detail in the conclusion.

2 Organization

We introduce notations and necessary preliminaries in section 3. We give a high level overview of our proof in section 4. In section 5 we compile some lemmas that will be needed in our main proof. Our main results are proven in section 6. Our main result is proven in section 6.2. It says that most random restrictions of a bounded low-degree function can be approximated by a small junta. In section 6.3 we prove that Aaronson-Ambainis conjecture is true for a non-negligible fraction of random restrictions.

3 Notations and preliminaries

Query algorithms

1. A classical query algorithm A (or equivalently, a decision tree) for computing a function $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ can access the input $x \in \{\pm 1\}^n$ by adaptively issuing queries to its bits. We assume internal computations have no cost. The depth of the query algorithm/decision tree is the maximum number of bit queries issued on an input. We say A ϵ -approximates f if $\|f - A\|_2^2 = \mathbf{E}_{x \in \{\pm 1\}^n} [(f(x) - A(x))^2] \leq \epsilon$. For a partial function $f : S(\subseteq \{\pm 1\}^n) \rightarrow \{0, 1\}$, its classical query complexity $D(f)$ is the smallest d for which there exists a decision tree T of depth d such that $T(x) = f(x)$ for all $x \in S$.
2. A quantum query algorithm can access the input $x \in \{\pm 1\}^n$ via an oracle O_x . The register has n qubits and some ancilla qubits. The oracle O_x applies the following unitary operation on the first n qubits:

$$O_x |s_1 s_2 \cdots s_n\rangle = (-1)^{\langle x, s \rangle} |s_1 s_2 \cdots s_n\rangle.$$

The quantum query algorithm applies a sequence of unitary operators, where each operator is either O_x or an input-independent unitary operator U . In the end, it measures the first qubit and outputs the measurement result. The number of queries issued is the number of times O_x is applied.

Notice that a quantum query algorithm Q naturally defines a function $P : \{\pm 1\}^n \rightarrow [0, 1]$:

$$P(x) = \Pr[Q \text{ outputs } 1 \text{ on input } x].$$

It is well-known that if Q makes q queries, then P is a degree $2q$ polynomial.

For a function $f : S(\subseteq \{\pm 1\}^n) \rightarrow \{0, 1\}$, we define its quantum query complexity $Q(f)$ to be the smallest q for which there exists a quantum query algorithm Q making q queries such that for all $x \in S$,

$$\Pr[Q \text{ outputs } 1 \text{ on input } x] \begin{cases} \geq 2/3 & \text{if } f(x) = 1 \\ \leq 1/3 & \text{if } f(x) = 0. \end{cases}$$

Analysis of boolean functions

In this section we recall some results from analysis of boolean functions. A good reference is O' Donnell's textbook [17].

1. Any function $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ has a unique representation as $f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x)$

where $\chi_S(x) = \prod_{i \in S} x_i$. The coefficients $\hat{f}(S)$ are the Fourier coefficients of f . The degree of f is $\max\{|S| \mid \hat{f}(S) \neq 0\}$.

2. The variance of f is

$$\text{Var}_{x \in \{\pm 1\}^n} [f(x)] = \sum_{S \neq \emptyset} \hat{f}(S)^2.$$

3. For a coordinate i , the influence of the i 'th coordinate is defined as

$$\begin{aligned} \text{Inf}_i[f] &= \mathbb{E}_{x \in \{\pm 1\}^n} \left[\left(\frac{f(x) - f(x^{(i)})}{2} \right)^2 \right] \\ &= \sum_{i \in S} \hat{f}(S)^2. \end{aligned}$$

The total influence of f is

$$\text{Inf}[f] = \sum_{i \in [n]} \text{Inf}_i[f] = \sum_S |S| \hat{f}(S)^2.$$

From the Fourier expansion it is clear that if $\deg(f) \leq d$, $\text{Inf}[f] \leq d \text{Var}[f]$.

4. Given two functions $f, g : \{\pm 1\}^n \rightarrow \mathbb{R}$, we say g ϵ -approximates f if $\|f - g\|_2^2 = \mathbb{E}_x [(f(x) - g(x))^2] \leq \epsilon$.
5. For a point $x \in \{\pm 1\}^n$ and a subset $S \subseteq [n]$ and $-1 \leq \rho \leq 1$, we define a distribution $N_{\rho, S}(x)$ on $\{\pm 1\}^n$ as follows:
 - The bits y_1, y_2, \dots, y_n are independent, and

$$\Pr[y_i = x_i] = \begin{cases} 1 & \text{if } i \notin S \\ (1 + \rho)/2 & \text{if } i \in S. \end{cases}$$

When $S = [n]$, we abbreviate $N_{\rho, S}(x)$ by $N_\rho(x)$.

6. For $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ and $-1 \leq \rho \leq 1$ define $T_\rho f : \{\pm 1\}^n \rightarrow \mathbb{R}$ by

$$T_\rho f(x) = \mathbb{E}_{z \leftarrow N_\rho(x)} [f(z)].$$

It is easy to see that the Fourier expansion of $T_\rho f$ is given by

$$T_\rho f(x) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S(x).$$

17:6 Random Restrictions of Bounded Low Degree Polynomials Are Juntas

7. A function $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ is a *junta of arity l* or *l -junta* if there exists a subset $S \subseteq [n], |S| \leq l$ such that f only depends on the coordinates in S . We say f is a (ϵ, l) junta if it can be ϵ -approximated by a l -junta, i.e., there exists a l -junta g such that $\|f - g\|_2^2 \leq \epsilon$.
8. A restriction $\rho = (S, y)$ of $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ is specified by a subset $S \subseteq [n]$ and an assignment $y \in \{\pm 1\}^{[n] \setminus S}$. Such a restriction naturally induces a function $f_\rho : \{\pm 1\}^S \rightarrow \mathbb{R}$. Sometimes we shall write f_y instead of f_ρ (note that S is determined by y since $y \in \{\pm 1\}^{[n] \setminus S}$).

By a random restriction with survival probability p , we mean sampling $\rho = (S, y \in \{\pm 1\}^{[n] \setminus S})$ where each coordinate $i \in [n]$ is included in S with probability p independently, and each bit of y is independently set to a uniformly random bit.

► **Remark 7.** Throughout the paper, all growing parameters (e.g., the degree d) will be assumed to be larger than some sufficiently big constant. This is to make the expressions look neat, as we will be replacing terms like $(C_1)^k \text{poly}(k)$ by $(C_2)^k$ where $C_2 > C_1$.

4 Proof Overview

The main result in this paper is a structural result for bounded low-degree functions similar in spirit to Hastad's switching lemma [10]. Roughly speaking, we show that for some constant $C > 0$, the following holds: let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a polynomial of degree d , and let ρ denote a random restriction with survival probability $\frac{\log(d)}{Cd}$. Then,

$$\Pr [f_\rho \text{ is a } (O(d^{-C}), O(d^C)) \text{ junta}] \geq 1 - \frac{1}{d^{\Omega(1)}}.$$

Once this is established, we can prove Aaronson-Ambainis conjecture for random restrictions as follows: it is easy to see that

$$\Pr \left[\text{Var}[f_\rho] \geq \frac{\text{Var}[f] \log(d)}{2Cd} \right] \geq \frac{\text{Var}[f] \log(d)}{2Cd}.$$

This probability will be significantly more than the failure probability of the main result $\left(\frac{1}{d^{\Omega(1)}}\right)$ (this is the only place where we need the lower bound on $\text{Var}[f]$). So for a $\Omega(\text{Var}[f] \log(d)/d)$ fraction of random restrictions, the variance of f_ρ is high and f_ρ can be well approximated by a junta with arity $\text{poly}(d)$. This means one of the coordinates of the junta must have high influence. This concludes the proof.

Now we give a brief overview of how we prove the main result (Theorem 1). The starting point is the work by Dinur, Friedgut et al. [7] which states the following:

► **Theorem 8** (Dinur et al. [7]). *For any $f : \{\pm 1\}^n \rightarrow [0, 1]$, if*

$$\sum_{|S| > k} \hat{f}(S)^2 \leq \exp(-O(k^2 \log(k)/\epsilon)),$$

then f is a $(\epsilon, 2^{O(k)}/\epsilon^2)$ junta.

In other words, if the Fourier tail above a certain level k is bounded, then f can be well approximated by juntas of arity roughly $2^{O(k)}$.

We start with the observation that random restrictions have bounded Fourier tails (Lemma 15): if the function has degree d and we make a random restriction with survival probability $\frac{\log(d)}{Cd}$ where $C > 1$ is a large constant, using Chernoff bound we can show that with very high probability the Fourier weight above level $\log(d)$ will be low; around the order of $\exp(-\Omega(C \log(d)))$. If we can manage to bring the Fourier weight above $\log(d)$ small enough so that Theorem 8 applies, then we will get that f_ρ can be well approximated by a $\text{poly}(d)$ junta. Unfortunately, if we try this, it turns out that we have to set the survival probability so low that on expectation the variance of f_ρ goes down significantly as well. In other words, while it is true that f_ρ can be well-approximated by juntas, it is for a trivial reason that its variance itself is very low. (And moreover, this is also true for functions $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ that are merely L^2 bounded i.e., $\mathbb{E}[f^2] \leq 1$, so we would not be using any additional structure that arises from the fact that f is *pointwise* bounded.)

In order to make this approach work, we need to improve the tail bound $\exp(-O(k^2 \log(k))/\epsilon)$ in Theorem 8. The problematic term is the quadratic $\exp(-O(k^2))$ in the exponential. If the dominant term were to the order of $\exp(-O(k))$ instead, the calculations would go through. Can we hope to increase the tail bound to $\exp(-O(k))$ while paying a cost by increasing the junta arity? Unfortunately, again, this is not possible: [7] constructs a function which shows that the tail bound is essentially tight upto the $\log(k)$ factor - their function has $\|f^{>k}\|_2^2 \approx \exp(-\Theta(k^2))$ but approximating it to even 1/3 accuracy requires reading $\Omega(n)$ coordinates. Our key observation is that the function constructed by [7] has full degree whereas we are working with random restrictions of a low degree function, so in addition to the fact that the Fourier tail of f_ρ above level $\log(d)$ is very small, we also know that f_ρ has degree d . Can we hope to improve the tail bound in Theorem 8 if we have the additional restriction that the function is of degree d ? Indeed, this turns out to be true. We prove the following result in section 6 (Theorem 28):

► **Theorem 9.** *There exists a constant C such that the following is true:*

If $f : \{\pm 1\}^n \rightarrow [0, 1]$ has degree d and $\sum_{|S|>k} \hat{f}(S)^2 \leq \frac{\epsilon}{C^k d^C}$, f is a $(\epsilon, \epsilon^{-2} d^C C^k)$ junta.

Below we briefly discuss how we are able to improve the tail bound under the additional degree assumption. Dinur et al [7] prove their tail bound (Theorem 8) by showing the following result (we are omitting the exact quantitative parameters here for reading convenience).

► **Theorem 10.** *Let $h : \{\pm 1\}^n \rightarrow \mathbb{R}$ be a degree k function with $\mathbb{E}[h^2] \leq 1$ (but not necessarily pointwise bounded) which cannot be approximated by $2^{O(k)}$ juntas to accuracy μ . Then, for any function $g : \{\pm 1\}^n \rightarrow [0, 1]$, $\mathbb{E}[(h - g)^2] \geq \epsilon$.*

Let's sketch how [7] proves Theorem 8 using Theorem 10. They assume f is not approximable by $2^{O(k)}$ juntas - their goal is then to show that the Fourier tail above level k , $\|f^{\geq k}\|_2^2$, is large. To do so, they use Theorem 10: they take h to be the truncated function $h = f^{\leq k}$ and g to be the original function $g = f$. This then lower bounds $\mathbb{E}[(f - f^{\leq k})^2]$ which is precisely the Fourier tail above weight k . Thus, the distance lower bound ϵ in Theorem 10 governs the Fourier tail lower bound in Theorem 8. Since in Theorem 28 we have the additional information that f is of degree d , for our purposes it will suffice to lower bound the distance of h from bounded degree d functions, not necessarily *all* bounded functions. In section 6.2 we prove a result of the following form (again, we are omitting the exact parameters for reading convenience; for the exact statement see Theorem 26).

► **Theorem 11** (Informal version of Theorem 26). *Let $h : \{\pm 1\}^n \rightarrow \mathbb{R}$ be a degree k function with $\mathbb{E}[h^2] \leq 1$ (but not necessarily pointwise bounded) which cannot be approximated by $2^{O(k)}$ juntas to accuracy μ . Then, for any degree d function $g : \{\pm 1\}^n \rightarrow [0, 1]$, $\mathbb{E}[(h - g)^2] \geq \tilde{\varepsilon}$.*

The parameter $\tilde{\varepsilon}$ in Theorem 11 is bigger than the corresponding ε parameter in Theorem 10 because we are only lower bounding the distance of h from bounded *low-degree* functions whereas Theorem 10 lower bounds the distance of h from *arbitrary* bounded functions. It turns out that the improvement in this parameter is sufficiently good for the calculations to go through.

In order to prove Theorem 11 we shall use the main idea of the proof of [7] along with a structural restriction for bounded low-degree functions proved first in Filmus et al. [8], and later used by Keller-Klein [13] and Lovett-Zhang [15]. Given any function $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ and $x \in \{\pm 1\}^n$ define the *block sensitivity* of f at x to be

$$\text{bs}(f, x) = \sup \left[\sum_{j \in [k]} \left| f(x) - f(x^{(B_j)}) \right| \right]$$

where the supremum ranges over all partitions (B_1, B_2, \dots, B_k) of the variables ($x^{(B_j)}$ denotes x with the coordinates in B_j flipped). Define the block sensitivity of f to be $\text{bs}(f) = \sup_x \text{bs}(f, x)$. We shall use the following fact about bounded low-degree functions:

► **Theorem 12.** *If $f : \{\pm 1\}^n \rightarrow [0, 1]$ has degree d , $\text{bs}(f) \leq 6d^2$.*

We give a high-level overview of how we are able to improve upon the bound of ε using the fact that the block sensitivity of a bounded low degree function is small. At one point in their proof, [7] lower bounds the probability of a linear form of Rademacher random variables $l(x_1, x_2, \dots, x_t) = a_1x_1 + \dots + a_tx_t$ exceeding a certain threshold times its standard deviation, i.e.,

$$\Pr \left[a_1x_1 + \dots + a_tx_t \geq \alpha \sqrt{a_1^2 + \dots + a_t^2} \right].$$

(In the proof, $l(x_1, \dots, x_t)$ is the linear part of a restriction of f : $\sum_{j \in U} \hat{f}_j(U)x_j$ for some $U \subseteq [n]$.)

For each such point x where this linear form is high, [7] shows that *many related points* x' must have $f(x') > 2$. (To be precise, these related points are obtained by applying an appropriate noise operator to x .) Using this they conclude that f must deviate from the interval $[0, 1]$ too often and therefore cannot be approximated by any bounded function.

We follow the proof of [7] up until this point. Instead of directly lower bounding the probability that $a_1x_1 + \dots + a_tx_t$ is high, we partition the set of variables $[t]$ into L blocks B_1, \dots, B_L (L is an appropriately chosen parameter) such that each block gets roughly same total weight: for all $j \in [L]$,

$$\sum_{i \in B_j} a_i^2 \geq \frac{a_1^2 + \dots + a_t^2}{2L}.$$

It will turn out that the a_i 's are sufficiently small for such a partition to exist. For each block we lower bound the probability that the linear form restricted to this block is high:

$$\Pr \left[\sum_{j \in B_i} a_jx_j \geq \tilde{\alpha} \sqrt{\sum_{j \in B_i} a_j^2} \right].$$

Now, on a random assignment z , the linear form restricted to many of these blocks will be high. Take such a block B_i : $\sum_{j \in B_i} a_j z_j \geq \tilde{\alpha} \sqrt{\sum_{j \in B_i} a_j^2}$. For each such block we will be able to find a large number of related points z_i such that $|f(z) - f(z_i)|$ is large. (In our case, these related points are obtained by applying an appropriate noise operator restricted to B_i .) Crucially, these related points will differ from z only at B_i . Thus, we will find many points which differ from z at disjoint sets and whose f differ from z significantly. This will show that f cannot be too close to a bounded low degree function, because those functions have low block sensitivity.

Our advantage is that we need to set α' so that we can conclude $|f(z) - f(z')|$ is only somewhat larger than $\Omega(d^2/L)$ (as opposed to $\Omega(1)$ in [7]) - by setting L large enough this allows us to set a much smaller α and get rid of the quadratic exponential dependence.

5 Tools

In this section we compile some lemmas that we shall use in our proof.

A reverse Markov inequality

We will use the following simple inequality throughout the proof.

► **Lemma 13.** *Let X be a random variable such that $X \leq M$ with probability 1. Let $E[X] = \mu > 0$. Then, $\Pr[X \geq \mu/2] \geq \frac{\mu}{2M}$.*

Proof. Assume $\Pr[X \geq \mu/2] < \frac{\mu}{2M}$. Then,

$$E[X] \leq \Pr[X \geq \mu/2]M + \Pr[X \leq \mu/2] \frac{\mu}{2} < \mu,$$

contradiction. ◀

An anticoncentration inequality for linear forms of Rademacher random variables

► **Lemma 14.** *There exists a universal constant K such that the following holds: let x_1, \dots, x_n be independent Rademacher random variables and let $l(x_1, \dots, x_n) = a_1 x_1 + \dots + a_n x_n$. Let $\sigma = \sqrt{a_1^2 + \dots + a_n^2}$. Suppose $|a_i| \leq \frac{\sigma}{Kt}$. Then,*

$$\Pr[l(x_1, \dots, x_n) \geq t\sigma] \geq \exp(-Kt^2).$$

Proof. Equation 4.2 in [14]. ◀

Random restrictions have small tail

► **Lemma 15.** *Let $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ have degree d , $C > 1$ be a sufficiently large constant, and let $\rho = (S, y \in \{\pm 1\}^{[n] \setminus S})$ be a random restriction with survival probability $\frac{\log(d)}{Cd}$. Let $k = \log(d)$. Then,*

$$E \left[\sum_{|T| > k} \hat{f}_y(T)^2 \right] \leq \exp(-C \log(d)/8) \text{Var}[f].$$

17:10 Random Restrictions of Bounded Low Degree Polynomials Are Juntas

Proof. First suppose $(S, y \in \{\pm 1\}^{[n] \setminus S})$ is a fixed restriction. Note that for $z \in \{\pm 1\}^S$,

$$f_y(z) = \sum_{U \subseteq [n]} \hat{f}(U) \chi_U(y, z)$$

so for $T \subseteq [S]$, $\hat{f}_y(T) = \sum_{U \subseteq S} \hat{f}(U \cup T) \chi_U(y)$. By Parseval's theorem, for a fixed S ,

$$\mathbb{E}_y \left[\hat{f}_y(T)^2 \right] = \sum_{U \subseteq [n] \setminus S} \hat{f}(U \cup T)^2.$$

Therefore, for a fixed S ,

$$\mathbb{E}_y \left[\sum_{|T| > k} \hat{f}_y(T)^2 \right] = \sum_{V \subseteq [n], |V \cap S| > k} \hat{f}(V)^2.$$

Randomizing over S again,

$$\mathbb{E}_{S,y} \left[\sum_{|T| > k} \hat{f}_y(T)^2 \right] = \sum_{V \subseteq [n]} \Pr[|V \cap S| > k] \hat{f}(V)^2.$$

Since f has degree d , we only need to worry about the terms where $|V| \leq d$. Also, for $|V| \leq k$ the relevant probability is 0. Since each element is included in S with probability $\frac{\log(d)}{Cd}$, by Chernoff bound, for each V with $|V| \leq d$,

$$\Pr[|V \cap S| > k] \leq \exp((C-1)^2 k / 4C) \leq \exp(-C \log(d) / 8).$$

Thus we get that

$$\mathbb{E}_{S,y} \left[\sum_{|T| > k} \hat{f}_y(T)^2 \right] \leq \exp(-C \log(d) / 8) \sum_{T \neq \emptyset} \hat{f}(T)^2 = \exp(-C \log(d) / 8) \text{Var}[f]. \quad \blacktriangleleft$$

Random restrictions don't have low variance

► **Lemma 16.** Let $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ be any function and let $\rho = (S, y \in \{\pm 1\}^{[n] \setminus S})$ be a random restriction with survival probability p . Then, $\mathbb{E}[\text{Var}[f_\rho]] \geq p \text{Var}[f]$.

Proof. Fix a restriction $(S, y \in \{\pm 1\}^{[n] \setminus S})$. For each $T \subseteq S$, $\hat{f}_y(T) = \sum_{U \subseteq [n] \setminus S} \hat{f}(T \cup U) \chi_U(y)$.

Thus, by Parseval's theorem, for a fixed S ,

$$\mathbb{E}_y[\text{Var}[f_y]] = \sum_{T: T \cap S \neq \emptyset} \hat{f}(T)^2.$$

Randomizing over S again,

$$\begin{aligned} \mathbb{E}_{S,y}[\text{Var}[f_y]] &= \sum_{T \neq \emptyset} \Pr[T \cap S \neq \emptyset] \hat{f}(T)^2 \\ &= \sum_{T \neq \emptyset} \left(1 - (1-p)^{|T|}\right) \hat{f}(T)^2 \\ &\geq \sum_{T \neq \emptyset} p \hat{f}(T)^2 \\ &= p \text{Var}[f]. \quad \blacktriangleleft \end{aligned}$$

Random restrictions with appropriate survival probability put large Fourier mass on the linear level

► **Lemma 17** (implicit in [7]). *Let $f : \{\pm 1\}^n \rightarrow \mathbb{R}$, $J \subseteq [n]$ and k be such that*

$$\sum_{2^k \leq |T \cap J^c| < 2^{k+1}} \hat{f}(T)^2 \geq \mu.$$

Consider a random restriction $\rho = (S, y \in \{\pm 1\}^{[n] \setminus S})$ where each $j \in J$ is fixed and given a uniformly random assignment, and each $i \in J^c$ is kept alive with probability $p = 2^{-k}$. Then,

$$\mathbb{E} \left[\sum_{i \in S} \hat{f}_y(\{i\})^2 \right] \geq \frac{\mu}{20}.$$

Proof. For a fixed $(S, y \in \{\pm 1\}^{[n] \setminus S})$ (note that $J \cap S = \emptyset$) and $j \in S$ we have

$$\hat{f}_y(\{j\}) = \sum_{T \subseteq [n], T \cap S = \{j\}} \hat{f}(T) \chi_{T \setminus \{j\}}(y).$$

By Parseval's theorem, for a fixed S ,

$$\mathbb{E}_y \left[\hat{f}_y(\{j\})^2 \right] = \sum_{T \subseteq [n], T \cap S = \{j\}} \hat{f}(T)^2.$$

Randomizing over S ,

$$\begin{aligned} \mathbb{E} \left[\sum_{j \in S} \hat{f}_y(\{j\})^2 \right] &= \sum_{T \subseteq [n]} \left[\sum_{j \in [n]} \Pr[T \cap S = \{j\}] \right] \hat{f}(T)^2 \\ &\geq \sum_{2^k \leq |T \cap J^c| < 2^{k+1}} |T| p (1-p)^{|T|-1} \hat{f}(T)^2. \end{aligned}$$

By standard inequalities, for $n \in [1/p, 2/p)$, $np(1-p)^{n-1} \geq 1/20$. It follows that

$$\mathbb{E}_y \left[\hat{f}_y(\{j\})^2 \right] \geq \frac{\mu}{20}. \quad \blacktriangleleft$$

Some hypercontractive inequalities for low degree functions

These lemmas and their proofs can be found in [7].

► **Lemma 18** (Corollary 2.4 in [7]). *There exists a universal constant $W > 0$ such that the following holds:*

Let $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ be a degree k function. Let $\mathbb{E}[f^2] = \sigma^2$. Then,

$$\mathbb{E} \left[f(z)^2 \mathbf{1}_{f(z)^2 \leq W^k \sigma^2} \right] \geq \frac{1}{2} \mathbb{E}[f(z)^2].$$

► **Lemma 19** (Lemma 2.5 in [7]). *There exists a universal constant $B > 0$ such that the following holds:*

Let $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ be a degree k function. Let $\rho \in [-1/2, 1/2]$ be a noise parameter, and $x_0 \in \{\pm 1\}^n$. Suppose $\mathbb{E}_{z \leftarrow N_\rho(x_0)}[f(z) - f(x_0)] = \mu \geq 0$. Then,

$$\Pr_{z \leftarrow N_\rho(x_0)} [f(z) - f(x_0) \geq \mu] \geq \frac{1}{B^k}.$$

17:12 Random Restrictions of Bounded Low Degree Polynomials Are Juntas

The noise lemma

This is the main result of [7]. We use a slight variant. First we recall some known results from approximation theory.

► **Lemma 20.** *For any k , there exist constants $\rho_1, \rho_2, \dots, \rho_{k+1} \in [-1/2, 1/2]$ with the following property: for any polynomial of degree k , $p(x) = a_0 + a_1x + \dots + a_kx^k$, there exists a $j \in [k+1]$ such that $p(\rho_j) \geq \frac{a_1}{2(k+1)}$.*

Proof. Page 112 in [20]. ◀

Now we state the lemma.

► **Lemma 21.** *There exists a universal constant $B > 0$ such that the following holds:
Consider a degree k polynomial $f : \{\pm 1\}^n \rightarrow \mathbb{R}$. Let $S \subseteq [n]$ and $\ell(x) = \sum_{i \in S} \hat{f}(\{i\})x_i$.*

Consider an input $x_0 \in \{\pm 1\}^n$ such that $\ell(x_0) \geq \gamma$. Sample a $z \leftarrow \{\pm 1\}^n$ by the following procedure:

1. *Sample $\rho \leftarrow \{\rho_1, \dots, \rho_{k+1}\}$ uniformly at random.*
2. *Sample $z \leftarrow N_{\rho, S}(x_0)$.*

Then,

$$\Pr \left[f(z) - f(x_0) \geq \frac{\gamma}{2(k+1)} \right] \geq \frac{1}{(k+1)B^k}.$$

► **Remark 22.** Observe that z differs from x only in the coordinates of S . This will be crucial later on.

Proof. Take B to be the same universal constant as in Lemma 19. By replacing f with an appropriate restriction if necessary, we can assume $S = [n]$. Consider the polynomial $p(\rho) = T_\rho f(x_0) - f(x_0)$. From the Fourier expansion of noise operator, we see that

$$p(\rho) = \sum_{S \neq \emptyset} \rho^{|S|} \hat{f}(S).$$

This is a degree k polynomial in ρ with linear coefficient $\ell(x_0)$. By Lemma 20, there exists a $h \in [k+1]$ such that $p(\rho_h) \geq \gamma/(2k+2)$. By Lemma 19,

$$\Pr_{z \leftarrow N_{\rho}(x_0)} \left[f(z) - f(x_0) \geq \frac{\gamma}{2(k+1)} \mid \rho = \rho_h \right] \geq \frac{1}{B^k}.$$

We choose $\rho = \rho_h$ in step (1) with probability $1/(k+1)$, so

$$\Pr_{z \leftarrow N_{\rho}(x_0)} \left[f(z) - f(x_0) \geq \frac{\gamma}{2(k+1)} \right] \geq \frac{1}{(k+1)B^k}. \quad \blacktriangleleft$$

Partitioning a set of numbers in a balanced manner

We need an easy lemma about partitioning a set of weights none of which is too large into disjoint buckets where each bucket gets roughly the same total weight. We will later use this lemma on the set of small linear Fourier coefficients of a function.

► **Lemma 23.** Let a_1, a_2, \dots, a_n be a set of non-negative real numbers and $1 \leq L \leq n$. Suppose $a_i \leq \frac{a_1 + a_2 + \dots + a_n}{2L}$ for all $1 \leq i \leq n$. Then, there exists a partition (B_1, B_2, \dots, B_L) of $[n]$ such that for all $1 \leq j \leq L$,

$$\sum_{i \in B_j} a_i \geq \frac{a_1 + \dots + a_n}{2L}.$$

Proof. Start with an arbitrary partition (B_1, B_2, \dots, B_L) . Then, refine it iteratively according to the following algorithm.

Refinement algorithm:

1. Locate a j such that the condition is violated for j , i.e.,

$$\sum_{i \in B_j} a_i < \frac{a_1 + \dots + a_n}{2L}.$$
 If no such j exists, terminate.
2. Locate a k such that

$$\sum_{i \in B_k} a_i \geq \frac{a_1 + \dots + a_n}{L}.$$
3. Take an arbitrary $l \in B_k$ such that $a_l \neq 0$ and place it in B_j ;

$$B_k \leftarrow B_k \setminus \{l\}$$

$$B_j \leftarrow B_j \cup \{l\}$$

An appropriate k always exists in step (2) by an averaging argument. Since $a_l \leq \frac{a_1 + \dots + a_n}{2L}$, the size of B_k does not go below $\frac{a_1 + \dots + a_n}{2L}$ after step (3). It is easy to see this procedure must terminate. Formally, notice that the quantity

$$\sum_{j \in [L]} \min \left(\frac{a_1 + \dots + a_n}{2L} - \sum_{i \in B_j} a_i, 0 \right)$$

reduces by $\frac{1}{2L} \min\{a_i | a_i \neq 0\}$ at each step, so at some point of time it must be 0 at which point the algorithm terminates and returns a valid partition. ◀

Bounded low-degree functions have low block sensitivity

► **Lemma 24.** Let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a polynomial of degree d . Then, $bs(f) \leq 6d^2$.

Proof. Let B_1, B_2, \dots, B_k be any partition of $[n]$. Our goal is to show that for all $x \in \{\pm 1\}^n$,

$$\sum_{i=1}^k |f(x) - f(x^{(B_i)})| \leq 6d^2.$$

Let $g : \{\pm 1\}^k \rightarrow [0, 1]$ be the function obtained from f obtained by identifying the variables of B_i with a single variable y_i for $1 \leq i \leq k$. By Proposition 3.7 in [8], we have, for every

17:14 Random Restrictions of Bounded Low Degree Polynomials Are Juntas

$$y \in \{\pm 1\}^k,$$

$$\sum_{i=1}^k |g(y) - g(y^{(i)})| \leq 6d^2.$$

This implies

$$\sum_{i=1}^k |f(x) - f(x^{(B_i)})| \leq 6d^2,$$

as desired. ◀

6 Main results

6.1 Improved tail bound for low degree functions

This section is the core technical part of our work. In Theorem 26 we show that if we have a function $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ (not necessarily bounded) with $\mathbb{E}[f^2] \leq 1$ which cannot be approximated by juntas, then f cannot be well-approximated by bounded low-degree functions. In Theorem 28 we use Theorem 26 on the truncated function $f^{\leq k}$ to conclude that the tail bound appearing in [7] (Theorem 8) can be improved under the additional assumption that f has degree d .

► **Remark 25.** For a subset $J \subseteq [n]$, consider the junta $u : \{\pm 1\}^n \rightarrow \mathbb{R}$ which reads the coordinates of J and outputs the average over the unqueried coordinates. It is easy to see that $u(x) = \sum_{S \subseteq J} \hat{f}(S) \chi_S(x)$, so $\|u - f\|_2^2 = \sum_{S \not\subseteq J} \hat{f}(S)^2$. Thus, u approximates f if and only if

$$\sum_{S \not\subseteq J} \hat{f}(S)^2 \text{ is small.}$$

In fact, it is easy to see that there exists a function u depending only on coordinates of J such that $\|f - u\|_2^2 \leq \epsilon$ if and only if $\sum_{S \not\subseteq J} \hat{f}(S)^2 \leq \epsilon$. This immediately follows from the

Fourier expansion of $f - u$.

► **Theorem 26.** *There exists a constant C such that the following holds:*

Let $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ be a degree k function (not necessarily bounded) with $\mathbb{E}[f^2] \leq 1$. Let $J = \{j \mid \text{Inf}_j[f] \geq \theta\}$ where $\theta = \frac{\mu^2}{C^k d^C}$. If $\sum_{S \not\subseteq J} \hat{f}(S)^2 \geq \mu$, then for any degree d function $g : \{\pm 1\}^n \rightarrow [0, 1]$, $\mathbb{E}[(f(x) - g(x))^2] \geq \delta = \frac{\mu}{C^k d^C}$.

► **Remark 27.** Notice here that although f is not pointwise bounded, g is.

Proof. Let W, B, K be the universal constants from Lemma 18, Lemma 21 and Lemma 14 respectively. We take C to be a constant sufficiently larger than B, K, W .

There exists a t such that

$$\sum_{2^t \leq |S \cap J^c| < 2^{t+1}} \hat{f}(S)^2 \geq \frac{\mu}{\log(k)}.$$

Let $\rho = (U, y \in \{\pm 1\}^{[n] \setminus U})$ ($U \subseteq J^c$) be a random restriction where each $j \in J$ is fixed to a uniformly random assignment, and survival probability for each $j \notin J$ is 2^{-t} . By Lemma 17,

$$\mathbb{E} \left[\sum_{j \in U} \hat{f}_y(\{j\})^2 \right] \geq \frac{\mu}{20 \log(k)}.$$

Fix a U such that

$$\mathbb{E}_{y \in \{\pm 1\}^{[n] \setminus U}} \left[\sum_{j \in U} \hat{f}_y(\{j\})^2 \right] \geq \frac{\mu}{20 \log(k)}.$$

By Parseval's theorem we have for all $j \in U$,

$$\mathbb{E} \left[\hat{f}_y(\{j\})^2 \right] = \sum_{S \cap U = \{j\}} \hat{f}(S)^2 \leq \text{Inf}_j[f].$$

For each $y \in \{\pm 1\}^{[n] \setminus U}$ define $\text{SMALL}_y = \{j \mid \hat{f}_y(\{j\})^2 \leq W^k \text{Inf}_j[f]\}$. Observe that for all y ,

$$\sum_{j \in \text{SMALL}_y} \hat{f}_y(\{j\})^2 \leq W^k \text{Inf}[f] \leq k \cdot W^k \leq (2W)^k.$$

For each $j \in U$ we have from Lemma 18

$$\mathbb{E} \left[\hat{f}_y(\{j\})^2 \mathbf{1}_{\hat{f}_y(\{j\})^2 \leq W^k \text{Inf}_j[f]} \right] \geq \frac{1}{2} \mathbb{E} \left[\hat{f}_y(\{j\})^2 \right].$$

Thus,

$$\mathbb{E}_{y \in \{\pm 1\}^{[n] \setminus U}} \left[\sum_{j \in \text{SMALL}_y} \hat{f}_y(\{j\})^2 \right] \geq \frac{\mu}{40 \log(k)},$$

so applying Lemma 13 ²

$$\Pr_{y \in \{\pm 1\}^{[n] \setminus U}} \left[\sum_{j \in \text{SMALL}_y} \hat{f}_y(\{j\})^2 \geq \frac{\mu}{80 \log(k)} \right] \geq \frac{\mu}{80 \log(k) (2W)^k} \geq \frac{\mu}{(3W)^k}.$$

Call $y \in \{\pm 1\}^{[n] \setminus U}$ for which $\sum_{j \in \text{SMALL}_y} \hat{f}_y(\{j\})^2 \geq \frac{\mu}{80 \log(k)}$ to be *good*. Let $\text{GOOD} = \{y \in \{\pm 1\}^{[n] \setminus U} \mid y \text{ is good}\}$. Let $L = \left\lceil \frac{(2B)^k d^8}{\text{Var}[f]} \right\rceil$. For each good y , choose a partition $\text{DIVIDE}(y) = (B_1, B_2, \dots, B_L)$ of SMALL_y such that for all $1 \leq i \leq L$,

$$\sum_{j \in B_i} \hat{f}_y(\{j\})^2 \geq \frac{\mu}{160L \log(k)}.$$

(If there are multiple such partitions, choose any one of them and call it $\text{DIVIDE}(y)$.)

Our choice of parameters ensures that for all $j \in \text{SMALL}_y$, $\hat{f}_y(\{j\})^2 \leq \frac{\mu}{160L \log(k)}$, so such a partition exists by Lemma 23. Let $\rho_1, \rho_2, \dots, \rho_{k+1}$ be the constants from Lemma 20.

Suppose, for the sake of contradiction, there exists a degree d polynomial $g : \{\pm 1\}^n \rightarrow [0, 1]$ such that $\mathbb{E}[(f(x) - g(x))^2] \leq \delta$. Throughout the rest of the proof, for a string $s_1 \in \{\pm 1\}^{[n] \setminus U}$ and a string $s_2 \in \{\pm 1\}^U$, the pair (s_1, s_2) denotes the string $s \in \{\pm 1\}^n$ which agrees with s_1 on $[n] \setminus U$ and with s_2 on U . Consider the following randomized procedure which returns a real number.

² See remark 7.

17:16 Random Restrictions of Bounded Low Degree Polynomials Are Juntas

Procedure 1:

1. Sample a $y \in \text{GOOD}$ uniformly at random.
2. Sample $\rho \leftarrow \{\rho_1, \rho_2, \dots, \rho_{k+1}\}$ uniformly at random.
3. Sample $z \leftarrow \{\pm 1\}^U$ uniformly at random.
4. Let $\text{DIVIDE}(y) = (B_1, B_2, \dots, B_L)$. Sample $\tilde{z}^{(i)} \leftarrow N_{B_i, \rho}(z)$ for $1 \leq i \leq L$.
5. Return $\sum_{i=1}^L |f(y, z) - f(y, \tilde{z}^{(i)})|$.

We estimate the probability that procedure 1 returns a number $> 15d^2$ in two different ways. First, we obtain a lower bound from the definition of GOOD. Then, we obtain an upper bound using the assumption that $\mathbb{E}[(f(x) - g(x))^2] \leq \delta$ and the fact that $\Pr[y \text{ is good}]$ is somewhat large. These two bounds will contradict each other - and that will prove the theorem.

Lower bound. Fix a $y \in \text{GOOD}$. Let $\text{DIVIDE}(y) = (B_1, B_2, \dots, B_L)$.

Let $w = \sqrt{\frac{\mu}{160L \log(k)}}$. For each $i \in [L]$ we have

$$\sqrt{\sum_{j \in B_i} \hat{f}_y(\{j\})^2} \geq w.$$

Choose α such that $\alpha w = \frac{100d^4(2B)^k}{L}$. Our choice of L ensures that $\alpha \leq 1$. Moreover, our choice for influence threshold θ ensures that $|\hat{f}_y(\{j\})| \leq \frac{w}{K\alpha}$ for all $j \in B_i$ where K is the universal constant from the Lemma 14.

Therefore, we can apply Lemma 14 to obtain that

$$\Pr_{z \in \{\pm 1\}^{B_i}} \left[\sum_{j \in B_i} z_j \hat{f}_y(\{j\}) \geq \frac{100d^4(2B)^k}{L} \right] \geq \exp(-K\alpha^2) \geq \frac{1}{K_1}.$$

Here $K_1 = \exp(K)$ is an absolute constant.

By Lemma 21 applied on the restriction $f_y : \{\pm 1\}^U \rightarrow \mathbb{R}$, as we sample $z \leftarrow \{\pm 1\}^U$ u.a.r, $\rho \leftarrow \{\rho_1, \dots, \rho_{k+1}\}$ u.a.r, $\tilde{z}^{(i)} \leftarrow N_{\rho, B_i}(z)$, we have that

$$\Pr \left[|f(y, z) - f(y, \tilde{z}^{(i)})| \geq \frac{30d^3(2B)^k}{L} \right] \geq \frac{1}{K_1(k+1)B^k} \geq \frac{1}{(2B)^k}.$$

By linearity of expectation,

$$\mathbb{E} \left[\left| \left\{ i \in [L] \mid |f(y, z) - f(y, \tilde{z}^{(i)})| \geq \frac{30d^3(2B)^k}{L} \right\} \right| \right] \geq \frac{L}{(2B)^k}.$$

Using Lemma 13,

$$\Pr \left[\left| \left\{ i \in [L] \mid |f(y, z) - f(y, \tilde{z}^{(i)})| \geq \frac{30d^3(2B)^k}{L} \right\} \right| \geq \frac{L}{2 \times (2B)^k} \right] \geq \frac{1}{2 \times (2B)^k}.$$

Observe that

$$\left| \left\{ i \in [L] \mid |f(y, z) - f(y, \tilde{z}^{(i)})| \geq \frac{30d^3(2B)^k}{L} \right\} \right| \geq \frac{L}{2 \times (2B)^k} \implies \sum_{i \in [L]} |f(y, z) - f(y, \tilde{z}^{(i)})| \geq 15d^3.$$

We conclude that for all $y \in \text{GOOD}$, as $z, \tilde{z}^{(1)}, \dots, \tilde{z}^{(L)}$ are sampled as in Procedure 1,

$$\Pr \left[\sum_{i \in [L]} |f(y, z) - f(y, \tilde{z}^{(i)})| \geq 15d^3 \right] \geq \frac{1}{2 \times (2B)^k}.$$

Thus, with probability at least $\frac{1}{2 \times (2B)^k}$, procedure 1 returns a number greater than $15d^2$ (as $15d^3 > 15d^2$).

Upper bound. Since $\mathbb{E}[(f(x) - g(x))^2] \leq \delta$ and $\Pr_{y \in \{\pm 1\}^{[n] \cup U}} [y \text{ is good}] \geq \mu/(3W)^k$, we have that

$$\mathbb{E}[(f(x) - g(x))^2 | x_{[n] \cup U} \text{ is good}] \leq \frac{\delta}{\mu} (3W)^k.$$

Now consider a uniformly sampled $y \in \text{GOOD}$. Observe that as we sample $z \leftarrow \{\pm 1\}^U$ u.a.r, $\rho \leftarrow \{\rho_1, \dots, \rho_{k+1}\}$ u.a.r and $\tilde{z}^{(i)} \leftarrow N_{\rho, B_i}(z)$, the marginal distribution of $\tilde{z}^{(i)}$ is uniform on $\{\pm 1\}^U$. By Markov's inequality, we have

$$\Pr \left[(f(y, z) - g(y, z))^2 \geq \frac{1}{L^2} \right] \leq \frac{L^2 \delta}{\mu} (3W)^k$$

and for all $i \in [L]$,

$$\Pr \left[(f(y, \tilde{z}^{(i)}) - g(y, \tilde{z}^{(i)}))^2 \geq \frac{1}{L^2} \right] \leq \frac{L^2 \delta}{\mu} (3W)^k.$$

By union bound, the probability that $(f(y, z) - g(y, z))^2 \geq \frac{1}{L^2}$ or for some i , $(f(y, \tilde{z}^{(i)}) - g(y, \tilde{z}^{(i)}))^2 \geq \frac{1}{L^2}$ is at most $(L+1) \frac{L^2 \delta}{\mu} (3W)^k \leq \frac{2L^3 \delta}{\mu} (3W)^k$. Our choice of δ ensures that this quantity is less than $< \frac{1}{2 \times (2B)^k}$. Observe that if none of these bad events holds, since the block sensitivity of g is bounded above by $6d^2$ (Theorem 24), we have that

$$\sum_{i \in [L]} |g(y, z) - g(y, \tilde{z}^{(i)})| \leq 6d^2 \implies \sum_{i \in [L]} |f(y, z) - f(y, \tilde{z}^{(i)})| \leq 6d^2 + 1 < 15d^2.$$

Thus, we conclude

$$\Pr \left[\sum_{i \in [L]} |f(y, z) - f(y, \tilde{z}^{(i)})| > 15d^2 \right] < \frac{2L^3 \delta}{\mu} (3W)^k < \frac{1}{2 \times (2B)^k}.$$

As promised, we get conflicting lower and upper bounds for the probability that procedure 1 returns a number $> 15d^2$. This is our desired contradiction. \blacktriangleleft

Now we show that we can improve the tail bound of Theorem 8 under the additional assumption that f has low degree. This follows straightforwardly from Theorem 26.

► Theorem 28. *There exists a universal constant $C > 0$ such that the following is true:*

Let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d function. Let $\theta = \frac{\text{Var}[f]^2}{d^C C^k}$ and $J = \{j | \text{Inf}_j[f] \geq \theta\}$. If

$$\sum_{S \not\subseteq J} \hat{f}(S)^2 \geq \mu, \text{ then } \sum_{|S| > k} \hat{f}(S)^2 \geq \frac{\mu}{d^C C^k}.$$

17:18 Random Restrictions of Bounded Low Degree Polynomials Are Juntas

Proof. Assume $\sum_{|S|>k} \hat{f}(S)^2 < \mu/2$ (otherwise we are done). Let \tilde{C} be the universal constant from Theorem 26.

The idea is to apply Theorem 26 to the truncated function

$$f^{\leq k}(x) = \sum_{|S|\leq k} \hat{f}(S)\chi_S(x).$$

Note that while $f^{\leq k}$ is not pointwise bounded, it satisfies $\mathbb{E}[(f^{\leq k})^2] \leq 1$ and $\text{Inf}_j[f^{\leq k}] \leq \text{Inf}_j[f]$ for all j (this is clear from the Fourier expressions). Let $H = \{j | \text{Inf}_j[f^{\leq k}] \geq \theta\}$. We have $H \subseteq J$, so

$$\sum_{S \notin H} \hat{f}^{\leq k}(S)^2 \geq \sum_{S \notin J} \hat{f}(S)^2 - \frac{\mu}{2} \geq \frac{\mu}{2}.$$

Applying Theorem 26, we get that for any bounded degree d $g : \{\pm 1\}^n \rightarrow [0, 1]$, $\mathbb{E}[(f(x) - g(x))^2] \geq \frac{\mu}{2d^{\tilde{C}}\tilde{C}^k}$. Taking g to be our original function f , we get the desired tail lower bound:

$$\mathbb{E}[(f - f^{\leq k})^2] \geq \frac{\mu}{2d^{\tilde{C}}\tilde{C}^k} \implies \sum_{|S|>k} \hat{f}(S)^2 > \frac{\mu}{2d^{\tilde{C}}\tilde{C}^k}.$$

Taking C to be a slightly larger constant than \tilde{C} , we get that

$$\sum_{|S|>k} \hat{f}(S)^2 \geq \frac{\mu}{d^C C^k}. \quad \blacktriangleleft$$

6.2 Random restrictions can be approximated by juntas

In this section we use the fact that random restrictions have bounded tails to show that they can be approximated by juntas. This is our main result.

► **Theorem 29 (Restatement of Theorem 1).** *For any constants $\tilde{C}_1, \tilde{C}_2 > 0$, there exist constants $\tilde{C}_3, \tilde{C}_4, \tilde{C}_5 > 0$ such that the following holds:*

Let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d polynomial and let ρ be a random restriction with survival probability $\frac{\log(d)}{\tilde{C}_3 d}$. With probability at least $1 - d^{-\tilde{C}_2}$, f_ρ is a $(d^{-\tilde{C}_1} \text{Var}[f], \text{Var}[f]^{-2} d^{\tilde{C}_4})$ junta. Moreover, if J denotes the set of coordinates on which the junta depends, for each $j \in J$ we have $\text{Inf}_j[f] \geq \text{Var}[f]^2 d^{-\tilde{C}_5}$.

Proof. We consider a random restriction with survival probability $\frac{\log(d)}{\tilde{C}_3 d}$.

By Lemma 15, the expected Fourier tail of f_ρ above level $\log(d)$ is at most $\exp(-\tilde{C}_3 \log(d)/8) \text{Var}[f] = \frac{\text{Var}[f]}{d^{\tilde{C}_3/8}}$. By Markov's inequality, with probability at least

$1 - d^{\tilde{C}_3/16}$, the Fourier tail above $\log(d)$ is $\leq \frac{\text{Var}[f]}{d^{\tilde{C}_3/16}}$. Let C be the constant from Theorem

28. Let $\mu = \frac{\text{Var}[f]}{d^{\tilde{C}_3/16}} d^C C^{\log(d)} = \frac{\text{Var}[f]}{d^{\tilde{C}_3/16}} d^{2C}$, $\theta = \frac{\mu^2}{d^C C^{\log(d)}} = \frac{\mu^2}{d^{2C}}$ and $J = \{j | \text{Inf}_j[f_\rho] \geq \theta\}$.

Let $u_\rho : \{\pm 1\}^n \rightarrow [0, 1]$ be the function which reads the coordinates in J and outputs the average of f_ρ over the coordinates in J^c . Choose \tilde{C}_3 large enough so that $\mu \leq d^{-\tilde{C}_1} \text{Var}[f]$. Applying Theorem 28, we see that u_ρ approximates f_ρ to accuracy $d^{-\tilde{C}_1} \text{Var}[f]$. Using the fact that total influence is bounded by d , we see that u_ρ has arity $\leq \text{Var}[f]^{-2} d^{C'} \tilde{C}_3$ for a universal constant C' . Taking $(\tilde{C}_4, \tilde{C}_5) = (C' \tilde{C}_3, \tilde{C}_3/32 - 2C)$, we are done. \blacktriangleleft

6.3 Aaronson-Ambainis Conjecture is true for random restrictions

► **Theorem 30 (Restatement of Theorem 5).** *There exist constants $C_1, C_2 > 0$ such that the following holds: let $f : \{\pm 1\}^n \rightarrow [0, 1]$ be a degree d polynomial ($d \geq 2$) with $\text{Var}[f] \geq 1/d$. Let ρ denote a random restriction with alive probability $\frac{\log(d)}{C_1 d}$. Then,*

$$\Pr \left[f_\rho \text{ has a coordinate with influence } \geq \frac{\text{Var}[f]^2}{d^{C_2}} \right] \geq \frac{\text{Var}[f] \log(d)}{50 C_1 d}.$$

Proof. Let M be a large constant. Apply Theorem 29 with $(\tilde{C}_1, \tilde{C}_2) = (M, M)$ to get constants $\tilde{C}_3, \tilde{C}_4, \tilde{C}_5$. Let ρ be a random restriction with survival probability $\frac{\log(d)}{\tilde{C}_3 d}$. By Lemma 16,

$$\mathbb{E}[\text{Var}[f_\rho]] \geq \frac{\text{Var}[f] \log(d)}{\tilde{C}_3 d}$$

so by Lemma 13,

$$\Pr \left[\text{Var}[f_\rho] \geq \frac{\text{Var}[f] \log(d)}{2\tilde{C}_3 d} \right] \geq \frac{\text{Var}[f] \log(d)}{2\tilde{C}_3 d}.$$

Since $\text{Var}[f] \geq 1/d$, $d^{-M} \leq \frac{\text{Var}[f] \log(d)}{10\tilde{C}_3 d}$. By Theorem 29 and Remark 25, with probability at least $1 - d^{-M}$, there exists a $J_\rho \subseteq [n]$ such that every coordinate in J_ρ has influence $\geq \text{Var}[f_\rho]^2 d^{-\tilde{C}_5}$ and

$$\sum_{S \not\subseteq J_\rho} \hat{f}_\rho(S)^2 \leq d^{-M} \text{Var}[f].$$

So with probability at least $\frac{\text{Var}[f] \log(d)}{2\tilde{C}_3 d} - d^{-M} \geq \frac{\text{Var}[f] \log(d)}{4\tilde{C}_3 d}$, both these events (high variance of f_ρ and existence of J_ρ) hold and we have that

$$\sum_{S \subseteq J_\rho} \hat{f}_\rho(S)^2 \geq \text{Var}[f_\rho] - d^{-M} \text{Var}[f] \geq \frac{\text{Var}[f] \log(d)}{4\tilde{C}_3 d}.$$

In particular, we have that $J_\rho \neq \emptyset$. Since for each $j \in J_\rho$ we have $\text{Inf}_j[f_\rho] \geq \text{Var}[f_\rho]^2 d^{-\tilde{C}_5}$, we are done by taking $(C_1, C_2) = (\tilde{C}_3, 2 + \tilde{C}_5)$. ◀

7 Conclusions and further directions

In this paper, we showed that if $f : \{\pm 1\}^n \rightarrow [0, 1]$ is a degree d polynomial, a large fraction of its random restrictions have an influential coordinate.

It would be interesting to see if we can extend this to the full Aaronson-Ambainis conjecture. We describe a few potential approaches here.

- Given a degree d polynomial $f : \{\pm 1\}^n \rightarrow [0, 1]$, we can lift it with a Boolean function $g : \{\pm 1\}^m \rightarrow \{\pm 1\}^n$ each of whose coordinates g_i is unbiased and given by a low degree function. Then, the lifted polynomial $f \circ g : \{\pm 1\}^m \rightarrow [0, 1]$ will be a low degree polynomial. As long as the g_i 's are pairwise independent, the variance of f will be preserved as well. Our result shows that a large fraction of random restrictions of

- $f \odot g$ have an influential coordinate. Can we construct g_1, g_2, \dots, g_n appropriately such that this allows us to conclude f must have an influential coordinate as well? The g_i 's should introduce correlations between the different input bits of f so that most random restrictions of $f \odot g^m$ look the same in some appropriate sense.
- Our result shows that for a non-negligible fraction of $y \in \{\pm 1\}^{[n] \setminus U}$, there exists a j such that $\text{Inf}_j[f_y] \geq \text{Var}[f]^2/d^{O(1)}$. Can we construct a decision tree of height $\text{poly}(d)$ which, on input y , outputs one of the influential coordinates of f_y with high probability? If this is possible, this will imply Aaronson-Ambainis via the approach of Keller and Klein. We may additionally assume that for all j , $\text{Inf}_j[f] = \mathbb{E}_y[\text{Inf}_j[f_y]] \leq \text{Var}[f]^2/d^{10000}$ (otherwise the result holds for f anyway). So basically the situation looks like this: for a non-negligible fraction of y , for some j , $\text{Inf}_j[f_y]$ is significantly higher than its expectation ($\text{Inf}_j[f]$) – and we want to probe y in a few bits to locate such a j .

References

- 1 Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *Electron. Colloquium Comput. Complex.*, TR14-155, 2014. URL: <https://eccc.weizmann.ac.il/report/2014/155>.
- 2 Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *Theory Comput.*, 10:133–166, 2014. doi:10.4086/TOC.2014.V010A006.
- 3 Nikhil Bansal, Makrand Sinha, and Ronald de Wolf. Influence in completely bounded block-multilinear forms and classical simulation of quantum algorithms. In Shachar Lovett, editor, *37th Computational Complexity Conference, CCC 2022, July 20-23, 2022, Philadelphia, PA, USA*, volume 234 of *LIPICs*, pages 28:1–28:21. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICs.CCC.2022.28.
- 4 Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. In *39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA*, pages 352–361. IEEE Computer Society, 1998. doi:10.1109/SFCS.1998.743485.
- 5 Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, Pavel Pudlák, and Alan R. Woods. Exponential lower bounds for the pigeonhole principle. In S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis, editors, *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 200–220. ACM, 1992. doi:10.1145/129712.129733.
- 6 Ethan Bernstein and Umesh V. Vazirani. Quantum complexity theory. *SIAM J. Comput.*, 26(5):1411–1473, 1997. doi:10.1137/S0097539796300921.
- 7 Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O’Donnell. On the fourier tails of bounded functions over the discrete cube. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing, STOC '06*, pages 437–446, New York, NY, USA, 2006. Association for Computing Machinery. doi:10.1145/1132516.1132580.
- 8 Yuval Filmus and Hamed Hatami. Bounds on the sum of L1 influences. *CoRR*, abs/1404.3396, 2014. arXiv:1404.3396.
- 9 Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Math. Syst. Theory*, 17(1):13–27, 1984. doi:10.1007/BF01744431.
- 10 Johan Håstad. Almost optimal lower bounds for small depth circuits. In Juris Hartmanis, editor, *Proceedings of the 18th Annual ACM Symposium on Theory of Computing, May 28-30, 1986, Berkeley, California, USA*, pages 6–20. ACM, 1986. doi:10.1145/12130.12132.
- 11 Johan Håstad. On small-depth frege proofs for PHP. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 37–49. IEEE, 2023. doi:10.1109/FOCS57990.2023.00010.

- 12 Russell Impagliazzo, Raghu Meka, and David Zuckerman. Pseudorandomness from shrinkage. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 111–119. IEEE Computer Society, 2012. doi:10.1109/FOCS.2012.78.
- 13 Nathan Keller and Ohad Klein. Quantum speedups need structure, 2019. arXiv:1911.03748.
- 14 Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Springer Berlin Heidelberg, 1991. doi:10.1007/978-3-642-20212-4.
- 15 Shachar Lovett and Jiapeng Zhang. Fractional Certificates for Bounded Functions. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 84:1–84:13, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2023.84.
- 16 Ashley Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12), December 2012. doi:10.1063/1.4769269.
- 17 Ryan O’Donnell. Analysis of boolean functions, 2021. arXiv:2105.10386.
- 18 Ryan O’Donnell, Michael E. Saks, Oded Schramm, and Rocco A. Servedio. Every decision tree has an influential variable. *CoRR*, abs/cs/0508071, 2005. arXiv:cs/0508071.
- 19 Ryan O’Donnell and Yu Zhao. Polynomial bounds for decoupling, with applications. *CoRR*, abs/1512.01603, 2015. arXiv:1512.01603.
- 20 Theodore J. Rivlin. *Chebyshev polynomials – From approximation theory to algebra and number theory: 2nd edition*. John Wiley & Sons Limited, 1990.
- 21 Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An average-case depth hierarchy theorem for boolean circuits. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1030–1048. IEEE Computer Society, 2015. doi:10.1109/FOCS.2015.67.