Polynomial Bounds for Decoupling, with Applications

Ryan O'Donnell[*1] **and Yu Zhao**[†2]

1 **Computer Science Department, Carnegie Mellon University, Pittsburgh, USA**
odonnell@cs.cmu.edu@cs.cmu.edu
2 **Computer Science Department, Carnegie Mellon University, Pittsburgh, USA**
yuzhao1@cs.cmu.edu

Abstract

Let $f(x) = f(x_1, \ldots, x_n) = \sum_{|S| \le k} a_S \prod_{i \in S} x_i$ be an n-variate real multilinear polynomial of degree at most k, where $S \subseteq [n] = \{1, 2, \ldots, n\}$. For its *one-block decoupled* version,

$$\breve{f}(y, z) = \sum_{|S| \le k} a_S \sum_{i \in S} y_i \prod_{j \in S \setminus \{i\}} z_j,$$

we show tail-bound comparisons of the form

$$\mathbf{Pr}\Big[\Big|\breve{f}(\boldsymbol{y}, \boldsymbol{z})\Big| > C_k t\Big] \le D_k \, \mathbf{Pr}\Big[|f(\boldsymbol{x})| > t\Big].$$

Our constants C_k, D_k are significantly better than those known for "full decoupling". For example, when $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ are independent Gaussians we obtain $C_k = D_k = O(k)$; when $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ are ± 1 random variables we obtain $C_k = O(k^2)$, $D_k = k^{O(k)}$. By contrast, for full decoupling only $C_k = D_k = k^{O(k)}$ is known in these settings.

We describe consequences of these results for query complexity (related to conjectures of Aaronson and Ambainis) and for analysis of Boolean functions (including an optimal sharpening of the DFKO Inequality).

1998 ACM Subject Classification G.2 Discrete Mathematics

Keywords and phrases Decoupling, Query Complexity, Fourier Analysis, Boolean Functions

Digital Object Identifier 10.4230/LIPIcs.CCC.2016.24

1 Introduction

Broadly speaking, *decoupling* refers to the idea of analyzing a complicated random sum involving dependent random variables by comparing it to a simpler random sum where some independence is introduced between the variables. For perhaps the simplest example, if $(a_{ij})_{i,j=1}^n \in \mathbb{R}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are independent uniform ± 1 random variables, we might ask how the moments of

$$\sum_{i,j=1}^n a_{ij} \boldsymbol{x}_i \boldsymbol{x}_j, \text{ and its "decoupled version" } \sum_{i,j=1}^n a_{ij} \boldsymbol{x}_i \boldsymbol{y}_j$$

* Supported in part by NSF grant CCF-1319743.
† Supported in part by NSF grant CCF-1319743.

31st Conference on Computational Complexity (CCC 2016).
Editor: Ran Raz; Article No. 24; pp. 24:1–24:18
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
COMPUTATIONAL COMPLEXITY CONFERENCE

compare. The theory of decoupling inequalities developed originally in the study of Banach spaces, stochastic processes, and U-statistics, mainly between the mid-'80s and mid-'90s; see [10] for a book-length treatment.

The powerful tool of decoupling seems to be relatively under-used in theoretical computer science. ([7] proves a variant of Hanson-Wright Inequality using decoupling inequalities with degree two; a recent work of Makarychev and Sviridenko [31] provides another exception, though they use a much different kind of decoupling than the one studied in this paper.) In this work we will observe several places where decoupling can be used in a "black-box" fashion to solve or simplify problems quite easily.

The main topic of the paper, however, is to study a partial form decoupling that we call "one-block decoupling". The advantage of one-block decoupling is that for degree-k polynomials we can achieve bounds with only *polynomial* dependence on k, as opposed to the exponential dependence on k that arises for the standard full decoupling. Although one-block decoupling does not introduce as much independence as full decoupling does, we show several applications where one-block decoupling is sufficient.

The applications we describe in this paper are the following:

- (Theorem 2.5.) Aaronson and Ambainis's conjecture concerning the generality of their [5, Theorem 4] holds. I.e., there is a sublinear-query algorithm for estimating any bounded, constant-degree Boolean function.
- (Theorem 2.8.) The Aaronson–Ambainis Conjecture [2, 4] holds if and only if it holds for one-block decoupled functions. We also show how the best known result towards the conjecture can be proven extremely easily (1) in the case of one-block decoupled functions.
- (Corollary 3.5.) An optimal form of the DFKO Fourier Tail Bound [13]: any bounded Boolean function f that is far from being a junta satisfies $\sum_{|S|>k} \widehat{f}(S)^2 \geq \exp(-O(k^2))$. Relatedly (Corollary 3.4), any degree-k real-valued Boolean function with $\Omega(1)$ variance and small influences must exceed 1 in absolute value with probability at least $\exp(-O(k^2))$; this can be further improved to $\exp(-O(k))$ if f is homogeneous.

1.1 Definitions

Throughout this section, let f denote a multilinear polynomial of degree at most k in n variables $x = (x_1, \ldots, x_n)$, with coefficients a_S from a Banach space:

$$f(x) = \sum_{\substack{S \subseteq [n] \\ |S| \leq k}} a_S x_S,$$

where we write $x_S = \prod_{i \in S} x_i$ for brevity. (The coefficients a_S will be real in all of our applications; however we allow them to be from a Banach space since the proofs are no more complicated.)

We begin by defining our notion of partial decoupling:

▶ **Definition 1.1.** The *one-block decoupled* version of f, denoted $\breve{f}$, is the multilinear polynomial over $2n$ variables $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$ defined by

$$\breve{f}(y, z) = \sum_{\substack{S \subseteq [n] \\ 1 \leq |S| \leq k}} a_S \sum_{i \in S} y_i z_{S \setminus i}.$$

In other words, each monomial term like $x_1 x_3 x_7$ is replaced with $y_1 z_3 z_7 + z_1 y_3 z_7 + z_1 z_3 y_7$. In case f is homogeneous we have the relation $\breve{f}(x, x) = k f(x)$.

Let us also recall the traditional notion of decoupling:

▶ **Definition 1.2.** The *(fully) decoupled* version of f, which we denote by $\widetilde{f}$, is a multilinear polynomial over k *blocks* $x^{(1)}, \dots, x^{(k)}$ of n variables; each $x^{(i)}$ is $x^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})$. It is formed as follows: for each monomial x_S in f, we replace it with the average over all ways of assigning its variables to different blocks. More formally,

$$\widetilde{f}(x^{(1)}, \dots, x^{(k)}) = a_\emptyset + \sum_{\substack{S \subseteq [n] \\ 1 \le |S| \le k}} \frac{(k - |S|)!}{k!} \cdot a_S \sum_{\substack{\text{injective} \\ b : S \to [k]}} \prod_{i \in S} x_i^{(b(i))}.$$

The definition is again simpler if f is homogeneous. For example, if f is homogeneous of degree 3, then each monomial in f like $x_1 x_3 x_7$ is replaced in $\widetilde{f}$ with

$$\frac{1}{6}\left(w_1 y_2 z_3 + w_1 z_2 y_3 + y_1 w_2 z_3 + y_1 z_2 w_3 + z_1 w_2 y_3 + z_1 y_2 w_3\right).$$

(Here we wrote w, y, z instead of $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, for simplicity.) Note that $\widetilde{f}(x, x, \dots, x) = f(x)$ always holds, even if f is not homogeneous.

We conclude by comparing the two kinds of decoupling. Assume for simplicity that f is homogeneous of degree k. The fully decoupled version $\widetilde{f}(x^{(1)}, \dots, x^{(k)})$ is in "block-multilinear form"; i.e., each monomial contains exactly one variable from each of the k "blocks". This kind of structure has often been recognized as useful in theoretical computer science; see, e.g., [24, 29, 21, 5]. By contrast, the one-block decoupling $\breve{f}(y, z)$ does not have such a simple structure; we only have that each monomial contains exactly one y-variable. Nevertheless we will see several examples in this paper where having one-block decoupled form is just as useful as having fully decoupled form. And as mentioned, we will show that it is possible to achieve one-block decoupling with only poly(k) parameter losses, whereas full decoupling in general suffers exponential losses in k.

▶ Remark 1.1. We have also chosen different "scalings" for the two kinds of decoupling. For example, in the homogeneous case, we have $\widetilde{f}(y, z, z, \dots, z) = \frac{1}{k} \cdot \breve{f}(y, z)$ and also $\mathbf{Var}[\widetilde{f}] = \frac{1}{k \cdot k!} \mathbf{Var}[\breve{f}]$ for $f : \{\pm 1\}^n \to \mathbb{R}$.

1.2 A useful inequality

Several times we will use the following basic inequality from analysis of Boolean functions, which relies on hypercontractivity; see [33, Theorems 9.24, 10.23].

▶ **Theorem 1.3.** *Let $f(x) = \sum_{|S| \le k} a_S x_S$ be a nonconstant n-variate multilinear polynomial of degree at most k, where the coefficients a_S are real. Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be independent uniform ± 1 random variables. Then*

$$\mathbf{Pr}\big[f(\boldsymbol{x}) > \mathbf{E}[f]\big] \ge \tfrac{1}{4} e^{-2k}.$$

This also holds if some of the $\boldsymbol{x}_i$'s are standard Gaussians.[1] *Finally, if the $\boldsymbol{x}_i$'s are not uniform ± 1 random variables, but they take on each value ± 1 with probability at least λ, then we may replace $\frac{1}{4} e^{-2k}$ by $\frac{1}{4}(e^2/2\lambda)^{-k}$.*

[1] Although it is not stated in [33], an identical proof works since Gaussians have the same hypercontractivity properties as uniform ± 1 random variables.

2 Decoupling theorems, and query complexity applications

2.1 Classical decoupling inequalities, and an application in query complexity

Traditional decoupling inequalities compare the probabilistic behavior of f and $\widetilde{f}$ under independent random variables (usually symmetric ones; e.g., standard Gaussians). The easier forms of the inequalities compare expectations under a convex test function; e.g., they can be used to compare p-norms. The following was essentially proved in [9]; see [10, Theorem 3.1.1,(3.4.23)–(3.4.27)]:

▶ **Theorem 2.1.** *Let $\Phi : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ be convex and nondecreasing. Let $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ consist of independent real random variables with all moments finite, and let $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}$ denote independent copies of $\boldsymbol{x}$. Then*

$$\mathbf{E}\Big[\Phi\Big(\Big\|\widetilde{f}\left(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}\right)\Big\|\Big)\Big] \leq \mathbf{E}\Big[\Phi\Big(C_k \left\|f\left(\boldsymbol{x}\right)\right\|\Big)\Big],$$

where $C_k = k^{O(k)}$ is a constant depending only on k.

▶ Remark. A reverse inequality also holds, with worse constant $C_k = k^{-O(k^2)}$.

Another line of research gave comparisons between tail bounds for f and $\widetilde{f}$. This culminated in the following theorem from [11, 18]; see also [10, Theorem 3.4.6]:

▶ **Theorem 2.2.** *In the setting of Theorem 2.1, for all $t > 0$,*

$$\mathbf{Pr}\Big[\Big\|\widetilde{f}\left(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(k)}\right)\Big\| > C_k t\Big] \leq D_k \, \mathbf{Pr}\Big[\left\|f\left(\boldsymbol{x}\right)\right\| > t\Big],$$

where $C_k = D_k = k^{O(k)}$. The analogous reverse bound also holds.

▶ Remark 2.1. Kwapień [28] showed that when the $\boldsymbol{x}_i$'s are α-stable random variables, the constant C_k in Theorem 2.1, can be improved to $k^{k/\alpha}/k!$; this is $k^{k/2}/k!$ for standard Gaussians. Furthermore, for uniform ± 1 random variables Kwapień's proof goes through as if they were 1-stable; thus in this case one may take $C_k = k^k/k! \leq e^k$. In the Gaussian setting with homogeneous f, Kwapień obtains $C_k = k^{k/2}/k!$ and $D_k = 2^k$ for Theorem 2.2.

For function $f(\boldsymbol{x}) = \sum_{|S| \leq k} a_S \boldsymbol{x}_S$ where coefficients a_S are real, we denote its p-norm $\|f\|_p = \mathbf{E}[f(\boldsymbol{x})^p]^{1/p}$. Furthermore if f is a bounded function with input $\boldsymbol{x}$, we denote the infinity norm

$$\|f\|_\infty = \lim_{p \to \infty} \|f\|_p = \sup_{\boldsymbol{x}} |f(\boldsymbol{x})|.$$

▶ **Corollary 2.3.** *In the setting of Theorem 2.1, it holds that $\|\widetilde{f}\|_\infty \leq k^{O(k)} \|f\|_\infty$. Further, if $f : \{\pm 1\}^n \to \mathbb{R}$ then $\|\widetilde{f}\|_\infty \leq (2e)^k \|f\|_\infty$.*

Proof. The first statement is an immediate corollary of either Theorem 2.1 (taking $\Phi(u) = u^p$ and $p \to \infty$) or Theorem 2.2 (taking $t = \|f\|_\infty$). The second statement is immediate from Remark 2.1, with the better constant $k^k/k!$ in case f is homogeneous. In the general case, we use the fact that if $f^{=j}$ denotes the degree-j part of f, then $\|f^{=j}\|_\infty \leq 2^j \|f\|_\infty$; this is also proved by Kwapień [28, Lemma 2]. Then

$$\left\|\widetilde{f}\right\|_\infty = \left\|\sum_{j=0}^{k} \widetilde{f^{=j}}\right\|_\infty \leq \sum_{j=0}^{k} \left\|\widetilde{f^{=j}}\right\|_\infty \leq \sum_{j=0}^{k} (j^j/j!) \left\|f^{=j}\right\|_\infty \leq \sum_{j=0}^{k} (j^j/j!) 2^j \left\|f\right\|_\infty$$
$$\leq (2e)^k \|f\|_\infty. \qquad \blacktriangleleft$$

▶ Remark. Classical decoupling theory has not been too concerned with the dependence of constants on k, and most statements like Theorem 2.2 in the literature simply write $D_k = C_k$ to conserve symbols. However there are good reasons to retain the distinction, since making C_k small is usually much more important than making D_k small. For example, we can deduce Corollary 2.3 from Theorem 2.2 regardless of D_k's value.

Let us give an example application of these fundamental decoupling results. In a recent work comparing quantum query complexity to classical randomized query complexity, Aaronson and Ambainis [5] proved[2] the following:

▶ **Theorem 2.4.** *Let f be an N-variate degree-k homogeneous block-multilinear polynomial with real coefficients. Assume that under uniformly random ± 1 inputs we have $\|f\|_\infty \leq 1$. Then there is a randomized query algorithm making $2^{O(k)}(N/\epsilon^2)^{1-1/k}$ nonadaptive queries to the coordinates of $x \in \{\pm 1\}^N$ that outputs an approximation to $f(x)$ that is accurate to within $\pm\epsilon$ (with high probability).*

The authors "strongly conjecture[d]" that the assumption of block-multilinearity could be removed, and gave a somewhat lengthy proof of this conjecture in the case of $k = 2$, using [13] . We note that the full conjecture follows almost immediately from full decoupling:

▶ **Theorem 2.5.** *Aaronson and Ambainis's Theorem 2.4 holds without the assumption of block-multilinearity or homogeneity.*

Proof. Given a non-block-multilinear f on N variables ranging in $\{\pm 1\}$, consider its full decoupling $\widetilde{f}$ on kN variables. By Corollary 2.3 we have $\|\widetilde{f}\|_\infty \leq (2e)^k$. Let $g = (2e)^{-k}\widetilde{f}$, so that $g : \{\pm 1\}^{kN} \to [-1, +1]$ is a degree-k block-multilinear polynomial with $f(x) = (2e)^k g(x, x, \ldots, x)$. Now given query access to $x \in \{\pm 1\}^N$ and an error tolerance ϵ, we apply Theorem 2.4 to $g(x, x, \ldots, x)$ with error tolerance $\epsilon_1 = (2e)^{-k}\epsilon$; note that we can simulate queries to $(x, x, \ldots, x)$ using queries to x. This gives the desired query algorithm, and it makes $2^{O(k)}(kN/\epsilon_1^2)^{1-1/k} = 2^{O(k)}(N/\epsilon^2)^{1-1/k}$ queries as claimed. There is one more minor point: Theorem 2.4 requires its function to be homogeneous in addition to block-multilinear. However this assumption is easily removed by introducing k dummy variables treated as $+1$, and padding the monomials with them. ◀

2.2 Our one-block decoupling theorems, and the AA Conjecture

We now state our new versions of Theorems 2.1, 2.2 which apply only to one-block decoupling, but that have *polynomial* dependence of C_k on k. Proofs are deferred to Section 4.

As before, let $f(x) = \sum_{|S| \leq k} a_S x_S$ be an n-variate multivariate polynomial of degree at most k with coefficients a_S in a Banach space; let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ consist of independent real random variables with all moments finite, and let $\boldsymbol{y}$, $\boldsymbol{z}$ be independent copies. We consider three slightly different hypotheses:

- **H1:** $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \sim \mathrm{N}(0, 1)$ are standard Gaussians.
- **H2:** $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are uniformly random ± 1 values.
- **H3:** $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are uniformly random ± 1 values and f is homogeneous.

[2] Actually, there is a small gap in their proof. In the line reading "By the concavity of the square root function…", they claim that $\|\boldsymbol{X}\|_1 \geq \|\boldsymbol{X}\|_2$ when $\boldsymbol{X}$ is a degree-k polynomial of uniformly random ± 1 bits. In fact the inequality goes the other way in general. But the desired inequality does hold up to a factor of e^k by [33, Theorem 9.22], and this is sufficient for their proof.

▶ **Theorem 2.6.** *If $\Phi : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ is convex and nondecreasing, then*

$$\mathbf{E}\Big[\Phi\Big(\Big\|\breve{f}(\boldsymbol{y},\boldsymbol{z})\Big\|\Big)\Big] \leq \mathbf{E}\Big[\Phi\Big(C_k \|f(\boldsymbol{x})\|\Big)\Big].$$

Also, if $t > 0$ (and we assume f's coefficients a_S are real under **H2**, **H3***), then*

$$\mathbf{Pr}\Big[\Big\|\breve{f}(\boldsymbol{y},\boldsymbol{z})\Big\| > C_k t\Big] \leq D_k \,\mathbf{Pr}\Big[\|f(\boldsymbol{x})\| > t\Big].$$

Here

$$C_k = \begin{cases} O(k) & \text{under } \mathbf{H1}, \\ O(k^2) & \text{under } \mathbf{H2}, \\ O(k^{3/2}) & \text{under } \mathbf{H3}, \end{cases} \qquad D_k = \begin{cases} O(k) & \text{under } \mathbf{H1}, \\ k^{O(k)} & \text{under } \mathbf{H2, H3}. \end{cases}$$

▶ Remark. It may seem that for the Φ-inequality in the Gaussian case, Kwapień's result mentioned in Remark 2.1 is better than ours, since he achieves full decoupling with a better constant than we get for one-block decoupling. But actually they are incomparable; the reason is the different scaling mentioned in Remark 1.1.

▶ Remark. As we will explain later in Remark 3.1, the bound $C_k = O(k)$ under **H1** is best possible (assuming that $D_k \leq \exp(O(k^2))$).

An immediate consequence of the above theorem, as in Corollary 2.3, is the following:

▶ **Corollary 2.7.** *If $f : \{\pm 1\}^n \to \mathbb{R}$ then $\|\breve{f}\|_\infty \leq O(k^2)\|f\|_\infty$.*

Let us now give an example of how one-block decoupling can be as useful as full decoupling, and why it is important to obtain $C_k = \mathrm{poly}(k)$. A very notable open problem in analysis of Boolean functions is the *Aaronson–Ambainis (AA) Conjecture*, originally proposed in 2008 [2, 4]:

AA Conjecture. *Let $f : \{\pm 1\}^n \to [-1, +1]$ be computable by a multilinear polynomial of degree at most k, $f(x) = \sum_{|S| \leq k} a_S x_S$. Then $\mathbf{MaxInf}_i[f] \geq \mathrm{poly}(\mathbf{Var}[f]/k)$.*

Here we use the standard notations for influences and variance:

$$\mathbf{MaxInf}_i[f] = \max_{i \in [n]} \{\mathbf{Inf}_i[f]\}, \quad \mathbf{Inf}_i[f] = \sum_{S \ni i} a_S^2, \qquad \mathbf{Var}[f] = \sum_{S \neq \emptyset} a_S^2, \qquad \|f\|_2^2 = \sum_S a_S^2.$$

The AA Conjecture is known to imply (and was directly motivated by) the following folklore conjecture concerning the limitations of quantum computation, dated to 1999 or before [4]:

Quantum Conjecture. *Any quantum query algorithm solving a Boolean decision problem using T queries can be correctly simulated on a $1 - \epsilon$ fraction of all inputs by a classical query algorithm using $\mathrm{poly}(T/\epsilon)$ queries.*

Because of their importance for quantum computation, Aaronson has twice listed these conjectures as "semi-grand challenges for quantum computing theory" [1, 3].

The best known result in the direction of the AA Conjecture [4] obtains an influence lower bound of $\mathrm{poly}(\mathbf{Var}[f])/\exp(O(k))$, using the DFKO Inequality [13]. Here we observe that there is a "one-line" deduction of this bound under the assumption that f is one-block

decoupled.[3] To see this, suppose that f is indeed one-block decoupled, so it can be written as $f(y,z) = \sum_{i=1}^{n} y_i g_i(z)$, where $g_i(z) = \sum_{S \ni i} a_S z_{S \setminus i}$ is the ith "derivative" of f. Observe that $\|g_i\|_2^2 = \mathbf{Inf}_i[f]$ and hence $\sum_{i=1}^{n} \|g_i\|_2^2 \geq \mathbf{Var}[f]$. Also note that for any $z \in \{\pm 1\}^n$ we must have $\sum_{i=1}^{n} |g_i(z)| \leq 1$, as otherwise we could achieve $|f(y,z)| > 1$ by choosing $y \in \{\pm 1\}^n$ appropriately. Taking expectations we get $\sum_{i=1}^{n} \|g_i\|_1 \leq 1$, and hence

$$e^{k-1} \geq e^{k-1} \sum_{i=1}^{n} \|g_i\|_1 \geq \sum_{i=1}^{n} \|g_i\|_2 \geq \frac{\sum_{i=1}^{n} \|g_i\|_2^2}{\max_{i=1}^{n} \|g_i\|_2} \geq \frac{\mathbf{Var}[f]}{\max_{i=1}^{n} \sqrt{\mathbf{Inf}_i[f]}}$$

$$\Rightarrow \qquad \mathbf{MaxInf}[f] \geq e^{2-2k} \mathbf{Var}[f]^2, \tag{1}$$

where the second inequality used the basic fact in analysis of Boolean functions [33, Theorem 9.22] that $\|g\|_2 \leq e^{k-1}\|g\|_1$ for $g : \{\pm 1\}^n \to \mathbb{R}$ of degree at most $k-1$.

The above gives a good illustration of how even one-block decoupling can already greatly simplify arguments in analysis of Boolean functions. We feel that (1) throws into sharp relief the challenge of improving $\exp(-O(k))$ to $1/\mathrm{poly}(k)$ for the AA Conjecture. We now use our results to show that the assumption that f is one-block decoupled is completely without loss of generality.

▶ **Theorem 2.8.** *The AA Conjecture holds if and only if it holds for one-block decoupled functions f.*

Proof. Suppose $f : \{\pm 1\}^n \to [-1,+1]$ has degree at most k. By Corollary 2.7 we get that $\|\breve{f}\|_\infty \leq C_k = O(k^2)$. Now $g = C_k^{-1} \breve{f}$ is one-block decoupled and has range $[-1,+1]$. Assuming the AA Conjecture holds for it, we get some $i \in [2n]$ such that $\mathbf{Inf}_i[g] \geq \mathrm{poly}(\mathbf{Var}[g]/k)$. Certainly this implies $\mathbf{Inf}_i[\breve{f}] \geq \mathrm{poly}(\mathbf{Var}[\breve{f}]/k)$. It is easy to see that $\mathbf{Inf}_i[f] = \mathbf{Inf}_i[\breve{f}]$ and $\mathbf{Inf}_i[f] \geq \mathbf{Inf}_{i+n}[\breve{f}]/(k-1)$ for all $i \in [n]$. Therefore letting $i' = \max\{i, i-n\} \in [n]$, we have $\mathbf{Inf}_{i'}[f] \geq \mathbf{Inf}_i[\breve{f}]/(k-1)$, and also $\mathbf{Var}[\breve{f}] \geq \mathbf{Var}[f]$. Thus $\mathbf{Inf}_{i'}[f] \geq \mathrm{poly}(\mathbf{Var}[f]/k)$, confirming the AA Conjecture for f. ◀

In particular, by combining this with (1) we recover the known $\mathrm{poly}(\mathbf{Var}[f])/\exp(O(k))$ lower bound for the AA Conjecture as applied to general f.

▶ Remark. Aaronson and Ambainis [5] recently observed that for the purposes of deriving the Quantum Conjecture, it suffices to prove the AA Conjecture for fully decoupled f. However the AA Conjecture is of significant interest in analysis of Boolean functions in and of itself, even independent of the Quantum Conjecture. Thus we feel Theorem 2.8 is worth knowing, especially in light of the simple argument (1).

3 Tight versions of the DFKO theorems

This section is concerned with analysis of Boolean functions $f : \{\pm 1\}^n \to \mathbb{R}$. We will use traditional Fourier notation, writing $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x_S$. A key theme in this field is the dichotomy between functions with "Gaussian-like" behavior and functions that are essentially "juntas". Recall that f is said to be an (ϵ, C)-junta if $\|f - g\|_2^2 \leq \epsilon$ for some $g : \{\pm 1\}^n \to \mathbb{R}$ depending on at most C input coordinates. Partially exemplifying this theme is a family of theorems stating that any Boolean function f which is not essentially a junta must have a large "Fourier tail" – something like $\sum_{|S|>k} \widehat{f}(S)^2 > \delta$. Examples of such results include

[3] This observation is joint with John Wright.

Friedgut's Average Sensitivity Theorem [15], the FKN Theorem [17] (sharpened in [19, 33]), the Kindler–Safra Theorem [27, 25], and the Bourgain Fourier Tail Theorem [8]. The last of these implies that any $f : \{\pm 1\}^n \to \{\pm 1\}$ which is not a $(.01, k^{O(k)})$-junta must satisfy $\sum_{|S|>k} \widehat{f}(S)^2 > k^{-1/2+o(1)}$. This $k^{-1/2+o(1)}$ bound was made more explicit in [23], and the optimal bound of $\Omega(k^{-1/2})$ was obtained in [26]. These "Fourier tail" theorems have had application in fields such as PCPs and inapproximability [22, 12], sharp threshold theory [16], extremal combinatorics [14], and social choice [17].

All of the aforementioned theorems concern Boolean-*valued* functions; i.e., those with range $\{\pm 1\}$. By contrast, the DFKO Fourier Tail Theorem [13] is a result of this flavor for *bounded* functions; i.e., those with range $[-1, +1]$.

DFKO Fourier Tail Theorem. *Suppose $f : \{\pm 1\}^n \to [-1, +1]$ is not an $(\epsilon, 2^{O(k)}/\epsilon^2)$-junta. Then*

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

Most applications do not use this Fourier tail theorem directly. Rather, they use a key intermediate result, [13, Theorem 3], which we will refer to as the "DFKO Inequality". This was the case, for example, in a recent work on approximation algorithms for the Max-kXOR problem [6].

DFKO Inequality. *Suppose $f : \{\pm 1\}^n \to \mathbb{R}$ has degree at most k and $\mathbf{Var}[f] \geq 1$. Let $t \geq 1$ and suppose that $\mathbf{MaxInf}[f] \leq 2^{-O(k)}/t^2$. Then $\mathbf{Pr}[|f(\boldsymbol{x})| > t] \geq \exp(-O(t^2 k^2 \log k))$.*

Returning to the theme of "Gaussian-like behavior" versus "junta" behavior, we may add that the DFKO results straightforwardly imply (by the Central Limit Theorem) analogous, simpler-to-state results concerning functions on Gaussian space and Hermite tails. We record these generic consequences here; see, e.g., [33, Sections 11.1, 11.2] for a general discussion of such implications, and the definitions of Hermite coefficients $\widehat{f}(\alpha)$.

▶ **Corollary 3.1.** *Any $f : \mathbb{R}^n \to [-1, +1]$ satisfies the Hermite tail bound*

$$\sum_{|\alpha|>k} \widehat{f}(\alpha)^2 > \exp(-O(k^2 \log k)/\mathbf{Var}[f]).$$

Furthermore, suppose $\boldsymbol{z}$ is a standard n-dimensional Gaussian random vector and $t \geq 1$. Then any n-variate polynomial f of degree at most k with $\mathbf{Var}[f(\boldsymbol{z})] \geq 1$ satisfies $\mathbf{Pr}[|f(\boldsymbol{z})| > t] \geq \exp(-O(t^2 k^2 \log k))$.

Even though the Gaussian results in Corollary 3.1 are formally easier than their Boolean counterparts, we are not aware of any way to prove them – even in the case $n = 1$ – except via DFKO.

Tightness of the bounds. In [13, Section 6] it is shown that the results in Corollary 3.1 are tight, up to the $\log k$ factor in the exponent; this implies the same statement about the DFKO Fourier Tail Theorem and the DFKO Inequality. The tight example in both cases is essentially the univariate, degree-k Chebyshev polynomial.[4] In the next section we will show

[4] Formally speaking, [13, Section 6] only argues tightness of the Boolean theorems, but their constructions are directly based on the degree-k Chebyshev polynomial applied to a single standard Gaussian.

how to use our one-block decoupling result to remove the $\log k$ in the exponential from both DFKO theorems. The results immediately transfer to the Gaussian setting, and we therefore obtain the tight $\exp(-\Theta(k^2))$ bound for all versions of the inequality.

Our method of proof is actually to *first* prove the results in the Gaussian setting, where the one-block decoupling makes the proofs quite easy. Then we can transfer the results to the Boolean setting by using the Invariance Principle [32]. This methodology – proving the more natural Gaussian tail bound first, then transferring the result to the Boolean setting via Invariance – is quite reminiscent of how the optimal form of Bourgain's Fourier Tail Theorem was recently obtained [26].

There is actually an additional, perhaps unexpected, bonus of our proof methodology; we show that the bound in the DFKO Inequality can be improved from $\exp(-O(t^2k^2))$ to $\exp(-O(t^2k))$ whenever f is *homogeneous*.

3.1 Proofs of the tight DFKO theorems

We begin with a tail-probability lower bound for one-block decoupled polynomials of Gaussians.

▶ **Lemma 3.2.** *Suppose $f(y,z) = \sum_{i=1}^n y_i g_i(z)$ is a one-block decoupled polynomial on $n+n$ variables, with real coefficients and degree at most k. Let $\boldsymbol{y}, \boldsymbol{z} \in \mathrm{N}(0,1)^n$ be independent standard n-dimensional Gaussians and write*

$$\sigma^2 = \mathbf{Var}[f(\boldsymbol{y},\boldsymbol{z})] = \sum_{i=1}^n \|g_i\|_2^2. \tag{2}$$

Then for $u > 0$ we have $\mathbf{Pr}[|f(\boldsymbol{y},\boldsymbol{z})| > u] \geq \exp(-O(k + u^2/\sigma^2))$.

Proof. Let $v(z) = \sum_{i=1}^n g_i(z)^2$, a polynomial of degree at most $2(k-1)$ in $z_1, \dots, z_n$. By (2) we have $\mathbf{E}[v(\boldsymbol{z})] = \sigma^2$. We now use Theorem 1.3 to get

$$\mathbf{Pr}[v(\boldsymbol{z}) > \sigma^2] \geq \frac{1}{4} e^{-2(2k-1)} = \exp(-O(k)).$$

On the other hand, for any outcome $\boldsymbol{z} = z$ we have that $f(\boldsymbol{y}, z) \sim \mathrm{N}(0, v(z))$. Thus

$$v(z) > \sigma^2 \quad \Longrightarrow \quad \mathbf{Pr}[|f(\boldsymbol{y},z)| > u] \geq \Omega(e^{-u^2/2\sigma^2}).$$

Combining the previous two statements completes the proof, since $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent. ◀

We can now prove an optimal version of the DFKO Inequality in the Gaussian setting. It is also significantly better in the homogeneous case.

▶ **Theorem 3.3.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree at most k, and let $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ be a standard n-dimensional Gaussian vector. Assume $\mathbf{Var}[f(\boldsymbol{x})] \geq 1$. Then for $t \geq 1$ it holds that $\mathbf{Pr}[|f(\boldsymbol{x})| > t] \geq \exp(-O(t^2k^2))$. Furthermore, if f is multilinear and homogeneous then the lower bound may be improved to $\exp(-O(t^2k))$.*

Proof. For any n-variate polynomial of Gaussians, we can find an N-variate multilinear polynomial of Gaussians of no higher degree that is arbitrarily close in Lévy distance (see, e.g., [20, Lemma 15], or use the CLT to pass to ± 1 random variables, then Invariance to pass back to Gaussians). Note, however, that this transformation does not preserve homogeneity. In any case, we can henceforth assume f is multilinear, $f(x) = \sum_{|S| \leq k} a_S x_S$.

For independent $\boldsymbol{y}, \boldsymbol{z} \sim \mathrm{N}(0,1)^n$, observe that

$$\mathbf{Var}[\breve{f}(\boldsymbol{y},\boldsymbol{z})] = \sum_{j=1}^{k} j \sum_{|S|=j} a_S^2 \geq \sum_{S \neq \emptyset} a_S^2 = \mathbf{Var}[f(\boldsymbol{x})] \geq 1,$$

and if f is homogeneous we get the better bound $\mathbf{Var}[\breve{f}(\boldsymbol{y},\boldsymbol{z})] \geq k$. By our Theorem 2.6 on one-block decoupling, we have

$$\mathbf{Pr}\Big[\Big|f(\boldsymbol{x})\Big| > t\Big] \geq D_k^{-1}\,\mathbf{Pr}\Big[\Big|\breve{f}(\boldsymbol{y},\boldsymbol{z})\Big| > C_k t\Big],$$

where $C_k = D_k = O(k)$. The theorem is now an immediate consequence of Lemma 3.2. ◀

▶ Remark 3.1. A consequence of this proof is that – assuming $D_k \leq \exp(O(k^2))$ – it is impossible to asymptotically improve on our $C_k = O(k)$ in Theorem 2.6 in the Gaussian setting **H1**. Otherwise, we would achieve a bound of $\exp(-o(k^2))$ in Theorem 3.3, contrary to the example in [13, Section 6].

We can now obtain the sharp DFKO Inequality in the Boolean setting by using the Invariance Principle.

▶ **Corollary 3.4.** *Theorem 3.3 holds when $\boldsymbol{x} \sim \{\pm 1\}^n$ is uniform and we additionally assume that $\mathbf{MaxInf}[f] \leq \exp(-Ct^2k^2)$, or just $\exp(-Ct^2k)$ in the homogeneous case. Here C is a universal constant.*

Proof. This follows immediately from the Lévy distance bound in [32, Theorem 3.19, Hypothesis 4]. We only need to ensure that the Lévy distance is noticeably less than the target lower bound we're aiming for. (We also remark that the Invariance Principle transformation preserves variance and homogeneity.) ◀

Next, we obtain the sharp DFKO Fourier Tail Theorem. Its deduction from the DFKO Inequality in [13] is unfortunately not "black-box", so we will have to give a proof.

▶ **Corollary 3.5.** *Suppose $f : \{\pm 1\}^n \to [-1,+1]$ is not an $(\epsilon, 2^{O(k^2/\epsilon)})$-junta. Then*

$$\sum_{|S|>k} \widehat{f}(S)^2 > \exp(-O(k^2)/\epsilon). \tag{3}$$

Proof. We use notation and basic results from [33]. Given $f : \{\pm 1\}^n \to [-1,+1]$, let $J = \{i \in [n] : \mathbf{Inf}_i^{\leq k}[f] > \exp(-Ak^2/\epsilon)\}$, where A is a large constant to be chosen later. Since $\|f\|_2^2 \leq 1$ it follows easily that $|J| \leq 2^{O(k^2/\epsilon)}$. Now define $g = f - f^{\subseteq J}$; note that g has range in $[-2,+2]$ since $f^{\subseteq J}$ has range in $[-1,+1]$, being an average of f over the coordinates outside J. If $\|g\|_2^2 < \epsilon/2$ then f is $\epsilon/2$-close to the $2^{O(k^2/\epsilon)}$-junta $f^{\subseteq J}$ and we are done. Otherwise, $\|g\|_2^2 \geq \epsilon/2$ and we let $h = g^{\leq k}$. If $\|h-g\|_2^2 > \epsilon/4$ then we immediately conclude that $\sum_{|S|>k} \widehat{f}(S)^2 > \epsilon/4$, which is more than enough to be done. Otherwise $\|h-g\|_2^2 \leq \epsilon/4$, from which we conclude $\|h\|_2^2 \geq \epsilon/4$. Now h has degree at most k and satisfies $\mathbf{Inf}_i[h] \leq \exp(-Ak^2/\epsilon)$ for all $i \notin J$. Let $\widetilde{h}$ denote the mixed Boolean/Gaussian function which has the same multilinear form as h, but where we think of the coordinates in J as being ± 1 random variables and the coordinates not in J as being standard Gaussians. We now "partially" apply the Invariance Principle [32, Theorem 3.19] to h, in the sense that we only hybridize over the coordinates not in J. We conclude that the Lévy distance between h and $\widetilde{h}$ is at most $\exp(-\Omega(Ak^2/\epsilon))$. Our goal is now to show that

$$\mathbf{Pr}[|\widetilde{h}| > 3] \geq \exp(-O(k^2/\epsilon)), \tag{4}$$

where the constant in the $O(\cdot)$ does not depend on A. Having shown this, by taking A large enough the Lévy distance bound lets us deduce (4) for h as well. But then since $|g| \le 2$ always, we may immediately deduce $\|g-h\|_2^2 \ge \exp(-O(k^2)/\epsilon)$ and hence (3).

It remains to verify (4). For each restriction x_J to the J-coordinates, the function $\widetilde{h}_{x_J}$ is a multilinear polynomial in independent Gaussians with some variance $\sigma_{x_J}^2$. From Theorem 3.3 we can conclude that $\mathbf{Pr}[|\widetilde{h}_{x_J}| > 3] \ge \exp(-O(k^2/\sigma_{x_J}^2))$. Thus if we can show $\sigma_{\boldsymbol{x}_J}^2 \ge \Omega(\epsilon)$ with probability at least $2^{-O(k)}$ when $\boldsymbol{x}_J \in \{\pm 1\}^J$ is uniformly random, we will have established (4). But this follows similarly as in Lemma 3.2. Note that $\sigma_{x_J}^2 = \mathbf{E}[\widetilde{h}_{x_J}^2]$, since h has no constant term. Now $\sigma_{x_J}^2$ is a degree-$2k$ polynomial in x_J, and its expectation is simply $\|h\|_2^2 \ge \epsilon/4$, so Theorem 1.3 indeed implies that $\mathbf{Pr}[\sigma_{\boldsymbol{x}_J}^2 \ge \epsilon/4] \ge 2^{-O(k)}$ and we are done. ◀

▶ Remark. We comment that the dependence of $\mathbf{MaxInf}[f]$ on t in Corollary 3.4, and the junta size in Corollary 3.5, are not as good as in [13]. This seems to be a byproduct of the use of Invariance.

A similar (but easier) proof can be used to derive the following Gaussian version of Corollary 3.5; alternatively, one can use a generic CLT argument, noting that the only "junta" a Gaussian function can be close to is a constant function:

▶ **Corollary 3.6.** *Any $f : \mathbb{R}^n \to [-1,+1]$ satisfies the Hermite tail bound*

$$\sum_{|\alpha|>k} \widehat{f}(\alpha)^2 > \exp(-O(k^2)/\mathbf{Var}[f]).$$

This strictly improves upon Corollary 3.1.

4 Proofs of our one-block decoupling theorems

In this section we prove Theorem 2.6. The key idea of the proof is to express $\breve{f}(y,z)$ as a "small" linear combination of expressions of the form $f(\alpha_i x + \beta_i y)$, where $\alpha_i^2 + \beta_i^2 = 1$ (in the Gaussian case) or $|\alpha_i| + |\beta_i| = 1$ (in the Boolean case). The following is the central lemma.

▶ **Lemma 4.1.** *In the setting of Theorem 2.6, there exists $m = O(k)$ and $\alpha, \beta, c \in \mathbb{R}^m$ such that*

- $\breve{f}(y,z) = \sum_{i=1}^m c_i f(\alpha_i y + \beta_i z)$;
- $\sum_{i=1}^m |c_i| \le C_k$;
- $\alpha_i^2 + \beta_i^2 = 1$ *for all* $i \in [m]$ *under* **H1**, *and* $|\alpha_i| + |\beta_i| = 1$ *for all* $i \in [m]$ *under* **H2**, **H3**;
- $|\alpha_i|, |\beta_i| \ge 1/O(C_k)$ *for all* $i \in [m]$.

With Lemma 4.1 in hand, the proof of Theorem 2.6 is quite straightforward in the Gaussian case, and not much more difficult in the Boolean case. We show these deductions first.

Proof of Theorem 2.6 under Hypothesis H1. By Lemma 4.1, for any convex nondecreasing function $\Phi : \mathbb{R}^{\ge 0} \to \mathbb{R}^{\ge 0}$ we have

$$\begin{aligned}\mathbf{E}\Big[\Phi\Big(\Big\|\breve{f}(\boldsymbol{y},\boldsymbol{z})\Big\|\Big)\Big] &= \mathbf{E}\Big[\Phi\Big(\Big\|\sum_{i=1}^m c_i f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\Big\|\Big)\Big] \\ &\le \mathbf{E}\Big[\Phi\Big(\sum_{i=1}^m |c_i| \Big\| f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\Big\|\Big)\Big]\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^{m} \frac{|c_i|}{C_k} \mathbf{E}\left[\Phi\left(C_k \|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\|\right)\right] \\
&= \sum_{i=1}^{m} \frac{|c_i|}{C_k} \mathbf{E}[\Phi(C_k \|f(\boldsymbol{x})\|)] \\
&\leq \mathbf{E}[\Phi(C_k \|f(\boldsymbol{x})\|].
\end{aligned}$$

Here the inequalities follow from the convexity and monotonicity of Φ, and the second equality holds because $\alpha_i \boldsymbol{y} + \beta \boldsymbol{z} \sim \mathrm{N}(0,1)^n$ due to $\alpha_i^2 + \beta_i^2 = 1$.

As for the tail-bound comparison, by Lemma 4.1, whenever y, z are such that $\|\breve{f}(y,z)\| > C_k t$, the triangle inequality implies that there must exist at least one $i \in [m]$ with $\|f(\alpha_i y + \beta_i z)\| > t$. It follows that there must exist at least one $i \in [m]$ such that

$$\mathbf{Pr}[\|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\| > t] \geq \frac{1}{m} \mathbf{Pr}[\|\breve{f}(\boldsymbol{y}, \boldsymbol{z})\| > C_k t].$$

This completes the proof, since $\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z} \sim \mathrm{N}(0,1)^n$ and $m = O(k)$. ◀

Proof of Theorem 2.6 under Hypotheses H2, H3. We define ± 1 random variables as follows:

$$\boldsymbol{x}_j^{(i)} = \begin{cases} \mathrm{sgn}(\alpha_i)\boldsymbol{y}_j & \text{with probability } |\alpha_i|, \\ \mathrm{sgn}(\beta_i)\boldsymbol{z}_j & \text{with probability } |\beta_i|, \end{cases}$$

for all $i \in [m]$ and $j \in [n]$ independently. Notice that each $\boldsymbol{x}^{(i)}$ is distributed uniformly on $\{\pm 1\}^n$, though they are not independent. To prove the desired inequality concerning Φ, we can repeat the proof in the Gaussian case, except that we no longer have the identity

$$\mathbf{E}\left[\Phi\left(C_k \|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\|\right)\right] = \mathbf{E}[\Phi(C_k \|f(\boldsymbol{x})\|)].$$

In fact we will show that the left-hand side is at most the right-hand side. Notice that for all fixed $y, z \in \{\pm 1\}^n$, the multilinearity of f implies that

$$f(\alpha_i y + \beta_i z) = \mathbf{E}[f(\boldsymbol{x}^{(i)}) \mid (\boldsymbol{y}, \boldsymbol{z}) = (y, z)]. \tag{5}$$

Thus

$$\begin{aligned}
\mathbf{E}\left[\Phi\left(C_k \|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\|\right)\right] &= \mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{z}}\left[\Phi\left(C_k \left\| \mathop{\mathbf{E}}_{\boldsymbol{x}^{(i)}|\boldsymbol{y},\boldsymbol{z}}\left[f(\boldsymbol{x}^{(i)})\right]\right\|\right)\right] \\
&\leq \mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{z}} \mathop{\mathbf{E}}_{\boldsymbol{x}^{(i)}}\left[\Phi\left(C_k \|f(\boldsymbol{x}^{(i)})\|\right)\right] = \mathbf{E}\left[\Phi\left(C_k \|f(\boldsymbol{x})\|\right)\right],
\end{aligned}$$

as claimed, where we used convexity again.

As for the tail-bound comparison, recall that we are now assuming f has real coefficients. As in the Gaussian case there is at least one $i \in [m]$ with

$$\mathbf{Pr}[|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})| > t] \geq \frac{1}{O(k)} \mathbf{Pr}[|\breve{f}(\boldsymbol{y}, \boldsymbol{z})| > C_k t].$$

Now suppose y, z are such that $|f(\alpha_i y + \beta_i z)| > t$ and consider the conditional distribution on $\boldsymbol{x}^{(i)}$. If we can show that, conditionally, $\mathbf{Pr}[|f(\boldsymbol{x}^{(i)})| > t] \geq k^{-O(k)}$ then we are done. But from (5) we have that $\left|\mathbf{E}[f(\boldsymbol{x}^{(i)})]\right| > t$; therefore the desired result follows from Theorem 1.3 and the fact that $\min(|\alpha_i|, |\beta_i|) \geq 1/O(C_k) = 1/\mathrm{poly}(k)$. ◀

4.1 Proof of Lemma 4.1

The proof of Lemma 4.1 involves minimizing $\sum_{i=1}^{m} |c_i|$ by carefully setting the ratios of α_i and β_i to be a hyperharmonic progression.

Proof of Lemma 4.1. The main work involves treating the homogeneous case.

Homogeneous case. Our goal for homogeneous f is to write

$$\breve{f}(y,z) = \sum_{i=1}^{k+1} c_i f(\alpha_i y + \beta_i z).$$

Comparing the expressions term by term, it is equivalent to say that for any $S \subseteq [n]$ with $|S| = k$,

$$\sum_{j \in S} y_j z_{S/j} = \sum_{i=1}^{k+1} c_i \prod_{j \in S} (\alpha_i y_j + \beta_i z_j).$$

We can further simplify this to the conditions

$$\sum_{i=1}^{k+1} c_i \alpha_i^{k-t} \beta_i^t = \begin{cases} 1 & \text{if } t = k-1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

for all integers $0 \le t \le k$. Let us write $\Delta_i = \frac{\beta_i}{\alpha_i}$ and introduce the Vandermonde matrix

$$V = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \Delta_1 & \Delta_2 & \cdots & \Delta_{k+1} \\ \cdots & \cdots & \cdots & \cdots \\ \Delta_1^{k-1} & \Delta_2^{k-1} & \dots & \Delta_{k+1}^{k-1} \\ \Delta_1^{k} & \Delta_2^{k} & \cdots & \Delta_{k+1}^{k} \end{bmatrix}.$$

We will also write A for the diagonal matrix $\mathrm{diag}(\alpha_1^k, \alpha_2^k, \dots, \alpha_{k+1}^k)$, and write e_k for the indicator vector of the kth coordinate, $e_k = (0, 0, \dots, 0, 1, 0)$. Then the necessary conditions (6) are equivalent to the matrix equation $VAc = e_k$. Assuming all the Δ_i's are different, V is invertible and there is an explicit formula for its inverse [30]. This yields the following expression for the $c_1, \dots, c_{k+1}$ in terms of α and β:

$$c_i = (A^{-1}V^{-1}e_k)_i = \frac{1}{\alpha_i^k} \cdot \frac{\Delta_i - \sum_{j=1}^{k+1} \Delta_j}{\prod_{j=1, j\neq i}^{k+1} (\Delta_i - \Delta_j)}. \tag{7}$$

The main illustrative case: Hypothesis H1 and k odd. We will now assume that k is odd; this assumption will be easily removed later. It will henceforth be convenient to replace our indices $1, \dots, k+1$ with the following slightly peculiar but symmetric set of indices:

$$I = \left\{\pm 1, \pm 2, \dots, \pm \tfrac{k-1}{2}, \pm \tfrac{1}{2}\right\}.$$

Now under Hypothesis **H1**, we will choose

$$\alpha_i = \frac{i}{\sqrt{k^2 + i^2}}, \quad \beta_i = \frac{k}{\sqrt{k^2 + i^2}} \quad \Longrightarrow \quad \Delta_i = \frac{k}{i}$$

for all $i \in I$. These choices satisfy $\alpha_i^2 + \beta_i^2 = 1$ and $|\alpha_i|, |\beta_i| \geq 1/O(C_k)$, so it remains to prove that for c defined by (7) we have $\sum |c_i| \leq O(k)$.

Let us upper-bound all $|c_i|$. Since it easy to see that $|c_i| = |c_{-i}|$ for all $i \in I$, it will suffice for us to consider the positive $i \in I$. For $1 \leq i \leq \frac{k-1}{2}$, we have

$$\begin{aligned}
\left| \prod_{j \in I, j \neq i} (\Delta_i - \Delta_j) \right| &= (\Delta_{1/2} - \Delta_i)(\Delta_i - \Delta_{-1/2}) \cdot \prod_{j=1, j\neq i}^{(k-1)/2} |\Delta_i - \Delta_j| \cdot \prod_{j=-(k-1)/2}^{-1} (\Delta_i - \Delta_j) \\
&= \left(2k - \frac{k}{i}\right)\left(2k + \frac{k}{i}\right) \cdot \prod_{j=1, j\neq i}^{(k-1)/2} \left|\frac{k}{i} - \frac{k}{j}\right| \cdot \prod_{j=1}^{(k-1)/2} \left(\frac{k}{i} + \frac{k}{j}\right) \\
&= k^k \left(4 - \frac{1}{i^2}\right) \cdot \prod_{j=1, j\neq i}^{(k-1)/2} \frac{|j-i|}{ij} \cdot \prod_{j=1}^{(k-1)/2} \frac{j+i}{ij} \\
&= \frac{k^k}{i^{k-2}} \left(4 - \frac{1}{i^2}\right) \frac{\left(\frac{k-1}{2} + i\right)! \left(\frac{k-1}{2} - i\right)!}{\left(\frac{k-1}{2}\right)!^2}.
\end{aligned}$$

Thus from (7),

$$\begin{aligned}
|c_i| &= \left(\frac{\sqrt{k^2 + i^2}}{i}\right)^k \cdot \frac{k}{i} \cdot \frac{i^{k-2}}{k^k} \cdot \frac{1}{4 - 1/i^2} \cdot \frac{\left(\frac{k-1}{2}\right)!^2}{\left(\frac{k-1}{2} + i\right)! \left(\frac{k-1}{2} - i\right)!} \\
&= \frac{k}{i^3} \left(1 + \frac{i^2}{k^2}\right)^{k/2} \frac{1}{4 - 1/i^2} \frac{\left(\frac{k-1}{2}\right)!^2}{\left(\frac{k-1}{2} + i\right)! \left(\frac{k-1}{2} - i\right)!}.
\end{aligned}$$

When $1 \leq i \leq \sqrt{k}$, we have

$$|c_i| = \frac{k}{i^3} \left(1 + \frac{i^2}{k^2}\right)^{k/2} \frac{1}{4 - 1/i^2} \frac{\left(\frac{k-1}{2}\right)!^2}{\left(\frac{k-1}{2} + i\right)! \left(\frac{k-1}{2} - i\right)!} \leq \frac{k}{i^3} \left(1 + \frac{1}{k}\right)^{k/2} \leq \frac{\sqrt{e}k}{i^3}.$$

For $\sqrt{k} \leq i \leq \frac{k-1}{2}$, consider the ratio between $(i+1)^3|c_{i+1}|$ and $i^3|c_i|$; it satisfies

$$\begin{aligned}
\frac{(i+1)^3|c_{i+1}|}{i^3|c_i|} &\leq \frac{(k^2 + (i+1)^2)^{k/2}}{(k^2 + i^2)^{k/2}} \cdot \frac{\frac{k-1}{2} - i}{\frac{k-1}{2} + i + 1} \\
&= \left(1 + \frac{2i+1}{k^2 + i^2}\right)^{k/2} \cdot \frac{k - 1 - 2i}{k + 1 + 2i} \\
&\leq \left(1 + \frac{2i+1}{k^2}\right)^{k/2} \cdot \frac{k - 1 - 2i}{k} \\
&\leq e^{\frac{2i+1}{2k}} \left(1 - \frac{2i+1}{k}\right) \leq 1.
\end{aligned}$$

The last inequality holds since $e^{x/2}(1-x) \leq 1$ for all $0 \leq x \leq 1$. Thus we have $(i+1)^3|c_{i+1}| \leq i^3|c_i|$, and hence by induction that

$$|c_i| \leq \frac{\sqrt{e}k}{i^3} \quad \forall\ 1 \leq i \leq \tfrac{k-1}{2}. \tag{8}$$

We now need to bound $c_{1/2}$. Similarly to the above, we have

$$\begin{aligned}\left|\prod_{j\in I, j\neq\frac{1}{2}}(\Delta_{1/2}-\Delta_j)\right| &= (\Delta_{\frac{1}{2}}-\Delta_{-1/2})\cdot\prod_{j=1}^{(k-1)/2}(\Delta_{1/2}-\Delta_j)\cdot\prod_{j=-(k-1)/2}^{-1}(\Delta_{\frac{1}{2}}-\Delta_j)\\ &= 4k\cdot\prod_{j=1}^{(k-1)/2}\left(2k-\frac{k}{j}\right)\cdot\prod_{j=1}^{(k-1)/2}(2k+\frac{k}{j})\\ &= 4k^k\cdot\prod_{j=1}^{(k-1)/2}\frac{2j-1}{j}\cdot\prod_{j=1}^{(k-1)/2}\frac{2j+1}{j}\\ &= 4k^k\frac{(k-2)!!k!!}{\left(\frac{k-1}{2}\right)!^2}\end{aligned}$$

Thus from (7) we get

$$\begin{aligned}|c_{1/2}| &= \frac{(\sqrt{k^2+(1/2)^2})^k}{(1/2)^k}\cdot 2k\cdot\frac{1}{4k^k}\cdot\frac{\left(\frac{k-1}{2}\right)!^2}{(k-2)!!k!!}\\ &= \left(1+\frac{1}{4k^2}\right)^{k/2}\left(\frac{(k-1)!!}{(k-2)!!}\right)^2 \le 4k. \qquad (9)\end{aligned}$$

Now combining (8), (9), we obtain

$$\sum_i |c_i| = 2\sum_{i=1}^{(k-1)/2}|c_i| + 2|c_{1/2}| \le 2\sqrt{e}\sum_{i=1}^{(k-1)/2}\frac{k}{i^3} + 8k \le 20k,$$

as needed.

Handling even k. If k is even, we define our index set to be

$$I = \left\{0, \pm1, \pm2, \ldots, \pm\tfrac{k-2}{2}, \pm\tfrac{1}{2}\right\}.$$

For $i \in I\setminus\{0\}$ we define α_i and β_i as before; we also define $\alpha_0 = 1$, $\beta_0 = 0$, and hence $\Delta_0 = 0$. It is easy to check that $c_0 = 0$ (and hence we haven't actually violated $|\beta_i| \ge 1/O(C_k)$), and the upper bounds for the other $|c_i|$ still hold. This completes the proof of the homogeneous case under Hypothesis **H1**.

Hypotheses H3. We explain the case of k odd; the same trick as before can be used for even k. For Hypothesis **H3** we use

$$\alpha_i = \frac{i}{k^{3/2}+|i|}, \quad \beta_i = \frac{k^{3/2}}{k^{3/2}+|i|} \quad\Longrightarrow\quad \Delta_i = \frac{k^{3/2}}{i},$$

which satisfy $|\alpha_i|+|\beta_i| = 1$ and $|\alpha_i|, |\beta_i| \ge 1/O(k^{3/2})$. Analysis similar to before shows that $\sum_i |c_i| \le O(k^{3/2})$. This completely finishes the proof under Hypothesis **H3**.

Hypothesis H2, the homogeneous case. Here we do something slightly different. For even or odd k we let the index set be $I = \{1, 2, \ldots, k, \frac{1}{2}\}$ and then define

$$\alpha_i = \frac{i^2}{k^2+i^2}, \quad \beta_i = \frac{k^2}{k^2+i^2} \quad\Longrightarrow\quad \Delta_i = \frac{k^2}{i^2}.$$

Now we have $|\alpha_i|+|\beta_i| = \alpha_i+\beta_i = 1$ and $|\alpha_i|, |\beta_i| \ge 1/O(k^2)$. Again, similar analysis shows that $\sum_i |c_i| \le O(k^2)$.

Extending to the non-homogeneous case under H2. Now we need to be concerned with the terms at degree $k' < k$. Here a key observation is that, since $\alpha_i + \beta_i = 1$ for all i, the following holds for all $k' < k$:

$$\sum_i c_i \alpha_i^{k'-t} \beta_i^t = \sum_i c_i \alpha_i^{k'-t} \beta_i^t (\alpha_i + \beta_i) = \sum_i c_i \alpha_i^{k'-t+1} \beta_i^t + \sum_i c_i \alpha_i^{k'-t} \beta_i^{t+1}.$$

Thus an induction shows that in fact

$$\sum_i c_i \alpha_i^{k'-t} \beta_i^t = \begin{cases} k - k' & \text{if } t = k' \\ 1 & \text{if } t = k' - 1 \\ 0 & \text{otherwise} \end{cases}$$

for all $k' \le k$. This is almost exactly what we need to treat the non-homogeneous case using all the same choices for c, α, β, except for the $t = k'$ case. But we can use a simple trick to fix this:

$$\frac{1}{2} \sum_i c_i \alpha_i^{k'-t} \beta_i^t - \frac{1}{2} \sum_i c_i (-\alpha_i)^{k'-t} \beta_i^t = \frac{1 - (-1)^{k'-t}}{2} \sum_i c_i \alpha_i^{k'-t} \beta_i^t = \begin{cases} 1 & \text{if } t = k' - 1 \\ 0 & \text{otherwise} \end{cases}$$

From this we get

$$\breve{f}(y, z) = \sum_{i=1}^{m} c_i f(\alpha_i y + \beta_i z)$$

even in the non-homogeneous case, with all the desired conditions and $m = 2(k + 1)$.

Extending to the non-homogeneous case under H1. The trick here for handling degree $k' < k$ is similar. Using the fact that $\alpha_i^2 + \beta_i^2 = 1$ for all i, we get that for all $k' < k$,

$$\sum_i c_i \alpha_i^{k'-t} \beta_i^t = \sum_i c_i \alpha_i^{k'-t} \beta_i^t (\alpha_i^2 + \beta_i^2) = \sum_i c_i \alpha_i^{k'-t+2} \beta_i^t + \sum_i c_i \alpha_i^{k'-t} \beta_i^{t+2}.$$

Then by induction, the we conclude that

$$\sum_{i=1}^{k+1} c_i \alpha_i^{k'-t} \beta_i^t = \begin{cases} 1 & \text{if } t = k' - 1 \\ 0 & \text{otherwise} \end{cases}$$

holds for all $0 \le k' \le k$ such that $k - k'$ is even. We are therefore almost done: we have established the **H1** case of Lemma 4.1 for all polynomials with only odd-degree terms or only even-degree terms. Finally, for a general polynomial f we can decompose it as $f = f_{\text{odd}} + f_{\text{even}}$, where f_{odd} (respectively, f_{even}) contains all the terms in f with odd (respectively, even) degree. We know that there exist some vectors α, β, c and α', β', c' satisfying

$$\breve{f}_{\text{odd}}(y, z) = \sum_{i=1}^{k+1} c_i f_{\text{odd}}(\alpha_i y + \beta_i z), \quad \breve{f}_{\text{even}}(y, z) = \sum_{i=1}^{k+1} c'_i f_{\text{even}}(\alpha'_i y + \beta'_i z),$$

and $\sum_i |c_i|, \sum_i |c'_i| \le 20k$. Thus

$$\begin{aligned}
\breve{f}(y,z) &= \breve{f}_{\text{odd}}(y,z) + \breve{f}_{\text{even}}(y,z) \\
&= \sum_{i=1}^{k+1} c_i f_{\text{odd}}(\alpha_i y + \beta_i z) + \sum_{i=1}^{k+1} c'_i f_{\text{even}}(\alpha'_i y + \beta'_i z) \\
&= \sum_{i=1}^{k+1} \frac{1}{2} c_i (f(\alpha_i y + \beta_i z) - f(-\alpha_i y - \beta_i z)) + \sum_{i=1}^{k+1} \frac{1}{2} c'_i (f(\alpha'_i y + \beta'_i z) + f(-\alpha'_i y - \beta'_i z)) \\
&= \sum_{i=1}^{4(k+1)} c''_i f(\alpha''_i y + \beta''_i z),
\end{aligned}$$

where $c'' = (c/2, -c/2, c'/2, c'/2), \alpha'' = (\alpha, -\alpha, \alpha', -\alpha'), \beta'' = (\beta, -\beta, \beta', -\beta')$ and $\sum_i |c''_i| \le 40k$. ◀

Acknowledgments. The authors would like to thank Oded Regev for helpful discussions, and John Wright for permission to include (1).

References

1. Scott Aaronson. Ten semi-grand challenges for quantum computing theory, 2005. `http://www.scottaaronson.com/writings/qchallenge.html`.
2. Scott Aaronson. How to solve longstanding open problems in quantum computing using only Fourier Analysis. Lecture at Banff International Research Station, 2008. `http://www.scottaaronson.com/talks/openqc.ppt`.
3. Scott Aaronson. Updated version of "ten semi-grand challenges for quantum computing theory", 2010. `http://www.scottaaronson.com/blog/?p=471`.
4. Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *Theory Of Computing*, 10(6):133–166, 2014.
5. Scott Aaronson and Andris Ambainis. Forrelation: a problem that optimally separates quantum from classical computing. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 307–316, 2015.
6. Boaz Barak, Ankur Moitra, Ryan O'Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright. Beating the random assignment on constraint satisfaction problems of bounded degree. In *Proceedings of the 18th Annual International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, 2015.
7. Franck Barthe and Emanuel Milman. Transference principles for log-sobolev and spectral-gap with applications to conservative spin systems. *Communications in Mathematical Physics*, 323(2):575–625, 2013.
8. Jean Bourgain. On the distribution of the Fourier spectrum of Boolean functions. *Israel Journal of Mathematics*, 131(1):269–276, 2002.
9. Victor de la Peña. Decoupling and Khintchine's inequalities for U-statistics. *Annals of Probability*, 20(4):1877–1892, 1992.
10. Víctor de la Peña and Evarist Giné. *Decoupling: from dependence to independence*. Springer, 1999.
11. Victor de la Peña and Stephen Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Annals of Probability*, 23(2):806–816, 1995.
12. Irit Dinur. The PCP Theorem by gap amplification. *Journal of the ACM*, 54(3):1–44, 2007.

13 Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel Journal of Mathematics*, 160(1):389–412, 2007.

14 David Ellis, Yuval Filmus, and Ehud Friedgut. Triangle-intersecting families of graphs. *Journal of the European Mathematical Society*, 14(3):841–885, 2012.

15 Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–36, 1998.

16 Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124(10):2993–3002, 1996.

17 Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean functions whose Fourier transform is concentrated on the first two levels and neutral social choice. *Advances in Applied Mathematics*, 29(3):427–437, 2002.

18 Evarist Giné. A consequence for random polynomials of a result of de la Peña and Montgomery-Smith. In *Probability in Banach Spaces 10*, volume 43 of *Progress in Probability*. Birkhäuser–Verlag, 1998.

19 Jacek Jendrej, Krzysztof Oleszkiewicz, and Jakub Wojtaszczyk. On some extensions of the FKN theorem. Manuscript, 2012. To appear in *Theory of Computation*.

20 Daniel Kane. k-independent Gaussians fool polynomial threshold functions. In *Proceedings of the 26th Annual Computational Complexity Conference*, pages 252–261, 2011.

21 Daniel Kane and Raghu Meka. A PRG for Lipschitz functions of polynomials with applications to Sparsest Cut. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 1–10, 2013.

22 Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002.

23 Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. *Mathematische Annalen*, 334(4):821–852, 2006.

24 Subhash Khot and Assaf Naor. Linear equations modulo 2 and the L_1 diameter of convex bodies. *SIAM Journal on Computing*, 38(4):1448–1463, 2008.

25 Guy Kindler. *Property Testing, PCP, and juntas*. PhD thesis, Tel Aviv University, 2002.

26 Guy Kindler and Ryan O'Donnell. Gaussian noise sensitivity and Fourier tails. In *Proceedings of the 27th Annual Computational Complexity Conference*, pages 137–147, 2012.

27 Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas. Manuscript, 2002.

28 Stanisław Kwapień. Decoupling inequalities for polynomial chaos. *Annals of Probability*, 15(3):1062–1071, 1987.

29 Shachar Lovett. An elementary proof of anti-concentration of polynomials in Gaussian variables. Technical Report 182, Electronic Colloquium on Computational Complexity, 2010.

30 Nathaniel Macon and Abraham Spitzbart. Inverses of Vandermonde matrices. *The American Mathematical Monthly*, 65:95–100, 1958.

31 Konstantin Makarychev and Maxim Sviridenko. Solving optimization problems with diseconomies of scale via decoupling. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science*, pages 571–580, 2014.

32 Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010.

33 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.