Electronic Colloquium on Computational Complexity, Report No. 15 (2018)

Pseudorandom Generators from Polarizing Random Walks

Eshan Chattopadhyay[*]
Cornell University and IAS
eshanc@ias.edu

Pooya Hatami[†]
University of Texas at Austin
pooyahat@gmail.com

Kaave Hosseini[‡]
University of California, San Diego
skhossei@ucsd.edu

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

January 25, 2018

Abstract

We propose a new framework for constructing pseudorandom generators for n-variate Boolean functions. It is based on two new notions. First, we introduce fractional pseudorandom generators, which are pseudorandom distributions taking values in $[-1,1]^n$. Next, we use a fractional pseudorandom generator as steps of a random walk in $[-1,1]^n$ that converges to $\{-1,1\}^n$. We prove that this random walk converges fast (in time logarithmic in n) due to polarization. As an application, we construct pseudorandom generators for Boolean functions with bounded Fourier tails. We use this to obtain a pseudorandom generator for functions with sensitivity s, whose seed length is polynomial in s. Other examples include functions computed by branching programs of various sorts or by bounded depth circuits.

1 Introduction

Pseudorandom generators (PRG) are widely studied in complexity theory. There are several general frameworks used to construct PRGs. One is based on basic building blocks, such as small bias generators [NN93, AGHP92], k-wise independence, or expander graphs [HLW06]. Another approach is based on hardness vs randomness paradigm, which was introduced by Nisan and Wigderson [NW88] and has been very influential. Many of the hardness results used in the latter framework are based on random restrictions, and the analysis of how they simplify the target class of functions. The number of papers in these lines of work is on the order of hundreds, so we do not even attempt to give a comprehensive survey of them all.

The purpose of this paper is to introduce a new framework for constructing PRGs based on polarizing random walks. We develop the theory in this paper and give a number of applications; perhaps the most notable one is a PRG for functions of sensitivity s whose seed length is polynomial in s. But, as this is a new framework, there are many questions that arise, both technical and conceptual, and we view this paper as mostly preliminary, with the hope that many more applications would follow.

1.1 PRGs and fractional PRGs

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. The standard definition of a PRG for f with error $\varepsilon > 0$, is a random variable $X \in \{-1,1\}^n$ such that

$$|\mathbb{E}_X[f(X)] - \mathbb{E}_U[f(U)]| \le \varepsilon,$$

[*] Supported by NSF grant CCF-1412958 and the Simons foundation.
[†] Supported by a Simons Investigator Award (#409864, David Zuckerman).
[‡] Supported by NSF grant CCF-1614023.
[§] Supported by NSF grant CCF-1614023.

ISSN 1433-8092

where U denotes a random variable with the uniform distribution in $\{-1,1\}^n$. We relax this definition by introducing a new object called a *fractional PRG*, defined in the next paragraph.

To prepare the notation for the definition, identify f with a real multi-linear polynomial, namely its Fourier expansion. This extends f to $f : \mathbb{R}^n \to \mathbb{R}$, although, we would only be interested in inputs from $[-1,1]^n$. Observe that if $x \in [-1,1]^n$ then $f(x) = \mathbb{E}_X[f(X)]$ where $X \in \{-1,1\}^n$ is a random variable sampled as follows: for every $i \in [n]$ sample $X_i \in \{-1,1\}$ independently with $\mathbb{E}[X_i] = x_i$. In particular, f on $[-1,1]^n$ is bounded, namely $f : [-1,1]^n \to [-1,1]$. Also, $f(\overline{0}) = \mathbb{E}_U[f(U)]$. The following is a key definition.

Definition 1.1 (Fractional PRG)**.** *Let $f : [-1,1]^n \to [-1,1]$ be multilinear. A fractional PRG for f is a random variable $X \in [-1,1]^n$ such that*

$$|\mathbb{E}_X[f(X)] - f(\overline{0})| \leq \varepsilon.$$

One trivial construction of a fractional PRG is $X \equiv \overline{0}$ but this is not going to be useful for our purpose of constructing PRGs. To disallow such examples, we require each coordinate of X to be far from zero with some noticeable probability. Formally, $X \in [-1,1]^n$ is called p-noticeable if $\mathbb{E}[X_i^2] \geq p$ for all $i = 1, \ldots, n$.

A good example to keep in mind is the following. Let $G : \{-1,1\}^r \to \{-1,1\}^n$ be a (Boolean valued) function, and set $X = pG(U)$, where $U \in \{-1,1\}^r$ is uniform. Notice that X takes values in $\{-p,p\}^n$ and in hence is p^2-noticeable. In this case we say X has seed length r. More generally, X has seed length r if $X = G(U)$ where $G : \{-1,1\}^r \to [-1,1]^n$.

Fractional PRGs are easier to construct than standard PRGs, as they can take values in $[-1,1]^n$. For example, assume that f has Fourier tails bounded in L_1. That is, there exist parameters $a, b \geq 1$ for which

$$\sum_{S \subset [n]:|S|=k} |\hat{f}(S)| \leq a \cdot b^k \qquad \forall k = 1, \ldots, n.$$

We show (in Lemma 4.4) that if $X \in \{-1,1\}^n$ is small-biased, then pX is a fractional PRG for f with $p \approx 1/b$. The reason is that this choice of p controls all the Fourier coefficients of f with large Hamming weight, while X controls the ones with small weight. (In fact, to optimize parameters one can choose X to be almost k-wise independent; see Lemma 4.4 for details). In any case, note that pX is p^2-noticeable as pX takes values in $\{-p,p\}^n$.

1.2 Fractional PRG as steps in a random walk

Let $X \in [-1,1]^n$ be a fractional PRG for f with error ε. That is,

$$|\mathbb{E}_X[f(X)] - f(\overline{0})| \leq \varepsilon.$$

The goal is to construct a random variable $Y \in \{-1,1\}^n$ such that $\mathbb{E}_Y[f(Y)] \approx f(\overline{0})$, where the fractional PRG X provides a "small step" towards this approximation. If we can combine these small steps in a way that they converge fast to $\{-1,1\}^n$, then we would be done. To be a bit more precise, consider a random walk starting at $\overline{0}$ with the following properties:

1. The value of f at each step typically does not change too much.
2. The random walk converges fast to $\{-1,1\}^n$.

Observe that if we take X as the first step, then property 1 is satisfied for the first step. Considering later steps leads to the following question: Given a point $\alpha \in [-1,1]^n$, can we find a random variable $A = A(\alpha, X)$ such that

$$|\mathbb{E}[f(A)] - f(\alpha)| \leq \varepsilon,$$

and such that A takes values closer to Boolean values? We show that this is indeed the case if we assume that X not only fools f, but also fools any possible restriction of f.

To formalize this, let $\mathcal{F}$ be a family of n-variate Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$. We say that $\mathcal{F}$ is closed under restrictions if for any $f \in \mathcal{F}$, if we fix some inputs of f to constants $\{-1,1\}$, then the new restricted function is still in $\mathcal{F}$. Most natural families of Boolean functions studied satisfy this condition. Some examples are functions computed by small-depth circuits, functions computed by bounded width branching programs, and functions of low sensitivity.

We show that if X is a fractional PRG for such $\mathcal{F}$, then it can be used to approximate $f(\alpha)$ for any $\alpha \in [-1,1]^n$. Define $\delta_\alpha \in [0,1]^n$ by $(\delta_\alpha)_i = 1 - |\alpha_i|$. For $x, y \in [-1,1]^n$ define $x \circ y \in [-1,1]^n$ to be their coordinate-wise product, $(x \circ y)_i = x_i y_i$. Note that under this definition, the sub-cube $\{\alpha + \delta_\alpha \circ y : y \in [-1,1]^n\}$ is the largest symmetric sub-cube of $[-1,1]^n$ centered at α.

We show (Claim 3.3) that if $X \in [-1,1]^n$ is a fractional PRG for $\mathcal{F}$ which is closed under restrictions, then for any $f \in \mathcal{F}$ and any $\alpha \in [-1,1]^n$ it holds that

$$|\mathbb{E}[f(\alpha + \delta_\alpha \circ X)] - f(\alpha)| \leq \varepsilon.$$

Technically, we need to also assume that X is *symmetric*, which means that $\Pr[X = x] = \Pr[X = -x]$ for all x. This is easy to achieve from any X which is not symmetric, for example by multiplying X with a uniform bit (thus, increasing its seed length by 1 bit).

1.3 Polarization and fast convergence

Our next goal is to show fast convergence of the random walk to $\{-1,1\}^n$. To that end, we need to analyze the following martingale:

$$\begin{aligned} Y_1 &= X_1 \\ Y_i &= Y_{i-1} + \delta_{Y_{i-1}} \circ X_i \end{aligned}$$

where $X_1, X_2, \ldots$ are independent copies of a fractional PRG. We show that for some t not too large, Y_t is close to a point in $\{-1,1\}^n$. But why would that be true? This turns out to be the result of *polarization* in the random walk. It suffices to show this for every coordinate individually.

So, let $Z_1, Z_2, \ldots \in [-1,1]$ be independent random variables (which are the i-th coordinate of $X_1, X_2, \ldots$ for some fixed i), and define the following one-dimensional martingale:

$$\begin{aligned} W_1 &= Z_1 \\ W_i &= W_{i-1} + (1 - |W_{i-1}|) Z_i. \end{aligned}$$

Claim 3.5 shows that if (i) Z_i is symmetric, and (ii) $\mathbb{E}[Z_i^2] \geq p$ (which follows from our assumption that the fractional PRG is p-noticeable), then it holds that

$$\Pr[|W_t| \geq 1 - \delta] \geq 1 - \delta$$

for $t = O(\log(1/\delta)/p)$. Setting $\delta = \varepsilon/n$ guarantees that with probability $1 - \varepsilon$ all the coordinates of Y_t are ε/n close to $\{-1,1\}$. Then a simple argument shows that rounding the coordinates gives a PRG with error $O(\varepsilon)$, as desired.

We now state our main theorem.

Theorem 1.2 (Main theorem, informal version of Theorem 2.5)**.** *Let $\mathcal{F}$ be a family of n-variate Boolean functions that is closed under restrictions. Let $X \in [-1,1]^n$ be a symmetric p-noticeable fractional PRG for $\mathcal{F}$ with error ε. Set $t = O(\log(n/\varepsilon)/p)$ and let $X_1, \ldots, X_t$ be i.i.d. copies of X. Define the following random variables taking values in $[-1,1]^n$:*

$$Y_0 = \bar{0}; \qquad Y_i = Y_{i-1} + \delta_{Y_{i-1}} \circ X_i \qquad i = 1, \ldots, t.$$

Let $G = sign(Y_t) \in \{-1,1\}^n$ obtained by taking the sign of the coordinates in Y_t. Then G is a PRG for $\mathcal{F}$ with error $(t+1)\varepsilon$.

1.4 PRG for functions with bounded Fourier tails

As mentioned above, the families of Boolean functions that are fooled by our PRG include ones that satisfy the following two properties: (i) being closed under restrictions; (ii) having bounded L_1 Fourier tails. Tal [Tal17] showed that the latter condition follows from a widely studied condition, that of bounded L_2 Fourier tails. Thus, using existing bounds for L_2 Fourier tails, we get that our PRG fools several classes of Boolean functions. Below we list the results for error $\varepsilon = O(1)$, and refer the reader to the corresponding claims for the details of the full range of parameters:

1. **Functions of sensitivity** s: seed length $O(s^3 \log\log n)$. The best previous construction [HT17] required seed length sub-exponential in s (concretely, their dependence on s is $\exp(\sqrt{s})$). See Corollary 4.6 for details.

2. **Unordered read-once branching programs of width** w: seed length $O(\log^{2w+1} n \cdot \log\log n)$. This is quadratically worse than the best known PRG [CHRT17]. However, unlike [CHRT17] our PRG construction does not utilize the branching program structure at all, except to obtain the Fourier tail bounds. See Corollary 4.7 for details.

3. **Permutation unordered read-once branching programs of width** w: seed length $O(w^4 \log n \cdot \log\log n)$. This improves the dependence on n quadratically compared to the previous best PRG [RSV13]. See Corollary 4.8 for details.

4. **Bounded depth circuits**: if f is computed by AC^0 circuits of depth d and size $\mathrm{poly}(n)$, our PRG has seed length $O(\log^{2d-1} n \cdot \log\log n)$. This is quadratically worse than the best known PRG [Tal17]. See Corollary 4.9 for details.

Other than the PRG for functions of low sensitivity, all the other PRGs are comparable to the best known tailored PRG. However, the main message is that **they are all the same PRG**. Our general theorem is the following.

Theorem 1.3 (PRG for functions of bounded L_1 Fourier tail, informal version of Theorem 4.5)**.** *Let $\mathcal{F}$ be a family of n-variate Boolean functions closed under restrictions. Assume that there exist $a, b \geq 1$ such that for every $f \in \mathcal{F}$,*

$$\sum_{S \subset [n] : |S| = k} |\hat{f}(S)| \leq a \cdot b^k.$$

Then, for any $\varepsilon > 0$ there exists an explicit PRG $X \in \{-1,1\}^n$ which fools $\mathcal{F}$ with error $\varepsilon > 0$, whose seed length is $O(\log(n/\varepsilon)(\log\log n + \log(a/\varepsilon))b^2)$.

We note again that by [Tal17], Theorem 1.3 holds also if we instead assume a bound on the L_2 Fourier tails (which are more common), namely if we assume that for every $f \in \mathcal{F}$ it holds that

$$\sum_{S \subset [n] : |S| \geq k} \hat{f}(S)^2 \leq a \cdot 2^{-k/b}.$$

1.5 PRG for functions which simplify under random restriction

A major component in prior constructions of PRGs that are based on random restrictions is finding a much smaller set of 'pseudorandom retrictions'. Ajtai and Wigderson [AW85] proposed such a PRG for low depth circuits based on Håstad's switching lemma [Has86]. Many follow-up works are based on this framework to build PRGs for various classes of functions including low depth circuits, branching programs, low-sensitivity functions [TX13, GMR$^+$12, RSV13, CHRT17, HT17], and a major component of the analysis is proving that the derandomized random restrictions work.

Our framework for constructing PRGs directly applies to function families that simplify under random restrictions without the need to derandomize the restrictions. Let $\mathcal{F}$ be a family of functions $f : \{-1,1\}^n \to$

$\{-1,1\}$ which are extended multilinearly to $[-1,1]^n$. Fix a parameter $0 < p < 1$ and define the p-averaged function of f, denoted $f_p : \{-1,1\}^n \to [-1,1]$, as follows: sample $A \subset [n]$ where $\Pr[i \in A] = p$ independently for $i \in [n]$, and define

$$f_p(x) = \mathbb{E}_{A,U}[f(x_A, U_{A^c})]$$

where $x_A \in \{-1,1\}^A$ is the restriction of the input x to the coordinates in A, and $U \in \{-1,1\}^n$ is independently and uniformly chosen. The crucial observation (Claim 5.1) is that for every $x \in \{-1,1\}^n$ it holds that

$$f(px) = f_p(x).$$

Suppose now we have a standard PRG X for the class of p-averaged functions $\mathcal{F}_p = \{f_p : f \in \mathcal{F}\}$. Note a PRG for the p-random restriction of functions in $\mathcal{F}$ would do, as f_p is a convex combination of p-random restrictions of f (namely, averaging over U). Then, using our observation above, this implies that $X' = pX$ is a fractional PRG for the class $\mathcal{F}$. Now by using our framework of viewing this fractional PRG as a random walk step, one can derive a standard PRG for $\mathcal{F}$ using $O(\log(1/\epsilon)/p^2)$ independent copies of X.

1.6 Fourier tails of low degree $\mathbb{F}_2$ polynomials

Viola [Vio09] gave a construction of a pseudorandom generator which fools n-variate polynomials over $\mathbb{F}_2$. The construction is the XOR of d independent small-bias generators. We wonder whether our framework can be used to achieve similar bounds. In particular, we raise the following problem: does the class of low-degree polynomials over $\mathbb{F}_2$ have bounded L_1 Fourier tails? It's trivially true for $d = 1$ and it can be shown to hold for $d = 2$. However, to the best of our knowledge nothing was known for $d \geq 3$.

We show (see *Theorem* 6.1 for more details) that for any Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ computed by a $\mathbb{F}_2$-polynomial of degree at most d, the following L_1 Fourier tail bound holds:

$$\sum_{|S|=k} |\widehat{f}(S)| \leq k^k 2^{3dk} \qquad \forall k = 1,\dots,n.$$

This bound however falls short of implying a PRG using our techniques, and we conjecture that the correct bound is c_d^k, for some constant $c_d = 2^{O(d)}$.

1.7 PRGs with respect to arbitrary product distributions

We note the following interesting generalization of our results that is almost direct from our techniques. Consider the problem of 'fooling' a family of functions with respect to an arbitrary product distribution D on $\{-1,1\}^n$ (the uniform distribution being a special case). More formally, given a distribution D on $\{-1,1\}^n$ and a family of functions $\mathcal{F}$, we say that a random variable X is a PRG for $\mathcal{F}$ (with respect to D) if $|\mathbb{E}[f(D)] - \mathbb{E}[f(X)]| \leq \epsilon$.

We show a way to fool functions with respect to arbitrary product distributions.

Corollary 1.4. *Let $\mathcal{F}$ be a family of n-variate Boolean functions which is closed under restrictions and let D be any product distribution on $\{-1,1\}^n$. Let $X \in [-1,1]^n$ be a symmetric p-noticeable fractional PRG for $\mathcal{F}$ with error ε and seed length ℓ. Let $t = O(\log(n/\epsilon)/p)$. Then there exists an explicit PRG for $\mathcal{F}$ with respect to D with error $t\varepsilon$ and seed length $t\ell$.*

Proof sketch. If D is a product distribution on $\{-1,1\}^n$, then $\mathbb{E}[f(D)] = f(\alpha)$, where $\alpha = \mathbb{E}[D] \in [-1,1]^n$. Thus, we now start our random walk (defined by the fractional PRG) from the point α instead of from $\overline{0}$, and the convergence follows from polarization in exactly the same way. □

Thus all our PRG results in fact generalize to PRGs with respect to arbitrary product distributions. To the best of our knowledge, we are not aware of any non-trivial PRGs against arbitrary product distributions for the classes of functions we study. We wonder if this notion of fooling arbitrary product distributions has interesting applications.

1.8 Related works

The line of research closest in spirit to our work, and which motivated our work, is that of using random and pseudo-random restrictions to construct PRGs. A good example is [GMR$^+$12] which uses pseudo-random restrictions to construct PRGs. Our framework can be seen as extending this, as we do not need to analyze pseudo-random restrictions; instead, we analyze fractional PRGs, where the restriction happens automatically from the fractional PRG structure, and no derandomization is necessary.

Another line of work is the use of random walks in combinatorial optimization, for example in the algorithmic versions of Spencer's theorem [Ban10, LM12] and follow up works. It would be interesting to see if polarization can be used to speed up random walks in combinatorial optimization as well.

1.9 Open problems

As we give a new framework for constructing PRGs, there are many open problems that arise, both conceptual and technical.

Early termination. Our analysis requires a random walk with $t = O(\log(n/\varepsilon)/p)$ steps, each coming from a p-noticeable fractional PRG. We believe that for some natural families of functions shorter random walks might also suffice, but we do not know how to show this. We discuss this further in *Section* 7.

Open problem 1.5. *Find conditions on classes of Boolean functions so that short random walks can be used to construct PRGs. In particular, are there nontrivial classes where the number of steps is independent of n?*

Less independence. Our analysis of Theorem 2.5 currently requires to assume t independent copies of a fractional PRG X. It might be possible that they copies can be chosen in a less independent form, where the analysis still holds.

Open problem 1.6. *Can the fractional PRGs $X_1, \dots, X_t$ in Theorem 2.5 be chosen not independently, such that the conclusion still holds? Concrete examples to consider are k-wise independence for $k \ll t$, or using an expander random walk.*

More applications. Our current applications follow from the construction of a fractional PRG for functions with bounded Fourier tails. The fractional PRG itself follows from standard constructions in pseudo-randomness (almost k-wise independent) adapted to our scenario. It will be interesting to try and find other classes of Boolean functions for which different constructions of fractional PRG work.

Gadgets. We can view the random walk as a "gadget construction". Given independent p-noticeable fractional PRGs $X_1, \dots, X_t \in [-1,1]^n$, view them as the rows of a $t \times n$ matrix, and then apply a gadget $g : [-1,1]^t \to \{-1,1\}$ to each column to obtain the outcome in $\{-1,1\}^n$. We show that the random walk gives such a gadget which converges for $t = O(\log(n/\varepsilon)/p)$. Many constructions of PRGs can be viewed in this framework, where typically $X_i \in \{-1,1\}^n$. Ours is the first construction which allows X_i to take non-Boolean values. It is interesting whether other gadgets can be used instead of the random walk gadget, and whether there are general properties of gadgets that would suffice.

Low degree polynomials. As discussed above, we wonder if our techniques can be used to construct a PRG for low degree $\mathbb{F}_2$ polynomials. In particular, we ask if one could improve the bounds we obtain (see *Theorem* 6.1) on the L_1 Fourier tails of low degree $\mathbb{F}_2$ polynomials.

Open problem 1.7. *Let $f = (-1)^p$ where $p : \mathbb{F}_2^n \to \mathbb{F}_2$ is a polynomial of degree d. Is there a constant c_d such that $\sum_{S:|S|=k} |\hat{f}(S)| \leq c_d^k$, which is independent of n? In particular, we conjecture that $c_d = 2^{O(d)}$ should work.*

Note that the exponential dependence on k is needed, as witnessed from the following example: consider the quadratic $\mathbb{F}_2$ polynomial $q(x) = \sum_{i=1}^{n/2} x_{2i-1}x_{2i}$. Then $(-1)^q$ has Fourier L_1 weight $\binom{n}{n/2} \cdot 2^{-n/2} = 2^{\Omega(n)}$ on the $(n/2)$-th level.

AC^0 with parity in the inputs. The class of AC^0 with parity gates has been well studied, and although worst-case lower bounds for it were established by Razborov [Raz86] and Smolensky [Smo93], no strong average-case lower bounds are known. Related to this, no explicit construction of pseudorandom generator is known. A special subclass that was studied is that of AC^0 with parity gates at the inputs. The following conjecture, which generalizes Braverman's theorem [Bra10], would give a pseudorandom generator for this class using our framework.

Open problem 1.8. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be computed by an AC^0 circuit of size s and depth d. Is it true that, for any $\varepsilon > 0$, if D', D'' are two distributions on $\{-1,1\}^n$ whose Fourier coefficients are the same up to level $k = \log(s/\varepsilon)^{O(d)}$, then $\mathbb{E}[f(D')] = \mathbb{E}[f(D'')] \pm \varepsilon$?*

Note that Braverman's theorem proves this for the special case of D' being the uniform distribution. In fact, to build a pseudorandom generator for AC^0 with parities at the inputs, it suffices to prove the conjecture for D' uniform on a subspace of $\mathbb{F}_2^n$.

1.10 Paper organization

We describe the general framework in detail in Section 2. We prove Theorem 2.5 in Section 3. We describe applications in Section 4. Our framework also applies to function families that simplify under random restrictions. We describe this in Section 5. We prove L_1 Fourier tail bounds for low degree $\mathbb{F}_2$ polynomials in Section 6. We give a partial answer the question related to early termination of the random walk in Section 7.

2 General framework

Boolean functions. Let $f : \{-1,1\}^n \to [-1,1]$ be an n-variate Boolean function, identified with its multilinear extension, also known as its Fourier expansion. For $x \in [-1,1]^n$ define $f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i$. As f is multilinear, a convenient viewpoint is to view $f(x)$ as computing the expected value of f on a product distribution on $\{-1,1\}^n$. That is, let $W = W(x) \in \{-1,1\}^n$ be a random variable, where $W_1, \ldots, W_n$ are independently chosen so that $\mathbb{E}[W_i] = x_i$. Then $f(x) = \mathbb{E}f(W)$. In particular, $f(\overline{0}) = \mathbb{E}f(U)$, where $U \in \{-1,1\}^n$ is uniformly chosen.

A family $\mathcal{F}$ of n-variate Boolean functions is said to be *closed under restrictions* if for any $f \in \mathcal{F}$ and any function $f' : \{-1,1\}^n \to \{-1,1\}$ obtained from f by fixing some of its inputs to $\{-1,1\}$ it holds that also $f' \in \mathcal{F}$.

Pseudorandom generators. Let $\mathcal{F}$ be a family of n-variate Boolean functions. The following is the standard definition of a pseudorandom generator (PRG) for $\mathcal{F}$, adapted to our notation.

Definition 2.1 (PRG). *A random variable $X \in \{-1,1\}^n$ is a PRG for $\mathcal{F}$ with error ε, if for any $f \in \mathcal{F}$ it holds that $\left|f(\overline{0}) - \mathbb{E}f(X)\right| \leq \varepsilon$.*

We introduce the notion of a *fractional PRG*. It is the same as a PRG, except that the random variable is allowed to take values in $[-1,1]^n$, instead of only Boolean values.

Definition 2.2 (Fractional PRG). *A random variable $X \in [-1,1]^n$ is a fractional PRG for $\mathcal{F}$ with error ε, if for any $f \in \mathcal{F}$ it holds that $\left|f(\overline{0}) - \mathbb{E}f(X)\right| \leq \varepsilon$.*

Our main goal will be to "amplify" fractional PRGs for $\mathcal{F}$ in order to obtain PRGs for $\mathcal{F}$. To that end, we need to enforce some non-triviality conditions on the fractional PRG. For example, $X = \overline{0}$ is a fractional PRG for any function. We require that for any coordinate $i \in [n]$, the value of X_i is far from zero with noticeable probability. Formally, we require a noticeable second moment.

Definition 2.3 (p-noticeable random variable)**.** *A random variable $X \in [-1,1]^n$ is p-noticeable if for every $i \in [n]$, $\mathbb{E}[X_i^2] \geq p$.*

For technical reasons, we would also need X to be *symmetric*, which means that the distribution of $-X$ is the same as the distribution of X. This is easy to achieve, for example by multiplying all elements of X with a uniformly chosen sign.

Polarizing random walks. The main idea is to view a fractional PRG as steps in a random walk in $[-1,1]^n$ that converges to $\{-1,1\}^n$. To that end, we define a gadget that implements the random walk; and moreover, that allows for fast convergence. As we will see later, the fast convergence is an effect of polarization.

Definition 2.4 (Random walk gadget)**.** *For any $t \geq 1$ define the random walk gadget $g_t : [-1,1]^t \to [-1,1]$ as follows. Let $a_1, \ldots, a_t \in [-1,1]$. Define $g_1(a_1) := a_1$ and for $t > 1$,*

$$g_t(a_1, \ldots, a_t) := g_{t-1}(a_1, \ldots, a_{t-1}) + (1 - |g_{t-1}(a_1, \ldots, a_{t-1})|)a_t.$$

We extend the definition to act on bit-vectors. Define $g_t^n : ([-1,1]^n)^t \to [-1,1]^n$ as follows. For $x_1, \ldots, x_t \in [-1,1]^n$ define

$$g_t^n(x_1, \ldots, x_t) = (g_t(x_{1,1}, \ldots, x_{t,1}), \ldots, g_t(x_{1,n}, \ldots, x_{t,n}))\,.$$

Equivalently, we can view g_t^n as follows: construct a $t \times n$ matrix whose rows are $x_1, \ldots, x_t$; and then apply g_t to each column of the matrix to obtain a resulting vector in $[-1,1]^n$.

The following theorem shows how to "amplify" fractional PRGs using the random walk gadget to obtain a PRG. Below, for $x \in [-1,1]^n$ we denote by $\text{sign}(x) \in \{-1,1\}^n$ the Boolean vector obtained by taking the sign of each coordinate (the sign of 0 can be chosen arbitrarily).

Theorem 2.5 (Amplification Theorem)**.** *Let $\mathcal{F}$ be a family of n-variate Boolean functions which is closed under restrictions. Let $X \in [-1,1]^n$ be a symmetric p-noticeable fractional PRG for $\mathcal{F}$ with error ε. Set $t = O(\log(n/\varepsilon)/p)$ and let $X_1, \ldots, X_t$ be iid copies of X. Define a random variable $G \in \{-1,1\}^n$ as follows:*

$$G := G(X_1, \ldots, X_t) = sign(g_t^n(X_1, \ldots, X_t)).$$

Then G is a PRG for $\mathcal{F}$ with error $(t+1)\varepsilon$.

3 Proof of Amplification Theorem

We prove Theorem 2.5 in this section. From here onwards, we fix a family $\mathcal{F}$ of n-variate Boolean functions which is closed under restrictions. The proof is based on the following two lemmas. The first lemma amplifies a p-noticeable fractional PRG to a $(1-q)$-noticeable fractional PRG. The second lemma shows that setting $q = \varepsilon/n$, the latter fractional PRG can be rounded to a Boolean-valued PRG without incurring too much error.

Lemma 3.1 (Amplification lemma)**.** *Let $X_1, \ldots, X_t \in [-1,1]^n$ be independent symmetric p-noticeable fractional PRGs for $\mathcal{F}$ with error ε. Define a random variable $Y \in [-1,1]^n$ as*

$$Y := g_t^n(X_1, \ldots, X_t).$$

Then Y is a $(1-q)$-noticeable fractional PRG for $\mathcal{F}$ with error $t\varepsilon$, where $q = 2^{-\Omega(pt)}$.

Lemma 3.2 (Rounding lemma)**.** *Let $Y \in [-1,1]^n$ be a $(1-q)$-noticeable fractional PRG for $\mathcal{F}$ with error ε. Then sign$(Y) \in \{-1,1\}^n$ is a PRG for $\mathcal{F}$ with error $\varepsilon + qn$.*

Theorem 2.5 follows directly by applying Lemma 3.1 with $t = O(\log(n/\varepsilon)/p)$ to obtain $q = \varepsilon/n$ and then applying Lemma 3.2.

3.1 Proof of Lemma 3.1

We prove Lemma 3.1 in this section. We need to prove two claims: that $g_t^n(X_1,\dots,X_t)$ is a fractional PRG for $\mathcal{F}$ with error εt, and that it is $(1-q)$-noticeable. This is achieved in the following sequence of claims.

First we need some notations. For $y \in [-1,1]^n$ define $\delta_y \in [-1,1]^n$ by $(\delta_y)_i := 1 - |y_i|$. For two vectors $x, y \in [-1,1]^n$ define $x \circ y \in [-1,1]^n$ to be their pointwise product, namely $(x \circ y)_i := x_i y_i$. Observe that $\{y + \delta_y \circ x : x \in [-1,1]^n\}$ is the largest symmetric sub-cube in $[-1,1]^n$ centered at y.

Claim 3.3. *Let $X \in [-1,1]^n$ be a fractional PRG for $\mathcal{F}$ with error ε. Then for any $f \in \mathcal{F}$ and any $y \in [-1,1]^n$,*

$$|f(y) - \mathbb{E} f(y + \delta_y \circ X)| \le \varepsilon.$$

Proof. Consider a distribution over $F \in \mathcal{F}$ obtained from f by fixing the i-th input to $\mathrm{sign}(y_i)$ with probability $|y_i|$, independently for each i. That is,

$$F(x) := f(R(x)),$$

where $R(x) \in \{-1,1\}^n$ is a random variable obtained by sampling $R_1,\dots,R_n$ independently where $\Pr[R_i = \mathrm{sign}(y_i)] = |y_i|$ and $\Pr[R_i = x_i] = 1 - |y_i|$. By the multi-linearity of f, and as $R(x)$ is a product distribution,

$$\mathbb{E}_F[F(x)] = \mathbb{E}_R[f(R(x))] = f(\mathbb{E}_R[R(x)]) = f(y + \delta_y \circ x).$$

Setting $x = X$ and averaging over X gives

$$|f(y) - \mathbb{E}_X[f(y + \delta_y \circ X)]| = \left|\mathbb{E}_F F[(\overline{0})] - \mathbb{E}_{F,X}[F(X)]\right| \le \mathbb{E}_F \left|F(\overline{0}) - \mathbb{E}_X[F(X)]\right| \le \varepsilon,$$

since $F \in \mathcal{F}$ with probability one and X is a fractional PRG for $\mathcal{F}$ with error ε. □

Claim 3.4. *Let $X_1,\dots,X_t \in [-1,1]^n$ be independent fractional PRGs for $\mathcal{F}$ with error ε. Then for any $f \in \mathcal{F}$,*

$$\left|f(\overline{0}) - \mathbb{E}_{X_1,\dots,X_t}[f(g_t^n(X_1,\dots,X_t))]\right| \le t\varepsilon.$$

Proof. The proof is by induction on t. The base case $t = 1$ follows by definition as $g_1^n(X_1) = X_1$. For $t > 1$ we will show that

$$\left|\mathbb{E}[f(g_{t-1}^n(X_1,\dots,X_{t-1}))] - \mathbb{E}[f(g_t^n(X_1,\dots,X_t))]\right| \le \varepsilon,$$

from which the claim follows by the triangle inequality. In fact, we will show a stronger inequality: for any fixing of $x_1,\dots,x_{t-1} \in [-1,1]^n$, it holds that

$$\left|f(g_{t-1}^n(x_1,\dots,x_{t-1})) - \mathbb{E}_{X_t}[f(g_t^n(x_1,\dots,x_{t-1},X_t))]\right| \le \varepsilon.$$

The first inequality then follows by averaging over $x_1 = X_1,\dots,x_{t-1} = X_{t-1}$. To see why this latter inequality holds, set $y = g_{t-1}^n(x_1,\dots,x_{t-1})$. Then by definition,

$$g_t^n(x_1,\dots,x_{t-1},X_t) = y + \delta_y \circ X_t.$$

The claim now follows from Claim 3.3. □

We have so far proved that $g_t^n(X_1,\dots,X_t)$ is a fractional PRG for $\mathcal{F}$ with slightly worse error. Although we do not need it, it is worth noting that it is symmetric since $X_1,\dots,X_t$ are symmetric and $-g_t^n(X_1,\dots,X_t) = g_t^n(-X_1,\dots,-X_t)$. To conclude, we show that it converges fast to a value close to $\{-1,1\}^n$. This is the effect of *polarization*. It will be enough to analyze this for one-dimensional random variables.

Claim 3.5. *Let $A_1, \dots, A_t \in [-1,1]$ be independent symmetric random variables with $\mathbb{E}[A_i^2] \geq p$. For $i = 1, \dots, t$ define*

$$B_i := g_i(A_1, \dots, A_i) = B_{i-1} + (1 - |B_{i-1}|)A_i.$$

Then $\mathbb{E}[B_t^2] \geq 1 - q$ where $q = 3\exp(-tp/8)$.

Proof. Define $C_i := 1 - |B_i|$. It satisfies the following recursive definition:

$$C_i = \begin{cases} C_{i-1}(1 - A_i) & \text{if } C_{i-1}(1 - A_i) \leq 1 \\ 2 - C_{i-1}(1 - A_i) & \text{if } C_{i-1}(1 - A_i) > 1 \end{cases}.$$

In either case one can verify that $C_i \in [0,1]$ and that

$$C_i \leq C_{i-1}(1 - A_i).$$

As C_{i-1} and A_i are independent we obtain that

$$\mathbb{E}\left[\sqrt{C_i}\right] = \mathbb{E}\left[\sqrt{C_{i-1}}\right]\mathbb{E}\left[\sqrt{1 - A_i}\right].$$

We now use the assumption that the A_i are symmetric. The Taylor expansion of $\sqrt{1-x}$ in $[-1,1]$ is

$$\sqrt{1-x} = 1 - \frac{x}{2} - \frac{x^2}{8} - \frac{x^3}{16} - \dots$$

In particular, all the coefficients except for the constant term are negative. As A_i is symmetric, $\mathbb{E}[A_i^k] = 0$ for any odd k, so

$$\mathbb{E}\left[\sqrt{1 - A_i}\right] \leq 1 - \frac{\mathbb{E}[A_i^2]}{8} \leq 1 - \frac{p}{8} \leq \exp(-p/8).$$

Thus

$$\mathbb{E}\left[\sqrt{C_t}\right] \leq \prod_{i=1}^{t} \mathbb{E}\left[\sqrt{1 - A_i}\right] \leq \exp(-tp/8).$$

By Markov's inequality, $\Pr[C_t \geq \exp(-tp/2)] \leq \exp(-tp/8)$. If $C_t \leq \exp(-tp/2)$ then $1 - B_t^2 \leq 2\exp(-tp/2)$. If not, then we can trivially bound $1 - B_t^2 \leq 1$. Putting these together gives

$$\mathbb{E}[1 - B_t^2] \leq 2\exp(-tp/2) + \exp(-tp/8) \leq 3\exp(-tp/8).$$

□

Corollary 3.6. *Let $X_1, \dots, X_t \in [-1,1]^n$ be independent symmetric p-noticeable random variables. Define $Y = g_t^n(X_1, \dots, X_t)$. Then Y is $(1-q)$-noticeable for $q = 3\exp(-tp/8)$.*

Proof. Apply Claim 3.5 to each coordinate of Y. □

Lemma 3.1 follows by combining Claim 3.4 and Corollary 3.6.

3.2 Proof of Lemma 3.2

We prove Lemma 3.2 in this section. Let $x \in [-1,1]^n$ be a potential value obtained by X. Let $W := W(x) \in \{-1,1\}^n$ be a random variable, where $W_1, \dots, W_n$ are independent and $\mathbb{E}[W_i] = x_i$. Then $\mathbb{E}_W[f(W)] = f(x)$. As f takes values in $[-1,1]$, we can upper bound $|f(x) - f(\text{sign}(x))|$ by

$$|f(x) - f(\text{sign}(x))| = |\mathbb{E}_W[f(W)] - f(\text{sign}(x))| \leq \Pr[W \neq \text{sign}(x)].$$

The last term can be bounded by the union bound,

$$\Pr[W \neq \mathrm{sign}(x)] \leq \sum_{i=1}^{n} \Pr[W_i \neq \mathrm{sign}(x_i)] = \frac{1}{2}\sum_{i=1}^{n} 1 - |x_i|.$$

Setting $x = X$ and averaging over X gives

$$|\mathbb{E}_X[f(X)] - \mathbb{E}_X[f(\mathrm{sign}(X))]| \leq \mathbb{E}_X|f(X) - f(\mathrm{sign}(X))| \leq \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[1 - |X_i|].$$

As X is $(1-q)$-noticeable it satisfies $\mathbb{E}[X_i^2] \geq 1 - q$ for all i. As $1 - z \leq 1 - z^2$ for all $z \in [0,1]$ we have

$$\mathbb{E}[1 - |X_i|] \leq \mathbb{E}[1 - X_i^2] \leq q.$$

This concludes the proof as

$$|f(\overline{0}) - \mathbb{E}_X[f(\mathrm{sign}(X))]| \leq |f(\overline{0}) - \mathbb{E}_X[f(X)]| + |\mathbb{E}_X[f(X)] - \mathbb{E}_X[f(\mathrm{sign}(X))]| \leq \varepsilon + qn,$$

where the first inequality follows as X is a fractional PRG with error ε, and the second by the discussion above.

4 PRGs for functions with bounded Fourier tails

Several natural families of Boolean functions have bounded Fourier tails, such as: AC^0 circuits [LMN93, Man95]; functions with bounded sensitivity [GSW16, LTZ16]; and functions computed by branching programs of various forms [RSV13, CHRT17]. Our goal is to construct a universal PRG which fools any such function. We consider two variants: L_1 bounds and L_2 bounds.

Definition 4.1 (L_1 bounds)**.** *For $a, b \geq 1$, we denote by $\mathcal{L}_1^n(a,b)$ the family of n-variate Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ which satisfy*

$$\sum_{\substack{S \subset [n] \\ |S|=k}} |\widehat{f}(S)| \leq a \cdot b^k \qquad \forall k = 1, \ldots, n.$$

Definition 4.2 (L_2 bounds)**.** *For $a, b \geq 1$, we denote by $\mathcal{L}_2^n(a,b)$ the family of n-variate Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ which satisfy*

$$\sum_{\substack{S \subset [n] \\ |S| \geq k}} \widehat{f}(S)^2 \leq a \cdot 2^{-k/b} \qquad \forall k = 1, \ldots, n.$$

Tal [Tal17] showed that L_2 bounds imply L_1 bounds: if $f \in \mathcal{L}_2(a,b)$ then $f \in \mathcal{L}_1(a,b')$ for $b' = O(b)$. The reverse direction is false, as can be witnessed by the PARITY function. So, the class of functions with L_1 bounded Fourier tails is richer, and we focus on it.

In the following lemma, we construct a fractional PRG for this class, which we will then amplify to a PRG. We note that this lemma holds also for bounded functions, not just Boolean functions. The construction is based on a scaling of almost d-wise independent random variables, whose definition we now recall.

Definition 4.3 (Almost d-wise independence)**.** *A random variable $Z \in \{-1,1\}^n$ is ε-almost d-wise independent if, for any restriction of Z to d coordinates, the marginal distribution has statistical distance at most ε from the uniform distribution on $\{-1,1\}^d$.*

Naor and Naor [NN93] gave an explicit construction of an ε-almost d-wise random variable $Z \in \{-1,1\}^n$ with seed length $O(\log\log n + \log d + \log(1/\varepsilon))$. We note that this seed length is optimal, up to the hidden constants.

Lemma 4.4. *Fix $n, a, b \geq 1$ and $\varepsilon > 0$. There exists a fractional PRG $X \in [-1,1]^n$ that fools $\mathcal{L}_1^n(a,b)$ with error ε, such that*

(i) X is p-noticeable for $p = \frac{1}{4b^2}$.

(ii) The seed length of X is $O(\log \log n + \log(a/\varepsilon))$.

Proof. Fix $f \in \mathcal{L}_1^n(a,b)$. Set $d = \lceil \log 2a/\varepsilon \rceil, \delta = \varepsilon/2a, \beta = 1/2b$. Let $Z \in \{-1,1\}^n$ be an δ-almost d-wise independent random variable, and set $X = \beta Z$ which takes values in $\{-\beta, \beta\}^n$. We claim that X satisfies the requirements of the lemma. Claim (i) clearly holds, and claim (ii) holds by the Naor-Naor construction. We thus focus on proving that X fools $\mathcal{F}$ with error ε.

Fix $f \in \mathcal{F}$ and consider its Fourier expansion:

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x^S.$$

We need to show that $\mathbb{E}[f(X)]$ is close to $f(\overline{0})$. Averaging over X gives

$$|\mathbb{E}[f(X)] - f(\overline{0})| \leq \sum_{|S|>0} |\widehat{f}(S)| \cdot |\mathbb{E}[X^S]| = \sum_{|S|>0} |\widehat{f}(S)| \cdot \beta^{|S|} |\mathbb{E}[Z^S]|.$$

We next bound $|\mathbb{E}[Z^S]|$. If $|S| \leq d$ then by the definition of Z we have $|\mathbb{E}[Z^S]| \leq \delta$. If $|S| > d$ we bound trivially $|\mathbb{E}[Z^S]| \leq 1$. Let $W_k = \sum_{S:|S|=k} |\hat{f}(S)|$, where by assumption $W_k \leq a \cdot b^k$. Thus

$$|\mathbb{E}[f(X)] - f(\overline{0})| \leq \delta \sum_{k=1}^{d} W_k \beta^k + \sum_{k>d} W_k \beta^k \leq \delta a \sum_{k=1}^{d} (\beta b)^k + a \sum_{k>d} (\beta b)^k \leq \delta a + 2^{-d} a$$

where we used the choice of $\beta = 1/2b$. The claim follows as we set $\delta = \varepsilon/2a$ and $2^{-d} \leq \varepsilon/2a$. □

Applying Theorem 2.5 using the fractional PRG constructed in Lemma 4.4 gives the following PRG construction. Note that we still need to require that $\mathcal{F}$ is closed under restrictions.

Theorem 4.5. *Let $\mathcal{F}$ be a family of n-variate Boolean functions closed under restrictions. Assume that $\mathcal{F} \subset \mathcal{L}_1^n(a,b)$ or that $\mathcal{F} \subset \mathcal{L}_2^n(a,b)$. Then, for any $\varepsilon > 0$ there exists an explicit PRG $X \in \{-1,1\}^n$ which fools $\mathcal{F}$ with error $\varepsilon > 0$, whose seed length is $O(\log(n/\varepsilon)(\log \log n + \log(a/\varepsilon))b^2)$.*

4.1 Applications

We apply our PRG from Theorem 4.5 to several well studied classes of Boolean functions that are known to satisfy a Fourier tail bound.

Functions of bounded sensitivity. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Its sensitivity at an input $x \in \{-1,1\}^n$ is the number of neighbors x' of x (that is, x' and x differ at exactly one coordinate) such that $f(x') \neq f(x)$. The (max) sensitivity of f is $s(f) = \max_x s(f,x)$. The sensitivity conjecture speculates that functions of sensitivity s can be computed by decision trees of depth poly(s). A corollary would be that almost poly(s)-wise distributions fool functions of low sensitivity. So, one may ask to construct comparable PRGs for functions of low sensitivity.

This question was first considered by Hatami and Tal [HT17]. They constructed a PRG with sub-exponential seed length $\exp(O(\sqrt{s}))$. Theorem 4.5 gives an improved construction that essentially matches the consequence of the sensitivity conjecture. Our PRG uses the recent bounds of Gopalan et al. [GSW16] on the Fourier tail of functions of low sensitivity. Concretely, Gopalan et al. [GSW16] show that if $s(f) = s$ then $f \in \mathcal{L}_2(n,t)$ for $t = O(s)$. It is straightforward to verify that a restriction can only decrease the sensitivity of the function, so that the class of functions of sensitivity at most s is closed under restrictions. A direct application of Theorem 4.5 gives a PRG with seed length $O(s^2 \log(n/\varepsilon)(\log \log(n) + \log(1/\varepsilon)))$.

To get a somewhat improved bound, one can apply a result of Simon [Sim83] that shows that if $s(f) = s$ then f depends on at most $m = 4^s$ many inputs. In this case, the analysis of Theorem 2.5 can be applied with m variables instead of n variables , so that we only need $O(\log m/\varepsilon)$ iterations. Note that the fractional PRG still requires a seed length which depends on the original n. We obtain:

Corollary 4.6. *For any $n, s \geq 1$ and $\varepsilon > 0$, there exists an explicit PRG which fools n-variate Boolean functions with sensitivity s with error ε, whose seed length is $O(s^3 \log(1/\varepsilon)(\log\log n + \log(1/\varepsilon)))$.*

We note that the $\log\log n$ term cannot be removed. Indeed, even if we restrict attention to functions which are XOR of at most 2 bits (for which $s = 2$) the seed length required is $\Omega(\log\log n + \log(1/\varepsilon))$.

Unordered branching programs. An oblivious read-once branching program (abbrv ROBP) B of width w is a non-uniform model of computation, that captures randomized algorithms with space $\log w$. A branching program B maintains a state in the set $\{1, \ldots, w\}$ and reads the input bits in a known fixed order. At time step $i = 1, \ldots, n$, B reads a bit and based on the time step, the read bit and the current state it transitions to a new state. Thus, B can be thought of as a layered directed graph, with w nodes in each layer, and two edges going out of each node to the immediately next layer, one labeled with a 1 and the other labeled with a -1.

Let $\mathcal{B}^n(w)$ be the class of n-variate Boolean functions computed by read-once oblivious branching programs of width w, where the order of the inputs is arbitrary. A recent work of Chattopadhyay et al. [CHRT17] showed that these functions have L_1 bounded Fourier tails. Concretely, $\mathcal{B}^n(w) \subset \mathcal{L}_1^n(t)$ for $t = (\log n)^w$. They used this to construct a PRG with seed length $O(\log n)^{w-1} \log^2(n/\epsilon) \log\log n$. Using our PRG from Theorem 4.5 we get a comparable (although slightly worse) seed length. Note that $\mathcal{B}^n(w)$ is closed under restrictions.

Corollary 4.7. *Fix $n, w \geq 1$ and $\varepsilon > 0$. There is an explicit PRG which fools $\mathcal{B}^n(w)$ with error $\varepsilon > 0$, whose seed length is $O(\log(n/\varepsilon)(\log\log n + \log 1/\varepsilon)(\log n)^{2w})$.*

Permutation branching programs. A special case of read-once branching programs are permutation branching programs, where the transition function from level i to level $i + 1$ in the graph is a permutation for every choice of the input bit. We denote it by $\mathcal{B}^n_{\text{perm}}(w) \subset \mathcal{B}^n(w)$. Reingold et al. [RSV13] showed that if a Boolean function is computed by a permutation branching program of width w, then it has L_2 bounded Fourier tails with parameter $2w^2$. Note that permutation branching programs are also closed under restrictions. Thus we obtain the following result:

Corollary 4.8. *Fix $n, w \geq 1$ and $\varepsilon > 0$. There is an explicit PRG which fools $\mathcal{B}^n_{perm}(w)$ with error $\varepsilon > 0$, whose seed length is $O(\log(n/\varepsilon)(\log\log n + \log 1/\varepsilon)w^4)$.*

The dependence on n in our PRG is better than in the previous work of [RSV13], as they obtained seed length $O(w^2 \log(w) \log(n) \log(nw/\varepsilon) + w^4 \log^2(w/\varepsilon))$.

The work of [RSV13] actually shows the Fourier tail bounds for a more general class of branching programs, called regular branching programs. However, these are not closed under restriction, and hence our PRG construction fails to work (the same problem occurs also in the construction of [RSV13]).

Bounded depth circuits. The class of bounded-depth Boolean circuits AC^0 has been widely studied. In particular, Linial, Mansour and Nisan [LMN93] showed that it has bounded L_2 Fourier tails. Tal [Tal17] obtained improved bounds. If f is an n-variate Boolean function computed by an AC^0 circuit of depth d and size s, then $f \in \mathcal{L}_2(n, t)$ for $t = 2^{O(d)} \log^{d-1} s$. Theorem 4.5 provides a new PRG for AC^0 which is comparable with the existing PRGs of Nisan [Nis91] and Braverman [Bra10].

Corollary 4.9. *Fix $n, s \geq 1$ and $\varepsilon > 0$. There is an explicit PRG which fools n-variate functions which can be computed by AC^0 circuits of size s and depth d, with error $\varepsilon > 0$, whose seed length is $O(\log(n/\varepsilon)(\log\log n + \log 1/\varepsilon) \log^{2d-2} s)$.*

5 PRG for functions which simplify under random restriction

Another generic application of our framework is constructing PRGs for classes that simplify under random restriction. Let $\mathcal{F}$ be a family of functions $f : \{-1,1\}^n \to \{-1,1\}$ which are extended multilinearly to $[-1,1]^n$. Fix a parameter $0 < p < 1$ and define the p-averaged function of f, denoted $f_p : \{-1,1\}^n \to [-1,1]$ as follows: sample $A \subset [n]$ where $\Pr[i \in A] = p$ independently for $i \in [n]$, and define

$$f_p(x) = \mathbb{E}_A \mathbb{E}_U[f(x_A, U_{A^c})]$$

where $x_A \in \{-1,1\}^A$ is the restriction of the input x to the coordinates in A, and $U \in \{-1,1\}^n$ is independently and uniformly chosen.

Claim 5.1. *$f_p(x) = f(px)$.*

Proof. Let A, U be random variables as defined above. Define a random variable $Y \in \{-1,1\}^n$ as follows:

$$Y_i = \begin{cases} x_i & \text{if } A_i = 1 \\ U_i & \text{if } A_i = 0 \end{cases}.$$

Note that Y is a product distribution. By definition of f_p, $f_p(x) = \mathbb{E}[f(Y)]$. By multilinearity of f, $\mathbb{E}[f(Y)] = f(\mathbb{E}[Y]) = f(px)$. □

Suppose that we have a standard PRG X for the class of p-averaged functions $\mathcal{F}_p = \{f_p : f \in \mathcal{F}\}$. Claim 5.1 implies that $X' = pX$ is a fractional PRG for the class $\mathcal{F}$. Theorem 2.5 then constructs a PRG for $\mathcal{F}$ using $O(\log(1/\epsilon)/p^2)$ independent copies of X.

6 Spectral tail bounds for low degree $\mathbb{F}_2$-polynomials

In this section, we prove L_1 Fourier tail bounds for functions computed by low degree polynomials on $\mathbb{F}_2$. However, our bounds fall short of implying PRGs for the class of low-degree $\mathbb{F}_2$ polynomials in our framework.

Theorem 6.1. *Let $p : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree d, and let $f(x) = (-1)^{p(x)}$. Then*

$$\sum_{\substack{S \subset [n] \\ |S|=k}} |\widehat{f}(S)| \le (k2^{3d})^k \qquad \forall k = 1, \ldots, n.$$

We note that L_2 bounds do not hold for low-degree polynomials, as can be witnessed by taking a high-rank quadratic polynomial. We prove Theorem 6.1 in the remainder of this section.

We first introduce some notation to simplify the presentation. Define

$$W_k(f) := \sum_{|S|=k} |\hat{f}(S)|$$

denote the weight of the level-k Fourier coefficients of a Boolean function f, and let

$$W(d,k) := \max\{W_k(f) : f = (-1)^p, \ \deg(p) \le d\}$$

be the maximum of W_k over degree d polynomials. Note that we do not make any assumption on the number of variables n. We prove the following lemma from which Theorem 6.1 follows relatively easily.

Lemma 6.2. *For any $d, k \ge 1$,*

$$W(d,k)^2 \le 2^{2k} W(d-1, 2k) + W(d,k) \cdot \sum_{\ell=1}^{k} \binom{k}{\ell} W(d, k-\ell).$$

We first show that Theorem 6.1 follows easily from Lemma 6.2.

Proof of Theorem 6.1 given Lemma 6.2. The proof of Theorem 6.1 is by induction, first on d and then on k. The base case of $d = 1$ is straightforward, so assume $d \geq 2$. By Lemma 6.2 we have

$$\begin{aligned} W(d,k)^2 &\leq 2^{2k}\left(2k \cdot 2^{3(d-1)}\right)^{2k} + W(d,k)\sum_{\ell=1}^{k}\binom{k}{\ell}\left((k-\ell)2^{3d}\right)^{k-\ell} \\ &\leq \left(k \cdot 2^{3d-1}\right)^{2k} + W(d,k)\sum_{\ell=1}^{k}\binom{k}{\ell}\left((k-1)2^{3d}\right)^{k-\ell} \\ &= \left(k \cdot 2^{3d-1}\right)^{2k} + W(d,k)\left(\left((k-1)2^{3d}+1\right)^{k} - \left((k-1)2^{3d}\right)^{k}\right). \end{aligned}$$

Assume towards a contradiction that $W(d,k) > (k2^{3d})^k$. Dividing by $W(d,k)$ on both sides gives

$$W(d,k) \leq \left(k \cdot 2^{3d-1}\right)^{k} + \left((k-1)2^{3d}+1\right)^{k} - \left((k-1)2^{3d}\right)^{k}.$$

If $k = 1$ then we reach a contradiction as $2^{3d-1} + 1 \leq 2^{3d}$. If $k > 1$ then as $(k-1)2^{3d} \geq k2^{3d-1}$ the first term gets canceled by the third term, and the second term is at most $(k2^{3d})^k$. In either case, we reached a contradiction. □

From now on we focus on proving Lemma 6.2. To that end, fix f computed by a polynomial of degree d which maximizes $W_k(f)$. We shorthand $g(S) = |\hat{f}(S)|$. The following claims are used in the proof of Lemma 6.2.

Claim 6.3. *For any $0 \leq a < b \leq n$ and $A \subset [n]$ of size $|A| = a$,*

$$\sum_{B:|B|=b, A\subset B} g(B) \leq W(d, b-a).$$

Claim 6.4.

$$\sum_{S,T:|S|=|T|=k, S\cap T=\emptyset} g(S)g(T) \leq 2^{2k}W(d-1, 2k).$$

Claim 6.5. *For any $1 \leq \ell \leq k$,*

$$\sum_{S,T:|S|=|T|=k, |S\cap T|=\ell} g(S)g(T) \leq \binom{k}{\ell}W(d,k)W(d,k-\ell).$$

We first show how to prove Lemma 6.2 using the above claims.

Proof of Lemma 6.2. We have,

$$\begin{aligned} W(d,k)^2 &= \sum_{S,T:|S|=|T|=k} g(S)g(T) \\ &= \sum_{\ell=0}^{k}\sum_{S,T:|S|=|T|=k, |S\cap T|=\ell} g(S)g(T) \\ &= \sum_{S,T:|S|=|T|=k, S\cap T=\emptyset} g(S)g(T) + \sum_{\ell=1}^{k}\sum_{S,T:|S|=|T|=k, |S\cap T|=\ell} g(S)g(T) \\ &\leq 2^{2k}W(d-1,2k) + W(d,k)\cdot\sum_{\ell=1}^{k}\binom{k}{\ell}W(d,k-\ell), \end{aligned}$$

where the last inequality follows by using the bounds from Claim 6.4 and Claim Claim 6.5. □

We now proceed to prove the missing claims.

Proof of Claim 6.3. We use induction on a and b. The claim is direct for $a = 0$ and any $b > a$. Thus suppose $b > a > 0$ and let $i \in A$. Let $A' = A \setminus \{i\}$. We have

$$\begin{aligned}\sum_{B:|B|=k, A\subset B} g(B) &= \sum_{B'\subset[n]\setminus\{i\}:|B'|=b-1, A'\subset B'} g(B'\cup\{i\}) \\ &= \sum_{B'\subset[n]\setminus\{i\}:|B'|=b-1, A'\subset B'} |\widehat{f}(B'\cup\{i\})|.\end{aligned}$$

Let $f_{i\to 1}$ and $f_{i\to -1}$ be the functions obtained from f by setting the i'th bit to 1 and -1, respectively. It is easy to verify that $|\widehat{f}(B\cup\{i\})| \le \frac{1}{2}(\widehat{f_{i\to 1}}(B) + \widehat{f_{i\to -1}}(B))$. Thus, continuing with our estimate, we have

$$\begin{aligned}\sum_{B:|B|=b, A\subset B} g(B) &\le \frac{1}{2} \sum_{B':|B'|=b-1, A'\subset B'\cup\{i\}} \left(|\widehat{f_{i\to 1}}(B')| + |\widehat{f_{i\to -1}}(B')|\right) \\ &\le W(d, (b-1)-(a-1)) = W(d, b-a),\end{aligned}$$

where the last inequality follows from induction hypothesis. □

Proof of Claim 6.4. For any $S \subset [n]$, let $e_S \in \{-1, 1\}$ be the sign of $\widehat{f}(S)$, so that $g(S) = e_S \cdot \widehat{f}(S)$. Let X, Y, Z be independent uniform distributions on $\{-1, 1\}^n$. We have

$$\begin{aligned}\sum_{S,T:|S|=|T|=k, S\cap T=\emptyset} g(S)g(T) &= \sum_{S,T:|S|=|T|=k, S\cap T=\emptyset} e_S e_T \mathbb{E}_Y[f(Y)Y^S] \cdot \mathbb{E}_Z[f(Z)Z^T] \\ &= \sum_{S,T:|S|=|T|=k, S\cap T=\emptyset} e_S e_T \mathbb{E}_{Y,Z}[f(Y)Y^S f(Z)Z^T] \\ &= \sum_{S,T:|S|=|T|=k, S\cap T=\emptyset} e_S e_T \mathbb{E}_{X,Y,Z}[f(X\circ Y)f(X\circ Z)X^{S\cup T}Y^S Z^T].\end{aligned}$$

This follows as (Y, Z) and $(X \circ Y, X \circ Z)$ are identically distributed. Now consider any fixing of $Y = y$ and $Z = z$. Define the function $h_{y,z}(x) = f(x \circ y)f(x \circ z)$. Recall that $f = (-1)^p$ where p is a $\mathbb{F}_2$-polynomial of degree d. Thus $h = (-1)^q$ where q is the derivative of f in direction $y \circ z$. In particular, its degree is at most $d - 1$. Thus we have

$$\begin{aligned}\sum_{S,T:|S|=|T|=k, S\cap T=\emptyset} e_S e_T y^S z^T \mathbb{E}[f(X\circ y)f(X\circ z)X^{S\cup T}] &\le \binom{2k}{k} \sum_{R:|R|=2k} \left|\mathbb{E}[h(X)X^R]\right| \\ &\le 2^{2k} W(d-1, 2k).\end{aligned}$$

The proof follows now by noting that the above bound holds for any choice of y and z, and then averaging over $y = Y, z = Z$. □

Proof of Claim 6.5. We have,

$$\begin{aligned}\sum_{S,T:|S|=|T|=k, |S\cap T|=\ell} g(S)g(T) &\le \sum_{L:|L|=\ell} \Big(\sum_{S:|S|=k, L\subset S} g(S)\Big)^2 \\ &\le \Big(\max_{L:|L|=\ell} \sum_{S:|S|=k, L\subset S} g(S)\Big)\Big(\sum_{L,S:|L|=\ell, |S|=k, L\subset S} g(S)\Big) \\ &\le W(d, k-\ell) \cdot \Big(\sum_{S:|S|=k} \sum_{L:L\subset S, |L|=\ell} g(S)\Big) \qquad \text{(using Claim 6.3)} \\ &\le W(d, k-\ell) \cdot \binom{k}{\ell} \cdot W(d, k).\end{aligned}$$

□

7 Smoothness

In this section we provide a partial answer for Open problem 1.5, regarding early termination of the random walk. Let $Y_t \in [-1,1]^n$ be the location of the random walk at time t. We would like to guarantee that if Y_t is close enough to $\text{sign}(Y_t)$ then we can round Y_t to $\text{sign}(Y_t)$ without changing the value of f by too much. Therefore, given $f : [-1,1]^n \to [-1,1]$, it would be desirable to show f is "smooth" enough: there is a bound W such that

$$\forall \alpha, \beta \in [-1,1]^n, \ |f(\alpha) - f(\beta)| \le W \|\alpha - \beta\|_\infty. \tag{1}$$

Observe that should such W exists, then if at some step t we have $\|Y_t - \text{sign}(Y_t)\|_\infty \le \varepsilon/W$, then we can terminate the random walk immediately and guarantee that $\|f(Y_t) - f(\text{sign}(Y_t))\|_\infty \le \varepsilon$. We show that such smoothness property holds for functions with bounded sensitivity.

Bounded sensitivity functions. To show Equation (1), first consider the case that $\|\alpha - \beta\|_\infty$ is very small.

Claim 7.1. *Let $f : [-1,1]^n \to [-1,1]$ be of maximum sensitivity s. Let $\alpha, \beta \in [-1,1]^n$ such that $\|\alpha - \beta\|_\infty \le 1/n^2$. Then*

$$|f(\alpha) - f(\beta)| \le 4s\|\alpha - \beta\|_\infty.$$

Proof. Let $\delta = \|\alpha - \beta\|_\infty$. We first consider the easier case of $\alpha \in \{-1,1\}^n$. Pick $b \in \{-1,1\}^n$ randomly by flipping each coordinate of α independently with probability $|\alpha_i - \beta_i|/2$ so that $\mathbb{E}f(b) = f(\beta)$. Note that $f(b) \neq f(\alpha)$ if either exactly one sensitive coordinate of α is flipped, which occurs with probability at most $s\delta$, or if at least two coordinates get flipped, which occurs with probability at most $(n\delta)^2$. Therefore

$$|f(\alpha) - \mathbb{E}f(b)| \le s\delta + n^2\delta^2 \le 2s\delta$$

given our assumption on δ.

Next, consider the general case of $\alpha \in [-1,1]^n$. This case requires introducing an extra point $\gamma \in [-1,1]^n$ in a way that allows us to prove

$$|f(\alpha) - f(\gamma)| \le 2s \cdot \|\alpha - \gamma\|_\infty \tag{2}$$

and

$$|f(\beta) - f(\gamma)| \le 2s \cdot \|\beta - \gamma\|_\infty \tag{3}$$

separately. We choose γ in a way that $\forall i \in [n], \gamma_i = \alpha_i$ or $\gamma_i = \beta_i$. These equations altogether give the claim. To choose γ, let $S \subset [n]$ be the set of coordinates that $|\alpha_i| < |\beta_i|$ and pick $\gamma_i = \alpha_i$ if $i \in S$, and $\gamma_i = \beta_i$ otherwise.

We next prove Equation (2). The proof of Equation (3) is analogous. Consider a joint random variable (a, c) that satisfies the following properties:

1. $a \in \{-1,1\}^n, c \in [-1,1]^n$, $\mathbb{E}a = \alpha$, and $\mathbb{E}c = \gamma$.

2. The marginal distributions of a and c are product distributions.

3. $\|a - \mathbb{E}_c[c|a]\|_\infty \le \|\alpha - \gamma\|_\infty$ holds with probability one.

Observe that given such (a, c),

$$|f(\alpha) - f(\gamma)| = |\mathbb{E}_{a,c}[f(a) - f(c)]| \le \mathbb{E}_a \left|f(a) - \mathbb{E}_c[f(c)|a]\right| \le 2s \cdot \|\alpha - \gamma\|_\infty,$$

where the last inequality uses the first case in the proof, as $a \in \{-1,1\}^n$.

Now let us construct the joint random variable (a, c). Fix $i \in [n]$ and suppose without loss of generality that $\alpha_i \ge 0$. Note that by construction $-\alpha_i \le \gamma_i \le \alpha_i$. First sample a_i so that $\mathbb{E}[a_i] = \alpha_i$. If $a_i = -1$ then set $c_i = -1$, otherwise set $c_i = \frac{2\gamma_i + 1 - \alpha_i}{1 + \alpha_i}$. It's easy to check that this choice of (a, c) satisfies the required conditions, finishing the proof.

□

To prove the result for arbitrary $\alpha, \beta \in [-1,1]^n$ using Claim 7.1, consider the line segment from α to β and integrate f along that line segment. Thus we obtain the following lemma.

Lemma 7.2. *If $f : [-1,1]^n \to [-1,1]$ is multilinear and has maximum sensitivity s. Then for any $\alpha, \beta \in [-1,1]^n$ it holds that*

$$|f(\alpha) - f(\beta)| \leq 4s\|\alpha - \beta\|_\infty.$$

Acknowledgments

We thank Avi Wigderson and David Zuckerman for various stimulating discussions during the course of our work.

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[AW85] Miklós Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits. In *Foundations of Computer Science, 1985., 26th Annual Symposium on*, pages 11–19. IEEE, 1985.

[Ban10] Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 3–10. IEEE, 2010.

[Bra10] Mark Braverman. Polylogarithmic independence fools ac 0 circuits. *Journal of the ACM (JACM)*, 57(5):28, 2010.

[CHRT17] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *Electronic Colloquium on Computational Complexity (ECCC), pages TR17–171*, 2017.

[GMR$^+$12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 120–129. IEEE, 2012.

[GSW16] Parikshit Gopalan, Rocco A Servedio, and Avi Wigderson. Degree and sensitivity: tails of two distributions. In *Proceedings of the 31st Conference on Computational Complexity*, page 13. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016.

[Has86] Johan Hastad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 6–20. ACM, 1986.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[HT17] Pooya Hatami and Avishay Tal. Pseudorandom generators for low-sensitivity functions. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 24, page 25, 2017.

[LM12] Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 61–67. IEEE, 2012.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.

[LTZ16] Shachar Lovett, Avishay Tal, and Jiapeng Zhang. Robust sensitivity. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, page 161, 2016.

[Man95] Yishay Mansour. An $n^{O(\log\log n)}$ learning algorithm for dnf under the uniform distribution. *Journal of Computer and System Sciences*, 50(3):543–550, 1995.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[NW88] Noam Nisan and Avi Wigderson. Hardness vs. randomness. In *Foundations of Computer Science, 1988., 29th Annual Symposium on*, pages 2–11. IEEE, 1988.

[Raz86] AA Razborov. Lower bounds on the size of bounded depth networks over a complete basis with logical addition, mathematische zametki 41 pp. 598–607. *English Translation inMathematical Notes of the Academy of Sciences of the USSR*, 41:333–338, 1986.

[RSV13] Omer Reingold, Thomas Steinke, and Salil Vadhan. Pseudorandomness for regular branching programs via fourier analysis. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 655–670. Springer, 2013.

[Sim83] Hans-Ulrich Simon. A tight ω (loglog n)-bound on the time for parallel ram's to compute nondegenerated boolean functions. In *International Conference on Fundamentals of Computation Theory*, pages 439–444. Springer, 1983.

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

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of ac0. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 79. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[TX13] Luca Trevisan and Tongke Xue. A derandomized switching lemma and an improved derandomization of ac0. In *Computational Complexity (CCC), 2013 IEEE Conference on*, pages 242–247. IEEE, 2013.

[Vio09] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Computational Complexity*, 18(2):209–217, 2009.