Fooling Polytopes

Li-Yang Tan (Stanford)

Joint work with
Ryan O’Donnell (CMU) and Rocco Servedio (Columbia)
Polytopes over \( \{0,1\}^n \)

= Intersections of boolean halfspaces

\[
F(x) = h_1(x) \land \cdots \land h_m(x) \quad x \in \{0,1\}^n
\]

\( F(x) = \text{Halfspace: } \text{sign}(w \cdot x - \theta) \)

- Within CS: optimization (\( \{0,1\}\)-integer programs \( Ax \leq b \)), complexity theory, learning theory, ...
- Beyond CS: Large body of work in combinatorics, high-dimensional geometry, ...
Main complexity measure for this talk: \# of facets

\[ F(x) = h_1(x) \land \cdots \land h_m(x) \quad x \in \{0, 1\}^n \]

\( m \) = number of facets of polytope 
\( = \) number of halfspaces 
\( 1 \leq m \leq 2^n \)

This talk: think of \( m = \text{poly}(n) \), say \( n^{10} \)
This talk: **Discrepancy Sets for Polytopes**

Want **small** set of points $\bullet$ in $\{0,1\}^n$ such that:

For all $m$-facet polytopes $F$, if $F$ accepts $\Delta$ fraction of inputs in $\{0,1\}^n$ ...

... then $F$ accepts $(\Delta \pm 0.01)$ fraction of points $\bullet$

*Random* set of points works great. Want **explicit** set.
Pseudorandom Generators for Polytopes

Pseudorandom output: $n$ bits

Truly random input: $r$ bits

PRG
**Pseudorandom Generators**

**Definition:** An $\epsilon$-PRG for a class $C$ is an explicit function $G : \{0,1\}^r \rightarrow \{0,1\}^n$ such that: for every function $F$ in $C$,

$$\left| \mathbb{E}_{x \sim \{0,1\}^n} [F(x)] - \mathbb{E}_{s \sim \{0,1\}^r} [F(G(s))] \right| \leq \epsilon$$

**This work:**

$C = \{ m$-facet polytopes $\}$

**Goal:** minimize seed length $r(n,m,\epsilon)$

$n$ bits $\rightarrow$ $n$ bits

$G$

$r$ bits $\rightarrow$ $r$ bits
Our main result: PRG for polytopes

An $\varepsilon$-PRG for $m$-facet polytopes over $\{0,1\}^n$ with seed length:

$$\text{poly}(\log m, 1/\varepsilon) \cdot \log n$$

- Previous best seed length had linear dependence on $m$

Equivalently:
Discrepancy set of size $n^{\text{polylog}(m)}$
(Previous best: $n^{O(m)}$)
## Comparison with prior results

<table>
<thead>
<tr>
<th>Class of functions:</th>
<th>Seed length:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any function of $m$ general halfspaces</td>
<td>$\tilde{O}(m \log(1/\epsilon)) \cdot \log n$</td>
</tr>
<tr>
<td>[Gopalan, O’Donnell, Wu, Zuckerman 10]</td>
<td></td>
</tr>
<tr>
<td>Intersections of $m$ “regular” halfspaces</td>
<td>$\text{poly}(\log m, 1/\epsilon) \cdot \log n$</td>
</tr>
<tr>
<td>[Harsha, Klivans, Meka 10]</td>
<td></td>
</tr>
<tr>
<td>Intersections of $m$ low-weight halfspaces</td>
<td>$\text{poly}(\log m, 1/\epsilon) \cdot \text{polylog } n$</td>
</tr>
<tr>
<td>[Servedio, T. 17]</td>
<td></td>
</tr>
<tr>
<td>Intersections of $m$ general halfspaces</td>
<td>$\text{poly}(\log m, 1/\epsilon) \cdot \log n$</td>
</tr>
<tr>
<td>(This work)</td>
<td></td>
</tr>
</tbody>
</table>
PRGs for halfspaces and their generalizations

Halfspaces
- Diakonikolas, Gopalan, Jaiswal, Servedio, Viola 2009
- Meka, Zuckerman 2009
- Karnin, Rabani, Shpilka 2011
- Kothari, Meka 2015
- Gopalan, Kane, Meka 2015

Polynomial threshold functions
- Meka, Zuckerman 2009
- Diakonikolas, Kane, Nelson 2009
- Kane 2011
- Kane 2011
- Kane, Meka 2013
- Kane 2014

Intersections of halfspaces
- Harsha, Klivans, Meka 2010
- Gopalan, O’Donnell, Wu, Zuckerman 2010
- Servedio, T. 2017
- This work
Structure of this talk

- **Part I:** The connection to **Central Limit Theorems**
  - Versatile and powerful framework for designing PRGs
    - [Meka, Zuckerman 09] [Harsha, Klivans, Meka 10]
  - Especially effective for analyzing “regular” halfspaces

- **Part II:** Our work
  - Challenges in dealing with **general** halfspaces
  - New ideas and ingredients in our work
    - New Littlewood–Offord theorem for polytopes

Chalk talk tomorrow!
Part I: Background and Context

PRGs via Central Limit Theorems

Illustrative example: Fooling a single “regular” halfspace [MZ10]
Central Limit Theorems

The sum of many independent “reasonable” random variables converges to **Gaussian** (of same mean and variance).

\[ S = X_1 + \cdots + X_n \approx \mathcal{N}(\mu, \sigma^2) \]

**CDF distance** (= Kolmogorov distance):
For all \( \theta \) in \( \mathbb{R} \),

\[ \Pr[S \leq \theta] \approx \Pr[\mathcal{N} \leq \theta] \]
Regularity and the Berry–Esséen CLT

**Definition:** An “$\varepsilon$-regular” linear form $S : \mathbb{R}^n \rightarrow \mathbb{R}$

$$S(x) = w_1 x_1 + \cdots + w_n x_n$$

is one in which **no weight is too dominant:**

$$|w_i| \leq \varepsilon \cdot \|w\|_2$$

**Berry–Esséen CLT:** For $x$ uniform from $\{-1,1\}^n$,

$$S(x) \approx_\varepsilon \mathcal{N}(\mu, \sigma^2)$$

**Observation:**

Regularity crucial; consider $S(x) = x_1$

CDF of $x_1 \neq$ Gaussian
The connection between CLTs and pseudorandomness

Regular linear form: \( S(x) = w_1 x_1 + \cdots + w_n x_n \) (\( x \) uniform \( \{-1,1\}^n \))

**CLT:** \( S(x) \) converges to Gaussian in CDF distance

\[
\Pr_{\text{uniform } x \sim \{\pm 1\}^n} \left[ S(x) \leq \theta \right] \approx \Pr[ \mathcal{N} \leq \theta ] \\
\text{versus}
\]

**Pseudorandomness:** Fool the regular halfspace \( \text{sign}(S(x) - \theta) \)

\[
\Pr_{\text{uniform } x \sim \{\pm 1\}^n} \left[ S(x) \leq \theta \right] \approx \Pr_{\text{pseudo } y \sim \{\pm 1\}^n} \left[ S(y) \leq \theta \right]
\]

*Pseudorandom* version of Berry–Esséen CLT?
Meka–Zuckerman: PRGs for regular halfspaces via CLTs

Berry–Esséen CLT
Uniform $x \sim \{-1, 1\}^n$

MZ’s derandomization of BE
Pseudorandom $y \sim \{-1, 1\}^n$

CDF-close by $\Delta$-inequality
Meka–Zuckerman’s PRG for regular halfspaces

1. Pseudorandomly hash $n$ variables into $1/\varepsilon^2$ buckets
2. Assign values within each bucket according to $O(1)$-wise independent distribution (independently across buckets)

\[ \begin{align*}
&y_{10}, y_{12}, \ldots, y_{9}, y_4 \\
&y_6, y_3, y_7 \\
&y_2 \\
\end{align*} \]

$1/\varepsilon^2$ buckets

[Meka–Zuckerman 10]
- **Theorem:** Berry–Esséen CLT holds for this distribution
- **Corollary:** This is an $\varepsilon$-PRG for $\varepsilon$-regular halfspaces with seed length $O((\log n)/\varepsilon^2)$.

This talk: All PRGs = this [MZ] generator
(possibly with different parameters)
Next: Fooling **Intersections of** regular halfspaces

Same overall framework, but many cool new ideas and ingredients...

**Regular halfspaces**
[Meka, Zuckerman 09]

**Intersections of** regular halfspaces
[Harsha, Klivans, Meka 10]
1 regular halfspace
[Meka, Zuckerman 09]

Intersection of $m$
regular halfspaces
[Harsha, Klivans, Meka 10]

Berry-Esséen CLT
\[
\sum_{i=1}^{n} X_i \rightarrow \mathcal{N}(\mu, \sigma^2)
\]
Sum of real-valued r.v.’s,
none too dominant

[HKM10] multidimensional CLT
\[
\sum_{i=1}^{n} \vec{X}_i \rightarrow \mathcal{N}(\vec{\mu}, \Sigma)
\]
Sum of $\mathbb{R}^m$-valued r.v.’s,
none too dominant

convergence in
multidimensional CDF distance
HKM’s PRG via their multidimensional CLT

Let $A$ be an $m \times n$ matrix, where every row of $A$ is regular
(row = weights of a regular halfspace)

Uniform $x \sim \{-1, 1\}^n$

Pseudorandom $y \sim \{-1, 1\}^n$

$Ax \overset{CDF-close}{\sim} A y$

$A x \overset{CDF-close by \Delta-inequality}{\sim} A y$

$m$-dimensional Gaussian
Part I:

- Regular halfspaces
  [Meka, Zuckerman 09]
- Intersections of regular halfspaces
  [Harsha, Klivans, Meka 10]

Part II:

(Our work)

- Intersections of general halfspaces
Part I:

- Regular halfspaces
- Intersections of $m$ regular halfspaces

Part II:

- Intersections of $m$ general halfspaces

Other relevant works not discussed in Part I:

- 1 general halfspace [Meka, Zuckerman 09]
- Any function of $m$ general halfspaces, but seed length $\tilde{O}(m)$ [Gopalan, O’Donnell, Wu, Zuckerman 10]
- Intersection of $m$ low-weight halfspaces, seed length $\text{polylog}(m)$ [Servedio, T. 17]
Main challenge:

CLTs and **regularity** go hand in hand

Central Limit Theorem **false** without regularity assumption

\[ S(x) = w_1 x_1 + \cdots + w_n x_n \xrightarrow{\text{not}} N(\mu, \sigma^2) \]

if there are dominant \( w_i \)'s

Recall simple example: \( S(x) = x_1 \)

Not close to CDF of any Gaussian
This work: bypassing Gaussian middleperson (by necessity)

Let $A$ be general $m \times n$ matrix
(row = weights of general halfspace, not necessarily regular)

Uniform $x \sim \{-1, 1\}^n$

Pseudorandom $y \sim \{-1, 1\}^n$

$m$-dimensional Gaussian
But—will still employ CLT proof techniques (even though CLT does not hold!)

\[ A \mathbf{x} \leftrightarrow \text{CDF-close!} \Rightarrow A \mathbf{y} \]

- Uniform \( \mathbf{x} \sim \{-1,1\}^n \)
- Pseudorandom \( \mathbf{y} \sim \{-1,1\}^n \)

- Lindeberg replacement method
  - Powerful technique for proving CLTs [Lindeberg 22]
  - [MZ, HKM]'s strategy for the all-regular case

- Non-regularity necessitates new ideas and ingredients:
  - PRGs for CNF formulas [AW85, Nis92, Baz07, ... ]
  - New Littlewood–Offord theorem for polytopes
Outline of the rest of the talk
(= the structure of our proof)

1. A useful decomposition of polytopes
2. “Smooth version” of the problem
3. Proving the smooth version
4. Going from smooth version to actual version
Regularity lemma for a single halfspace

Weights of **regular** halfspace:

(No weight too dominant)

Weights of **general** halfspace:

**Halfspace Regularity Lemma**

[Servedio 07]

Every halfspace can be made **regular**\* by restricting a small number of variables

*or very close to constant
Halfspace Regularity Lemma as a picture

\( \tilde{O}(1/\varepsilon^2) \) “head” variables

\( \varepsilon \)-regular “tail”

Weights of a general halfspace sorted by magnitude
Applying the regularity lemma to $m$ halfspaces

Remark:
- Each head small, but union of all $m$ heads could cover $[n]$
- So the natural strategy of reducing to the all-regular case – by “restricting away” all head variables – does not work
A useful mental picture:

Goal: $Ax$ and $Ay$ are close in multidimensional CDF distance

- $x \sim \{-1,1\}^n$ uniform
- $y \sim \{-1,1\}^n$ pseudorandom:

  $y_{10} \ y_6 \ y_2 \ y_3 \ y_{12} \ \cdots \ y_9 \ y_4 \ y_7 \ y_{11}$

(Each halfspace has few head variables)
Outline of the rest of the talk
(= the structure of our proof)

1. A useful decomposition of polytopes
2. “Smooth version” of the problem
3. Proving the smooth version
4. Going from smooth version to actual version
A smooth version of CDF distance

Ax and Ay are CDF-close

\[ \iff \quad \Pr[ Ax \leq b ] \approx \Pr[ Ay \leq b ] \quad \text{for all } b \in \mathbb{R}^m \]

\[ \iff \quad E[ O_b(Ax) ] \approx E[ O_b(Ay) ] \quad \text{for all } b \in \mathbb{R}^m \]

\[ O_b = \{0,1\}-\text{indictor of orthant defined by } b \]

We will first show:

\[ E[ \tilde{O}_b(Ax) ] \approx E[ \tilde{O}_b(Ay) ] \]

where

\[ \tilde{O}_b : \mathbb{R}^m \rightarrow [0, 1] \]

is smooth approximator of \( O_b : \mathbb{R}^m \rightarrow \{0, 1\} \)

Discontinuous function!
Smooth approximators of orthants

Standard way of mollifying a function: adding Gaussian noise

$$\tilde{O}_b(v) = \mathbb{E}_{G \sim \mathcal{N}(0,1)^m} [O_b(v + \lambda G)]$$

Two important properties of $\tilde{O}_b$ [Bentkus 90]:

1. Good approximation of $O_b$
2. Small derivatives: for all $c > 1$,

$$\sup_{v \in \mathbb{R}^m} \left\{ \sum_{|\alpha| = c} |\partial_\alpha \tilde{O}_b(v)| \right\} \lesssim \frac{(\log m)^{c/2}}{\lambda^c}$$

$m^c$ many partial derivatives
Outline of the rest of the talk
(= the structure of our proof)

1. A useful decomposition of polytopes
2. “Smooth version” of the problem
3. Proving the smooth version
4. Going from smooth version to actual version
Proving the smooth version via a hybrid argument

Goal is to “fool” the orthant mollifier: \[ \mathbb{E}_{\mathbf{x}}[\tilde{O}_b(A\mathbf{x})] \approx \mathbb{E}_{\mathbf{y}}[\tilde{O}_b(A\mathbf{y})] \]

**Bucket-wise hybrid argument [MZ, HKM]:**

- Start will all buckets filled in uniformly (i.e. start with \( \mathbf{x} \))
- Bucket by bucket, “swap out” uniform bits for \( r \)-wise independent bits
- Argue that each swap incurs small error
**Single swap in the hybrid argument**

Fix bucket $B \subseteq [n]$ of variables. Let $A^B = A$ restricted to columns in $B$.

Want to show:  
\[
\mathbb{E}_{\mathbf{x}}[\tilde{O}_b(A^B \mathbf{x})] \approx \mathbb{E}_{\mathbf{y}}[\tilde{O}_b(A^B \mathbf{y})]
\]

\[
\iff \mathbb{E}_{\mathbf{x}}[\tilde{O}_b(H \mathbf{x} + T \mathbf{x})] \approx \mathbb{E}_{\mathbf{y}}[\tilde{O}_b(H \mathbf{y} + T \mathbf{y})]
\]

Write $A^B = H + T$, where:
- $H$ contains only the head variables
- $T$ contains only the tail variables

CLTs (e.g. [HKM]) deal with regular linear forms; Presence of $H = $ our main challenge
Multidimensional Taylor expansion

**Claim:** \( \mathbb{E}_{x}[\tilde{O}_b(Hx + Tx)] \approx \mathbb{E}_{y}[\tilde{O}_b(Hy + Ty)] \)

Equivalently, \( y \) fools the function \( z \mapsto \tilde{O}_b(Hz + Tz) \)

**Warmup:** Does \( y \) fool the zeroth-order term?

\[
\mathbb{E}_{x}[\tilde{O}_b(Hx)] \approx \mathbb{E}_{y}[\tilde{O}_b(Hy)]
\]

(Observation: trivial when \( H = 0 \))
**Claim:** $y$ fools the function $z \mapsto \tilde{O}_b(Hz)$

Recalling the definition of $\tilde{O}_b$,

$$\tilde{O}_b(Hz) = \mathbb{E}_{G}[O_b(Hz + \lambda G)]$$

$$\equiv \prod_{i=1}^{m} 1[(Hz + \lambda G)_i \leq b_i]$$

**Orthant $O_b$**

**Product structure of $\tilde{O}_b +$ Sparsity of $H$**

Suffices for $y$ to fool small-width CNFs

Simple but key idea: product of $k$-juntas = width-$k$ CNF
Back to the Taylor expansion

**Claim:** $y$ fools the function $z \mapsto \tilde{O}_b(Hz + Tz)$

We consider the multidimensional Taylor expansion:

$$\tilde{O}_b(Hz) + \sum_{1 \leq |\alpha| \leq c} \frac{1}{\alpha!} \partial_\alpha \tilde{O}_b(Hz)(Tz)^\alpha \pm \text{err}$$

More complicated, but same key ideas:

- Product structure of $\partial_\alpha \tilde{O}_b$ $\implies$ Suffices for $y$ to fool CNFs
- Sparsity of $H$

To bound error term, use fact that $\tilde{O}_b$ has small derivatives (same as [HKM]):

$$\sup_{v \in \mathbb{R}^m} \left\{ \sum_{|\alpha| = c} |\partial_\alpha \tilde{O}_b(v)| \right\} \lesssim \frac{(\log m)^{c/2}}{\lambda^c}$$

[Bentkus 90]
Recap

Goal is to “fool” the orthant mollifier: \( \mathbb{E}[\widetilde{O}_b(Ax)] \approx \mathbb{E}[\widetilde{O}_b(Ay)] \)

What we just sketched

Bounding error incurred by a single swap:
Outline of the rest of the talk
(= the structure of our proof)

1. A useful decomposition of polytopes
2. “Smooth version” of the problem
3. Proving the smooth version
4. Going from smooth version to actual version
What we’ve shown:

\[ \mathbb{E}_{x}[\tilde{O}_b(Ax)] \approx \mathbb{E}_{y}[\tilde{O}_b(Ay)] \]

What we’d like to show: (closeness in CDF distance)

\[ \mathbb{E}_{x}[O_b(Ax)] \approx \mathbb{E}_{y}[O_b(Ay)] \]

\[ \tilde{O}_b : \mathbb{R}^m \to [0, 1] \]

\[ O_b : \mathbb{R}^m \to \{0, 1\} \]
Another conceptual difference/challenge: Boolean vs. Gaussian anticoncentration

Since we are bypassing Gaussians:
Have to instead reason about **Boolean** anticoncentration

(Fact: Boolean anticoncentration \(\downarrow\) Gaussian anticoncentration)

Proofs of CLTs (e.g. [HKM]):
Gaussian anticoncentration

\[ AG, \quad G \sim N(0,1)^n \]
Littlewood–Offord anticoncentration inequality

Let $w \in \mathbb{R}^n$ where $|w_i| \geq 1$ for all $i$. For all open intervals $I \subset \mathbb{R}$ of radius 2,

$$\Pr_{x \sim \{\pm 1\}^n} [w \cdot x \in I] \lesssim \frac{1}{\sqrt{n}}.$$  

In fact:

$$\leq \left(\frac{n}{\lceil n/2 \rceil}\right) \cdot 2^{-n}$$

[Erdoes 45]

(Exactly tight for $w = 1^n$)
A high-dimensional Littlewood–Offord inequality
(LO: \(m=1\) case)

Let \(A \in \mathbb{R}^{m \times n}\) where \(|A_{ij}| \geq 1\) for all \(i, j\).

For all orthant boundaries \(B \subset \mathbb{R}^m\) of width 2,

\[
\Pr_{x \sim \{\pm 1\}^n} [Ax \in B] \lesssim ?
\]

- 1-dimensional LO + union bound: \(O(m/\sqrt{n})\)
- We prove \(O(\sqrt{\log m}/\sqrt{n})\), which we show is tight
- Need various technical extensions for our purposes

\{ x : Ax \leq b \} \setminus \{ x : Ax \leq b-2 \}

Chalk talk tomorrow!
Recap of proof structure

1. A useful decomposition of polytopes
2. “Smooth version” of the problem
3. Proving the smooth version
4. Going from smooth version to actual version
Summary

Our main result

An $\varepsilon$-PRG for $m$-facet polytopes over $\{0,1\}^n$ with seed length:

$$\text{poly}(\log m, \log(1/\varepsilon)) \cdot \log n$$

- Previous best seed length had linear dependence on $m$
- Many interesting future directions:
  - Seed length $\text{poly}(\log m, \log(1/\varepsilon)) \cdot \log n$
  - PRGs for other geometric sets?
    - PRG for all convex sets?

Discrepancy set of size $n^{\text{polylog}(m)}$
Thanks!
Let $G : \{0,1\}^r \rightarrow \{0,1\}^n$ be an $\varepsilon$-PRG for the class $m$-facet polytopes. Consider $\{ G(s) : s \in \{0,1\}^r \}$ → A set of $2^r$ points in $\{0,1\}^n$.

Discrepancy set for the class of $m$-facet polytopes
Analogous results for other domains

- Solid cube $[0, 1]^n$
- Hypergrid $\{0, 1, ..., k\}^n$
- Gaussian space ($\mathbb{R}^n$ under $N(0, 1)^n$) [HKM10]
- ...

$\{0, 1\}^n$
One algorithmic application:
Counting # of solutions of \{0,1\}-integer programs

Maximize \( c^T x \)
Subject to \( Ax \leq b \)
And \( x \in \{0,1\}^n \)

Given as input a \{0,1\}-IP with \( m \) constraints, there is a deterministic algorithm that runs in time
\( n^{\text{poly}(\log m, 1/\epsilon)} \)
and outputs an estimate of the fraction of feasible solutions, accurate to \( \pm \epsilon \).

PRG = input-oblivious algorithm

Average w.r.t. fixed discrepancy set works for all possible inputs (all possible \{0,1\}-IPs)