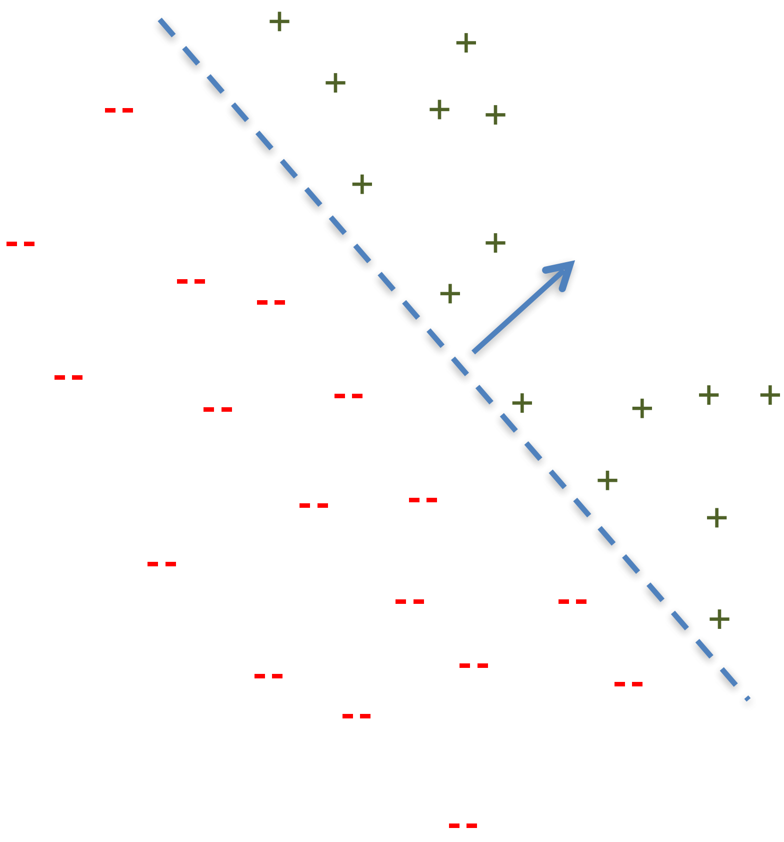
Simple and Efficient Pseudorandom generators from Gaussian Processes

Eshan Chattopadhyay
Cornell

Anindya De
U Penn

Rocco Servedio
Columbia

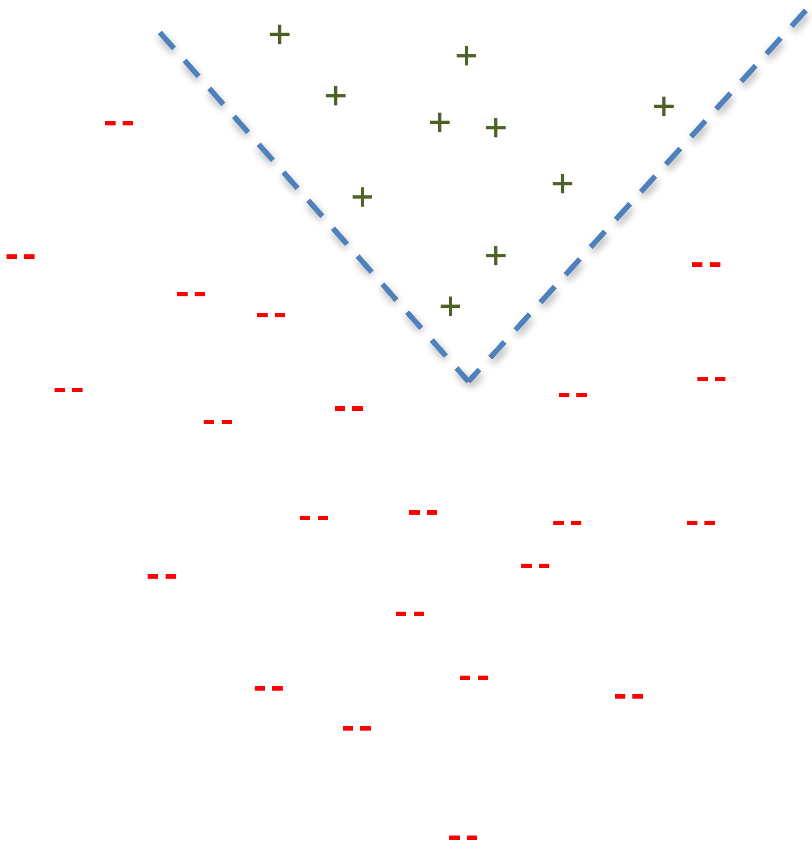
Halfspaces (aka LTFs)



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

$f(x) = \text{sign}(w \cdot x - \theta)$

Halfspaces (and their intersections)



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

$f(x) = \wedge_{j=1}^{k} \mathsf{sign}(w_j \cdot x - \theta_j)$

Intersection of k-halfspaces

$=$ polytope with k-facets

Intersections of k-halfspaces

- Fundamental for several areas of math and theory CS.

- Well investigated in terms of

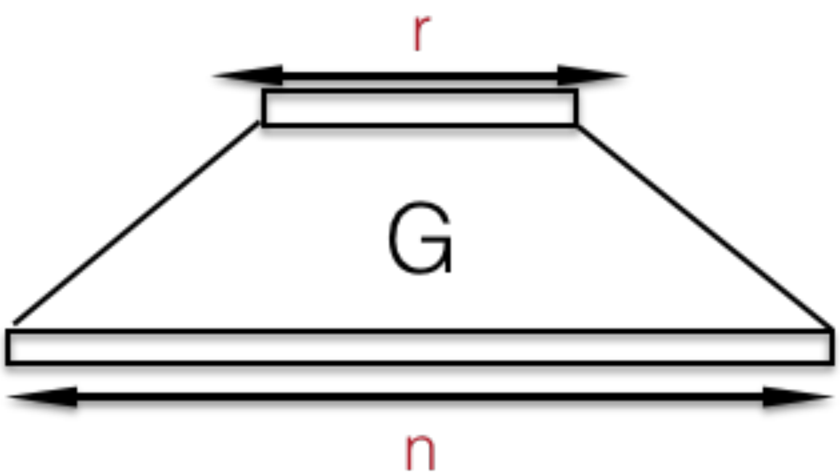
1. Learning – [Vempala ’10, Klivans-O’Donnell-Servedio ‘08]

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

$\forall f \in F,$

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

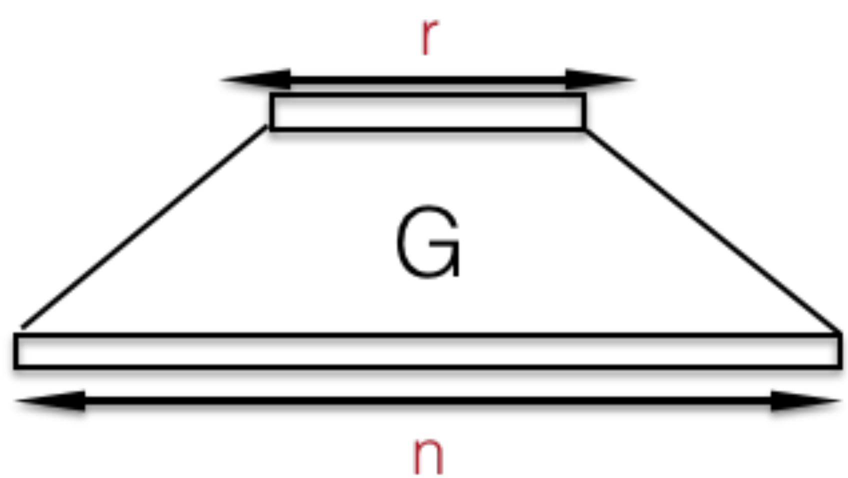
BPP

- Languages that admit an efficient randomized algorithm.

$x \in L$: $\Pr[A(x) = 1] > 2/3$

$x \notin L$: $\Pr[A(x) = 0] > 2/3$

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.
- Setup: $\mathbb{R}^n$ endowed with the standard normal measure.
- Task: Produce a small and explicit set of points $\mathcal{A} \subseteq \mathbb{R}^n$ such that for $f : \mathbb{R}^n \to \{\pm 1\}$ (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 $\mathcal{A} \subseteq \mathbb{R}^n$ such that for $f : \mathbb{R}^n \to \{\pm 1\}$ (intersection of k LTFs)
$$\left| \Pr_{x \sim \mathcal{A}}[f(x) = 1] - \Pr_{x \sim \gamma_n}[f(x) = 1] \right| \leq 0.01.$$

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

Best known explicit construction: Harsha-Klivans-Meka gave a construction of size $n^{\mathrm{poly}\log k}$

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 $\cancel{n^{\text{poly}\log k}}\ n^{O(1)} \cdot 2^{\text{poly}\log k}$
on the Gaussian measure whose size is

➢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.

$$f = g_1(x) \vee g_2(x) \ldots \vee g_k(x)$$

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

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

max/sup of Gaussian processes

Main idea

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

- We are interested in producing a small set $\mathcal{A} \subseteq \mathbb{R}^n$ so that

$$\Pr_{x\sim\gamma_n}[\mathbf{I}_{\geq 0}(\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\})]$$
$$\approx \Pr_{x\sim\mathcal{A}}[\mathbf{I}_{\geq 0}(\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\})]$$

Setting sights lower

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

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

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

- Recall: statistics of Gaussian process governed by mean and covariances -- determined by $\{\theta_j\}_{j=1}^k, \quad \{\langle w_i, w_j\rangle\}_{1\le i,j\le k}$

- 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 $\mathcal{A}$

Question:

(i) Mean / covariance of the distributions

$$\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}_{x \sim \gamma_n} \approx \{w_1 \cdot x' - \theta_1, \ldots, w_k \cdot x' - \theta_k\}_{x' \sim \mathcal{A}}$$

Does this imply

$$\mathbf{E}_{x \sim \gamma_n}[\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}]$$
$$\approx \mathbf{E}_{x' \sim \mathcal{A}}[\max\{w_1 \cdot x - \theta_1, \ldots, w_k \cdot x - \theta_k\}]$$

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 $\mathbf{E}[X_i] = \mathbf{E}[Y_i]$

a. $\mathbf{E}[(X_i - X_j)^2] \approx_\epsilon \mathbf{E}[(Y_i - Y_j)^2]$

b.

Then,

$$|\mathbf{E}[\sup X_i] - \mathbf{E}[\sup Y_i]| \leq \sqrt{\epsilon \cdot \log k}.$$

In a nutshell: To get non-trivial approximations, we only need $\epsilon = O(1/\log k)$. This can be achieved by random projections to $O(\log^3 k)$ 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 $\mathbf{E}[X_i] = \mathbf{E}[Y_i]$

a. $\mathbf{E}[(X_i - X_j)^2] \approx_\epsilon \mathbf{E}[(Y_i - Y_j)^2]$

b.

Then, $|\mathbf{E}[\sup X_i] - \mathbf{E}[\sup Y_i]| \leq \sqrt{\epsilon \cdot \log k}.$

Main thing we need to do: Prove the same for $\mathbf{I}_{\geq 0}[\sup X_i]$ vis-à-vis $\mathbf{I}_{\geq 0}[\sup Y_i]$

Quick proof sketch

Main trick: Consider smooth maxima function instead of maxima.

Define the function $g_\beta(x_1, \ldots, x_k) = \frac{1}{\beta} \log\big(\sum_{i=1}^{k} \exp(\beta x_i)\big)$

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_β

Stein's interpolation method

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

$$\frac{\partial \mathbf{E}[g_\beta(Z_{1,t},\ldots,Z_{k,t})]}{\partial t} \leq \beta \cdot \left| \max_{i,j} [\mathsf{Cov}(X_i, X_j) - \mathsf{Cov}(Y_i, Y_j)] \right|$$

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

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

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

$$\left|\mathbf{E}[g_\beta(Y_1,\ldots,Y_k)] - \mathbf{E}[g_\beta(X_1,\ldots,X_k)]\right| \leq \epsilon \cdot \beta.$$

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

Our goal

- Recall: We want to prove

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

- **Two step procedure:**

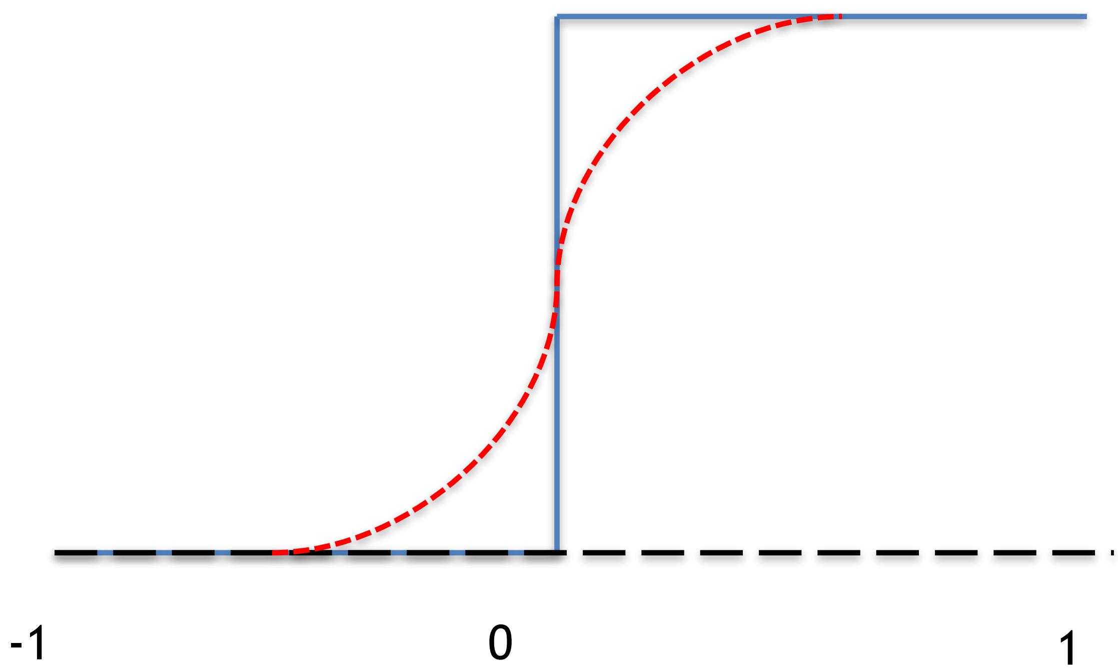
- Prove for smooth F

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

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

- How concentrated can $\sup(X_1, \ldots, X_k)$ be?

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

- 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 F to the test function $\mathbf{1}_{\geq 0}$.

- This proves that $|\mathbf{E}[\mathbf{1}_{\geq 0}(\sup(X_1, \ldots, X_k)] - \mathbf{E}[\mathbf{1}_{\geq 0}(\sup(Y_1, \ldots, Y_k)]| \leq \mathsf{poly}(\epsilon, \log k).$

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 $Y_1, \ldots, Y_k$

=> JL implies means and covariance preserved.

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

- Sudakov-Fernique:

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

- 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 ϵ - matching covariances,
$$\left|\mathbf{E}[g(\text{sign}(X_1), \ldots, \text{sign}(X_k))] - \mathbf{E}[g(\text{sign}(Y_1), \ldots, \text{sign}(Y_k))]\right| \leq \epsilon \cdot \text{poly}(k).$$

- Proof: closeness in covariance ➔ closeness in Wasserstein
➔ 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:
  - 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 ε.
  - 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 ε.
- 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?