A Pseudorandom Generator for Polynomial Threshold Functions with Subpolynomial Seed Length

Daniel M. Kane

Department of Mathematics
Stanford University
dankane@math.stanford.edu

June 13th, 2014
We briefly recall some basic definitions:

**Definition**

We call a function \( f : \mathbb{R}^n \to \mathbb{R} \) a (degree-\( d \)) **Polynomial Threshold Function** (or PTF) if it is of the form \( f(x) = \text{sgn}(p(x)) \) for \( p \) a (degree-\( d \)) polynomial in \( n \) variables.

**Definition**

For \( p : \mathbb{R}^n \to \mathbb{R} \) and define

\[
|p|_2 := \left( \mathbb{E}_{X \sim G^n} \left[ |p(X)|^2 \right] \right)^{1/2}.
\]
Pseudorandom Generators

Definition
Given a class $C$ of functions $f : \mathbb{R}^n \to \mathbb{R}$, and a probability distribution $D$ on $\mathbb{R}^n$ we say that another distribution $B$ on $\mathbb{R}^n$ $\epsilon$-fools $C$ with respect to $D$ if for every $f \in C$,

$$|\mathbb{E}_{X \sim D}[f(X)] - \mathbb{E}_{Y \sim B}[f(Y)]| \leq \epsilon.$$ 

Definition
We say that the probability distribution $B$ is a Pseudorandom Generator (PRG) for $C$ with respect to $D$ if it can be produced by a polynomial time randomized algorithm using few random bits.

We will produce a small-seed PRG to $\epsilon$-fool the class of degree-$d$ PTFs with respect to the $n$-dimensional Gaussian (or Bernoulli) distribution.
PRGs from $k$-independence

Recall that a random variable is $k$-wise independent if any $k$ of its coordinates are independent. There are good constructions of $k$-wise independent variables from small seeds, and they can be used as PRGs for PTFs.

- [Diakonikolas-Gopalan-Jaiswal-Servedio-Viola, 2010] $k = \tilde{O}(\epsilon^{-2})$-independence fools degree-1 PTFs
- [Diakonikolas-K.-Nelson 2010] $k = O(\epsilon^{-8})$-independence fools degree-2 PTFs
- [K. 2011] $k = O_d(\epsilon^{-2^{O(d)}})$-independence fools degree-$d$ PTFs
- Best lower bound that I know: $k = \Omega(d^2\epsilon^{-2})$.
- I suspect that the lower bound is tight.
Other PRGs

- [Meka-Zuckerman 2010] $O(d \log(n) + \log(1/\epsilon))$ (Existential)
- [Meka-Zuckerman 2010] $\log(n)2^{O(d)}\epsilon^{-8d-3}$ (Bernoulli)
- [K. 2011] $\log(n)2^{O(d)}\epsilon^{-4.1}$ (Gaussian)
- [K. 2012] $\log(n)O_d(\epsilon^{-11.1})$ (Bernoulli)
- [K. 2012] $\log(n)O_d(\epsilon^{-2.1})$ (Gaussian)

In this talk we discuss the structure and analysis of a generator with seed length subpolynomial in the error parameter. Namely, seed length

$$\log(n)O_{c,d}(\epsilon^{-c}).$$
General Construction Idea

Basic idea for the above: Combine a bunch of independent copies of a weak PRG.

Bernoulli case:
- Split coordinates into $M$ bins in a 2-independent fashion.
- Fill each bin using a $k$-independent generator.

Gaussian Case:
- $Y_i$ are $k$-independent Gaussians chosen independently.
- $Y = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} Y_i$. 

The Replacement Method

These generators can be analyzed using Lindeberg’s replacement method.

- Approximate $f$ by smooth function $g$
- Show $\mathbb{E}[g(X)] \approx \mathbb{E}[g(Y)]$
- Replacing $X_i$ by $Y_i$ introduces small error, since low-order moments agree, and $g$ is approximated by Taylor polynomial.
New Idea

- Replacement method loses out by needing to approximate by smooth function.
- Hope to show

$$\mathbb{E}_{X,Y} \left[ f \left( \sqrt{\frac{M-1}{M}} X + \frac{1}{\sqrt{M}} Y \right) \right]$$

is approximately determined by $k$-independence.
The Degree 1 Case

We begin by seeing how this works in the degree 1 case. Let

\[ f(x) = \text{sgn}(\nu \cdot x + \theta) \]

for some vector \( \nu \) with \( |\nu| = 1 \) and some \( \theta \in \mathbb{R} \). For fixed \( Y \), we have

\[
\mathbb{E}_X \left[ f \left( \sqrt{\frac{M-1}{M}} X + \frac{1}{\sqrt{M}} Y \right) \right] \\
= \mathbb{E}_X \left[ \text{sgn} \left( \nu \cdot X + \frac{1}{\sqrt{M-1}} \nu \cdot Y + \sqrt{\frac{M}{M-1}} \theta \right) \right] \\
= \text{erf} \left( \frac{1}{\sqrt{M-1}} \nu \cdot Y + \sqrt{\frac{M}{M-1}} \theta \right) \\
= T_k(\nu \cdot Y) + O(|\nu \cdot Y|^k (kM)^{-k/2}).
\]

Expectation is determined by \( k \)-independence up to an error of \( O(M^{-1})^{k/2} \).
The Degree 1 Case

Lemma

If $f$ is a degree 1 PTF, $X$ a random Gaussian, and $Y$ a $k$-independent Gaussian ($k$ even) and $\delta > 0$,

$$
\mathbb{E}_X[f(X)] = \mathbb{E}_{X,Y}[f(\sqrt{1-\delta^2}X + \delta Y)] + O(\delta)^k.
$$

For fixed $Y$, have another degree 1 PTF in $X$, so we can iterate:

$$
\mathbb{E}[f(X)] = \mathbb{E}[f(\sqrt{1-\delta^2}X + \delta Y_1)] + O(\delta)^k
= \mathbb{E}[f((1-\delta^2)X + \delta Y_1 + \delta(1-\delta^2)^{1/2}Y_2)] + 2O(\delta)^k
= \ldots
= \mathbb{E} \left[ f \left( (1-\delta^2)^{\ell/2}X + \delta \sum_{i=1}^{\ell} (1-\delta^2)^{(i-1)/2}Y_i \right) \right] + \ell O(\delta)^k.
$$
To get this generator to work for higher degree PTFs we need to show
\[ \mathbb{E}[f(X)] \approx \mathbb{E}[f(\sqrt{1 - \delta^2 X + \delta Y})]. \]

Show that
\[ \mathbb{E}_X[f(\sqrt{1 - \delta^2 X + \delta Y})] \]
is approximated by a polynomial in \( Y \).
Approximately Linear Polynomials

We first consider the case where \( p \) is approximately linear,

\[
p(x) = (1 - \delta^2)^{-1/2}x_{(1)} + \theta + q(x)
\]

with \( |q(x)|_2 \) small. Letting \( X = (X_{(1)}, X') \), we have that

\[
p(\sqrt{1 - \delta^2}X + \delta Y) = X_{(1)} + \theta + r(X_{(1)}, X', Y).
\]

Fixing the values of \( X' \) and \( Y \) we have that

\[
p = p(X_{(1)}) = X_{(1)} + \theta + R_{x', y}(X_{(1)}).
\]

With \( |R|_2 \) small with high probability.
Approximately Linear Polynomials

\[ p = p(X_{(1)}) = X_{(1)} + \theta + R_{x',y}(X_{(1)}) . \]

For small \( X \), \( p \) is invertible by the Inverse Function Theorem.

\[ \mathbb{E}[\text{sgn}(p(X_{(1)}))] \approx \text{erf}(p^{-1}(0)) . \]

We have \( p^{-1}(0) \) smooth in coefficients of \( R \), so Taylor expanding,

\[ \mathbb{E}[\text{sgn}(p(X_{(1)}))] = \text{Polynomial}(R) + \tilde{O}_{d,k}(|R|^k) . \]

Since the expectation of a degree-\( k \) polynomial in \( R \) is determined by \( dk \)-independence of \( Y \), we have that

\[ \mathbb{E}[\text{sgn}(p(X))] = \mathbb{E}[\text{sgn}(p(\sqrt{1 - \delta^2}X + \delta Y))] + \tilde{O}_{d,k}((|q|_2 + \delta)^k) . \]
Local Restrictions

- Problem: Most polynomials are not approximately linear
- Idea: A smooth function is approximately linear on small scales
  - Let \( p_Z(X) = p(\sqrt{1 - \delta^2}Z + \delta X) \).
  - With high probability over \( Z \),
    \[
    p_Z(X) = \text{Const.} + \delta p'(Z) \cdot X + \tilde{O}(\delta^2)
    \]
  - Need linear term not too small
  - Want \(|p'(Z)| > \delta^{1/2}\) with high probability

Definition

We say that \( p \) is \((\delta, c, N)\)-non-singular if

\[
\Pr_Z(|p'(Z)| \leq \delta^c |p|_2) \leq \delta^N.
\]
Non-Singular Polynomials

**Proposition**

If \( p \) is \((\delta, 1/2, k)\)-non-singular, and \( Y \) is \(4dk\)-wise independent, then for \( f(x) = \text{sgn}(p(x))\),

\[
\left| \mathbb{E}[f(X)] - \mathbb{E}[f(\sqrt{1 - \delta^4 X + \delta^2 Y})] \right| = \tilde{O}_{d,k}(\delta^k).
\]

**Proof.**

- Let \( \sqrt{1 - \delta^4 X} = \sqrt{1 - \delta^2 X_1} + \delta \sqrt{1 - \delta^2 X_2} \)
- With probability \( 1 - \delta^k \), \( p_{X_1}(-) \) is approximately linear
- When this happens,

\[
\left| \mathbb{E}_{X_2,Y}[f(\sqrt{1 - \delta^2 X_1} + \delta \sqrt{1 - \delta^2 X_2 + \delta^2 Y})] - \mathbb{E}_{X_2}[f(\sqrt{1 - \delta^2 X_1} + \delta X_2)] \right| = \tilde{O}_{d,k}(\delta^k)
\]
Getting Non-Singular Polynomials

- Most polynomials are non-singular
- Some aren’t.
  - $p(x) = L(x)^d$
  - $|p'(x)| = d|L'(x)||L(x)|^{d-1}$ often small
  - Suffices to study $L(x)$ instead
- Idea: Decompose arbitrary polynomial in terms of non-singular polynomials
Non-Singular Decomposition

Definition

We say that a sequence of polynomials, \((p_1, \ldots, p_m)\), is \((\delta, c, N)\)-non-singular if \(|p_i|_2 = 1\) for all \(i\) and except for with probability \(\delta^N\)

\[
\begin{bmatrix}
p'_1(X) & p'_2(X) & \ldots & p'_m(X)
\end{bmatrix}
\]

has product of singular values at least \(\delta^c\).

Definition

A degree \(d\) polynomial \(p\) has a \((\delta, c, N)\)-non-singular decomposition of size \(m\) if \(p(x)\) can be written as

\[
p(x) = Q(p_1(x), p_2(x), \ldots, p_m(x))
\]

for some \(Q\) and polynomials \(p_1, \ldots, p_m\) of degree at most \(d\) so that \((p_1, \ldots, p_m)\) is a \((\delta, c, N)\)-non-singular set.
The Decomposition Theorem

**Theorem**

For any $d, c, N > 0$ there exists a constant $s(d, c, N)$ so that for any degree-$d$ polynomial $p$, and any $\delta > 0$ sufficiently small, there exists a degree $d$ polynomial $p_0$ with $|p - p_0|_2 \leq \delta^{2dN}|p|_2$, so that $p_0$ has a $(\delta, c, N)$-non-singular decomposition of size at most $s(d, c, N)$.

In particular, we may take $s(1, c, N) = 1$ and $s(2, c, N) = O(N^2/c^2)$.

**Remark**

The proof for $d > 2$ is quite technical. Also the bounds on $s$ are quite bad. The best I can show is $s(d, c, N) \leq A(d + O(1), N/c)$, where $A$ is the Ackermann function.
Using the Decomposition

Proposition

Let $f$ be a degree $d$ PTF. Let $M = dks(d, 1/2, k)$. Let $X$ be a random Gaussian and $Y$ a $2kd$-independent Gaussian. Then for $\delta > 0$

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(\sqrt{1 - \delta^4 X} + \delta^2 Y)]| = O(M)^{O(M)}\tilde{O}(\delta^k).$$

Proof.

- Replace $p$ by $p_0$, which decomposes.
- Take random restriction, reduce to approximately linear case.
- Use smoothing argument and approximate by Taylor polynomial.
Putting it Together

**Theorem**

For $d, k$ positive integers and $\delta > 0$, there exists an explicit pseudorandom generator, $Y$ of seed length $O(d^2 k^2 \log(n) \delta^{-1})$ so that for $X$ an $n$-dimensional Gaussian, and $f$ any degree-$d$ polynomial threshold function in $n$ variables, and $M = dks(d, 1/2, 3k)$

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| = O(M^{O(M)}(\delta^k)).$$

In particular, such a generator is given by letting

$$Y = \sum_{i=1}^{\lceil \delta^{-2/3}(2d+1)k \rceil} (1 - \delta^{2/3})^{i/2} Y_i$$

$$\sqrt{\sum_{i=1}^{\lceil \delta^{-1} dk \rceil} (1 - \delta^{2/3})^i}$$

Where the $Y_i$ are independent of each other and approximate $10d(3k + 3)$-independent random Gaussians.
Applying this theorem, we get PRGs of error $\epsilon$ and seed length

- $O(\log(n) \log^2(1/\epsilon))$ for $d = 1$
- $\log(n) \exp(O(\log^{2/3}(1/\epsilon) \log \log^{1/3}(1/\epsilon)))$ for $d = 2$
- $\log(n) O_{c,d}(\epsilon^{-c})$ for $d > 2$
Improvements

- $d = 1$
  - Need to fool linear function of $Y_i$
  - Seed $Y_i$ with PRG for ROBPs
  - Seed length $O(\log(n) + \log^{3/2}(\epsilon))$

- $d = 2$ (Upcoming paper)
  - Need $Y_i$ to fool many LTFs and one approximately-linear PTF
  - Use $Y_i$ as a PRG for ROBPs
  - Seed length $\tilde{O}(\log(n) \log^6(\epsilon))$. 
Conclusions

We have thus made substantial improvements to the smallest known PRGs for PTFs in the Gaussian case. In particular, for degrees 1 and 2 we seem to be getting close to the lower bounds of $\Omega(\log(n/\epsilon))$. For degree $d > 2$, we are still hampered by the fact that our decompositions may be of enormous size. On the other hand if, as I conjecture to be the case, $s(d, 1/2, k) = \text{Poly}(d, k)$, we would have a generator of seed length $\log(n) \exp(O(d \log(1/\epsilon)^{1-a}))$. 

D. Kane (Stanford)
Acknowledgements

This work was done with the support of an NSF postdoctoral research fellowship.


Daniel M. Kane *A Pseudorandom Generator for Polynomial Threshold Functions of Gaussians with Subpolynomial Seed Length*, submitted to Conference on Computational Complexity.


Daniel M. Kane *k-Independent Gaussians Fool Polynomial Threshold Functions*, Conference on Computational Complexity (CCC 2011).
