An Exposition of "Fractional Pseudorandom Generators from Any Fourier Level"

Alek Westover, Addison

March 2024

Abstract

In this note we give an exposition of Chattopadhyay et al's research [Cha+20] on "Fractional Pseudorandom Generators from Any Fourier Level". A major goal in complexity theory is showing that certain computations utilizing randomness can be derandomized with small overhead. A PRG for a family of functions $\mathcal{F}$ with error ε is a random sequence of n bits X such that for any $f \in \mathcal{F}$ we have $\left|\mathbb{E}_{x\sim X} f(x) - \mathbb{E}_{y\sim\{-1,1\}^n} f(y)\right| \leq \varepsilon$. The quality of a PRG can be measured by its seed length: the number of bits of randomness needed to generate an output of the PRG.

The main result of [Cha+20] is to give a PRG for the class of functions which have bounded Fourier mass on *some* level. Previous work showed how to give a PRG for the smaller class of functions with bounded Fourier mass on *all* levels. The benefit of having a PRG for function classes satisfying this weaker condition is that, for a given function class of interest, it is easier to show that the function class has bounded Fourier mass on *some* level than it is to show it has bounded Fourier mass on *all* levels.

[Cha+20] apply their result to the problem of making a PRG for low-degree polynomials over $\mathbb{F}_2$, an important open question in complexity theory. Their PRG recovers a nearly state-of-the-art PRG for this class. [Cha+20] also show that to make further progress on the problem of creating PRGs for low-degree polynomials over $\mathbb{F}_2$ it suffices to analyze the correlation of such polynomials with "shifted majorities".

1 Introduction

Derandomization is a central topic in complexity theory. Famous conjectures such as $\mathsf{P} = \mathsf{BPP}$ and $\mathsf{L} = \mathsf{BPL}$ assert that certain classes of randomized algorithms can be derandomized with minimal overhead. Resolving these conjectures remains out of reach, but there has been lots of progress on derandomization simpler classes of functions.

An important idea for performing derandomization is a ***PRG***. Intuitively, a PRG is a low-entropy substitute X for a uniformly random string U_n such that an algorithm f from the class of functions performs similarly to f operating on U_n. Suppose that the PRG X can be deterministically generated from $s \ll n$ bits of true randomness; s is called the ***seed length*** of the PRG. Then, instead of running f on a random input we iterate over all 2^s possible seeds and run f on the output of the PRG with each of the seeds. If f succeeds with good probability when evaluated at a random output of the PRG, then enumerating over all outputs of the PRG ensures that f will succeed at least once. Formally a PRG is defined as follows:

Definition 1.1. A PRG for a class $\mathcal{F}$ of functions $f : \{-1,1\}^n \to \mathbb{R}$ with error $\varepsilon > 0$ is a random variable $X \in \{-1,1\}^n$ such that for all $f \in \mathcal{F}$:

$$|\mathbb{E}_{x\sim X} f(x) - \mathbb{E}_{y\sim U_n} f(y)| \leq \varepsilon.$$

We say X ***fools*** $\mathcal{F}$ with error ε. If $X = G(U_s)$ for a deterministic function G then we say X has ***seed length*** s.

In this paper we will be concerned with classes of functions $\mathcal{F}$ which have bounded Fourier mass on some level. This includes many natural families of functions such as small-depth circuits, low-sensitivity functions, read-once branching programs, and low-degree polynomials.

1.1 The Polarizing Random Walk Framework

The polarizing random walk framework is a technique introduced in [Cha+19a] for creating PRGs. The technique is to first create a "fractional PRG" (***FPRG***) and then convert the FPRG into a PRG. An FPRG is a generalization of a PRG that is allowed to take on values in $[-1,1]^n$. We say an FPRG X ***fools*** $\mathcal{F}$ with error ε if

$$|\mathbb{E}_{x\sim X} f(x) - \mathbb{E}_{y\sim U_n} f(y)| \leq \varepsilon,$$

where f now denotes the multilinear extension of the original boolean f. In order to convert an FPRG into a PRG it is also important that the coordinates of the FPRG have high variance. Towards this end we say a FPRG X is p-noticeable if, for each $i \in [n]$, we have $\mathbb{E}[X_i^2] \geq p$.

Intuitively, FPRGs are more convenient to construct than PRGs because we can directly use analytical tools such as Taylor's theorem to analyze multilinear polynomials used in constructing FPRGs.

The high level idea for converting an FPRG to a PRG is as follows: Let f be the function we are trying to fool, and X be an FPRG which fools f. Then, by definition $\mathbb{E}_{x\sim X} f(x)$ is close to $\mathbb{E}_{y\sim U_n} f(y)$. Now, we sample $z \sim X$ and consider the shifted function $f(x+z)$. It turns out that we can think of $f(x+z)$ as being a version of f where we have taken a restriction and then biased some coordinates. Then, we will repeat this process, accumulating more bias at each step until we get close to a corner of the hypercube. The fact that X is an FPRG means that we accumulate only a small amount of error at each step. If the FRPG X satisfies the p-noticable condition, i.e., each of its coordinates has large variance, then the number of steps that we must take before this process reaches the boundary of the cube is small, so we will accumulate only a small total amount of error.

1.2 Results

First we give a high-level overview of [Cha+20] 's results.

In Section 3 [Cha+20] prove Theorem 1.2: if we have a bound on the Fourier mass of f at some level k then $|f_{\leq k}(x) - f(x)|$ is small for all $x \in [-c,c]^n$ for appropriate $c \in (0,1)$. Their proof technique is to use Taylor's theorem to isolate the dependence of f on high degree Fourier coefficients. This is the main result that enables the construction of their FPRG.

In Section 4 [Cha+20] show how to use Theorem 1.2 to construct an FPRG, and then how to convert this FPRG into a PRG using the polarizing random walk technique. Their FPRG is simply a $(k-1)$-wise independent vector $X \in \{-c,c\}^n$ for suitable $c \in (0,1/2]$, which is successful because f is well approximated by a low-degree function on the subcube $[-c,c]^n$, and low-degree functions behave well on inputs that exhibit limited independence.

In Section 5 [Cha+20] demonstrate that their PRG based on a bound on Fourier mass at some level gives a new method for obtaining a nearly state-of-the-art PRG for the class of low-degree polynomials over $\mathbb{F}_2$. They conclude in Section 6 by showing that in order to obtain a major break-through in PRGs for low-degree polynomials it would suffice to bound the correlation between such polynomials and "shifted majorities". They pose proving this as an important open question.

Now we more formally state the theorems that [Cha+20] prove. Define the ***level-k absolute fourier sum*** of f as

$$M_k(f) = \max_{x\in\{-1,1\}^n} \left| \sum_{|S|=k} \hat{f}(S)\chi_S(x) \right|,$$

and define the ***level-k L_1 Fourier mass*** of f to be

$$L_{1,k}(f) = \sum_{|S|=k} |\hat{f}(S)|.$$

Define $L_{1,k}(\mathcal{F}), M_k(\mathcal{F})$ as the maximum over $f \in \mathcal{F}$ of $L_{1,k}(f), M_k(f)$ respectively. Observe that $M_k(\mathcal{F}) \leq L_{1,k}(\mathcal{F})$, and that $M_k(\mathcal{F})$ may be substantially smaller than $L_{1,k}(\mathcal{F})$. In other words, bounding $M_k(\mathcal{F})$ is potentially easier than bounding $L_{1,k}(\mathcal{F})$. In Section 3 [Cha+20] prove:

Theorem 1.2. Let $f \in \mathcal{F}$ with $\mathcal{F}$ closed under restrictions. Then for all $c \in (0,1)$, we have

$$\max_{x\in[-c,c]^n} |f_{\geq k}(x)| \leq (c/(1-c))^k M_k(\mathcal{F}).$$

This theorem means that, so long as c is bounded away from 1, $f_{<k}$ is a ***uniform*** (i.e., ℓ_∞) approximation to f on the subcube $[-c,c]^n$. They also give a lower bound, showing that Theorem 1.2 is essentially tight.

In Section 4 they apply Theorem 1.2 to construct a FPRG. The FPRG is simply a $(k-1)$-wise independent random vector $X \in \{-c, c\}^n$ for some appropriate $c \in (0, 1/2]$. That is, a random vector such that for any set $J \subseteq [n]$ of at most $(k-1)$ coordinates the set of random variables $\{X_j \mid j \in J\}$ is independent. Let χ_S be a monomial of degree $0 < |S| \leq k-1$. Then, $\mathbb{E}[\chi_S(X)] = 0$. This shows that $\mathbb{E} f_{<k}(X) = \mathbb{E} f(U_n)$. Thus, to show that X fools $\mathcal{F}$ with error ε it suffices to bound $f_{\geq k}(X)$, which is precisely what Theorem 1.2 does. Note that there is a tradeoff: smaller values for c make $f_{\geq k}(\pm c)$ smaller, but at the cost of causing X_i to have lower variance, making it harder to convert X into a PRG. Balancing these requirements they show the following theorem:

Theorem 1.3. Let $\mathcal{F}$ be a class of functions $f : \{-1, 1\}^n \to \mathbb{R}$ which is closed under restrictions. Suppose $\mathcal{F}$ satisfies $M_k(\mathcal{F}) \leq b^k$ for *some* $b \geq 1, k > 2$. Then, for any $\varepsilon > 0$ there exists an $\Omega(\varepsilon^{2/k}/b^2)$-noticable FPRG for $\mathcal{F}$ with error ε and seed length $\mathcal{O}(k \log n)$.

Next, using the random walk gadget from [Cha+19a] they convert this FPRG into a PRG:

Theorem 1.4. Let $\mathcal{F}$ be a class of functions $f : \{-1, 1\}^n \to \mathbb{R}$ which is closed under restrictions. Suppose $\mathcal{F}$ satisfies $M_k(\mathcal{F}) \leq b^k$ for some $b \geq 1, k > 2$. Then, for any $\varepsilon > 0$ there exists a PRG for $\mathcal{F}$ with error ε and seed length $kb^{2+4/(k-2)} \operatorname{polylog}(n/\varepsilon)/\varepsilon^{2/(k-2)}$

In Section 5 they apply Theorem 1.4 to give an alternate method for obtaining nearly state-of-the-art PRGs for low-degree polynomials over $\mathbb{F}_2$.

Theorem 1.5. Fix $d \in \mathbb{N}, \varepsilon > 0$. Let $\mathcal{F}$ be the class of degree-d polynomials over $\mathbb{F}_2$ in n-variables. There exists a PRG for $\mathcal{F}$ with error ε and seed length $2^{\mathcal{O}(d)} \operatorname{polylog}(n/\varepsilon)$.

Obtaining stronger bounds on $M_k(\mathcal{F})$, for $\mathcal{F}$ being low-degree polynomials over $\mathbb{F}_2$, would result in stronger PRGs for $\mathcal{F}$, a major open problem in complexity theory. In Section 6 they show that to prove bounds on $M_k(\mathcal{F})$ it would suffice to analyze the correlation between $f \in \mathcal{F}$ and "shifted majorities".

2 Preliminaries

In this section, we provide some preliminary tools that will be used quite often in our analysis. There are two main tools that we cover here, notations and lemmas for Fourier analysis on boolean functions and definitions and theorems about FPRGs. Furthermore, in this paper, the authors only focus on a family of functions $\mathcal{F}$ that are closed under restrictions.

2.1 Background on Fourier Analysis

Denote f_k as the level-k part of a polynomial (keeping only monomials of degree k), and for any $x, z \in \{\pm 1\}^n$, let $x \circ z$ be the entry-wise product. From here, we present the following lemma:

Lemma 2.1. Let $\mathcal{F}$ is a class of function that is closed under negation of the domain and image, i.e. if $f \in \mathcal{F}$ then $-f \in \mathcal{F}$, and if $f \in \mathcal{F}$ and $g(x) = f(x \circ z)$ for some $z \in \{\pm 1\}^n$ and any $x \in \{\pm 1\}^n$, then $g \in \mathcal{F}$. Then

$$M_k(\mathcal{F}) = \max_{f \in \mathcal{F}} \sum_{|S|=k} \hat{f}(S) = \max_{f \in \mathcal{F}} f_k(\mathbf{1}).$$

Proof of Lemma 2.1. For any $f \in \mathcal{F}$ and $z \in \{-1, 1\}^n$, let $f^z : \{-1, 1\}^n \to \{-1, 1\}^n$ be a function such that $f^z(x) = f(x \circ z)$ for all $x \in \{-1, 1\}^n$. By the assumption of this lemma, $f^z \in \mathcal{F}$. Therefore,

$$\begin{aligned}
M_k(\mathcal{F}) &= \max_{f \in \mathcal{F}} \max_{z \in \{-1,1\}^n} \left| \sum_{|S|=k} \hat{f}(S) z^S \right| \\
&= \max_{f \in \mathcal{F}} \max_{z \in \{-1,1\}^n} \left| \sum_{|S|=k} \hat{f^z}(S) \right| \\
&= \max_{f \in \mathcal{F}} \left| \sum_{|S|=k} \hat{f}(S) \right| \\
&= \max_{f \in \mathcal{F}} \sum_{|S|=k} \hat{f}(S) = f_k(\mathbf{1}).
\end{aligned}$$

□

Another notion we need to introduce is the convex closure of a function class:

Definition 2.2 (Convex Closure)**.** Let $\text{conv}(\mathcal{F})$ be the convex closure of $\mathcal{F}$, defined as follows,

$$\text{conv}(\mathcal{F}) := \left\{ \sum_{f \in \mathcal{F}} \lambda_f f \middle| \sum_{f \in \mathcal{F}} \lambda_f = 1, \lambda_f \geq 0, \forall f \in \mathcal{F} \right\}.$$

We address two important facts about the convex closure. The first is that if $\mathcal{F}$ is closed under restriction, then so is $\text{conv}(\mathcal{F})$, which is trivial. The second is the following lemma:

Lemma 2.3.

$$M_k(\mathcal{F}) = M_k(\text{conv}(\mathcal{F})).$$

Proof of Lemma 2.2. Trivially, we get that $\mathcal{F} \subset \text{conv}(\mathcal{F})$, so $M_k(\mathcal{F}) \leq M_k(\text{conv}(\mathcal{F}))$. On the other hand, for any collection of non-negative coefficients λ_f indexed by elements of $\mathcal{F}$ such that its sum is one, if we denote $g = \sum_{f \in \mathcal{F}} \lambda_f f \in \text{conv}(\mathcal{F})$, we would have

$$\begin{aligned}
M_k(g) = \max_{x \in \{-1,1\}^n} \left| \sum_{|S|=k} \hat{g}(S) x^S \right| & \\
&= \max_{x \in \{-1,1\}^n} \left| \sum_{|S|=k} \sum_{f \in \mathcal{F}} \lambda_f \hat{f}(S) x^S \right| \\
&\leq \max_{x \in \{-1,1\}^n} \sum_{f \in \mathcal{F}} \lambda_f \left| \sum_{|S|=k} \hat{f}(S) x^S \right| \\
&\leq \max_{f \in \mathcal{F}} \max_{x \in \{-1,1\}^n} \left| \sum_{|S|=k} \hat{f}(S) x^S \right| \\
&= M_k(\mathcal{F}),
\end{aligned}$$

so $M_k(\text{conv}(\mathcal{F})) \leq M_k(\mathcal{F})$, making $M_k(\mathcal{F}) = M_k(\text{conv}(\mathcal{F}))$. □

2.2 Background on FPRGs

We address the definition of PRG and FPRG as defined on [Cha+19a]:

Definition 2.4 (Pseudorandom Generator (PRG))**.** Let $\mathcal{F}$ be a class of n-variate boolean functions. A PRG on $\mathcal{F}$ with error $\varepsilon > 0$ is a random variable $\mathbf{X} \in \{-1,1\}^n$ such that

$$|\mathbb{E}_{\mathbf{X}}[f(\mathbf{X})] - \mathbb{E}_{U_n}[f(U_n)]| \leq \varepsilon, \forall f \in \mathcal{F},$$

where U_n is the uniform distribution over $\{-1,1\}^n$. If $\mathbf{X} = G(U_s)$ for some function $G : \{-1,1\}^s \to \{-1,1\}^n$, then $\mathbf{X}$ has seed length s.

Definition 2.5 (Fractional Pseudorandom Generator (FPRG))**.** Let $\mathcal{F}$ be a class of n-variate boolean functions. An FPRG on $\mathcal{F}$ with error $\varepsilon > 0$ is a random variable $\mathbf{X} \in [-1,1]^n$ such that

$$|\mathbb{E}_{\mathbf{X}}[f(\mathbf{X})] - f(\mathbf{0})| \leq \varepsilon, \forall f \in \mathcal{F}.$$

The definition of seed length is analogous to Definition 2.4. Furthermore, $\mathbf{X}$ is p-noticable if $\mathbb{E}[\mathbf{X}_i^2] \geq p$ for all $i \in [n]$.

Using these two definitions, we present two results from [Cha+19a] and [Cha+19b] about how to construct an explicit PRG given an FPRG, or given some bounds on the L_1 Fourier masses.

Theorem 2.6 (Theorem 7 of [Cha+19b])**.** Given a class of functions $\mathcal{F}$ closed under restrictions and a p-noticable FPRG X with error ε and seed length s, one can construct an explicit PRG for $\mathcal{F}$ with seed length $O(s \log(n/\varepsilon)/p)$ and error $O(\varepsilon \log(n/\varepsilon)/p)$.

Theorem 2.7 (From [Cha+19a])**.** Given a class of functions $\mathcal{F}$ closed under restrictions such that for some $b \geq 1$,

$$L_{1,k}(\mathcal{F}) \leq b^k, \forall 1 \leq k \leq n,$$

we have that for any $\varepsilon > 0$, there exists an explicit PRG for $\mathcal{F}$ with error ε and seed length $b^2 \operatorname{polylog}(n/\varepsilon)$.

In addition to these two theorems, [Cha+19b] proved the following result that relaxes the condition on the L_1 bounds, but worsens the guarantees on the seed length.

Theorem 2.8 (Lemma 8 on [Cha+19b])**.** Given a class of functions $\mathcal{F}$ closed under restrictions such that for some $b \geq 1$, we have $L_{1,2}(\mathcal{F}) \leq b^2$, we have that for any $\varepsilon > 0$, there exists an explicit PRG for $\mathcal{F}$ with error ε and seed length $O((b^2/\varepsilon)^{2+o(1)} \operatorname{polylog}(n))$.

3 proof of Main Result (Theorem 1.2)

Main result is theorem about $\max_{x \in [-c,c]^n} |f_{\geq k}(x)|$. Remark: this is equivalent to a statement about doing noisy $T_c f$ approximation.

Some definitions:

$$\varepsilon_{c,k}(f) = \inf_{g | \deg(g) < k} \max_{x \in [-c,c]^n} |f(x) - g(x)|$$

$$\varepsilon_{c,k}(\mathcal{F}) = \max_{f \in \mathcal{F}} \varepsilon_{c,k}(f)$$

$$c_k(\varepsilon, \mathcal{F}) = \max \{c \geq 0 \mid \varepsilon_{c,k}(\mathcal{F}) \leq \varepsilon\}.$$

That is, $c_k(\varepsilon, \mathcal{F})$ is how small need to take a cube so that for every function $f \in \mathcal{F}$ there is a degree $k-1$ approximation of f with error ε.

By multilinearity it suffices for the property to hold at the extreme points of the cube.

We will show that $\varepsilon_{c,k}(\mathcal{F})$ can be bounded by some function of c, k and $M_k(\mathcal{F})$. Specifically we will show:

Theorem 1.2. Let $f \in \mathcal{F}$ with $\mathcal{F}$ closed under restrictions. Then for all $c \in (0,1)$, we have

$$\max_{x \in [-c,c]^n} |f_{\geq k}(x)| \leq (c/(1-c))^k M_k(\mathcal{F}).$$

An immediate corollary is:

Corollary 3.1. Let $\mathcal{F}$ be closed under restrictions. Fix $\varepsilon > 0, k \leq n$. Then

$$c_k(\varepsilon, \mathcal{F}) \geq \Omega((\varepsilon / M_k(\mathcal{F}))^{1/k})$$

We now prove some lemmas that will be useful in proving Theorem 1.2.

Lemma 3.2. Let $f \in \mathsf{conv}(\mathcal{F})$. Then, for all $c \in (0,1)$ and $x \in [-c,c]^n$ we have $|f_k(x)| \leq c^k M_k(\mathcal{F})$.

Proof. $c^{-1}x \in [-1,1]^n$. Because f_k is a sum of degree-k monomials we have:

$$|f_k(x)| = c^k |f_k(c^{-1}x)| \leq c^k M_k(\mathsf{conv}(\mathcal{F})) = c^k M_k(\mathcal{F}).$$

The last equality is non-obvious. It was a lemma in their preliminaries that $M_k(\mathsf{conv}(\mathcal{F})) = M_k(\mathcal{F})$. It's fairly straightforward to show. □

Now we have a lemma about "recentering functions" . This was also a key tool in [Cha+19a].

Lemma 3.3. $f \in \mathsf{conv}(\mathcal{F})$, a, b $\tilde{f}(x) = f(a + b \circ x)$. $\tilde{f} \in \mathsf{conv}(\mathcal{F})$

Proof. □

We will bound higher-order terms of Fourier expansion via error term in Taylor's theorem.

Lemma 3.4. $f : \mathbb{R}^n \to \mathbb{R}$ multilinear. Fix $x \in \mathbb{R}^n$. $g : \mathbb{R} \to \mathbb{R}$ defined as $g(t) = f(tx)$. Then $g^{(k)}(0) = k! \cdot f_k(x)$.

Proof. just calculus it. □

Lemma 3.5. $f \in \mathsf{conv}(\mathcal{F})$, $c \in (0,1), x \in [-c,c]^n$. $g(t) = f(tx)$. Then

$$\max_{t \in [0,1]} |g^{(k)}(t)| \leq (c/(1-c))^k k! M_k(\mathcal{F}).$$

Proof. Fix $s \in [0,1]$. Let $\lambda = 1 - c \in [0,1]$. Define $\tilde{f}(y) = f(sx + \lambda y)$. Write $a = sx, b = \lambda \mathbb{1}$. There was some lemma from earlier that now applies so we have $\tilde{f} \in \mathsf{conv}(\mathcal{F})$. Write $\tilde{g}(t) = \tilde{f}(tx)$. Turns out $\tilde{g}(t) = g(s + t\lambda)$

Chain rule, use some previous lemmas conclude. □

Now we prove the theorem:

Theorem 1.2. Let $f \in \mathcal{F}$ with $\mathcal{F}$ closed under restrictions. Then for all $c \in (0,1)$, we have

$$\max_{x \in [-c,c]^n} |f_{\geq k}(x)| \leq (c/(1-c))^k M_k(\mathcal{F}).$$

Proof. Define $g(t) = f(tx)$. Do a Taylor expansion around $t = 0$ for g. Then we have

$$g(1) = \sum_{i<k} \frac{g^{(i)}(0)}{i!} + R_k$$

where R_k is error term $R_k = g^{(k)}(s)/k!$ for some $s \in (0,1)$.

By Lemma 3.4 where we computed derivatives of g we can deduce that the left part of this is $f_{<k}$. Using Lemma 3.5 we can bound the error term in terms of $M_k(\mathcal{F})$.

□

Remark: then they use "chebyshev polynomials" to give a lower bound. This seems not super related to the rest of the paper so we skip it. Basically they just cite some fact about Chebyshev polynomials.

4 From Polynomial Approximations to PRGs

4.1 From Polynomial Approximations to FPRGs

Now we use our understanding of f on the subcube $[-c,c]^n$ to construct an FPRG.

Theorem 1.3. Let $\mathcal{F}$ be a class of functions $f : \{-1,1\}^n \to \mathbb{R}$ which is closed under restrictions. Suppose $\mathcal{F}$ satisfies $M_k(\mathcal{F}) \leq b^k$ for *some* $b \geq 1, k > 2$. Then, for any $\varepsilon > 0$ there exists an $\Omega(\varepsilon^{2/k}/b^2)$-noticable FPRG for $\mathcal{F}$ with error ε and seed length $\mathcal{O}(k \log n)$.

Proof. Fix $\varepsilon > 0$. Let $c \in (0, 1/2]$ be $c_k(\varepsilon/2, \mathcal{F})$, i.e., the smallest size c of hypercube that we can take such that for each $f \in \mathcal{F}$ there exists a degree-$(k-1)$ polynomial whose ℓ_∞ distance over $[-c,c]^n$ to f is at most $\varepsilon/2$. Let X be a $(k-1)$-wise independent random vector in $\{-c,c\}^n$. That is, for any set $I \subset [n]$ of size at most $k-1$ the values $\{X_i \mid i \in I\}$ are independent and uniformly distributed on $\{-c,c\}^I$. X can be generated with seed length $\mathcal{O}(k \log n)$ (see e.g., [Vad+12]). As intuition for why this is possible, suppose that n is a power of two. Take $a_0, a_1, \ldots, a_{k-2}$ uniformly at random from $\mathbb{F}_n$ (the finite field of order n, which exists as we are assuming for the moment that n is a power of 2). Construct the polynomial $p(x) = \sum_{i=0}^{k-2} a_i x^i$. Then, the points $\{p(x) \mid x \in \mathbb{F}_n\}$ form a $(k-1)$-wise independent vector in $\mathbb{F}_n^n$. Finally, this can be "squished" to a $(k-1)$-wise independent vector in $\{-c,c\}^n$. For a more formal treatment see [Vad+12].

We claim that X thus constructed is a FPRG that fools $\mathcal{F}$ with error ε. To verify this claim, we fix $f \in \mathcal{F}$ and bound the difference

$$\Delta = |\mathbb{E}_{x \sim X} f(x) - \mathbb{E}_{y \sim U_n} f(y)| = |\mathbb{E}_{x \sim X} f(x) - f(0)|.$$

Let $\tilde{f}$ be a degree-$(k-1)$ multilinear polynomial on $[-c,c]^n$ such that

$$\max_{y \in [-c,c]^n} \left| f(y) - \tilde{f}(y) \right| \leq \varepsilon/2;$$

Such an $\tilde{f}$ exists by definition of our choice of c. Because $\tilde{f}$ uniformly approximates f on $[-c,c]^n$ we have:

$$\Delta \leq \varepsilon/2 + \left| \mathbb{E}_{x \sim X} f(x) - \tilde{f}(0) \right|.$$

For any non-constant monomial χ_S of degree at most $k-1$ we have:

$$\mathbb{E}_{x \sim X} \chi_S(x) = 0;$$

This is because X is $(k-1)$-wise independent. Thus we have

$$\Delta \leq \varepsilon/2 + \left| \mathbb{E}_{x \sim X} [f(X) - \tilde{f}(X)] \right|.$$

Applying the triangle inequality gives

$$\Delta \leq \varepsilon/2 + \mathbb{E}_{x \sim X} \left| f(X) - \tilde{f}(X) \right|.$$

Now, again using the fact that $\tilde{f}$ uniformly approximates f on $[-c,c]^n$, and using the fact that $X \in [-c,c]^n$ we obtain:

$$\Delta \leq \varepsilon.$$

That is, X fools f with error ε, as desired.

□

[Cha+20] also shows that if you have bounds on $L_{1,i}(\mathcal{F})$ for all $i \in [k]$ then we can get an FPRG with a smaller seed length.

4.2 From FPRGs to PRGs

Now using the polarizing random walk framework of [Cha+19a] we fairly immediately can construct a PRG from this FPRG. Before doing so, we give some intuition for how the polarizing random walk framework works. The authors intuition for this framework is from personal communication with Dor Minzer as well as the preliminaries of [Cha+19a]. We aim to construct a random walk $Y_1, Y_2, \ldots, \in [-1,1]^n$ based on repeated samples of the FPRG that quickly converges to a point in $\{-1,1\}^n$, and such that the error accumulated at each step is small. Let $X_1, X_2, \ldots$ be independent samples of the FPRG. As our first step we take $Y_1 = X_1$. By definition of f being a FPRG we have that $\mathbb{E}f(Y_1)$ is close to $\mathbb{E}f(U_n) = f(0)$. To describe the future steps, define $\delta(y) \in [0,1]^n$ by $(\delta(y))_i = 1 - |y_i|$. Then, our future steps are defined as:

$$Y_i = Y_{i-1} + \delta(Y_{i-1}) \circ X_i.$$

If we ignore the $\delta(Y_{i-1})$ term, then this simply says that we step by X_i each time. The p-noticability condition on the FPRG will then make this random walk quickly reach the boundary of the hypercube. The $\delta(Y_{i-1})$ term is to penalize coordinates which are already close ± 1, i.e., to make them update more slowly. The key property of this random walk is that we accumulate very error at each step of this random walk. More precisely, we have the following lemma:

Lemma 4.1. Fix $y \in [-1,1]^n$. Then,

$$|\mathbb{E}_{x \sim X} f(y + \delta(y) \circ x) - f(y)| \leq \varepsilon.$$

Proof. Define random variable $R(x) \in \{-1,1\}^n$ as follows: For each i independently, set $R_i(x) = \operatorname{sgn}(y_i)$ with probability $|y_i|$ and otherwise pick $R_i(x) = 1$ with probability $(x_i + 1)/2$ and $R_i(x) = -1$ otherwise. Then,

$$\mathbb{E}_R[R_i(x)] = |y_i| \operatorname{sgn}(y_i) + (1 - |y_i|)x_i.$$

Thus,

$$\mathbb{E}_R[R(x)] = y + \delta(y) \circ x.$$

Because f is multilinear we have that

$$\mathbb{E}_R f(R(x)) = f(\mathbb{E}_R R(x)) = f(y + \delta(y) \circ x).$$

Because $\mathcal{F}$ is closed under restrictions the function $x \mapsto f(R(x))$ is also in $\mathcal{F}$. Thus we have

$$|\mathbb{E}_X f(y + \delta(y) \circ X) - f(y)| = |\mathbb{E}_X \mathbb{E}_R f(R(X)) - \mathbb{E}_R f(R(0))| \leq \mathbb{E}_R |f(R(X)) - f(R(0))| \leq \varepsilon.$$

□

We now comment on the intuition behind this proof. We can interpret $x \mapsto f(y + x)$ as being a biased version of f, where coordinate i is biased by amount y_i. For instance, if $y_1 = 0.3$ and x is sampled randomly then $y_1 + x_1$ is likely closer to 1 than to -1. We can think of this biased version of f as an average of restrictions of f. To see why this is the case it is helpful to consider the following instructive example: Suppose we want to generate a biased distribution on $\{-1,1\}^n$ where the i-th coordinate is $+1$ with probability $1/2 + \delta$ and -1 otherwise. We can accomplish this as follows: start by randomly choosing a subset of the coordinates to set to $+1$, where each coordinate is included in this subset with probability δ. Now, assign the remaining coordinates uniformly at random to be ± 1. This shows how bias can be interpreted as an average of random restrictions.

The quick convergence of the random walk is guaranteed by the lower bound on the variance of each coordinate. Now we are ready to construct the PRG.

Theorem 1.4. Let $\mathcal{F}$ be a class of functions $f : \{-1,1\}^n \to \mathbb{R}$ which is closed under restrictions. Suppose $\mathcal{F}$ satisfies $M_k(\mathcal{F}) \leq b^k$ for some $b \geq 1, k > 2$. Then, for any $\varepsilon > 0$ there exists a PRG for $\mathcal{F}$ with error ε and seed length $kb^{2+4/(k-2)} \operatorname{polylog}(n/\varepsilon)/\varepsilon^{2/(k-2)}$

Proof. We apply Theorem 2.6 which was proven in [Cha+19a]. Specifically, Theorem 2.6 allows us to convert an FPRG with error δ, p-noticability, and seed length s into a PRG with error $\mathcal{O}(\delta \log(n/\delta)/p)$ and seed length $\mathcal{O}(s \log(n/\delta))/p)$. Setting

$$\delta = \Theta\left(\left(\frac{\varepsilon}{b^2 \log(n/\varepsilon)}\right)^{k/(k-2)}\right),$$

and applying Theorem 1.3 we achieve the desired conclusion. □

5 Low Degree Polys over $\mathbb{F}_2$

In Section 4 we have seen how to construct a PRG for a class of functions closed under restrictions just based on the function family having bounded fourier mass at some level. This gives us greater flexibility compared to prior work at applying this theorem to giving PRGs for natural function families of interest. In this section we apply the theorems of Section 4 to the class of low-degree polynomials over $\mathbb{F}_2$. This comes close to recovering the guarantees of Viola's state-of-the-art PRG [Vio09] for this family of functions, and the technique seems fundamentally new.

[Cha+19a] proved the following bound on the fourier mass of low-degree polynomials:

Lemma 5.1. Let $p : \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree-d polynomial, and let $f(x) = (-1)^{p(x)}$. Then

$$L_{1,k}(f) \leq (k \cdot 2^{3d})^k.$$

Applying this mass bound in a version of Theorem 1.4 that requires a mass bound up until a certain level rather than a mass bound at a specific level we obtain:

Theorem 5.2. Let $\mathcal{F}$ be the class of degree-d polynomials over $\mathbb{F}_2$ on n variables. There exists a PRG for $\mathcal{F}$ with error ε and seed length

$$2^{O(d)} \log^3(\log(n)/\varepsilon) \log(n/\varepsilon).$$

We omit the proof which is simply setting parameters appropriately as dictated by Lemma 5.1 in the version of Theorem 1.4 that we have also omitted.

One potential attempt at obtaining stronger PRGs for this class of functions is to prove stronger bounds on $M_k(\mathcal{F})$. Note that this is potentially easier than proving bounds for $L_{1,k}(\mathcal{F})$. The authors offer the following conjecture about a potential bound on $M_k(\mathcal{F})$:

Conjecture 5.3. Let $\mathcal{F}$ be the class of degree-d polynomials over $\mathbb{F}_2$ in n variables. Then

$$M_k(\mathcal{F}) \leq (\text{poly}(k, \log n) 2^{o(d)})^k.$$

In the next section the authors show that bounding $M_k(\mathcal{F})$ can be done by analyzing "correlations with shifted majorities".

6 Bounds on $M_k(\mathcal{F})$ via Correlation with Shifted Majorities

Now we are going to switch to $\{0,1\}$ notation. Sigh.

Define $e(f)(x) = (-1)^{f(x)}$. Let $F = e(f)$. Then $\hat{F}(S) = \mathbb{E}_x[F(x)e(\sum_{i \in S} x_i)]$.

Define covariance as follows:

$$\mathsf{cov}(f, g) = \mathbb{E}[e(f(x))e(g(x))] - \mathbb{E}[e(f(x))e(g(x))].$$

Let $|x|$ for $x \in \{0,1\}^n$ denote hamming weight: $\sum x_i$. Define $\mathsf{Maj}_a(x)$ as : $\mathbb{1}_{|x|>a}$. Define Thr_θ BLAH

The main result of this section is that a bound on covariance between f and shifted majorities would imply a bound on $M_k(\mathcal{F})$ (which we care about for PRG applications). The lemma is:

Lemma 6.1. Suppose BLHA BLAH we could show a bound on

$$\mathsf{cov}(f, \bigoplus_{i=1}^{k} \mathsf{Maj}_{a_i}).$$

Then we would get

$$M_k(\mathcal{F}) \leq \mathcal{O}(\sqrt{tk \log n})^k.$$

Proof. Some complicated computations. □

7 Conclusion

The main open problem that they're excited about is the Conjecture about correlations which would imply some cool PRG for low-degree polynomials.

References

[Cha+19a] Eshan Chattopadhyay et al. "Pseudorandom Generators from Polarizing Random Walks". In: *Theory of Computing* 15.10 (2019), pp. 1–26. DOI: 10.4086/toc.2019.v015a010. URL: https://theoryofcomputing.org/articles/v015a010.

[Cha+19b] Eshan Chattopadhyay et al. "Pseudorandom Generators from the Second Fourier Level and Applications to AC0 with Parity Gates". In: *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*. Ed. by Avrim Blum. Vol. 124. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019, 22:1–22:15. ISBN: 978-3-95977-095-8. DOI: 10.4230/LIPIcs.ITCS.2019.22. URL: https://drops-dev.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2019.22.

[Cha+20] Eshan Chattopadhyay et al. *Fractional Pseudorandom Generators from Any Fourier Level*. 2020. arXiv: 2008.01316 [cs.CC].

[Vad+12] Salil P Vadhan et al. "Pseudorandomness". In: *Foundations and Trends in Theoretical Computer Science* 7.1–3 (2012), pp. 1–336.

[Vio09] Emanuele Viola. "The sum of d small-bias generators fools polynomials of degree d". In: *Computational Complexity* 18.2 (2009), pp. 209–217.