Electronic Colloquium on Computational Complexity, Report No. 35 (2024)

AARONSON-AMBAINIS CONJECTURE IS TRUE FOR RANDOM RESTRICTIONS

SREEJATA KISHOR BHATTACHARYA

Abstract. In an attempt to show that the acceptance probability of a quantum query algorithm making q queries can be well-approximated almost everywhere by a classical decision tree of depth $\leq \mathsf{poly}(q)$, Aaronson and Ambainis proposed the following conjecture: let $f : \{\pm 1\}^n \to [0,1]$ be a degree d polynomial with variance $\geq \epsilon$. Then, there exists a coordinate of f with influence $\geq \mathrm{poly}(\epsilon, 1/d)$.

We show that for any polynomial $f : \{\pm 1\}^n \to [0,1]$ of degree d ($d \geq 2$) and variance $\mathsf{Var}[f] \geq 1/d$, if ρ denotes a random restriction with survival probability $\dfrac{\log(d)}{C_1 d}$,

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

where $C_1, C_2 > 0$ are universal constants. Thus, Aaronson-Ambainis conjecture is true for a non-negligible fraction of random restrictions of the given polynomial assuming its variance is not too low.

1. Introduction

One of the central open problems in the field of quantum query complexity is finding 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 seminal result of Beals, Burhman 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 [5]) 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 1.1. Let Q be a quantum query algorithm with Boolean output on n qubits making q queries. Let $P : \{\pm 1\}^n \to [0,1]$ 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 $\mathsf{E}[(A(x) - Q(x))^2] \leq \epsilon$ and A makes at most $\mathsf{poly}\left(q, \dfrac{1}{\epsilon}\right)$ queries.

ISSN 1433-8092

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

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

As a side remark, we mention that O'Donnell et al. [14] had shown previously that functions which can be approximated by decision trees have a coordinate with high influence. So conjectures 1.2 and 1.3 are equivalent.

Aaronson-Ambainis conjecture has received significant attention in the past few years. A 2006 result of Dinur, Friedgut et al. [7] shows that the conjecture is true if $\mathsf{poly}(d)$ is replaced by $\exp(d)$. In 2012, Montanaro [12] proved the conjecture in the special case of block-multilinear forms where all coefficients have the same magnitude. In 2016, O'Donnell and Zhao [15] 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 [9] 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 [11] initiated a new line of attack using the notions of *fractional block sensitivity* and *fractional certificate complexity*. In 2022, Bansal, Sinha, 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 $\mathsf{Var}[f]$ is not too low. We hope our result gives new insights to the Aaronson-Ambainis conjecture. In particular, this opens up a possible line of attack:

- Assuming a supposed counterexample $f : \{\pm 1\}^n \to [0,1]$, modify it appropriately (e.g., by composing it with some appropriate gadget or applying

a low noise operator) to get a function $\tilde{f} : \{\pm 1\}^{\tilde{n}} \to [0,1]$ 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.

Our main result is a new structural restriction about bounded low-degree polynomials over the hypercube. While several structural results are known about low-degree *boolean* functions $f : \{\pm 1\}^n \to \{0,1\}$, such results are rare for low-degree *bounded* functions $f : \{\pm 1\}^n \to [0,1]$. We show that if $f : \{\pm 1\}^n \to [0,1]$ has degree d and ρ is a random restriction with survival probability $O(\log(d)/d)$, then with very high probability f_ρ depends essentially on $\approx \mathsf{poly}(d)$ coordinates, even though there are $O(n \log(d)/d)$ alive coordinates on average.

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 technical tool is Theorem 6.3, which says that most random restrictions of a bounded low-degree function can be approximated by a small junta. In Theorem 6.4 we prove the result mentioned in the abstract (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 $f : \{\pm 1\}^n \to \mathbb{R}$ 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 $||f - A||_2^2 = \mathsf{E}_{x \in \{\pm 1\}^n}\left[(f(x) - A(x))^2\right] \leq \epsilon$.

For a partial function $f : S(\subseteq \{\pm 1\}^n) \to \{0,1\}$, 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 $P : \{\pm 1\}^n \to [0,1]$:

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

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

For a function $f : S(\subseteq \{\pm 1\}^n) \to \{0,1\}$, 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 $x \in S$,

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

Analysis of boolean functions. In this section we recall some results from analysis of boolean functions. A good reference is O' Donnell's textbook [13].

(1) Any function $f : \{\pm 1\}^n \to \mathbb{R}$ 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 $\max\{|S||\hat{f}(S) \neq 0\}$.

(2) The variance of f is

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

(3) For a cooordinate i, the influence of the i'th coordinate is defined as

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

The total influence of f is

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

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

(4) Given two functions $f, g\{\pm 1\}^n \to \mathbb{R}$, we say g $\epsilon-$approximates f if $||f - g||_2^2 = \mathsf{E}_x\left[(f(x) - g(x))^2\right] \leq \epsilon$.

(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 $y_1, y_2, \cdots, y_n$ are independent, and

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

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

(6) For $f : \{\pm 1\}^n \to \mathbb{R}$ and $-1 \le \rho \le 1$ define $T_\rho f : \{\pm 1\}^n \to \mathbb{R}$ by
$$T_\rho f(x) = \mathsf{E}_{z \leftarrow N_\rho(x)}[f(z)].$$
It is easy to see that the Fourier expansion of $T_\rho f$ is given by
$$T_\rho f(x) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S(x).$$

(7) A function $f : \{\pm 1\}^n \to \mathbb{R}$ is a *junta of arity l* or *l-junta* if there exists a subset $S \subseteq [n], |S| \le l$ 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 $||f - g||_2^2 \le \epsilon$.

(8) A restriction $\rho = (S, y)$ of $f : \{\pm 1\}^n \to \mathbb{R}$ 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 $f_\rho : \{\pm 1\}^S \to \mathbb{R}$. 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 3.1. *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 $(C_1)^k poly(k)$ by $(C_2)^k$ where $C_2 > C_1$.*

4. Proof Overview

The main technical tool in this paper is a structural result for bounded low-degree functions similar in spirit to Hastad's switching lemma [8]. Let $f : \{\pm 1\}^n \to [0, 1]$ be a polynomial of degree d, and let ρ denote a random restriction with survival probability $\dfrac{\log(d)}{Cd}$. We show that for some constant C,
$$\Pr\left[f_\rho \text{ is a } (O(d^{-C}), O(d^C)) \text{ junta}\right] \ge 1 - \frac{1}{d^{\Omega(1)}}.$$

Once this is established, we can prove Aaronson-Ambainis conjecture for random restrictions as follows: it is easy to see that
$$\mathsf{Pr}\left[\mathsf{Var}[f_\rho] \ge \frac{\mathsf{Var}[f]\log(d)}{2Cd}\right] \ge \frac{\mathsf{Var}[f]\log(d)}{2Cd}.$$
This probability will be significantly more than the failure probability of the switching lemma $\left(\dfrac{1}{d^{\Omega(1)}}\right)$ (this is the only place where we need the lower bound on $\mathsf{Var}[f]$). So for a $\approx O(\mathsf{Var}[f]\log(d)/d)$ fraction of random restrictions, the variance of f_ρ is high *and* f_ρ can be well approximated by a junta with arity

$\mathsf{poly}(d)$. 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 switching lemma. The starting point is the work by Dinur, Friedgut et al. [7] which states the following:

Theorem 4.1. *For any $f : \{\pm 1\}^n \to [0, 1]$, 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: if the function has degree d and we make a random restriction with survival probability $\frac{\log(d)}{Cd}$, 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 4.1 applies, then we will get that f_ρ can be well approximated by a $\mathsf{poly}(d)$ 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 $f : \{\pm 1\}^n \to \mathbb{R}$ that are merely L^2 bounded i.e, $\mathsf{E}[f^2] \leq 1$, 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 $\exp(-O(k^2 \log(k))/\epsilon)$. 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 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 4.1 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 Theorem 4.2:.

Theorem 4.2. *There exists a constant C such that the following is true:*

If $f : \{\pm 1\}^n \to [0,1]$ *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 by showing the following result (we are omitting the exact quantitative parameters here for reading convenience).

Theorem 4.3. *Let* $h : \{\pm 1\}^n \to \mathbb{R}$ *be a degree* k *function with* $\mathsf{E}[h^2] \leq 1$ *(but not necessarily pointwise bounded) which cannot be approximated by* $2^{O(k)}$ *juntas to accuracy* μ*. Then, for any function* $g : \{\pm 1\}^n \to [0,1]$, $E[(h-g)^2] \geq \varepsilon$.

To prove Theorem 4.1, [7] applies Theorem 4.3 on the truncated function $h = f^{\leq k} = \sum_{|S|\leq k} \hat{f}(S)\chi_S$ and takes g to be the original function f. This then lower bounds $E[(f - f^{\leq k})^2]$ which is precisely the Fourier tail above weight k. Thus, the distance lower bound ε in Theorem 4.3 governs the Fourier tail lower bound in Theorem 4.1. Since we have the additional information that f is of degree d, for our purposes it will suffice to bound the distance from bounded degree d functions, not necessarily *all* bounded functions. In Theorem 6.1 we prove a result of the following form (again, we are omitting the exact parameters for reading convenience)

Theorem 4.4. *Let* $h : \{\pm 1\}^n \to \mathbb{R}$ *be a degree* k *function with* $\mathsf{E}[h^2] \leq 1$ *(but not necessarily pointwise bounded) which cannot be approximated by* $2^{O(k)}$ *juntas to accuracy* μ*. Then, for any degree* d *function* $g : \{\pm 1\}^n \to [0,1]$, $E[(h-g)^2] \geq \tilde{\varepsilon}$.

The parameter $\tilde{\varepsilon}$ in Theorem 4.4 is bigger than the corresponding ε parameter in Theorem 4.3 because we are only lower bounding the distance of h from bounded *low-degree* functions whereas Theorem 4.3 lower bounds the distance of h from *arbitrary* bounded functions. It turns out that the improvement in this parameter is sufficiently good for the random restriction approach to go through.

In order to prove Theorem 4.4 we shall use the main idea of the proof of [7] along with a structural restriction for bounded low-degree functions discovered first in [4]. Given any function $f : \{\pm 1\}^n \to \mathbb{R}$ and $x \in \{\pm 1\}^n$ define the *block sensitivity* of f at x to be

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

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

Theorem 4.5. [4] *If* $f : \{\pm 1\}^n \to [0,1]$ *has degree* d, $bs(f) \leq 6d^2$.

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 $l(x_1, x_2, \cdots, x_t) = a_1x_1 + \cdots + a_tx_t$ exceeding a certain threshold times its standard deviation, i.e.,

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

For each such point x where this linear form is high, [7] shows that *many related points* x' must have $f(x') > 2$. 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 $a_1x_1 + \cdots + a_tx_t$ is high, we partition the set of variables $[t]$ into L blocks $B_1, \cdots, B_L$ (L is an appropriately chosen parameter) such that each block gets roughly same total weight: for all $j \in [L]$,

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

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:

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

Now, on a random assignment z, the linear form restricted to many of these blocks will be high. Take such a block B_i: $\sum_{j \in B_i} a_jz_j \geq \tilde{\alpha}\sqrt{\sum_{j \in B_i} a_j^2}$. 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. Crucially, these related points will differ from x 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 5.1. *Let X be a random variable such that $X \leq M$ with probability 1. Let $\mathsf{E}[X] = \mu > 0$. Then, $\mathsf{Pr}[X \geq \mu/2] \geq \dfrac{\mu}{2M}$*

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

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

contradiction. □

An anticoncentration inequality for linear forms of Rademacher random variables.

Lemma 5.2. *There exists a universal constant K such that the following holds: let $x_1, \cdots, x_n$ be independent Rademacher random variables and let $l(x_1, \cdots, x_n) = a_1x_1 + \cdots + a_nx_n$. Let $\sigma = \sqrt{a_1^2 + \cdots + a_n^2}$. Suppose $|a_i| \leq \dfrac{\sigma}{Kt}$. Then,*

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

Proof. Equation 4.2 in [10]. □

Random restrictions have small tail.

Lemma 5.3. *Let $f : \{\pm 1\}^n \to \mathbb{R}$ 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 $\dfrac{\log(d)}{Cd}$. Let $k = \log(d)$. Then,*

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

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,

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

Therefore, for a fixed S,

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

Randomizing over S again,

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

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 $\dfrac{\log(d)}{Cd}$, by Chernoff bound, for each V with $|V|\leq d$,

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

Thus we get that

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

□

Random restrictions don't have low variance.

Lemma 5.4. *Let $f:\{\pm1\}^n\to\mathbb{R}$ be any function and let $\rho=(S,y\in\{\pm1\}^{[n]\setminus S})$ be a random restriction with survival probability p. Then, $\mathsf{E}[\mathsf{Var}[f_\rho]]\geq p\mathsf{Var}[f]$.*

Proof. Fix a restriction $(S,y\in\{\pm1\}^{[n]\setminus S})$. 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,

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

Randomizing over S again,

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

□

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

Lemma 5.5. *Let $f:\{\pm1\}^n\to\mathbb{R}$, $J\subseteq[n]$ 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 an uniformly random assignment, and each $i \in J^c$ is kept alive with probability $p = 2^{-k}$. Then,

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

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

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

Randomizing over S,

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

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

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

□

Some hypercontractive inequalities for low degree functions. The proof of these lemmas can be found in [7].

Lemma 5.6. *There exists a universal constant $W > 0$ such that the following holds:*

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

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

Proof. Corollary 2.4 in [7]. □

Lemma 5.7. *There exists a universal constant $B > 0$ such that the following holds:*

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

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

Proof. Lemma 2.5 in [7] □

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

Proof. Page 112 in [6]. □

Now we state the lemma.

Lemma 5.9. *There exists a universal constant $B > 0$ such that the following holds:*

Consider a degree k polynomial $f : \{\pm 1\}^n \to \mathbb{R}$. 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 $\rho \leftarrow \{\rho_1, \cdots, \rho_{k+1}\}$ 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 5.1. *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 5.6. 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

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

This is a degree k polynomial in ρ with linear coefficient $l(x_0)$. By Lemma 5.8, there exists a $h \in [k+1]$ such that $p(\rho_h) \geq \gamma/(2k+2)$. By Lemma 5.6,

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

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

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

□

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

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

Proof. Start with an arbitrary partition $(B_1, B_2, \cdots, B_L)$. 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.,

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

If no such j exists, terminate.

(2) Locate a k such that

$$\sum_{i \in B_k} a_i \geq \frac{a_1 + \cdots + a_n}{L}.$$

(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 $a_l \leq \frac{a_1 + \cdots + a_n}{2L}$, the size of B_k does not go below $\frac{a_1 + \cdots + a_n}{2L}$ after step (3). It is easy to see this procedure must terminate. Formally, notice that the quantity

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

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

□

6. Main results

6.1. **Improved tail bound for low degree functions.** This section is the core technical part of our work: we show that if we have a function $f : \{\pm 1\}^n \to \mathbb{R}$ *(not necessarily bounded)* with $\mathsf{E}[f^2] \leq 1$ which cannot be approximated by juntas, then f cannot be well-approximated by bounded low-degree functions.

For a subset $J \subseteq [n]$, consider the junta $u : \{\pm 1\}^n \to \mathbb{R}$ 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 $\sum_{S \not\subseteq J} \hat{f}(S)^2$ is small.

Remark 6.1. *In fact, it is easy to see that there exists a junta 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 6.1. *There exists a constant C such that the following holds: Let $f : \{\pm 1\}^n \to \mathbb{R}$ be a degree k function (not necessarily bounded) with $\mathsf{E}[f^2] \leq 1$. Let $J = \{j | \mathsf{Inf}_j[f] \geq \theta\}$ 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 $g : \{\pm 1\}^n \to [0, 1]$, $\mathsf{E}[(f(x) - g(x))^2] \geq \delta = \frac{\mu}{C^k d^C}$.*

Remark 6.2. *Notice here that although f is not pointwise bounded, g is.*

Proof. Let W, B, K be the universal constants from Lemma 5.6, Lemma 5.9 and Lemma 5.2 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 $\rho = (U, y \in \{\pm 1\}^{[n] \setminus U})$ be a random restriction where each $j \in J$ is killed and given a uniformly random assignment, and survival probability for each $j \notin J$ is 2^{-t}. By Lemma 5.5,

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

Fix a U such that

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

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

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

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

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

For each $j \in U$ we have from Lemma 5.7

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

Thus,

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

so applying Lemma 5.1 [2]

$$\mathsf{Pr}_{y\in\{\pm 1\}^{[n]\setminus U}}\left[\sum_{j\in \mathsf{SMALL}_y} \hat{f}_y(\{j\})^2 \geq \frac{\mu}{40\log(k)}\right] \geq \frac{\mu}{40\log(k)(2W)^k} \geq \frac{\mu}{(3W)^k}.$$

Call $y \in \{\pm 1\}^{[n]\setminus U}$ for which $\sum_{j\in \mathsf{SMALL}_y} \hat{f}_y(\{j\})^2 \geq \frac{\mu}{40\log(k)}$ to be *good*. Let $\mathrm{GOOD} = \{y \in \{\pm 1\}^{[n]\setminus U} | y \text{ is good}\}$. Let $L = \left\lceil \frac{(2B)^k d^8}{\mathsf{Var}[f]} \right\rceil$. For each good y, choose a partition $\mathsf{DIVIDE}(y) = (B_1, B_2, \cdots, B_L)$ of SMALL_y such that for all $1 \leq i \leq L$,

$$\sum_{j\in B_i} \hat{f}_y(\{j\})^2 \geq \frac{\mu}{80L\log(k)}.$$

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

Our choice of parameters ensures that for all $j \in \mathsf{SMALL}_y$, $\hat{f}_y(\{j\})^2 \leq \frac{\mu}{80L\log(k)}$, so such a partition exists by Lemma 5.10. Let $\rho_1, \rho_2, \cdots, \rho_{k+1}$ be the constants from Lemma 5.8.

Suppose, for the sake of contradiction, there exists a degree d polynomial $g : \{\pm 1\}^n \to [0,1]$ such that $\mathsf{E}[(f(x)-g(x))^2] \leq \delta$. 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.

Procedure 1:

(1) Sample a $y \in \mathsf{GOOD}$ uniformly at random.
(2) Sample $\rho \leftarrow \{\rho_1, \rho_2, \cdots, \rho_{k+1}\}$ uniformly at random.
(3) Sample $z \leftarrow \{\pm 1\}^U$ uniformly at random.
(4) Let $\mathsf{DIVIDE}(y) = (B_1, B_2, \cdots, B_L)$. 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 our assumption that $\sum_{S\not\subseteq J} \hat{f}(S)^2 \geq$

[2] See remark 3.1.

μ. Then, we obtain an upper bound from the assumption that $\mathsf{E}[(f(x)-g(x))^2] \leq \delta$. These two bounds will contradict each other - and that will prove the theorem.

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

Let $w = \sqrt{\dfrac{\mu}{80L\log(k)}}$. For each $i \in [L]$ we have

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

Choose α such that $\alpha w = \dfrac{100d^4(2B)^k}{L}$. Our choice of L ensures that $\alpha \leq 1$. Moreover, our choice for influence threshold θ ensures that $|\hat{f}_y(\{j\})| \leq \dfrac{w}{K\alpha}$ for all $j \in B_i$ where K is the universal constant from the Lemma 5.2.

Therefore, we can apply Lemma 5.2 to obtain that

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

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

By Lemma 5.9 applied on the restriction $f_y : \{\pm 1\}^U \to [0,1]$, as we sample $z \leftarrow \{\pm 1\}^U$ u.a.r, $\rho \leftarrow \{\rho_1, \cdots, \rho_{k+1}\}$ u.a.r, $\tilde{z}^{(i)} \leftarrow N_{\rho, B_i}(z)$, we have that

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

By linearity of expectation,

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

Using Lemma 5.1,

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

Observe that

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

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

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

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

Upper bound: Since $\mathsf{E}[(f(x)-g(x))^2] \leq \delta$ and $\mathsf{Pr}_{y\in\{\pm1\}^{[n]\setminus U}}[y \text{ is good}] \geq \mu/(3W)^k$, we have that

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

Now consider a uniformly sampled $y \in \mathsf{GOOD}$. Observe that as we sample $z \leftarrow \{\pm1\}^U$ u.a.r, $\rho \leftarrow \{\rho_1, \cdots, \rho_{k+1}\}$ u.a.r and $\tilde{z}^{(i)} \leftarrow N_{\rho,B_i}(z)$, the marginal distribution of $\tilde{z}^{(i)}$ is uniform on $\{\pm1\}^U$. By Markov's inequality, we have

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

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

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

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 $(L+1)\frac{L^2\delta}{(3W)^k\mu} \leq \frac{2L^3\delta}{(3W)^k\mu}$. Our choice of δ ensures that this quantity is less than $< \frac{1}{2L \times (2B)^k}$. Observe that if none of these bad events holds, since the block sensitivity of g is bounded above by $6d^2$ (Theorem 4.5), 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

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

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. □

Now we show that we can improve the tail bound of [7] under the additional assumption that f has low degree. This follows straightforwardly from Theorem 6.1.

Theorem 6.2. *There exists a universal constant $C > 0$ such that the following is true:*
Let $f : \{\pm1\}^n \to [0,1]$ be a degree d function. Let $\theta = \frac{\mathsf{Var}[f]^2}{d^C C^k}$ and $J = \{j | \mathsf{Inf}_j[f] \geq \theta\}$. If $\sum_{S\not\subseteq J} \hat{f}(S)^2 \geq \mu$, then $\sum_{|S|>k} \hat{f}(S)^2 \geq \frac{\mu}{d^C C^k}$.

Proof. Assume $\sum_{|S|>k} \hat{f}(S)^2 < \mu/2$ (otherwise we are done). Let $\tilde{C}$ be the universal constant from Theorem 6.1.

The idea is to apply Theorem 6.1 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 $\mathsf{E}[(f^{\leq k})^2] \leq 1$ and $\mathsf{Inf}_j[f^{\leq k}] \leq \mathsf{Inf}_j[f]$ for all j (this is clear from the Fourier expressions). Let $H = \{j | \mathsf{Inf}_j[f^{\leq k}] \geq \theta\}$. We have $H \subseteq J$, so

$$\sum_{S \not\subseteq H} \hat{f^{\leq k}}(S)^2 \geq \sum_{S \not\subseteq J} \hat{f}(S)^2 - \frac{\mu}{2} \geq \frac{\mu}{2}.$$

Applying Theorem 6.1, we get that for any bounded degree d $g : \{\pm 1\}^n \to [0,1]$, $\mathsf{E}[(f(x) - g(x))^2] \geq \frac{\mu}{2d^{\tilde{C}}\tilde{C}^k}$. Taking g to be our original function f, we get the desired tail lower bound:

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

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

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

□

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.

Theorem 6.3. *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 $f : \{\pm 1\}^n \to [0,1]$ 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 $(d^{-\tilde{C}_1}\mathsf{Var}[f], \mathsf{Var}[f]^{-2}d^{\tilde{C}_4})$ junta. Moreover, if J denotes the set of coordinates on which the junta depends, for each $j \in J$ we have $\mathsf{Inf}_j[f] \geq \mathsf{Var}[f]^{-2}d^{-\tilde{C}_5}$.

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

By Lemma 5.3, the expected Fourier tail of f_ρ above level $\log(d)$ is at most $\exp(-\tilde{C}_3 \log(d)/8)\mathsf{Var}[f] = \frac{\mathsf{Var}[f]}{d^{\tilde{C}_3/8}}$. By Markov's inequality, with probability at least $1 - d^{\tilde{C}_3/16}$, the Fourier tail above $\log(d)$ is $\leq \frac{\mathsf{Var}[f]}{d^{\tilde{C}_3/16}}$. Let C be the constant

from Theorem 6.2. Let $\mu = \frac{\mathsf{Var}[f]}{d^{\tilde{C}_3/16}} d^C C^{\log(d)} = \frac{\mathsf{Var}[f]}{d^{\tilde{C}_3/16}} d^{2C}$, $\theta = \frac{\mu^2}{d^C C^{\log(d)}} = \frac{\mu^2}{d^{2C}}$ and $J = \{j | \mathsf{Inf}_j[f_\rho] \geq \theta\}$. Let $u : \{\pm 1\}^n \to [0,1]$ be the junta which reads the coordinates in J and outputs the average over the coordinates in J^c. Choose $\tilde{C}_3$ large enough so that $\mu \leq d^{-\tilde{C}_1}\mathsf{Var}[f]$. Applying Theorem 6.2, we see that u approximates f_ρ to accuracy $d^{-\tilde{C}_1}\mathsf{Var}[f]$. Using the fact that total influence is bounded by d, we see that u has arity $\leq \mathsf{Var}[f]^{-2} d^{C'\tilde{C}_3}$ for a universal constant C'. Taking $(\tilde{C}_4, \tilde{C}_5) = (C'\tilde{C}_3, \tilde{C}_3/32 - 2C)$, we are done.

□

6.3. **Aaronson-Ambainis conjecture is true for random restrictions.**

Theorem 6.4. *There exist constants $C_1, C_2 > 0$ such that the following holds: let $f : \{\pm 1\}^n \to [0,1]$ be a degree d polynomial ($d \geq 2$) with $\mathsf{Var}[f] \geq 1/d$. Let ρ denote a random restriction with alive probability $\frac{\log(d)}{C_1 d}$. Then,*

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

Proof. Let M be a large constant. Apply Theorem 6.3 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 5.4,

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

so by Lemma 5.1,

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

Since $\mathsf{Var}[f] \geq 1/d$, $d^{-M} \leq \frac{\mathsf{Var}[f]\log(d)}{10\tilde{C}_3 d}$. By Theorem 6.3 and Remark 6.1, with probability at least $1 - d^{-M}$, there exists a $J_\rho \subseteq [n]$ such that every coordinate in J_ρ has influence $\geq \mathsf{Var}[f_\rho]^{-2} d^{-\tilde{C}_5}$ and

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

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

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

In particular, we have that $J_\rho \neq \phi$. Since for each $j \in J_\rho$ we have $\mathsf{Inf}_j[f_\rho] \geq \mathsf{Var}[f_\rho]^{-2} d^{-\tilde{C}_5}$, 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 $f : \{\pm 1\}^n \to \{0, 1\}$ is a degree d polynomial, a large fraction of its random restrictions have an influential coordinate. We observe that this implies one of the results proven in [11] about the existence of small sensitive blocks with a slightly different set of parameters.

Let $f : \{\pm 1\}^n \to [0, 1]$. 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 $|f(x) - f(y)| \geq \epsilon$. [11] proves the following:

Theorem 7.1. *If $f : \{\pm 1\}^n \to [0, 1]$ has degree d, then at least $\Omega(\mathsf{Var}[f])$ fraction of the inputs are (r, ϵ) sensitive where $\epsilon = poly(\mathsf{Var}[f]/d), r = poly(d, 1/\epsilon, \log(n))$*

An immediate consequence of our result is that at least $\Omega(\mathsf{Var}[f]/d^{O(1)})$ fraction of inputs are $(1, \epsilon)$ sensitive where $\epsilon = \text{poly}(\mathsf{Var}[f]/d)$. 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 would be interesting to see if we can extend this to the full Aaronson-Ambainis conjecture. We describe a potential approach here.

- Given a degree d polynomial $f : \{\pm 1\}^n \to [0, 1]$, we can lift it with a Boolean function $g : \{\pm 1\}^m \to \{\pm 1\}^n$ each of whose coordinates g_i is unbiased and given by a low degree function. Then, the lifted polynomial $f \odot g : \{\pm 1\}^m \to [0, 1]$ 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 $g_1, g_2, \cdots, g_n$ 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.

References

[1] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. 2014.

[2] Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups, 2014.

[3] Nikhil Bansal, Makrand Sinha, and Ronald de Wolf. Influence in completely bounded block-multilinear forms and classical simulation of quantum algorithms, 2022.

[4] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials, 1998.

[5] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.

[6] Abdellatif Charafi. Chebyshev polynomials—from approximation theory to algebra and number theory: 2nd edition. theodore j. rivlin, john wiley & sons limited, 1990. pp. 249, hardcover. £45.50. *Engineering Analysis With Boundary Elements*, 9, 1992.

[7] Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the fourier tails of bounded functions over the discrete cube. New York, NY, USA, 2006. Association for Computing Machinery.

[8] J Hastad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, STOC '86, page 6–20, New York, NY, USA, 1986. Association for Computing Machinery.

[9] Nathan Keller and Ohad Klein. Quantum speedups need structure, 2019.
[10] Michel Ledoux and Michel Talagrand. Probability in banach spaces : isoperimetry and processes.
[11] Shachar Lovett and Jiapeng Zhang. Fractional certificates for bounded functions. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2023.
[12] Ashley Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12), December 2012.
[13] Ryan O'Donnell. Analysis of boolean functions, 2021.
[14] Ryan O'Donnell, Michael Saks, Oded Schramm, and Rocco A. Servedio. Every decision tree has an influential variable, 2005.
[15] Ryan O'Donnell and Yu Zhao. Polynomial bounds for decoupling, with applications, 2015.

School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai
Email address: sreejata.bhattacharya@tifr.res.in

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il