Simple and Efficient Pseudorandom generators from Gaussian Processes

Eshan Chattopadhyay
Cornell

Anindya De
U Penn

Rocco Servedio
Columbia
Halfspaces (aka LTFs)

\[ f : \mathbb{R}^n \to \{-1, 1\} \text{ of the form} \]

\[ f(x) = \text{sign}(w \cdot x - \theta) \]
Halfspaces (and their intersections)

\[ f : \mathbb{R}^n \rightarrow \{-1, 1\} \text{ of the form} \]
\[ f(x) = \bigwedge_{j=1}^{k} \text{sign}(w_j \cdot x - \theta_j) \]

Intersection of k-halfspaces
\[ = \text{polytope with k-facets} \]
Intersections of k-halfspaces

• Fundamental for several areas of math and theory CS.

• Well investigated in terms of

1. Learning - [Vempala ’10, Klivans-O’Donnell-Servedio ‘08]

2. Derandomization - [Harsha-Klivans-Meka ‘10, Servedio-Tan ‘17]

3. Noise sensitivity - [Nazarov ‘03, Kane ’14]

4. Sampling - [Dyer-Frieze-Kannan ’89, Lovasz-Vempala 04, ...]
Pseudorandom Generator (PRG)

Let $F$ be a class of Boolean functions

$$\forall f \in F,$$

$$|E[f(U_n)] - E[f(G(U_r))]| < \varepsilon$$
BPP

- Languages that admit an efficient randomized algorithm.

\[ x \in L: \quad \Pr[A(x) = 1] > \frac{2}{3} \]

\[ x \notin L: \quad \Pr[A(x) = 0] > \frac{2}{3} \]
Derandomization via PRGs

Suppose seed length is $O(\log n)$, $\varepsilon = 1/10$.

- $x \in L$: $\Pr[A(x, G(U_r)) = 1] > 1/2$
- $x \notin L$: $\Pr[A(x, G(U_r)) = 0] > 1/2$

To prove $P=BPP$, construct a PRG for efficient randomized algorithms with seed length $O(\log n)$. 
Our focus: derandomization

• This talk: focus on derandomization in the Gaussian space.

\[ \mathbb{R}^n \]

• Setup: endowed with the standard normal measure.

\[ \mathcal{A} \subseteq \mathbb{R}^n \]

• Task: Produce a small and explicit set of points such that for (intersection of k LTFs)

\[
\left| \Pr_{x \sim \mathcal{A}} [f(x) = 1] - \Pr_{x \sim \gamma_n} [f(x) = 1] \right| \leq 0.01.
\]
Our focus: derandomization

Task: Produce a small and *explicit* set of points \( A \subseteq \mathbb{R}^n \) such that for \( f : \mathbb{R}^n \to \{ \pm 1 \} \) (intersection of \( k \) LTFs)

\[
\left| \Pr_{x \sim A} [f(x) = 1] - \Pr_{x \sim \gamma_n} [f(x) = 1] \right| \leq 0.01.
\]

Non-constructively: A of size \( \operatorname{poly}(n, k) \) exists.

Best known explicit construction: Harsha-Klivans-Meka gave a construction of size \( n \log k \).

O’Donnell-Servedio-Tan 2019: matching construction w.r.t uniform on Boolean cube
Our main result

An explicit construction for fooling intersection of $k$-halfspaces on the Gaussian measure whose size is $\frac{n^{\text{poly log } k} O(1) \cdot 2^{\text{poly log } k}}{n^k}$. 

➢ Our construction has polynomial size for $k = 2^{(\log n)^\delta}$.

➢ Arguably much simpler construction.
Connection to Gaussian processes

- Connection is an overstatement -- it’s a simple rephrasing.

- Instead of looking at AND of halfspaces, let us look at OR of halfspaces.

\[ f = g_1(x) \lor g_2(x) \ldots \lor g_k(x) \]

where \( g_i(x) = \text{sign}(w_i \cdot x - \theta_i) \)

\[ f = \mathbf{1}_{\geq 0}(\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}) \]

max/sup of Gaussian processes
Main idea

- We are interested in studying a non-smooth function of the supremum of Gaussian processes.

\[ \mathcal{A} \subseteq \mathbb{R}^n \]

- We are interested in producing a small set so that

\[
\Pr \left[ \mathbb{I}_{\geq 0} \left( \max \{ w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k \} \right) \right] \\
\approx \Pr \left[ \mathbb{I}_{\geq 0} \left( \max \{ w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k \} \right) \right]
\]
Setting sights lower

• What if we want to produce $A \subseteq \mathbb{R}^n$ such that

$$E_{x \sim \gamma_n}[\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}]$$

$$\approx E_{x' \sim A}[\max\{w_1 \cdot x' - \theta_1, \ldots, w_k \cdot x' - \theta_k\}]$$

• Recall: statistics of Gaussian process governed by mean and covariances -- determined by

$$\{\theta_j\}_{j=1}^k, \{\langle w_i, w_j \rangle\}_{1 \leq i, j \leq k}$$

• Johnson-Lindenstrauss can preserve covariances approximately

by projecting on to random subspaces.
Johnson-Lindenstrauss

- Strategy: Sample a random low-dimensional subspace \( H \).

- Sample \( x' \) from \( H \). Call this distribution

**Question:**

(i) Mean / covariance of the distributions
\[
\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}_{x \sim \gamma_n} \approx \{w_1 \cdot x' - \theta_1, \ldots, w_k \cdot x' - \theta_k\}_{x' \sim A}
\]

Does this imply
\[
E_{x \sim \gamma_n}[\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}] \\
\approx E_{x' \sim A}[\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}]
\]
Preserving expected maxima

• Yes - Sudakov-Fernique lemma (quantitative version by Sourav Chatterjee)

• Randomness complexity of sampling from a random low-dimensional subspace $H$?

• JL can be derandomized (Kane, Meka, Nelson - 2011) - in particular, random projection from $n$ to $m$ dimensions can be replaced by a set of size $\approx \text{poly}(n) \cdot 2^{\tilde{O}(m)}$. 
Preserving expected maxima

**Lemma:** Let \( \{X_i\}_{i=1}^{k} \) and \( \{Y_i\}_{i=1}^{k} \) be two sets of normal random variables with \( \mathbb{E}[X_i] = \mathbb{E}[Y_i] \).

a. \( \mathbb{E}[(X_i - X_j)^2] \approx \epsilon \mathbb{E}[(Y_i - Y_j)^2] \)
b. \( |\mathbb{E}[\sup X_i] - \mathbb{E}[\sup Y_i]| \leq \sqrt{\epsilon \cdot \log k}. \)

Then,

In a nutshell: To get non-trivial approximations, we only need \( (1/\log k) \) \( O(\log^3 k) \). This can be achieved by random projections to dimensions.
Preserving expected maxima

Lemma: Let \( \{X_i\}_{i=1}^k \) and \( \{Y_i\}_{i=1}^k \) be two sets of normal random variables with \( \mathbb{E}[X_i] = \mathbb{E}[Y_i] \).

a. \( \mathbb{E}[(X_i - X_j)^2] \approx \epsilon \mathbb{E}[(Y_i - Y_j)^2] \)
b. \( |\mathbb{E}[\sup X_i] - \mathbb{E}[\sup Y_i]| \leq \sqrt{\epsilon \cdot \log k}. \)

Then, the main thing we need to do: Prove the same for \( \mathbb{E}[\sup X_i] \) vis-à-vis \( \mathbb{E}[\sup Y_i] \).
Quick proof sketch

Main trick: Consider smooth maxima function instead of maxima.

Define the function

$$g_\beta(x_1, \ldots, x_k) = \frac{1}{\beta} \log \left( \sum_{i=1}^{k} \exp(\beta x_i) \right)$$

Fact: $$|g_\beta(x_1, \ldots, x_k) - \max(x_1, \ldots, x_k)| \leq \frac{\log k}{\beta}$$

Much easier to work with the smooth function $g_\beta$.
Stein’s interpolation method

• Comparing the quantities \( \mathbb{E}[g_\beta(X_1, \ldots, X_k)] \) and \( \mathbb{E}[g_\beta(Y_1, \ldots, Y_k)] \):

• Condition: \( \{X_i\}, \{Y_i\} \) have matching means and nearly matching covariances.

• For \( t \in [0, 1] \), define \( Z_{i,t} = \sqrt{t}X_i + \sqrt{1-t}Y_i \).
Key statement

Lemma:
\[
\frac{\partial \mathbb{E}[g_\beta(Z_{1,t}, \ldots, Z_{k,t})]}{\partial t} \leq \beta \cdot \max_{i,j} |\text{Cov}(X_i, X_j) - \text{Cov}(Y_i, Y_j)|
\]

Proof is based on Stein’s formula (integration by parts) and some algebraic manipulations.

One useful fact:
\[
\sum_{i=1}^{k} \frac{\partial g_\beta(x_1, \ldots, x_k)}{\partial x_i} = 1
\]
Putting things together

\[ |E[g_\beta(X_1, \ldots, X_k)] - E[\sup(X_1, \ldots, X_k)]| \leq \frac{\log k}{\beta} \]

\[ |E[g_\beta(Y_1, \ldots, Y_k)] - E[\sup(Y_1, \ldots, Y_k)]| \leq \frac{\log k}{\beta} \]

\[ |E[g_\beta(Y_1, \ldots, Y_k)] - E[g_\beta(X_1, \ldots, X_k)]| \leq \epsilon \cdot \beta. \]

\[ |E[\sup(Y_1, \ldots, Y_k)] - E[\sup(X_1, \ldots, X_k)]| \leq \sqrt{\epsilon \cdot \log k}. \]
Our goal

• Recall: We want to prove

\[ |\mathbb{E}[1_{\geq 0}(\sup(Y_1, \ldots, Y_k))] - \mathbb{E}[1_{\geq 0}(\sup(X_1, \ldots, X_k))]| \leq \sqrt{\epsilon \cdot \log k}. \]

• Two step procedure:

• Prove for smooth F

\[ |\mathbb{E}[F(g_\beta(X_1, \ldots, X_k))] - \mathbb{E}[F(g_\beta(Y_1, \ldots, Y_k))]| \leq \|F'\|_\infty \beta \cdot \epsilon + \|F''\|_\infty \cdot \epsilon \]

The error bound depends on derivatives of F.
Going from smooth to non-smooth

• To go from smooth test functions to non-smooth test functions, the random variable \( \Phi(X_1, \ldots, X_k) \) should not be very concentrated.
Going from smooth to non-smooth

- Suppose $X_1, \ldots, X_k$ are (potentially correlated) normal random variables with variance 1.

$$\sup(X_1, \ldots, X_k)$$

- How concentrated can it be?

$$\Pr[|\sup(X_1, \ldots, X_k) - \theta| \leq \epsilon] \leq O(\epsilon \cdot k).$$

- Easy to show:

- **Much harder [Nazarov]:**

  $$\Pr[|\sup(X_1, \ldots, X_k) - \theta| \leq \epsilon] \leq O(\epsilon \cdot \sqrt{\log k}).$$
Putting it together

• Anti-concentration bound allows us to transfer bounds from smooth test function to the test function $1_{\geq 0}$.

• This proves that
  \[|\mathbb{E}[1_{\geq 0}(\sup(X_1, \ldots, X_k)) \mathbb{E}[1_{\geq 0}(\sup(Y_1, \ldots, Y_k))] \leq \text{poly}(\epsilon, \log k)|.\]
Summary

• If we start with a set of jointly Gaussian random variables $X_1, \ldots, X_k$, and do a (pseudo) random projection to obtain $Y_1, \ldots, Y_k$,

$$\Rightarrow \text{JL implies means and covariance preserved.}$$

$$\mathbb{E}[\sup(X_1, \ldots, X_k)] \approx_{\text{poly}(\epsilon, \log k)} \mathbb{E}[\sup(Y_1, \ldots, Y_k)]$$

• Sudakov-Fernique:

$$\mathbb{E}[1_{\geq 0}(\sup(X_1, \ldots, X_k))] \approx_{\text{poly}(\epsilon, \log k)} \mathbb{E}[1_{\geq 0}(\sup(Y_1, \ldots, Y_k))]$$

• This work, we exploit:
Other results

• What other statistics of Gaussians can be preserved by using random projections?

• If \( (X_1, \ldots, X_k) \) and \( (Y_1, \ldots, Y_k) \) have - matching covariances,

\[
|E[g(\text{sign}(X_1), \ldots, \text{sign}(X_k))] - E[g(\text{sign}(Y_1), \ldots, \text{sign}(Y_k))]| \leq \epsilon \cdot \text{poly}(k).
\]

• Proof: closeness in covariance \(\rightarrow\) closeness in Wasserstein

\(\rightarrow\) closeness in union of orthants distance (Chen-Servedio-Tan)

• PRG for arbitrary functions of LTFs on Gaussian space with seed \(O(\log n + \text{poly}(k, 1/\epsilon))\).
Other results

• Deterministic Approximate Counting:
  – $\text{poly}(n) 2^{\text{poly}(\log k, \varepsilon)}$ time algorithm for counting fraction of Boolean points in a $k$-face polytope, up to additive error $\varepsilon$.
  – $\text{poly}(n) 2^{\text{poly}(k, \varepsilon)}$ time algorithm for counting fraction of Boolean points satisfied by an arbitrary function of $k$ halfspaces, up to additive error $\varepsilon$.

• Technique based on invariance principles and regularity lemmas.
  – Beats vanilla use of a PRG that brute-forces over all seeds!
Open questions

• PRGs for fooling DNFs of halfspaces using similar techniques?

• Extending techniques to the Boolean setting?