Electronic Colloquium on Computational Complexity, Report No. 145 (2019)

XOR Lemmas for Resilient Functions Against Polynomials

Eshan Chattopadhyay[*]
Cornell University
eshanc@cornell.edu

Pooya Hatami
Ohio State University
pooyahat@gmail.com

Kaave Hosseini
Carnegie Melon University
kaave.hosseini@gmail.com

Shachar Lovett[†]
University of California, San Diego
slovett@cs.ucsd.edu

David Zuckerman[‡]
University of Texas at Austin
diz@cs.utexas.edu

October 30, 2019

Abstract

A major challenge in complexity theory is to explicitly construct functions that have small correlation with low-degree polynomials over $\mathbb{F}_2$. We introduce a new technique to prove such correlation bounds with $\mathbb{F}_2$ polynomials. Using this technique, we bound the correlation of an XOR of Majorities with constant degree polynomials. In fact, we prove a more general XOR lemma that extends to arbitrary resilient functions. We conjecture that the technique generalizes to higher degree polynomials as well.

A key ingredient in our new approach is a structural result about the Fourier spectrum of low degree polynomials over $\mathbb{F}_2$. We show that for any n-variate polynomial p over $\mathbb{F}_2$ of degree at most d, there is a small set $S \subset [n]$ of variables, such that almost all of the Fourier mass of p lies on Fourier coefficients that intersect with S. In fact our result is more general, and finds such a set S for any low-dimensional subspace of polynomials. This generality is crucial in deriving the new XOR lemmas.

[*] Supported by NSF grant CCF-1849899

[†] Supported by NSF grant CCF-1614023

[‡] Supported by NSF Grant CCF-1705028 and a Simons Investigator Award (#409864).

ISSN 1433-8092

1 Introduction

Understanding the power and limitations of multivariate polynomials over the field $\mathbb{F}_2$ as a model of computation is of fundamental interest in complexity theory. A natural measure of complexity of a Boolean function is the degree of the unique multilinear polynomial that exactly computes it. However, under this measure the very simple AND_n function has maximum possible degree n, since it is computed by the monomial $x_1x_2\dots x_n$. A robust measure that is more indicative of complexity is the minimum degree of a polynomial that correlates with a function. Following tradition in the computer science community, we define the correlation of two Boolean functions f and g as:

$$\mathrm{corr}(f,g) := \big|\Pr[f(x) = g(x)] - \Pr[f(x) \neq g(x)]\big|.$$

Hence, $\mathrm{corr}(f,g)$ is a real number in the interval $[0,1]$, and the closer $\mathrm{corr}(f,g)$ is to 1 the better g approximates f. Let $\mathrm{Poly}_{n,d}$ denote the class of n-variate polynomials of degree at most d over $\mathbb{F}_2$. Given any n-variate Boolean function f, the following is a natural quantity that measures how well degree d polynomials approximate the function f:

$$\mathrm{corr}(f,d) := \max_{p\in\mathrm{Poly}_{n,d}} \mathrm{corr}(f,p).$$

One motivation for studying the above quantity stems from seminal works of Razborov [Raz87] and Smolensky [Smo87], who proved that small-depth circuits can be well approximated by low-degree polynomials over $\mathbb{F}_2$. This leads to a natural way of proving circuit lower bounds by explicitly constructing functions that cannot be approximated by low-degree polynomials. Another motivation for constructing average-case hard functions comes from the seminal hardness vs. randomness paper of Nisan and Wigderson [NW94], where they proved that an explicit average-case hard function for a complexity class can be used to construct pseudorandom generators for the same complexity class. Pseudorandom generators provide a generic way of removing or reducing the use of randomness, and have a host of applications in computer science.

A third motivation to study correlation bounds for polynomials stems from connections to obtaining certain bounds on the Fourier spectrum of polynomials. A recent line of work [CHHL18, CHLT19] suggests a new way of constructing efficient pseudorandom generators for polynomials and the class $\mathrm{AC}^0[\oplus]$ based on bounds on the Fourier mass on the level 2 Fourier coefficients. Such Fourier bounds look intimately connected with the problem of proving new correlation bounds. In fact, a tight bound on the Fourier mass on level 1 Fourier coefficients for polynomials was obtained in [CHLT19] using correlation bounds proved by Razborov and Smolensky [Raz87, Smo93]. Finally, recent work of Golovnev, Kulikov, and Williams [GKW19] shows that even mild improvements in known correlation bounds against polynomials will result in new circuit lower bounds.

Known correlation bounds for polynomials. The best known correlation bounds against polynomials can be described in two different regimes. For constant degree polynomials, there exists explicit functions [BNS92, Bou05, GRS05, VW07] that have exponentially small correlation. More precisely, we know examples of explicit functions that have correlation $2^{-\Omega(n/2^d)}$ with degree d polynomials. Note that this gives nothing meaningful when $d = \log n$.

For the regime of large d, we have much weaker correlation bounds. In particular, techniques introduced by Razborov [Raz87] and Smolensky [Smo93] can be used to show that the n-variate MAJORITY function (which we shorthand as Maj) has correlation $O(d/\sqrt{n})$ with degree d polynomials, and this is in fact the best known correlation bound in this regime. In particular, even constructing an explicit function $f : \{0,1\}^n \to \{0,1\}$ with $\operatorname{corr}(f, n^{0.51}) = o(1)$ or $\operatorname{corr}(f, \log n) = O(1/n)$ remains an outstanding challenge. See the survey by Viola [Vio09] for more discussion on the current state of the art in correlation bounds for polynomials.

One way of constructing hard functions is to start with a mildly hard function, and use some type of hardness amplification. A general template of hardness amplification is based on XOR lemmas. Typically one starts with a function f such that $\operatorname{corr}(f, g) \leq \varepsilon$, for all g in some function class $\mathcal{G}$, and hopes to prove that the function computing the XOR of f applied on k independent inputs has correlation ε^k with any $g \in \mathcal{G}$. Some examples of such XOR lemmas are known in complexity theory and cryptography [Yao82,Vaz86,VV84,GL89], with the most well known being Yao's XOR lemma. The only XOR lemmas for the class of $\mathbb{F}_2$ polynomials is from the work of Viola and Wigderson [VW07], who proved the following theorem.

Given any Boolean function f, define $f^{\oplus k}$ to be the function that outputs the XOR of f applied on k independent inputs.

Theorem 1.1 ([VW07]). *Suppose that for some Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have $\operatorname{corr}(f, d) \leq 1 - 1/2^d$. Then, for any integer $k \geq 1$,*

$$\operatorname{corr}(f^{\oplus k}, d) \leq 2^{-\Omega(k/d \cdot 4^d)}.$$

This was used in [VW07] to give a unified way of proving correlation bounds obtained in [BNS92, Bou05, GRS05, VW07]. However, the above theorem does not give improved bounds if the base function f has much smaller correlation with degree d polynomials (e.g., when applied to Maj).

1.1 Our results

One of the motivations for this work is a conjecture of [CHLT19] on the second level of the Fourier spectrum of degree d polynomials, which would imply explicit pseudorandom generators for $\mathrm{AC}^0[\oplus]$ with polylogarithmic seed-length. Such bounds on the Fourier spectrum of polynomials are related to bounding $\operatorname{corr}(\mathrm{Maj} \oplus \mathrm{Maj}, d)$ by a quantity of the form $\operatorname{poly}(\log n, d)/n$. We discuss these connections in more details towards the end of this section and in Section 6. Unfortunately, none of the known techniques for proving correlation bounds with polynomials is capable of proving such bounds for Maj$\oplus$Maj. The Razborov-Smolensky technique cannot prove correlation bounds smaller than $\frac{1}{\sqrt{n}}$ for any function. We discuss why the Viola-Wigderson approach of using Gowers norm doesn't work in Appendix A. See the survey by Viola [Vio09] for more discussion on limitations of existing techniques.

In this paper, we introduce a new technique for proving correlation bounds against $\mathbb{F}_2$ polynomials that is based on structural results for low-degree polynomials. Our new technique allows us to prove the desired bounds for correlation of $\mathrm{Maj} \oplus \mathrm{Maj}$ with constant degree polynomials. That is, we prove that

$$\mathrm{corr}(\mathrm{Maj} \oplus \mathrm{Maj}, O(1)) \leq \frac{\mathrm{poly}(\log n)}{n}.$$

In fact our method allows proving correlation bounds for XOR of several copies of Maj.

Proposition 1.2 (informal). *Let* Maj *be the Majority function on n bits. Then for any $k \geq 1$ it holds that*

$$\mathrm{corr}(\mathrm{Maj}^{\oplus k}, O(1)) \leq \left(\frac{\mathrm{poly}(k, \log n)}{\sqrt{n}}\right)^k.$$

As discussed above, even the $k = 2$ case of the above result was not known prior to our work. In fact, we obtain nontrivial correlation bounds for degrees that can grow slowly with n. See Section 5 for the precise statements of our results.

Correlation of polynomials with resilient functions. Our bounds for Maj are a special case of more general correlation bounds that we prove for *resilient functions*. The notion of resilience of Boolean functions, first introduced by Ben-Or and Linial [BL85], is well studied in distributed computing with applications to collective coin flipping. Roughly speaking, a function is highly resilient if no small coalition of its variables are able to bias the outcome of the function.

Definition 1.3 (Resilient functions). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and $S \subset [n]$ be an arbitrary subset of coordinates. The* influence of S on f, *denoted by $I_S(f)$, is the probability that randomly fixing the coordinates outside S does not fix the value of f. The function f is called* (q, ε)-resilient *if for every subset $S \subset [n]$ of size at most q, we have $I_S(f) \leq \varepsilon$.*

The PARITY function is not even $(1, .99)$-resilient because any single coordinate can change the value of the function (irrespective of the values of the other variables). On the other hand, Maj is $(\Theta(\sqrt{n}), .01)$-resilient, which is fairly good.

In particular, our correlation bounds hold for resilient functions with the stronger property that the influence of a set S scales proportional to its cardinality. We call such functions *strong resilient functions*.

Definition 1.4 (Strong resilient functions). *A function $f : \{0,1\}^n \to \{0,1\}$ is called strong r-resilient if for all $q \leq r$, f is a $(q, q/r)$-resilient function.*

Fact 1.5. Maj *is a strong $\Theta(\sqrt{n})$-resilient function.*

There are better strong resilient functions than Maj. Ben-Or and Linial [BL85] showed that the recursive majority function, defined on $n = 3^k$ bits as $\mathrm{Maj}^k(x^1, x^2, x^3) = \mathrm{Maj}(\mathrm{Maj}^{k-1}(x^1), \mathrm{Maj}^{k-1}(x^2), \mathrm{Maj}^{k-1}(x^3))$ and $\mathrm{Maj}^1 = \mathrm{Maj}$, is a strong n^β-resilient function, where $\beta = \log_3 2 \approx 0.63$. Ajtai and Linial [AL93] proved the existence of Boolean functions that are strong $\Omega(n/(\log n)^2)$-resilient function. Recent works [Mek17, CZ19] explicitly constructed such strong resilient functions that match the probabilistic construction of [AL93]. A consequence of the KKL theorem [KKL88] is that this is almost tight. In particular, the KKL theorem implies that for any Boolean function f, there exists a set S of size $O(n/\log n)$ such that $I_S(f) = \Omega(1)$.

We need one more definition before stating our result for resilient functions. The *bias* of a function $f : \{0,1\}^n \to \{0,1\}$ is its correlation with constant functions, namely $\mathrm{bias}(f) = \mathrm{corr}(f, 0)$. Given $d \geq 1$ we will take $D = O(d)^{O(d)}$ in the theorems below. Our main result for resilient functions is the following.

Theorem 1.6 (informal version of Theorem 4.1)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a strong r-resilient function. Then for $d \geq 1$, there is a choice of $D = O(d)^{O(d)}$ such that*

$$\mathrm{corr}(f, d) \leq \mathrm{bias}(f) + \frac{\log(r)^D}{r}.$$

New XOR lemmas for polynomials. Our technique for proving correlation bounds for resilient functions allows us to prove XOR lemmas for resilient functions. We state a slightly informal version of our XOR lemma below.

Theorem 1.7 (informal version of Theorem 5.2)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a strong r-resilient function. Assume furthermore that f is unbiased. Then for $d \geq 1$, there is a choice of $D = O(d)^{O(d)}$ such that*

$$\mathrm{corr}(f^{\oplus k}, d) \leq \left(\frac{(k \log r)^D}{r}\right)^k.$$

Our correlation bounds for $\mathrm{Maj} \oplus \mathrm{Maj}$ and more generally $\mathrm{Maj}^{\oplus k}$ follow directly from the above theorem by setting $r = \Theta(\sqrt{n})$. We sketch a proof of this theorem in Section 1.3.

A structure theorem for polynomials. Our correlation bounds are based on a new structural result for polynomials. To state our result precisely, we introduce a notion that we call *local correlation*, first studied by Lovett, Mukhopadhyay and Shpilka [LMS13].

Define $\mathrm{e} : \mathbb{F}_2 \to \mathbb{R}$ as $\mathrm{e}(x) = (-1)^x$. For $f : \mathbb{F}_2^n \to \mathbb{F}_2$ define $\mathrm{e}(f) : \mathbb{F}_2^n \to \mathbb{R}$ by $\mathrm{e}(f)(x) = \mathrm{e}(f(x))$.

Definition 1.8 (Local correlation)**.** *Given a function $F : \mathbb{F}_2^n \to \mathbb{R}$ and $S \subset [n]$ define the S-local correlation of F as*

$$\Delta_S(F) := \mathbb{E}_{x,y}\left[F(x)F(y) | x_{S^c} = y_{S^c}\right] - \mathbb{E}[F]^2.$$

For $f : \mathbb{F}_2^n \to \mathbb{F}_2$ we abbreviate $\Delta_S(f) = \Delta_S(e(f))$.

To develop some intuition about local correlation, we record the following simple fact. Denote by U_S the uniform distribution over $\mathbb{F}_2^S$. Given $x_1 \in \mathbb{F}_2^{[n]\setminus S}$ and $x_2 \in \mathbb{F}_2^S$, we denote by $F(x_1, x_2)$ the function F applied to the concatenated input $(x_1, x_2) \in \mathbb{F}_2^n$, where x_1 appears in the coordinates given by $[n] \setminus S$ and x_2 in the coordinates given by S.

Fact 1.9. *For any $F : \mathbb{F}_2^n \to \mathbb{R}$, and any set $S \subseteq [n]$,*

$$\Delta_S(F) = \mathbb{E}_{x_1 \sim U_{[n]\setminus S}} \left[(\mathbb{E}_{x_2 \sim U_S}[F(x_1, x_2)] - \mathbb{E}[F])^2 \right].$$

Thus, if S-local correlation of F is small, then for most fixings of the variables outside S, the average of the resulting restricted function is close to the global average of F. In other words, for most fixings of the coordinates outside S, the restricted function on the resulting affine subspace has bias that is close to the global bias of F.

Our method involves finding, for any low-degree polynomial p, a small set S with small local correlation. For intuition, we discuss degrees 1 and 2 first. For degree 1, we can take S to be any variable appearing in p, and we get $|S| = 1$ and $\Delta_S(p) = 0$.

For degree 2, we can write $p = \ell_0 + \sum_{i=1}^r \ell_{2i-1}\ell_{2i}$, for p a rank r quadratic form, where the linear terms $\ell_j = \langle v_j, x\rangle$, $j = 1, \ldots, 2r$, are linearly independent linear functions. Construct an $n \times 2r$ matrix M over $\mathbb{F}_2$, with v_j as the j'th column. Then we can take $S \subset [n]$ to be any subset of size $\min\{\log(1/\epsilon), 2r\}$ such that the corresponding rows of M are linearly independent. Note that such a set S indeed exists since the rank of M is $2r$. The proof that this indeed works follows from the Fourier interpretation of local correlation (Fact 1.12). Briefly, Fact 1.12 implies that the S-local correlation of p equals the Fourier mass on the subspace $W = \operatorname{span}\{e_i : i \in [n] \setminus S\}$ (where e_i's denote the standard unit vectors in $\mathbb{F}_2^n$). It can be shown that the Fourier mass of p is evenly spread on an affine shift V' of the subspace $V = \operatorname{span}\{v_j : j \in [2r]\}$. It is then not difficult to prove that the affine dimension of $V' \cap W$ is at most $\dim(V) - \log(1/\varepsilon)$, and hence the Fourier mass on W is bounded by ε.

Finding small S for larger degrees is much harder. Our main technical contribution is the following theorem.

Theorem 1.10 (Informal version, special case of Theorem 3.1)**.** *For $d \geq 1$, there is a choice of $D = O(d)^{O(d)}$ such that the following holds. For any polynomial $p \in \mathrm{Poly}_{n,d}$ and any $\varepsilon > 0$, there exists a set $S \subset [n]$ of size $|S| \leq \log(1/\varepsilon)^D$ such that $\Delta_S(p) \leq \varepsilon$.*

This is an exponential improvement in terms of the parameter ε over the bound of $(1/\varepsilon)^{O(4^d)}$ proved in [LMS13]. This improvement on the dependence of ε in the size of the set S is crucial for our approach to prove correlation bounds, as it allows us to choose $\varepsilon = 1/\mathrm{poly}(n)$, which is needed to analyze Majority or more general resilient functions. We conjecture that this can be improved even further, and discuss this in Section 1.2.

In fact, our result is more general. We show in Theorem 3.1 that for any subspace $V \subset \mathrm{Poly}_{n,d}$, there exists a set $S \subset [n]$, $|S| \leq (\dim(V)\log(1/\varepsilon))^D$, such that for all $p \in V$, $\Delta_S(p) \leq \varepsilon$. This generality is crucial in our application to proving the new XOR lemmas for correlation against polynomials, which we explain below. Again, we conjecture that this bound can be improved.

We give a proof overview of Theorem 1.10 in Section 1.4. Combining Theorem 1.10 with Fact 1.9, we obtain the following useful result.

Corollary 1.11. *For any polynomial $p \in \mathrm{Poly}_{n,d}$, there exists a set of size $S \subset [n]$, $|S| \leq (\log(1/\varepsilon))^D$, such that*

$$\Pr_{x_1 \sim U_{[n]\setminus S}}[|\mathbb{E}_{x_2 \sim U_S}[e(p(x_1, x_2))] - \mathbb{E}[e(p)]| > \varepsilon] \leq \varepsilon.$$

Proof sketch of Theorem 1.6. Given this structural result, it is simple to prove Theorem 1.6. Roughly, Corollary 1.11 says that for any low degree polynomial p, there is a small set S such that for most fixings of variables outside S, restricting p to S leaves its bias almost unchanged. On the other hand, resilient functions f have the property that for most such fixings, the restriction of f is constant. Therefore, for most fixings the restrictions of f and p are uncorrelated.

In a little more detail, recall that we are trying to bound the quantity

$$|\mathbb{E}[\mathrm{e}(p(x)) \cdot \mathrm{e}(f(x))]|,$$

where $p \in \mathrm{Poly}_{n,d}$ and f is a strong r-resilient function. For simplicity, we assume here that $\mathrm{bias}(f) = 0$ (see Section 4 for the general case). Recall that $D = O(d)^{O(d)}$ and define the function $C(d, \varepsilon) = (\log(1/\varepsilon))^D$. Let ε be a parameter that we set later. By Corollary 1.11, there exists a set $S \subset [n]$, $|S| \leq C(d, \varepsilon^4)$, such that with probability $1 - \varepsilon$ over $x_1 \sim U_{[n]\setminus S}$, we have $|\mathbb{E}_{x_2 \sim U_S}[\mathrm{e}(p(x_1, x_2)] - \mathbb{E}[\mathrm{e}(p)]| \leq \varepsilon$. Further, since f is a strong r-resilient function, with probability at least $1 - \frac{C(d,\varepsilon^4)}{r}$ over the sampling of x_1, $f(x_1, \cdot)$ is a constant function. Thus, with probability $1 - \frac{C(d,\varepsilon^4)}{r} - \varepsilon$ over the sampling of $x_1 \sim U_{[n]\setminus S}$,

$$\mathbb{E}_{x_2 \sim U_S}[\mathrm{e}(f(x_1, x_2)) \cdot (\mathrm{e}(p(x_1, x_2)) - \mathbb{E}[\mathrm{e}(p)])] \leq \varepsilon.$$

Thus, it follows that $|\mathbb{E}_x[\mathrm{e}(p(x)) \cdot \mathrm{e}(f(x))]| \leq \frac{C(d,\varepsilon^4)}{r} + 2\varepsilon$. We set $\varepsilon = O(1/r)$. This completes the proof sketch of Theorem 1.6.

On the Fourier spectrum of low degree polynomials. The S-local correlation of F is related to the Fourier spectrum of F as the following fact shows.

Fact 1.12. $\Delta_S(F) = \sum_{T \subseteq [n]: T \neq \emptyset, T \cap S = \emptyset} \widehat{F}(T)^2$.

Thus, $\Delta_S(F) \leq \varepsilon$ if most of the Fourier mass of F is on sets that intersect S. Hence, an immediate consequence of Theorem 1.10 is that for any polynomial $p \in \mathrm{Poly}_{n,d}$, there is a small set $S \subset [n]$ such that almost all of the Fourier mass of $\mathrm{e}(p)$ lies on Fourier coefficients that intersect with S.

Corollary 1.13. *For any polynomial $p \in \mathrm{Poly}_{n,d}$, there exists a set $S \subset [n]$ of size at most $(\log(1/\varepsilon))^D$ such that*

$$\sum_{T \subseteq [n]: T \neq \emptyset, T \cap S = \emptyset} \widehat{e(p)}(T)^2 \leq \varepsilon.$$

The previous bound on the smallest such set S was $|S| \leq (1/\varepsilon)^{O(4^d)}$ [LMS13].

In a different direction, we show in Section 6 that bounding the Fourier mass of polynomials on degree 2 coefficients is related to bounds on $\mathrm{corr}(\mathrm{Maj} \oplus \mathrm{Maj}, d)$. The motivation for studying this quantity arises from recent works [CHHL18, CHLT19], where bounds on Fourier tails of classes of Boolean functions have been been exploited to construct pseudorandom generators. In particular, it is proved in [CHLT19] that if for a class of Boolean functions $\mathcal{F}$ that is closed under restrictions, all $f \in \mathcal{F}$ satisfy $\sum_{1 \leq i < j \leq n} |\widehat{f}(i,j)| \leq t$, then there exists an efficient pseudorandom generator for $\mathcal{F}$ with seed-length $\mathrm{poly}(\log n, t)$.

It is conjectured in [CHLT19] that for any $p \in \mathrm{Poly}_{n,d}$, $\sum_{1 \leq i < j \leq n} |\widehat{p}(i,j)| \leq d^2$. Proving this will immediately imply explicit pseudorandom generators for the class $\mathrm{AC0}[\oplus]$ with polylogarithmic seed-length, which is an outstanding open question in complexity theory. However, currently the best known bound is

$$\sum_{1 \leq i < j \leq n} |\widehat{p}(i,j)| \leq \min \left\{ O(2^d), O(d\sqrt{n \log n}) \right\},$$

while it was proved in [CHLT19] that $\sum_i |\widehat{p}(i)| \leq 4d$. The proof of the latter bound crucially used bounds on $\mathrm{corr}(\mathrm{Maj}, d)$ due to Razborov and Smolensky [Raz87, Smo87]. A natural step towards proving the above conjecture is to first bound the weaker quantity $\left| \sum_{1 \leq i < j \leq n} \widehat{p}(i,j) \right|$. We show in Section 6 this is related to proving bounds for $\mathrm{corr}(\mathrm{Maj} \oplus \mathrm{Maj}, d)$.

1.2 Discussion and future directions

New correlation bounds. We view this paper as a proof of concept, showing that structural results such as Theorem 1.10 and Theorem 3.1 can be used to obtain correlation bounds and XOR lemmas for resilient functions. As a result, any improvement to the bounds in Theorem 1.10 immediately gives stronger correlation bounds for explicit functions that are highly resilient. Furthermore, any improvement on Theorem 3.1 would lead to stronger XOR type theorems. While our bounds are exponentially smaller than prior results in terms of the error parameter ε, we believe that the dependence on degree d is far from optimal and can be drastically improved.

Conjecture 1.14. *Theorem 1.10 holds with* $|S| \leq \mathrm{poly}(d, \log(1/\varepsilon))$.

More generally, we believe that the following holds.

Conjecture 1.15. *Theorem 3.1 holds with* $|S| \leq \mathrm{poly}(d, k, \log(1/\varepsilon))$.

Conjecture 1.15 allows one to replace Theorem 1.7 with much stronger correlation bounds. For example, in the case of $\mathrm{Maj}^{\oplus k}$ we would get $\mathrm{corr}(\mathrm{Maj}^{\oplus k}, \log n) \leq (\log n)^{O(k)}/n^{k/2}$, which would be a major breakthrough for any $k = \omega(1)$.

Pseudorandom generators. Another natural direction of research is to see if one can extend techniques from [CHLT19] to construct efficient pseudorandom generators with seed-length $\mathrm{poly}(\log n, t)$ for any classes $\mathcal{F}$ of Boolean functions that is closed under restrictions, such that for all $f \in \mathcal{F}$ we have $|\sum_{1\le i<j\le n} \widehat{f}(i,j)| \le t$. (To recall, [CHLT19] makes the stronger assumption $\sum_{1\le i<j\le n} |\widehat{f}(i,j)| \le t$). We show in Section 6 that the quantity $|\sum_{1\le i<j\le n} \widehat{f}(i,j)|$ is closely related to the correlation of f with $\mathrm{Maj} \oplus \mathrm{Maj}$. Thus if one can construct pseudorandom generators for such $\mathcal{F}$, then along with Conjecture 1.15, this will immediately imply polylogarithmic seed-length pseudorandom generators for $\mathrm{AC0}[\oplus]$.

1.3 Proof sketch of Theorem 1.7

We now sketch the main ideas that are used to prove Theorem 1.7, for the case $k = 2$. The general case can be proved using similar ideas based on an inductive strategy. Our goal is to bound

$$|\,\mathbb{E}_{x,y}[\mathrm{e}(p(x,y))\mathrm{e}(f(x))\mathrm{e}(f(y))|,$$

for any polynomial $p(x,y) \in \mathrm{Poly}_{2n,d}$, where we assume that f is unbiased (namely, $\mathbb{E}[\mathrm{e}(f(x))] = 0$). Towards this, define the following functions:

$$\begin{aligned} H(x,y) &= \mathrm{e}(p(x,y)) \cdot \mathrm{e}(f(y)), \\ K(x) &= \mathbb{E}_y[H(x,y)] = \mathbb{E}_y[\mathrm{e}(p(x,y)) \cdot \mathrm{e}(f(y))]. \end{aligned}$$

We have

$$\mathbb{E}_{x,y}\left[\mathrm{e}(p(x,y))\mathrm{e}(f(x))\mathrm{e}(f(y))\right] = \mathbb{E}_x\left[K(x)\cdot \mathrm{e}(f(x))\right] = \mathbb{E}_x[(K(x) - \mathbb{E}[K])\cdot \mathrm{e}(f(x))],$$

where the second equality holds because $\mathbb{E}_x[\mathrm{e}(f(x))] = 0$.

Define the function $C(d,k,\varepsilon) = (k\log(1/\varepsilon))^D$, where to recall $D = D(d) = O(d)^{O(d)}$. We claim that there is a set $S \subset [n]$, $|S| \le C(d, k = \log^{O(d)}(r), \varepsilon = 1/r^4)$, such that $\Delta_S(K) \le \varepsilon$. Before proving this, we first assume such a set S and show how to obtain the required correlation bound.

Let $\mathcal{E}_1$ be the event that on sampling $x_1 \sim U_{[n]\setminus S}$, the restricted function $f(x_1,\cdot)$ is a constant function. Since f is a strong r-resilient function, it follows that $\Pr[\mathcal{E}_1] > 1 - |S|/r$. Thus, we have

$$\begin{aligned} |\,\mathbb{E}_x[(K(x) - \mathbb{E}[K])\cdot \mathrm{e}(f(x))]| \le |\,\mathbb{E}_x[(K(x) - \mathbb{E}[K])\cdot \mathrm{e}(f(x))|\mathcal{E}_1]+ \\ |\,\mathbb{E}_x[(K(x) - \mathbb{E}[K])\cdot e[f(x)]|\neg\mathcal{E}_1]|\cdot \Pr[\neg\mathcal{E}_1] \end{aligned} \qquad (1)$$

We now individually bound the two terms appearing on right hand side of Equation (1). Using Fact 1.9 and an application of Markov's inequality, it follows that with probability $1 - \varepsilon^{1/2}$ over $x_1 \sim U_{[n]\setminus S}$, we have $|\,\mathbb{E}_{x_2\sim U_S}[K(x_1,x_2) - \mathbb{E}[K]| \le \varepsilon^{1/2}$. Thus, the first term can be bounded by $O(\varepsilon^{1/2}) = O(1/r^2)$, by our choice of ε.

Next, we bound the term $|\,\mathbb{E}_x[(K(x) - \mathbb{E}[K])\cdot e[f(x)]|\neg\mathcal{E}_1]|$. Consider any fixing of x. Recalling that $K(x) = \mathbb{E}_y[\mathrm{e}(p(x,y)\cdot f(y)]$, it follows that for a fixed x, $|K(x)|$ is just the correlation of a resilient function (namely, $f(y)$) with a polynomial of degree at most d (namely, $q(y) = p(x, y)$). This is exactly the quantity we bound in Theorem 1.6. Thus, for any x, $|K(x)| \leq (\log r)^D/r$. Hence, we have

$$\begin{aligned} |\mathbb{E}_x[(K(x) - \mathbb{E}[K]) \cdot \mathrm{e}(f(x))|\neg\mathcal{E}_1]| &\leq |\mathbb{E}_x[K(x) \cdot e[f(x)]|\neg\mathcal{E}_1]| + |\mathbb{E}[K]| \cdot |\mathbb{E}_x[\mathrm{e}(f(x))|\neg\mathcal{E}_1]| \\ &\leq \max_x |K(x)| + |\mathbb{E}[K]| \\ &\leq 2\max_x |K(x)| \leq (\log r)^D/r. \end{aligned}$$

Since $\Pr[\mathcal{E}_1] > 1 - |S|/r$, we can now bound the second term appearing on the right hand side of Equation (1). Using the above estimate, we have

$$|\mathbb{E}_x[(K(x) - \mathbb{E}[K]) \cdot e[f(x)]|\neg\mathcal{E}_1]| \cdot \Pr[\neg\mathcal{E}_1] \leq \frac{(\log r)^D \cdot |S|}{r^2}.$$

Combining the above estimates, and continuing from Equation (1), we have

$$|\mathbb{E}_x[(K(x) - \mathbb{E}[K]) \cdot \mathrm{e}(f(x))]| \leq \frac{(\log r)^D \cdot |S|}{r^2} + O(1/r^2).$$

Thus, we obtain the required correlation bound.

We now sketch the proof to show the existence of the set S such that $\Delta_S(K) \leq \varepsilon$. Recall that $K(x) = \mathbb{E}_y[H(x, y)] = \mathbb{E}_y[\mathrm{e}(p(x, y)) \cdot \mathrm{e}(f(y))]$. For a randomly sampled subspace A of dimension ℓ (to be fixed later), we have

$$\begin{aligned} K(x) &= 2^{-n} + (1 - 2^{-n})\,\mathbb{E}_{y \in \mathbb{F}_2^n \setminus \{0\}}[\mathrm{e}(p(x, y)) \cdot \mathrm{e}(f(y))] \\ &= 2^{-n} + \mathbb{E}_A\, \mathbb{E}_{a \in A \setminus \{0\}}[\mathrm{e}(p(x, a) + f(a))]. \end{aligned}$$

Further, for distinct $a, a' \in \mathbb{F}_2^n \setminus \{0\}$, the events $a \in A$ and $a' \in A$ are pairwise independent. Using this, one can show the existence of a subspace A of dimension $\ell = O(\log(1/\varepsilon))$, such that

$$\mathbb{E}_x[(K(x) - K_A(x))^2] \leq \varepsilon,$$

where $K_A(x) = 2^{-n} + \mathbb{E}_{a \in A \setminus \{0\}}[\mathrm{e}(p(x, a) + f(a))]$.

Thus, up to an additional error of ε, it is enough to find a set S such that for any $a \in A$, $\Delta_S(q_a) \leq \varepsilon$, where $q_a(x) = p(x, a) + f(a)$. We prove that the dimension of the span of the polynomials $\{q_a : a \in A\}$ can be bounded by $\binom{\ell}{\leq d}$. This follows from a more general bound that we prove on the dimension of derivatives of polynomials (Claim 2.5). Now, we can appeal to the more general version of Theorem 1.10 that works for low dimensional subspaces of polynomials (Theorem 3.1) to finish the proof.

1.4 Proof overview of Theorem 1.10

At a high level, the proof of Theorem 1.10 goes via induction on d, using the well known *structure vs randomness paradigm.* We sketch a more general version of Theorem 1.10 (see

Theorem 3.1) and prove the following: Let $V \subset \mathrm{Poly}_{n,d}$ be a subspace of dimension k. We prove the existence of a set $S \subset [n]$, $|S| \leq C(d,k,\varepsilon) = (k\log(1/\varepsilon))^D$ for $D = O(d)^{O(d)}$, such that for all $f \in V$, $\Delta_S(f) \leq \varepsilon$.

We first introduce a definition. Let W be a linear space of functions $g : \mathbb{F}_2^n \to \mathbb{F}_2$. Define the ball of radius r around W, denoted $\mathcal{B}(W,r)$, as the set of all functions $G : \mathbb{F}_2^n \to \mathbb{R}$ of the form

$$\mathcal{B}(W,r) := \left\{ G = \sum_{g \in W} c_g \mathrm{e}(g) : \sum |c_g| \leq r \right\}.$$

Our key result that fits into the "structure vs randomness" paradigm, stated as Lemma 3.3, is the following: we prove the existence of a low dimensional subspace $W \subset \mathrm{Poly}_{n,d-1}$, such that each $f \in V$ either has no large Fourier coefficient or is close to a function in a small ball around W.

More precisely we prove the following result. For any $F : \{0,1\}^n \to \mathbb{R}$, define $\|\widehat{F}\|_\infty = \max_\gamma |\widehat{F}(\gamma)|$.[1] Let $V \subset \mathrm{Poly}_{n,d}$ be a subspace of dimension k. For any parameters $\varepsilon_0, \delta > 0$, we prove the existence of subspaces $W \subset \mathrm{Poly}_{n,d-1}$ of dimension $\ell = O(k\log(1/\varepsilon_0\delta))^{O(d)}$ and $U \subset V$, such that

(i) Each $f \in V \setminus U$ satisfies $\|\widehat{\mathrm{e}(f)}\|_\infty \leq \varepsilon_0$.

(ii) Each $f \in U$ can be expressed as $\mathrm{e}(f) = G + H$, where $G \in \mathcal{B}(W, 1/\varepsilon_0^k)$ and $\|H\|_2 \leq \delta/\varepsilon_0^k$.

Before sketching the proof of this structure result, we first show how this can be used to construct the required set S.

Let U, W_1 be the subspaces that we get on applying the above structure result to V with parameters δ, ε_0 (to be fixed later). Since our proof is by induction on d (the base case, for $d = 1$, is direct), we can assume that, for some parameter ε_1 to be fixed later, there exists a set $S_1 \subset [n]$, $|S_1| \leq (k\log(1/\varepsilon_1))^D$ such that for any $g \in W_1$, $\Delta_{S_1}(g) \leq \varepsilon_1$. For any $f \in U$, we use the fact that it is close (in L_2 distance) to a function in a small ball around W, to show that $\Delta_{S_1}(f) \leq O((\varepsilon_1 + \delta)/\varepsilon_0^{2k})$.

Now suppose $f \in V \setminus U$. Here appealing to a result proved by Lovett et al. [LMS13] (see Lemma 3.4 and Lemma 3.5) we show that: for any set $S \subset [n]$, any parameter $m > 0$, there exists a subspace $A \subset \mathbb{F}_2^n$ of dimension m such that for any $f \in V \setminus U$, we have

$$\Delta_S(f)^2 \leq 2^k \cdot \mathbb{E}_{a \sim A}[\Delta_S(f_a)] + 2^{-m} + \|e(f)\|_\infty^2,$$

where $f_a(x) := f(x+a) - f(x)$ is derivative of f in the direction a. Now, since $f \in V \setminus U$, we know that $\|e(f)\|_\infty^2 \leq \varepsilon_0^2$. Further, note that $W_2 = \{f_a : a \in A\} \subset \mathrm{Poly}_{n,d-1}$. We prove that the dimension of the space $\mathrm{span}\{W_2\}$ can be bounded by $\binom{m}{\leq d}$ (see Claim 2.5), where m is the dimension of A. Thus, again using induction, we get a set $S_2 \subset [n]$, such that $\Delta_{S_2}(g) \leq \varepsilon_2$, for all $g \in W_2$. It is now straightforward to show that for each $f \in V \setminus U$, $\Delta_{S_2}(f)^2 \leq \varepsilon_2 + 2^k(\varepsilon_0^2 + 2^{-m})$. Setting $S = S_1 \cup S_2$, with appropriate choices of parameters finishes the proof.

[1] see Section 2.2 for a quick recap of Boolean Fourier analysis.

We now briefly sketch the proof of Lemma 3.3. A crucial ingredient (see Lemma 3.2) is the following result about biased functions: Let $f \in \mathrm{Poly}_{n,d}$ and $\eta = \mathbb{E}[\mathrm{e}(f)]$, such that $|\eta| \geq \varepsilon_0$. Then there exists a subspace $W \subset \mathrm{Poly}_{n,d-1}$ of dimension $\binom{c' \log(1/\epsilon_0 \delta)}{\leq d}$, for some constant c', such that $\mathrm{e}(f) = G + H$ where $G, H : \mathbb{F}_2^n \to \mathbb{R}$ satisfy $G \in \mathcal{B}(W, 1/\varepsilon_0)$ and $\|H\|_2 \leq \delta$.

Given this, we prove in Lemma 3.3, that the following iterative procedure can be used to obtain the required subspaces U and W: Initialize $U = W = \{0\}$, and as long as there exists $f \in V \setminus U$ for which $|\widehat{f}(\gamma)| > \varepsilon_0$ for some γ, apply the following update step. First, we add f to U. Next, let $l = \langle x, \gamma \rangle$ be a linear function, and define $f' = f - l$ so that $\mathrm{bias}(f') \geq \varepsilon_0$. Now using the Lemma 3.2, $\mathrm{e}(f')$ can be expressed as $\mathrm{e}(f') = G' + H'$ where $W' \subset \mathrm{Poly}_{n,d-1}$ is a subspace of dimension at most $\binom{\ell}{\leq d}$, $G' \in \mathcal{B}(W', 1/\varepsilon_0)$, and $\|H'\|_2 \leq \delta$. We add to W both W' and l.

1.5 Organization

We introduce some preliminaries in Section 2. We use Section 3 to present the proof of Theorem 3.1, which is our main structure result for polynomials. We prove correlation bounds of polynomials with resilient function in Section 4, and prove the XOR lemmas in Section 5. In Section 6 we discuss applications to the Fourier spectrum of polynomials. We analyze the order-2 Gowers norm of the Majority function in Appendix A.

2 Preliminaries

2.1 Notation

We identify $\mathbb{F}_2^n$ with $\{0,1\}^n$. We use lower case letters f, g, h to denotes functions $\mathbb{F}_2^n \to \mathbb{F}_2$, and upper case letters F, G, H to denotes functions $\mathbb{F}_2^n \to \mathbb{R}$. We use U_n to denote the uniform distribution on n bits. The XOR of two bits b_1, b_2 is denoted by $b_1 \oplus b_2$. For any $x \in \{0,1\}^n$, $\mathrm{Maj}(x)$ denotes the usual MAJORITY function. For a function $F : \{0,1\}^n \to \mathbb{R}$, we use the shorthand $\mathbb{E}[F]$ or $\mathbb{E}_x[F(x)]$ to denote the average of F, namely the quantity $\frac{1}{2^n} \sum_{x \in \{0,1\}^n} F(x)$. For a function $f : \{0,1\}^n \to \{0,1\}$, any set $S \subset [n]$, $x_1 \in \{0,1\}^{[n] \setminus S}$, and $x_2 \in \{0,1\}^S$, we use $f(x_1, x_2)$ to denote applying f to the concatenated inputs, with x_1 appearing in the coordinates corresponding to $[n] \setminus S$ and x_2 appearing in the coordinates corresponding to S. We use similar notations for functions $F : \{0,1\}^n \to \mathbb{R}$.

2.2 Fourier analysis of Boolean functions

We briefly review the basics of Boolean Fourier analysis and refer the reader to the excellent book by O'Donnell [O'D14] for more details. Given $F : \mathbb{F}_2^n \to \mathbb{R}$ and $\gamma \in \mathbb{F}_2^n$, let $\widehat{F}(\gamma) = \mathbb{E}_x[F(x)(-1)^{\langle \gamma, x \rangle}]$ denote its Fourier coefficients. Define $\|\widehat{F}\|_\infty = \max_\gamma |\widehat{F}(\gamma)|$. Further, for any subset $S \subset [n]$, define $\widehat{F}(S) = \widehat{F}(\gamma)$, where $\gamma \in \mathbb{F}_2^n$ is the indicator vector for S. For a

Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, we shorthand $\widehat{f}(\gamma) = \widehat{\mathrm{e}(f)}(\gamma)$. Any function $F : \mathbb{F}_2^n \to \mathbb{R}$ has a unique Fourier representation given by

$$F(x) = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{F}(\gamma) \cdot (-1)^{\langle \gamma, x \rangle}.$$

For any function $F : \mathbb{F}_2^n \to \mathbb{R}$, Parseval's identity states that

$$\mathbb{E}_x[F(x)^2] = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{F}(\gamma)^2.$$

2.3 Definitions of bias, correlation, covariance, and local correlation

Define $\mathrm{e} : \mathbb{F}_2 \to \mathbb{R}$ by $\mathrm{e}(x) = (-1)^x$. For $f : \mathbb{F}_2^n \to \mathbb{F}_2$ define $\mathrm{e}(f) : \mathbb{F}_2^n \to \mathbb{R}$ by $\mathrm{e}(f)(x) = \mathrm{e}(f(x))$. The bias of $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is $\mathrm{bias}(f) = |\mathbb{E}_x \mathrm{e}(f(x))|$. Define the correlation of two Boolean functions $f, g : \{0,1\}^n \to \{0,1\}$ as:

$$\mathrm{corr}(f, g) := |\mathbb{E}[\mathrm{e}(f(x))\mathrm{e}(g(x)))| = \big|\Pr[f(x) = g(x)] - \Pr[f(x) \neq g(x)]\big|.$$

Define the covariance of two functions F, G with codomain $\mathbb{R}$ as:

$$\mathrm{cov}(F, G) := \mathbb{E}[F(x)G(x)] - \mathbb{E}[F(x)]\,\mathbb{E}[G(x)].$$

For Boolean f, g, define $\mathrm{cov}(f, g) = \mathrm{cov}(e(f), e(g))$. Thus, we have the following simple fact:

Fact 2.1. *For Boolean functions f, g, we have:*

$$\mathrm{corr}(f, g) \leq |\mathrm{cov}(f, g)| + \mathrm{bias}(f).$$

Definition 2.2 (Local correlation). *Given a function $F : \mathbb{F}_2^n \to \mathbb{R}$ and $S \subset [n]$ define the S-local correlation of F as*

$$\Delta_S(F) = \mathbb{E}[F(x)F(y)|x_{S^c} = y_{S^c}] - \mathbb{E}[F]^2.$$

For $f : \mathbb{F}_2^n \to \mathbb{F}_2$ we abbreviate $\Delta_S(f) = \Delta_S(\mathrm{e}(f))$. The following is straightforward to show.

Fact 2.3. *For any $F : \mathbb{F}_2^n \to \mathbb{R}$, and any set $S \subseteq [n]$,*

$$\Delta_S(F) = \sum_{\gamma \in \mathbb{F}_2^n, \gamma \neq 0, \gamma_S = 0} \widehat{F}(\gamma)^2.$$

2.4 Polynomials and their derivatives

Let $\mathrm{Poly}_{n,d}$ denote the linear space of polynomials over $\mathbb{F}_2^n$ of total degree at most d. The following quantity measures how well degree d polynomials approximate a Boolean function f:

$$\mathrm{corr}(f,d) := \max_{p \in \mathrm{Poly}_{n,d}} \mathrm{corr}(f,p).$$

Define $\mathrm{cov}(f,d)$ analogously, and note that it is nonnegative. Observe that $\mathrm{bias}(f) = \mathrm{corr}(f,0)$.

Definition 2.4. *Given $f \in \mathrm{Poly}_{n,d}$ define $f_a(x) = f(x+a) + f(x)$ to be its derivative in direction a. Note that $f_a \in \mathrm{Poly}_{n,d-1}$.*

For a sequence $S = (a_1, \ldots, a_k)$ of directions, define $f_S(x)$ to be the iterated derivative. It satisfies

$$f_S(x) = \sum_{\phi \neq T \subseteq S} f_{\Sigma T}(x),$$

where $\Sigma T = \sum_{t \in T} t$. In addition, if $f \in \mathrm{Poly}_{n,d}$ and $|S| > d$ then $f_S \equiv 0$. The following claim bounds the dimension of the space of derivative polynomials. We use the shorthand $\binom{\ell}{\le d} = \sum_{i=0}^{d} \binom{\ell}{i}$.

Claim 2.5. *Let $f \in \mathrm{Poly}_{n,d}$, and let A be an arbitrary subspace in $\mathbb{F}_2^n$ of dimension $\ell \ge 1$. Let $W = \mathrm{span}\{f_a : a \in A\}$. Then $\dim(W) \le \binom{\ell}{\le d}$.*

Proof. Fix a basis $B = \{a_1, \ldots, a_\ell\}$ of the subspace A. Let

$$R = \{f_{\Sigma S} : S \subset B, |S| \le d\}$$

and let $V = \mathrm{span}(R)$. We prove the following statement: for any $S \subseteq B$, it holds that $f_{\Sigma S} \in V$. This proves the claim as $\dim(V) \le |R| \le \binom{\ell}{\le d}$, and any element $a \in A$ can be expressed as $a = \Sigma S$ for some $S \subseteq B$. The proof is by induction on $|S|$. Clearly, this holds if $|S| \le d$, so assume $|S| > d$. As f is a degree d polynomial, it holds that $f_S \equiv 0$. This implies that $f_{\Sigma S} = \sum_{T \subsetneq S} f_{\Sigma T}$. We know by induction that $f_{\Sigma T} \in V$ for all $T \subsetneq S$ since $|T| < |S|$. Thus also $f_{\Sigma S} \in V$. □

3 Subspace of polynomials re-randomization

The main result that we prove in this section is that for any low-dimensional subspace V of polynomials, there exists a small set S such that the S-correlation of any $f \in V$ is small. We now state our result more formally.

Theorem 3.1. *Let $V \subset \mathrm{Poly}_{n,d}$ be a subspace of dimension $k \ge 1$ and let $\varepsilon \in (0, 1/2)$. Then there exists a set $S \subset [n]$ of size $|S| \le C(d,k,\varepsilon) = (k \log(1/\varepsilon))^{(cd)^d}$, for some absolute constant $c > 0$, such that*

$$\Delta_S(f) \le \varepsilon \qquad \forall f \in V.$$

To prove Theorem 3.1 we start with a useful lemma regarding the structure of one biased low-degree polynomial. Let W be a linear space of functions $g : \mathbb{F}_2^n \to \mathbb{F}_2$. Define the ball of radius r around W, denoted $\mathcal{B}(W, r)$, as the set of all functions $G : \mathbb{F}_2^n \to \mathbb{R}$ of the form

$$\mathcal{B}(W, r) := \left\{ G = \sum_{g \in W} c_g \mathrm{e}(g) : \sum |c_g| \le r \right\}.$$

Lemma 3.2. *For $\varepsilon, \delta > 0$, there exists $\ell = O(\log(1/\varepsilon\delta))$ such that the following holds. Let $f \in \mathrm{Poly}_{n,d}$ be a polynomial with $|\mathrm{bias}(f)| \ge \varepsilon$. Then there exists a subspace $W \subset \mathrm{Poly}_{n,d-1}$ of dimension $\binom{\ell}{\le d}$, such that we can express*

$$e(f) = G + H$$

where $G, H : \mathbb{F}_2^n \to \mathbb{R}$ satisfy $G \in \mathcal{B}(W, 1/\varepsilon)$ and $\|H\|_2 \le \delta$.

Proof. For $d = 1$, if $\mathrm{bias}(f) \ne 0$ then f is a constant function, in which case we can take $W = \{0\}$, $G = f$ and $H = 0$. So assume $d > 1$ from now on.

Let $\eta = \mathbb{E}[\mathrm{e}(f)]$ where $|\eta| = \mathrm{bias}(f) \ge \varepsilon$. We have

$$\eta \cdot \mathrm{e}(f(x)) = \mathbb{E}_{y \in \mathbb{F}_2^n} [\mathrm{e}(f_y(x))] \qquad \forall x \in \mathbb{F}_2^n.$$

Let $\ell = O(\log(1/\varepsilon\delta))$ to be determined later. If $n \le \ell$ then we can take $W = \{f_y : y \in \mathbb{F}_2^n\}$, $G = \eta^{-1} \mathbb{E}_{y \in \mathbb{F}_2^n} [\mathrm{e}(f_y(x))]$ and $H = 0$, where the dimension of W is bounded by $\binom{n}{\le d} \le \binom{\ell}{\le d}$ by Claim 2.5. Thus, we assume from now on that $n > \ell$.

Let $A \subset \mathbb{F}_2^n$ be a uniform linear subspace of dimension ℓ. Then as

$$\mathbb{E}_{y \in \mathbb{F}_2^n \setminus \{0\}} [\mathrm{e}(f_y(x))] = \mathbb{E}_A \mathbb{E}_{a \in A \setminus \{0\}} [\mathrm{e}(f_a(x))]$$

and as $f_0 \equiv 0$, we obtain

$$\begin{aligned} \eta \cdot \mathrm{e}(f)(x) &= 2^{-n} + (1 - 2^{-n}) \mathbb{E}_{y \in \mathbb{F}_2^n \setminus \{0\}} [\mathrm{e}(f_y(x))] \\ &= 2^{-n} + (1 - 2^{-n}) \mathbb{E}_A \mathbb{E}_{a \in A \setminus \{0\}} [\mathrm{e}(f_a(x))] . \end{aligned}$$

For a fixed subspace A define

$$\begin{aligned} &W_A = \{f_a : a \in A\}, \\ &G_A(x) = 2^{-n} + (1 - 2^{-n}) \mathbb{E}_{a \in A \setminus \{0\}} [\mathrm{e}(f_a(x))] , \\ &H_A(x) = \eta \cdot \mathrm{e}(f(x)) - G_A(x). \end{aligned}$$

Observe that $G_A \in \mathcal{B}(W_A, 1)$. Fix $x \in \mathbb{F}_2^n$. For a random choice of A we get $\mathbb{E}_A[H_A(x)] = 0$.

We now want to bound $\mathbb{E}_A[H_A(x)^2]$. To do this, we specify that we choose A by choosing a random full rank linear map $L : \mathbb{F}_2^\ell \to \mathbb{F}_2^n$ and setting A to be the range of L. For each $v \in \mathbb{F}_2^\ell \setminus \{0\}$, define the random variable

$$X_v = \eta \cdot \mathrm{e}(f(x)) - 2^{-n} - (1 - 2^{-n}) e(f_{L(v)}(x)) \in [-2, 2].$$

Then

$$H_A(x) = \frac{1}{|A|-1} \sum_{v \in \mathbb{F}_2^\ell \setminus \{0\}} X_v.$$

Note that for any $v \in \mathbb{F}_2^\ell \setminus \{0\}$, we have $\mathbb{E}[X_v] = 0$.

Further, we claim that for any distinct $v, w \in \mathbb{F}_2^\ell \setminus \{0\}$, $\mathbb{E}[X_v X_w] \leq 0$. This can be seen in the following way. Condition on the random variable $L(v)$. Thus, $L(w)$ is uniform over $\mathbb{F}_2^n \setminus \{0, L(v)\}$. If we chose $L(w)$ uniform on $\mathbb{F}_2^n \setminus \{0\}$, then clearly we would have $\mathbb{E}[X_v X_w] = 0$ (since X_v is fixed, and $\mathbb{E}[X_w] = 0$). But, we are removing one point (namely, $L(v)$) from the support of $L(w)$, where $X_w = X_v$ and hence $X_w X_v = X_v^2 \geq 0$. Thus it follows that

$$\left(1 - \frac{1}{2^n - 1}\right) \cdot \mathbb{E}[X_v X_w] + \frac{1}{2^n - 1} \cdot X_v^2 = 0,$$

which implies $\mathbb{E}[X_v X_w] \leq 0$. Since this is true for any conditioning of $L(v)$, the claim follows.

Thus, we have

$$\mathbb{E}_A[H_A(x)^2] \leq \frac{1}{(|A|-1)^2} \sum_{v \in \mathbb{F}_2^\ell \setminus \{0\}} \mathbb{E}[X_v^2] \leq \frac{4}{|A|-1}.$$

By averaging, there exists a subspace A for which,

$$\mathbb{E}_x[H_A(x)^2] \leq \frac{4}{|A|-1}.$$

The claim follows by taking $W = W_A$, $G = \eta^{-1} G_A$ and $H = \eta^{-1} H_A$. The dimension bound on W follows from Claim 2.5. Observe that $\|H\|_2 \leq 2/(\varepsilon\sqrt{|A|-1})$. To obtain the bound $\|H\|_2 \leq \delta$ we can set $\ell = \dim(A) \geq 1 + 2\log(2/\varepsilon\delta)$. □

Now we continue to study the structure of a subspace of low-degree polynomials, by iteratively applying Lemma 3.2.

Lemma 3.3. *For $\varepsilon \in (0, 1/2), \delta > 0$, there exists $\ell = O(\log(1/\varepsilon\delta))$ such that the following holds. Let $V \subset \mathrm{Poly}_{n,d}$ be a subspace of polynomials of dimension k. Then there exists a subspace $U \subset V$ and a subspace $W \subset \mathrm{Poly}_{n,d-1}$ of dimension $k\binom{\ell}{\leq d}$ such that the following holds:*

(i) Each $f \in V \setminus U$ satisfies $\|\widehat{e(f)}\|_\infty \leq \varepsilon$.

(ii) Each $f \in U$ can be expressed as $e(f) = G + H$, where $G \in \mathcal{B}(W, 1/\varepsilon^k)$ and $\|H\|_2 \leq \delta/\varepsilon^k$.

Proof. We describe a procedure for obtaining U and W. Initialize $U = W = \{0\}$. As long as there exists $f \in V \setminus U$ for which $|\widehat{f}(\gamma)| > \varepsilon$ for some γ, apply the following update step:

1. $U \leftarrow \mathrm{span}(U \cup \{f\})$.
2. Let $l = \langle x, \gamma \rangle \in \mathrm{Poly}_{n,1}$ so that $f' = f - l$ satisfies $\mathrm{bias}(f') \geq \varepsilon$.

3. Apply Lemma 3.2 to f'. Let $W' \subset \mathrm{Poly}_{n,d-1}$ be the resulting subspace so that we can express $\mathrm{e}(f') = G' + H'$ where $G' \in \mathcal{B}(W', 1/\varepsilon)$, $\|H'\|_2 \leq \delta$.

4. $W \leftarrow \mathrm{span}(W \cup W' \cup \{l\})$.

5. We have $\mathrm{e}(f) = G + H$ where $G = G' \cdot \mathrm{e}(l), H = H' \cdot \mathrm{e}(l)$ satisfy $G \in \mathcal{B}(W, 1/\varepsilon)$ and $\|H\|_2 = \|H'\|_2 \leq \delta$.

Assume that we applied the procedure $m \leq k$ times, so that at the end $\dim(U) = m$. Let $f_1, \ldots, f_m$ be the polynomials for which we applied the update step, and hence they form a basis for U. For each f_i we can express $\mathrm{e}(f_i) = G_i + H_i$ where $G_i \in \mathcal{B}(W, 1/\varepsilon)$ and $\|H_i\|_2 \leq \delta$. Any $f \in U$ can be expressed as $f = \sum_{i \in S} f_i$ for some $S \subseteq [m]$. Assume for simplicity of notation that $S = \{1, \ldots, s\}$. Then

$$\mathrm{e}(f) = \prod_{i=1}^{s} \mathrm{e}(f_i) = \prod_{i=1}^{s} (G_i + H_i) = G + H$$

where $G = \prod_{i=1}^{s} G_i$ and H is the remaining terms. We get $G \in \mathcal{B}(W, 1/\varepsilon^k)$. To bound $\|H\|_2$ we write

$$H = \sum_{i=1}^{s} \left(\prod_{j=1}^{i-1} G_i \right) H_i \left(\prod_{j=i+1}^{s} \mathrm{e}(f_i) \right).$$

As $\|G_i\|_\infty \leq 1/\varepsilon$ and $\|\mathrm{e}(f_i)\|_\infty = 1$ and we assume $\varepsilon < 1/2$ we can bound

$$\|H\|_2 \leq \sum_{i=1}^{s} (1/\varepsilon)^i \|H_i\|_2 \leq \delta/\varepsilon^k.$$

□

We will need the following lemmas from [LMS13].

Lemma 3.4 (Claim 31 in [LMS13]). *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$, $S \subseteq [n]$, and $A \subset \mathbb{F}_2^n$ be a linear subspace. Then*

$$\Delta_S(f)^2 \leq \mathbb{E}_{a \in A}\left[\Delta_S(f_a)\right] + \mathbb{E}_{a \in A}\left[\mathrm{bias}(f_a)^2\right].$$

Lemma 3.5 (Claim 32 in [LMS13]). *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$, $S \subseteq [n]$, and $A \subset \mathbb{F}_2^n$ be a random linear subspace of dimension m. Then*

$$\mathbb{E}_A \, \mathbb{E}_{a \in A}\left[\mathrm{bias}(f_a)^2\right] \leq 2^{-m} + \|\widehat{e(f)}\|_\infty^2.$$

We now prove Theorem 3.1.

Proof of Theorem 3.1. The proof is by induction on d. If $d = 1$ then it suffices to take $|S| = k$ for any $\varepsilon > 0$, so assume $d \geq 2$ from now on. Let $V \subset \mathrm{Poly}_{n,d}$ be a space of polynomials of dimension k. Apply Lemma 3.3 with parameters ε_0, δ to be determined later. Let $U \subset V$ and $W_1 \subset \mathrm{Poly}_{n,d-1}$ be as obtained by the lemma, where $k_1 = \dim(W_1) = k\binom{\ell}{\leq d}$

for $\ell = O(\log(1/\varepsilon_0\delta))$. Apply Theorem 3.1 inductively to W_1 and error parameter ε_1 to obtain a set $S_1 \subset [n]$ such that

$$\Delta_{S_1}(g) \leq \varepsilon_1 \qquad \forall g \in W_1.$$

We first analyze $f \in U$. Let $F = \mathrm{e}(f)$. We can express $F = G + H$ with $G \in \mathcal{B}(W_1, r)$ for $r = 1/\varepsilon_0^k$ and $\|H\|_2 \leq \delta/\varepsilon_0^k$. Let $G = \sum_{g \in W_1} c_g \mathrm{e}(g)$ where $\sum |c_g| \leq r$. Let $\Gamma = \{\gamma \in \mathbb{F}_2^n : \gamma \neq 0, \gamma_{S_1} = 0\}$. We have

$$\Delta_{S_1}(f) = \sum_{\gamma \in \Gamma} \widehat{F}(\gamma)^2 \leq \sum_{\gamma \in \Gamma} \left(\widehat{G}(\gamma) + \widehat{H}(\gamma)\right)^2 \leq 2\left(\sum_{\gamma \in \Gamma} \widehat{G}(\gamma)^2 + \sum_{\gamma \in \Gamma} \widehat{H}(\gamma)^2\right).$$

We analyze each term separately. We can bound $\widehat{G}(\gamma)^2$ by the Cauchy-Schwartz inequality

$$\widehat{G}(\gamma)^2 = \left(\sum_{g \in W_1} c_g \widehat{\mathrm{e}(g)}(\gamma)\right)^2 \leq \left(\sum_{g \in W_1} |c_g|\right)\left(\sum_{g \in W_1} |c_g| \widehat{\mathrm{e}(g)}(\gamma)^2\right) \leq r \sum_{g \in W_1} |c_g| \widehat{\mathrm{e}(g)}(\gamma)^2.$$

Summing over all $\gamma \in \Gamma$ gives

$$\sum_{\gamma \in \Gamma} \widehat{G}(\gamma)^2 \leq r \sum_{g \in W_1} |c_g| \sum_{\gamma \in \Gamma} \widehat{\mathrm{e}(g)}(\gamma)^2 \leq r \sum_{g \in W_1} |c_g| \Delta_{S_1}(g) \leq r^2 \varepsilon_1.$$

Moving to bound the term for H, we have

$$\sum_{\gamma \in \Gamma} \widehat{H}(\gamma)^2 \leq \sum_{\gamma \in \mathbb{F}_2^n} \widehat{H}(\gamma)^2 = \|H\|_2^2 \leq \delta^2/\varepsilon_0^{2k}.$$

Combining these, and plugging in $r = 1/\varepsilon_0^k$, we can bound

$$\Delta_{S_1}(f) \leq 2(\varepsilon_1 + \delta^2)/\varepsilon_0^{2k} \qquad \forall f \in U.$$

Next, we handle $f \in V \setminus U$. Let $A \subset \mathbb{F}_2^n$ be a random subspace of dimension m to be determined later. As $\|\widehat{\mathrm{e}(f)}\|_\infty \leq \varepsilon_0$ for all $f \in V \setminus U$, Lemma 3.5 gives

$$\mathbb{E}_A \sum_{f \in V \setminus U} \mathbb{E}_{a \in A}\left[\mathrm{bias}(f_a)^2\right] \leq |V \setminus U|(2^{-m} + \varepsilon_0^2) \leq 2^k(2^{-m} + \varepsilon_0^2).$$

By averaging there exists a choice of A such that

$$\mathbb{E}_{a \in A}\left[\mathrm{bias}(f_a)^2\right] \leq 2^k(2^{-m} + \varepsilon_0^2) \qquad \forall f \in V \setminus U.$$

Let $W_2 = \{f_a : f \in V, a \in A\}$ where $k_2 = \dim(W_2) \leq k\binom{m}{\leq d}$ by applying Claim 2.5 to a basis of V. Apply inductively Theorem 3.1 to W_2 with error parameter ε_2, to obtain a set $S_2 \subset [n]$ such that

$$\Delta_{S_2}(g) \leq \varepsilon_2 \qquad \forall g \in W_2.$$

Next, applying Lemma 3.4 for the subspace A and any $f \in V \setminus U$ gives

$$\Delta_{S_2}(f)^2 \leq \mathbb{E}_{a \in A} \Delta_{S_2}(f_a) + \mathbb{E}_{a \in A}\left[\mathrm{bias}(f_a)^2\right] \leq \varepsilon_2 + 2^k(2^{-m} + \varepsilon_0^2).$$

We take $S = S_1 \cup S_2$. Thus for any $f \in V$ we have that

$$\Delta_S(f) \leq \max\left(2(\varepsilon_1 + \delta^2)/\varepsilon_0^{2k}, \sqrt{\varepsilon_2 + 2^k(2^{-m} + \varepsilon_0^2)}\right).$$

We next set the parameters to obtain error ε. To simplify the calculations, we will assume without loss of generality that $\varepsilon = 2^{-k}$, by either decreasing ε or increasing k. Denote

$$D(d, k) := C(d, k, \varepsilon = 2^{-k}).$$

We set

$$\varepsilon_0 = \varepsilon_2 = 2^{-O(k)}, \varepsilon_1 = \delta = 2^{-O(k^2)}, m = O(k).$$

In addition we have $k_1 = k\binom{\ell}{\leq d}$ and $k_2 = k\binom{m}{\leq d}$ where $\ell = O(\log(1/\varepsilon_0\delta)) = O(k^3)$. Thus

$$|S| \leq |S_1| + |S_2| \leq C(d-1, k_1, \varepsilon_1) + C(d-1, k_2, \varepsilon_2) \leq 2D(d-1, k^{\lambda d})$$

for some absolute constant $\lambda > 0$. We obtain the recursion

$$D(d, k) \leq 2D(d-1, k^{\lambda d})$$

which solves to

$$D(d, k) \leq 2^d k^{(\lambda d)^d}.$$

Thus for any $k \geq 1$ and $\varepsilon \in (0, 1/2)$, for a large enough absolute constant $c > 0$ we have

$$C(d, k, \varepsilon) \leq (k \log(1/\varepsilon))^{(cd)^d}.$$

□

4 Correlation bounds for resilient functions

Our first main result is the following theorem. The function $C(d, k, \varepsilon)$ is as given in Theorem 3.1. For any function h, the quantity $\mathrm{cov}(h, d)$ is defined in Section 2.4. Combining this theorem with Fact 2.1 gives Theorem 1.6.

Theorem 4.1. *Let $h : \{0,1\}^n \to \{0,1\}$ be a strong r-resilient function, and $d \geq 1$. There is $D = O(d)^{O(d)}$ such that*

$$|\mathrm{cov}(h, d)| \leq \frac{C(d, 1, r^{-4}) + 2}{r} = \frac{\log(r)^D}{r}.$$

Proof. Fix any $f \in \mathrm{Poly}_{n,d}$. Let $F = \mathrm{e}(f)$, $G = F - \mathbb{E}[F]$ and $H = \mathrm{e}(h)$. Observe that

$$\mathrm{cov}(f,h) = \mathbb{E}[F(x)H(x)] - \mathbb{E}[F]\,\mathbb{E}[H] = \mathbb{E}[G(x)H(x)].$$

We proceed to bound $|\,\mathbb{E}[GH]|$.

Let $\varepsilon > 0$ be a parameter that we will fix later. Using Theorem 3.1, there exists a set $S \subset [n]$, $|S| = C(d,1,\varepsilon)$ such that $\Delta_S(F) \leq \varepsilon$. Observe that

$$\mathbb{E}_{x_1 \sim U_{[n]\setminus S}}[(\mathbb{E}_{x_2 \sim U_S}[G(x)])^2] = \Delta_S(F) \leq \varepsilon.$$

Thus, by an application of Markov's inequality, it follows that

$$\Pr_{x_1 \sim U_{[n]\setminus S}}[(\mathbb{E}_{x_2 \sim U_S}[G(x)])^2 > \varepsilon^{1/2}] \leq \varepsilon^{1/2}.$$

For $x_1 \sim U_{[n]\setminus S}$, let $\mathcal{E}_1$ denote the event that $|\,\mathbb{E}_{x_2 \sim U_S}[G(x)]| \leq \varepsilon^{1/4}$, and $\mathcal{E}_2$ denote the event that $H(x_1, \cdot)$ is a constant function. Observe that if x_1 is such that both $\mathcal{E}_1$ and $\mathcal{E}_2$ hold, then

$$|\,\mathbb{E}_{x_2 \sim U_S}[G(x)H(x)]| = |\,\mathbb{E}_{x_2 \sim U_S}[G(x)]| \leq \varepsilon^{1/4}.$$

We thus have

$$|\,\mathbb{E}[G(x)H(x)]| \leq \Pr[\neg\mathcal{E}_1] + \Pr[\neg\mathcal{E}_2] + \varepsilon^{1/4} \leq \varepsilon^{1/2} + \frac{|S|}{r} + \varepsilon^{1/4}.$$

Setting $\varepsilon = 1/r^4$ gives

$$|\mathrm{cov}(h,f)| \leq \frac{|S|+2}{r}.$$

□

The above theorem can be applied to Maj to obtain an alternate proof of the Razborov-Smolensky correlation bound [Raz87, Smo87] for $d = O(1)$, up to logarithmic factors.

Corollary 4.2. *$\mathrm{corr}(\mathrm{Maj}, \mathrm{Poly}_{n,O(1)}) \leq \mathrm{poly}(\log n)/\sqrt{n}$.*

Proof. Follows directly from Theorem 4.1 using Fact 1.5. □

In the next section we extend these techniques to prove correlation bounds with XOR of Maj, or more generally XOR of resilient functions. As we discussed in the introduction, these results do not follow from the Razborov-Smolensky techniques.

5 Correlation bounds for XORs of resilient functions

We extend the techniques in the previous section to prove correlation bounds for XORs of resilient functions. To simplify the bounds and proofs, we assume that the resilient functions in question are unbiased. With a bit more work, we can obtain similar bounds for the covariance even if the functions are biased.

Recall that the function $C(d,k,\varepsilon)$ is as defined in Theorem 3.1. Given functions $h_1 : \{0,1\}^n \to \{0,1\}$ and $h_2 : \{0,1\}^m \to \{0,1\}$ we let $h_1 \oplus h_2 : \{0,1\}^{n+m} \to \{0,1\}$ be given by $(h_1 \oplus h_2)(x,y) = h_1(x) \oplus h_2(y)$.

Theorem 5.1. *Let $h_1 : \{0,1\}^n \to \{0,1\}, h_2 : \{0,1\}^m \to \{0,1\}$. Assume that h_1, h_2 are unbiased, and that h_1 is a strong r-resilient function. For $\varepsilon > 0$, there is $m = O(\log 1/\varepsilon)^d$ such that*

$$\operatorname{corr}(h_1 \oplus h_2, d) \le \operatorname{corr}(h_2, d) \cdot \frac{2C(d, m, \varepsilon)}{r} + 2\varepsilon^{1/4}.$$

We record some immediate consequences. For $h : \{0,1\}^n \to \{0,1\}$ we denote by $h^{\oplus k} : \{0,1\}^{nk} \to \{0,1\}$ the function obtained by taking the direct sum of h iteratively k times, namely $h^{\oplus 1} = h$ and $h^{\oplus k} = h^{\oplus (k-1)} \oplus h$.

Theorem 5.2. *Let $h : \{0,1\}^n \to \{0,1\}$ be an unbiased strong r-resilient function. For $d \ge 1$, there is $D = O(d)^{O(d)}$ such that*

$$\operatorname{corr}(h^{\oplus k}, d) \le \left(\frac{(k \log r)^D}{r} \right)^k.$$

Proof. Let $\alpha_k = \operatorname{corr}(h^{\oplus k}, d)$. Assume we already have a bound on α_{k-1}, and we next derive a bound on α_k. Let ε to be determined later and set $m = O(\log 1/\varepsilon)^d$. Theorem 5.1 gives that

$$\alpha_k \le \alpha_{k-1} \cdot \frac{2C(d, m, \varepsilon)}{r} + 2\varepsilon^{-1/4}.$$

In order to bound $C(d, m, \varepsilon)$ apply Theorem 3.1. Let $D = d^{cd}$ for a large enough constant $c > 0$ so that

$$C(d, m, \varepsilon) \le (\log 1/\varepsilon)^D.$$

Set $\varepsilon = r^{-4k}$. Then we obtain that

$$\alpha_k \le \alpha_{k-1} \cdot \frac{(4k(\log r))^D}{r} + \frac{2}{r^k}.$$

This gives the bound

$$\alpha_k \le \left(\frac{O(k \cdot \log r)^D}{r} \right)^k.$$

□

The above result gives us a way to obtain new correlation bounds for XORs of Maj.

Corollary 5.3. *Let* Maj *denote the Majority function on n bits. For $d \ge 1$, there is $D = O(d)^{O(d)}$ such that*

$$\operatorname{corr}(\mathrm{Maj}^{\oplus k}, d) \le \left(\frac{(k \log n)^D}{\sqrt{n}} \right)^k.$$

We use the rest of the section to prove Theorem 5.1. Fix any $f \in \mathrm{Poly}_{n,d}$. First, we define some useful functions: $H_1(x) = \mathrm{e}(h_1(x))$, $H_2(y) = \mathrm{e}(h_2(y))$, $F(x,y) = \mathrm{e}(f(x,y))$, $G(x,y) = F(x,y)H_2(y)$, $K(x) = \mathbb{E}_y[G(x,y)]$. Our goal is to bound

$$\operatorname{corr}(h_1(x) \oplus h_2(y), f(x,y)) = |\,\mathbb{E}[F(x,y)H_1(x)H_2(y)]| = |\,\mathbb{E}[K(x)H_1(x)]|.$$

The following lemma is a key ingredient in our proof.

Claim 5.4. *For $\varepsilon > 0, d \geq 1$, there is $k = O(\log 1/\varepsilon)^d$, such that there exists a set $S \subset [n]$, $|S| \leq C(d,k,\varepsilon)$ for which $\Delta_S(K) \leq \varepsilon$.*

Proof. For any subspace A in $\mathbb{F}_2^n$, define W_A to be subspace spanned by the polynomials $\{f(x,a) : a \in A\}$ and the constant function 1. We show that there exists a subspace A in $\mathbb{F}_2^n$ of dimension $\ell = O(\log(1/\varepsilon))$ such that $K(x) = V(x) + Z(x)$, where $V \in \mathcal{B}(W_A, 1)$ and $\|Z\|_2 \leq \varepsilon/2$. This will directly yield the claim using Theorem 3.1, noting the fact that by an application of Claim 2.5, the dimension of W_A is at most $\binom{\ell}{\leq d} + 1 \leq O(\ell^d)$.

We now prove that K can be written as $V + Z$. The argument to show this is very similar to the one used in the proof of Lemma 3.2. For a random subspace A of dimension ℓ, we have

$$\begin{aligned} K(x) &= 2^{-n} + (1-2^{-n})\,\mathbb{E}_{y \in \mathbb{F}_2^n \setminus \{0\}}[\mathrm{e}(f(x,y)) \cdot \mathrm{e}(h_2(y))] \\ &= 2^{-n} + \mathbb{E}_A\,\mathbb{E}_{a \in A\setminus\{0\}}[\mathrm{e}(f(x,a) + h_2(a))]. \end{aligned}$$

For any subspace A in $\mathbb{F}_2^n$, define

$$\begin{aligned} &V_A(x) = 2^{-n} + \mathbb{E}_{a \in A\setminus\{0\}}[\mathrm{e}(f(x,a) + h_2(a))], \\ &Z_A(x) = K(x) - V_A(x). \end{aligned}$$

Note that $V_A \in \mathcal{B}(W_A, 1)$. For a fixed $x \in \mathbb{F}_2^n$, and a random subspace A of dimension ℓ, observe that $\mathbb{E}_A[Z_A(x)] = 0$. Proceeding similarly to Lemma 3.2, we can write $Z_A(x)$ as the sum of zero-mean random variables with non-positive covariance, and conclude that

$$\mathbb{E}_A[Z_A(x)^2] \leq \frac{4}{|A|-1}.$$

Since the above holds for all $x \in \mathbb{F}_2^n$, by an averaging argument there exists a subspace A such that $\mathbb{E}_x[Z_A(x)^2] \leq \frac{4}{|A|-1}$. The claim now follows by setting $V = V_A$ and $Z = Z_A$. □

We now proceed to prove the required correlation bound.

Proof of Theorem 5.1. Let S be the set from Claim 5.4 such that $\Delta_S(K) \leq \varepsilon$. Let $\beta = \mathbb{E}[K] = \mathbb{E}_{x,y}[\mathrm{e}(f(x,y) + h_2(y))]$ where $\beta \leq \mathrm{corr}(h_2, d)$. As $\mathbb{E}[H_1] = 0$ we have

$$\mathrm{corr}(h_1(x) \oplus h_2(y), f(x,y)) = |\mathbb{E}_x[(K(x) - \beta)H_1(x)]|$$

Let $x = (x_1, x_2)$ with $x_1 \sim U_{[n]\setminus S}$ and $x_2 \sim U_S$. Let $\mathcal{E}_1 = \mathcal{E}_1(x_1)$ denote the event that $h_1(x_1, \cdot)$ is a constant function. By our assumption on h_1 it holds that $\Pr[\neg\mathcal{E}_1] \leq |S|/r$. Thus we can bound

$$\begin{aligned} \mathrm{corr}(h_1(x) \oplus h_2(y), f(x,y)) \leq\ &|\mathbb{E}_x[(K(x) - \beta)H_1(x)|\mathcal{E}_1]| &(2)\\ &+ |\mathbb{E}_x[(K(x) - \beta)H_1(x)|\neg\mathcal{E}_1]| \cdot \frac{|S|}{r}. &(3) \end{aligned}$$

We will first upper bound the term. Let $\mathcal{E}_2 = \mathcal{E}_2(x_1)$ denote the event that $|\mathbb{E}_{x_2}[K(x_1, x_2)] - \beta| \leq \varepsilon^{1/4}$. We claim that $\Pr[\neg\mathcal{E}_2] \leq \varepsilon^{1/2}$. To see this, let $x_2, x_2' \sim U_S$ be independent. Then

$$\begin{aligned} \Delta_S(K) &= \mathbb{E}_{x_1} \mathbb{E}_{x_2, x_2'}[(K(x_1, x_2) - \beta)(K(x_1, x_2') - \beta)] \\ &= \mathbb{E}_{x_1}[(\mathbb{E}_{x_2}[K(x_1, x_2)] - \beta)^2] \\ &= \mathbb{E}_{x_1}[(\mathbb{E}_{x_2, y}[F(x, y)H_2(y)] - \beta)^2]. \end{aligned}$$

The bound on $\Pr[\neg\mathcal{E}_2]$ follows from Markov's inequality using the fact that $\Delta_S(K) \leq \varepsilon$. Thus we can bound the first term in Equation (2) by

$$|\mathbb{E}_x[(K(x) - \beta)H_1(x)|\mathcal{E}_1] \leq \mathbb{E}_{x_1} |\mathbb{E}_{x_2}[K(x_1, x_2) - \beta)]|\mathcal{E}_2| + \Pr[\neg\mathcal{E}_2] \leq 2\varepsilon^{1/4}.$$

We now proceed to bound the second term. We have

$$|\mathbb{E}_x[(K(x) - \beta)H_1(x)|\neg\mathcal{E}_1]| \leq \beta + \mathbb{E}_{x,y}[H_1(x)H_2(y)F(x, y)|\neg\mathcal{E}_1].$$

Fix x such that $\mathcal{E}_1$ holds. Averaging the second term over y gives

$$\mathbb{E}_y[H_1(x)H_2(y)F(x, y)] \leq \text{corr}(h_2, d).$$

Thus

$$|\mathbb{E}_x[(K(x) - \beta)H_1(x)|\neg\mathcal{E}_1]| \leq 2\text{corr}(h_2, d).$$

Combining the bounds for the two terms in Equation (2) gives

$$\text{corr}(h_1(x) \oplus h_2(y), f(x, y)) \leq 2\varepsilon^{1/4} + \text{corr}(h_2, d) \cdot \frac{2C(d, k, \varepsilon)}{r}$$

for $k = O(\log 1/\varepsilon)^d$. □

6 Level 2 Fourier bounds from correlation with XOR of shifted majority

For $x \in \{0, 1\}^n$, let $|x|$ denote the Hamming weight of x. We define the class of shifted majority functions, $\text{Maj}_a : \{0, 1\}^n \to \{0, 1\}$ for $a \in \{0, 1, \ldots, n\}$ as

$$\text{Maj}_a(x) := \begin{cases} 1 & \text{if } |x| > a \\ 0 & \text{otherwise} \end{cases}$$

The main result of this section is a bound on the level two Fourier mass of functions from correlation bounds with XOR of shifted majority functions.

Lemma 6.1. *Let $\mathcal{F}$ be a family of $2n$-variate Boolean functions that is closed under relabelling variables. Suppose that for all integers a, b such that $\frac{n}{2} - 2\sqrt{n \log n} \le a, b \le \frac{n}{2} + 2\sqrt{n \log n}$, and any $f \in \mathcal{F}$, it holds that*

$$\operatorname{corr}\left(f(x,y), \mathrm{Maj}_a(x) \oplus \mathrm{Maj}_b(y)\right) \le t/n,$$

for some $t \ge 1$. Then,

$$\left| \sum_{1 \le i < j \le 2n} \widehat{f}(\{i,j\}) \right| \le O(t \log n).$$

We use the rest of the section to prove the above lemma. It is convenient to define the following functions. For $\theta \in [n/2]$, let

$$\mathrm{Thr}_\theta(x) := \begin{cases} (-1)^{\mathrm{Maj}(x)} & \text{if } |\sum x_i - n/2| \ge \theta \\ 0 & \text{otherwise} \end{cases}$$

Observe that for any $\theta \in [n/2]$,

$$2 \cdot \mathrm{Thr}_\theta(x) = e(\mathrm{Maj}_{n/2+\theta-1}(x)) + e(\mathrm{Maj}_{n/2-\theta}(x)).$$

We record a straightforward consequence of the above claim and the hypothesis of Lemma 6.1.

Claim 6.2. *For all integers $1 \le a, b \le 2\sqrt{n \log n}$, and any $f \in \mathcal{F}$, it holds that*

$$|\mathbb{E}_{x,y}[e(f(x,y))\mathrm{Thr}_a(x)\mathrm{Thr}_b(y)]| \le t/n.$$

We use a couple of useful observations that appeared in [CHLT19].

Claim 6.3 ([CHLT19])**.** *For any $x \in \{0,1\}^n$, $\sum_{1 \le i \le n} e(x_i) = 2 \sum_{1 \le a \le n/2} \mathrm{Thr}_a(x)$.*

Claim 6.4 ([CHLT19])**.** *For any Boolean function $f : \{0,1\}^n \to \{0,1\}$, there exists an equi-partition of $[2n]$ into disjoint sets S, T such that*

$$\left| \sum_{1 \le i < j \le 2n} \widehat{f}(\{i,j\}) \right| \le 2 \left| \sum_{i \in S, j \in T} \widehat{f}(\{i,j\}) \right|.$$

We also note that for large enough a, the support of Thr_a is small. This is a straightforward consequence of the Chernoff bound.

Claim 6.5. *For any $a \ge 2\sqrt{n \log n}$, we have $\mathbb{E}[|\mathrm{Thr}_a|] \le O(1/n^8)$.*

Proof of Lemma 6.1. Let $f \in \mathcal{F}$. Using Claim 6.4, it is enough to bound $\sum_{i \in S, j \in T} \widehat{f}(\{i,j\})$ for some equipartition of $[2n]$. Without loss of generality suppose that $S = [n]$ and $T =$

$[2n] \setminus S$ since we can always relabel variables without changing the Fourier spectrum. Let $F = \mathrm{e}(f)$. We have

$$\begin{aligned}
\left| \sum_{1 \le i \le n, n+1 \le j \le 2n} \widehat{f}(i,j) \right| &= \left| \sum_{1 \le i,j \le n} \mathbb{E}[F(x,y)\mathrm{e}(x_i)\mathrm{e}(y_j)] \right| \\
&= \left| \mathbb{E}\left[F(x,y) \left(\sum_{1 \le i \le n} \mathrm{e}(x_i) \right) \left(\sum_{1 \le j \le n} \mathrm{e}(y_j) \right) \right] \right| \\
&\le 4 \sum_{1 \le a \le n/2, 1 \le b \le n/2} |\mathbb{E}[F(x,y)\mathrm{Thr}_a(x)\mathrm{Thr}_b(y)]| && \text{(using Claim 6.3)} \\
&\le 4 \sum_{1 \le a,b \le 2\sqrt{n \log n}} |\mathbb{E}[F(x,y)\mathrm{Thr}_a(x)\mathrm{Thr}_b(y)]| + O(1/n^6) && \text{(using Claim 6.5)} \\
&\le \frac{4t}{n} \cdot (4n \log n) && \text{(using Claim 6.2)} \\
&= O(t \log n).
\end{aligned}$$

□

References

[AL93] Miklós Ajtai and Nathan Linial. The influence of large coalitions. *Combinatorica*, 13(2):129–145, 1993.

[BL85] Michael Ben-Or and Nathan Linial. Collective coin flipping, robust voting schemes and minima of Banzhaf values. In *26th Annual Symposium on Foundations of Computer Science (SFCS 1985)*, pages 408–416. IEEE, 1985.

[BNS92] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, 1992.

[Bou05] Jean Bourgain. Estimation of certain exponential sums arising in complexity theory. *Comptes Rendus Mathematique*, 340(9):627–631, 2005.

[CHHL18] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. In *33rd Computational Complexity Conference (CCC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[CHLT19] Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom generators from the second Fourier level and applications to AC0 with parity gates. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[CZ19] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. *Annals of Mathematics*, 189(3):653–705, 2019.

[GKW19] Alexander Golovnev, Alexander S. Kulikov, and Ryan Williams. Circuit depth reductions. *CoRR*, abs/1811.04828, 2019.

[GL89] Oded Goldreich and Leonid A Levin. A hard-core predicate for all one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 25–32. ACM, 1989.

[GRS05] Frederic Green, Amitabha Roy, and Howard Straubing. Bounds on an exponential sum arising in Boolean circuit complexity. *Comptes Rendus Mathematique*, 341(5):279–282, 2005.

[GT08] Ben Green and Terence Tao. An inverse theorem for the Gowers U3(G) norm. *Proceedings of the Edinburgh Mathematical Society*, 51(1):73–153, 2008.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *29th Annual Symposium on Foundations of Computer Science*, pages 68–80. IEEE, 1988.

[LMS13] Shachar Lovett, Partha Mukhopadhyay, and Amir Shpilka. Pseudorandom generators for CC0[p] and the Fourier spectrum of low-degree polynomials over finite fields. *Computational Complexity*, 22(4):679–725, 2013.

[Mek17] Raghu Meka. Explicit resilient functions matching Ajtai-Linial. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1132–1148. SIAM, 2017.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs Randomness. *Journal of computer and System Sciences*, 49(2):149–167, 1994.

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions.* Cambridge University Press, 2014.

[Raz87] Alexander A Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki*, 41(4):598–607, 1987.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 77–82. ACM, 1987.

[Smo93] Roman Smolensky. On representations by low-degree polynomials. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 130–138. IEEE, 1993.

[Vaz86] Umesh Vazirani. *Randomness, adversaries and computation.* PhD thesis, University of California, Berkeley, 1986.

[Vio09] Emanuele Viola. Guest column: correlation bounds for polynomials over $\{0\ 1\}$. *ACM SIGACT News*, 40(1):27–44, 2009.

[VV84] Umesh V Vazirani and Vijay V Vazirani. Efficient and secure pseudo-random number generation. In *Workshop on the Theory and Application of Cryptographic Techniques*, pages 193–202. Springer, 1984.

[VW07] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 141–154. IEEE, 2007.

[Yao82] Andrew C Yao. Theory and application of trapdoor functions. In *23rd Annual Symposium on Foundations of Computer Science (SFCS 1982)*, pages 80–91. IEEE, 1982.

A 2-order Gowers norm of Majority

In this section, we discuss the XOR lemma proved by Viola and Wigderson [VW07], and why it doesn't seem to work in our setting. Their approach uses the Gowers norm of a function as the measure of correlation of the function with polynomials. It is known that the order-d Gowers norm of a function f (see below for definitions) is an upper bound on the correlation of f with degree $d-1$ polynomials [GT08]. To derive XOR lemmas for polynomials, Viola and Wigderson use the fact that the Gowers norm of the product of functions defined on disjoint inputs is multiplicative. However, it is not clear how to use this approach to prove correlation bounds for XOR of Maj since the Gowers norm of Maj is too big. Specifically, in this section we compute the order-2 Gowers norm of Maj and show that it is $\Theta(1/n^{1/4})$, while the correlation of Maj with linear functions is $\Theta(1/n^{1/2})$. Thus, the bound on the order-2 Gowers norm of Maj yields a sub-optimal bound on the correlation of Maj with polynomials of degree 1 (i.e. linear functions). As the Gowers norms are increasing, this implies that the order-d Gowers norm of Maj is also at least $\Theta(1/n^{1/4})$. This indicates that the Gowers norm is not the right measure to bound correlation of polynomials with Maj.

Let G be any Abelian group. For any positive integer $d > 0$, the Gowers norm of order d of a function $f : G \to \mathbb{C}$ is defined as

$$\|f\|_{U_d} = |\,\mathbb{E}_{x,y_1,y_2,\ldots,y_d \in G}[D_{y_1} D_{y_2} \ldots D_{y_d}[f(x)]]|^{1/2^d},$$

where $D_y(f(x)) := f(x+y)\overline{f(x)}$, is the multiplicative derivative of f in the direction y.

Let Maj denote the Majority function on n inputs, where we assume n is odd. We shorthand $F(x) = (-1)^{\mathrm{Maj}(x)}$, and so $F : \mathbb{F}_2^n \to \{-1, 1\}$. The following is the main result of this section.

Lemma A.1. $\|F\|_{U_2} = \Theta(1/n^{1/4})$.

We first record some useful facts for proving the above lemma. For any $f : G \to \mathbb{C}$, define $\|f\|_p = [\sum_{x \in G} |f(x)|^p]^{1/p}$.

Fact A.2. *For any function $f : G \to \mathbb{C}$, we have $\|f\|_{U_2} = \|\widehat{f}\|_4$.*

Fact A.3 (Theorem 5.19 in [O'D14])**.** *For any $S \subseteq [n]$, $|S| = k$, $\widehat{F}(S) = F_k$ where*

$$F_k = \begin{cases} 0 & \textit{if } k \textit{ is even} \\ \frac{1}{2^{n-1}} \cdot \binom{n-1}{(n-1)/2} \cdot \frac{\binom{(n-1)/2}{(k-1)/2}}{\binom{n-1}{k-1}} & \textit{if } k \textit{ is odd} \end{cases}$$

Thus we have the following identity:

$$\|F\|_{U_2}^4 = \|\widehat{F}\|_4^4 = \sum_{S \subseteq [n]} |\widehat{F}(S)|^4 = \sum_{k=1}^{n} \binom{n}{k} F_k^4.$$

We next bound this expression. Let $\gamma = \frac{1}{2^{n-1}} \cdot \binom{n-1}{(n-1)/2}$ where it is known that $\gamma = \Theta(1/\sqrt{n})$. Hence $F_1 = \gamma = \Theta(1/\sqrt{n})$ and we can lower bound $\|F\|_{U_2}^4$ by using the terms for $k = 1$:

$$\|F\|_{U_2}^4 \geq \binom{n}{1} F_1^4 = n\gamma^4 = \Omega(1/n).$$

Next we upper bound $\|F\|_{U_2}^4$. Observe that

$$\frac{\binom{(n-1)/2}{(k-1)/2}^2}{\binom{n-1}{k-1}} = \frac{\binom{k-1}{(k-1)/2}\binom{n-k}{(n-k)/2}}{\binom{n-1}{(n-1)/2}} \leq 1.$$

Thus

$$\binom{n}{k} F_k^4 \leq \binom{n}{k} \frac{\gamma^4}{\binom{n-1}{k-1}^2} = \frac{n}{k} \cdot \frac{\gamma^4}{\binom{n-1}{k-1}} = \frac{\Theta(1/n)}{k\binom{n-1}{k-1}}.$$

The sum $\sum_{k \geq 1} \frac{1}{k\binom{n-1}{k-1}}$ is bounded by $O(1)$, and hence $\|F\|_{U_2}^4 = O(1/n)$.