Positive spectrahedra:
Invariance principles and PRGs

Srinivasan Arunachalam (IBM Research)

joint with Penghui Yao (Nanjing University)
Linear programs

- Optimizing a linear function over a polytope
- A general LP has the form: \( w^1, \ldots, w^k, c \in \mathbb{R}^n \) and \( \theta^i \in \mathbb{R} \)

\[
\text{OPT} = \max_{x \in \mathbb{R}^n} \{ c^T x : w^1 \cdot x \leq \theta^1, \ldots, w^k \cdot x \leq \theta^k \}
\]

- Efficiently solvable!

- **Halfspace** in \( \mathbb{R}^n \) is a constraint that divides the space, i.e., \( h_1 : \mathbb{R}^n \to \{0, 1\} \)
- Let \( w \in \mathbb{R}^n \) and \( x \in \{-1, 1\}^n \), then a halfspace \( h_1(x) = 1 \) iff \( w \cdot x \leq \theta \), or \( h_1(x) = [w \cdot x \leq \theta] \)

- **Polytope** is an intersection of halfspaces
- Let \( w^i \in \mathbb{R}^n, \theta^i \in \mathbb{R} \), a \( k \)-facet polytope is

\[
P = \{ x : h_1(x) \land h_2(x) \land \cdots \land h_k(x) \}
\]
where \( h_i = [w^i \cdot x \leq \theta^i] \)

**Applications:** Optimization, combinatorics, geometry, computational complexity, \ldots
What is an SDP?

- Optimizing a linear function over a spectrahedron
- A general SDP has the form: \( W^1, \ldots, W^k, B \in \text{Sym}_n \)

\[
\text{OPT} = \max_{x \in \mathbb{R}^n} \{ c^T x : x_1 W^1 + \cdots + x_n W^n \preceq B \},
\]

where \( C \preceq D \) means \( D - C \) is PSD (i.e., all eigenvalues are \( \geq 0 \))

How does it look?

- Generalizes linear programs and still efficiently solvable!
- Unfortunately, spectrahedra are not very well understood!
- But SDPs have found applications in approximation theory, SoS heirarchy, quantum computing
Pseudorandom generators

A PRG is a function that “expands” randomness

\[ G : \{0, 1\}^r \rightarrow \{0, 1\}^n \]

PRGs for a class of functions

An \( \varepsilon \)-PRG for \( C \subseteq \{ F : \{0, 1\}^n \rightarrow \{0, 1\} \} \) is a function \( G : \{0, 1\}^r \rightarrow \{0, 1\}^n \) such that

\[
\left| \mathbb{E}_{x \sim \mathcal{U}_r} [F(G(x))] - \mathbb{E}_{u \sim \mathcal{U}_n} [F(u)] \right| \leq \varepsilon.
\]

The seed length of \( G \) is \( r \). **Goal** is to have \( r = \text{polylog}(\cdot) \) in all relevant parameters.

**Holy grail.** Can we design a PRG against the class of polynomial sized circuits unconditionally? If so, would imply \( \text{BPP} = \text{P} \).

**PRGs for geometric objects.** Constructing PRGs using geometric properties has been a rich area of study this work.
Some PRGs for geometric objects

**Halfspaces**
- Diakonikolas et al.’09
- Meka, Zuckerman’09
- Karnin, Rabani, Shpilka’11
- Kothari, Meka’15
- Gopalan, Kane, Meka’15

**Polytopes**
- Harsha, Klivans, Meka’13
- Gopalan et al.’13
- Servedio, Tan’17
- O’Donnell, Servedio, Tan’19

**Polynomial Threshold function**
\[ T(x) = \text{sign}(p(x_1, \ldots, x_n)) \] where \( p \) is a polynomial
- Meka, Zuckerman’09
- Diakonikolas’10
- Kane’11, Kane’12, Kane’13
- Kane, Meka’14
- O’Donnell, Servedio, Tan’20

Spectrahedra: generalization of halfspaces, polytopes and PTFs in one framework

**In this work:** Can we construct PRGs for spectrahedra?
Recall. Spectrahedron is the set \( S = \{ x \in \{ -1, 1 \}^n : \sum_i x_i A_i \preceq B \} \).

1. **Positive**: \( A^1, \ldots, A^n, B \) are \( k \times k \) PSD matrices
2. **Bounded width**: \( \mathbb{I} \leq \sum_i (A^i)^2 \leq M \cdot \mathbb{I} \)
3. **Regular**: \( (A^i) \preceq \tau \cdot \mathbb{I} \) for every \( i \)

**Main Theorem**

There exists a PRG \( G : \{0, 1\}^r \to \{ -1, 1 \}^n \) with seed length

\[
r = (\log n) \cdot \text{poly} \left( \log k \cdot M \cdot 1/\delta \right),
\]

that \( \delta \)-fools the class of positive bounded width regular spectrahedron \( S \), i.e.,

\[
\left| \mathbb{E}_{x \sim \mathcal{U}_r} [G(x) \in S] - \mathbb{E}_{u \sim \mathcal{U}_n} [u \in S] \right| \leq \delta.
\]

**Main technical contributions**: Rest of this talk

**An invariance principle for positive regular spectrahedra**
Punchline: Invariance principles give pseudorandom generators.

What is an invariance principle? Generalization of Berry-Esseen theorem

Standard central limit theorem states: suppose \( x_1, \ldots, x_n \) are random variables satisfying \( \mathbb{E}[x] = 0 \) and \( \text{Var}[x^2] = 1 \), then

\[
\frac{x_1 + \cdots + x_n}{\sqrt{n}} \rightarrow g(0, 1),
\]

where \( g(0, 1) \) is a Gaussian

But what about convergence? Berry-Esseen states that for every \( u \in \mathbb{R} \)

\[
\Pr \left[ \frac{x_1 + \cdots + x_n}{\sqrt{n}} \leq u \right] - \Pr \left[ g(0, 1) \leq u \right] \leq \frac{C}{\sqrt{n}},
\]

for "C-nice" \( x_1, \ldots, x_n \). Proved using the Lindeberg method’22 (aka hybrid method)

Invariance principles: understanding in the Gaussian space is similar to Boolean space
How was M-Z used so far?

**Halfspace** (Meka-Zuckerman'09)
- Halfspace is
  \[ \{ x \in \{-1, 1\}^n : \sum_i w_i x_i \leq \theta \} \]
- For smooth \( w \in \mathbb{R}^n \)
  \[ \sum_i w_i x_i \to g(0, 1) \]

**Polytope** (Harsha-Klivans-Meka’13)
- Polytope is
  \[ \{ x \in \{-1, 1\}^n : w^1 \cdot x \leq \theta_1, \ldots, w^k \cdot x \leq \theta_k \} \]
- Let \( w^1, \ldots, w^k \in \mathbb{R}^n \) all be smooth, then
  \[
  \begin{bmatrix}
    w^1 & \cdots & w^k \\
    \vdots & \ddots & \vdots \\
    w^k & \cdots & w^1 \\
  \end{bmatrix}
  \begin{bmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_n \\
  \end{bmatrix}
  \to
  \begin{bmatrix}
    g_1 \\
    g_2 \\
    \vdots \\
    g_k \\
  \end{bmatrix}
  \]
- Recently OST’19 removed regularity
Establishing the invariance principle

Recall: Polytope \( F(x) = [w^1 \cdot x \leq \theta_1 \land \cdots \land w^k \cdot x \leq \theta_k] \) or \( W \cdot x \leq \vec{\theta} \)

Main result of HKM’13 Invariance principle for \( \tau \)-regular polytopes (i.e., \( \|w^i\| \leq \tau \))

\[
\left| \mathbb{E}_{x \sim \mathcal{U}_n} [Wx \leq \vec{\theta}] - \mathbb{E}_{g \sim \mathcal{G}_n} [Wg \leq \vec{\theta}] \right| \leq \tau \text{ polylog } k \tag{1}
\]

How to prove this?

1. Smooth invariance. Establish (1) for smooth functions \( \mathcal{O} : \mathbb{R}^k \to \mathbb{R} \), i.e.,

\[
\left| \mathbb{E}_{x \sim \mathcal{U}_n} [\mathcal{O}(Wx)] - \mathbb{E}_{g \sim \mathcal{G}_n} [\mathcal{O}(Wg)] \right| \leq \tau \log k \cdot \|\mathcal{O}^{(3)}\|_1
\]

- **Lindeberg method**: Write out Taylor series for \( \mathcal{O} : \mathbb{R}^k \to \mathbb{R} \), since \( \mathcal{U}_n \) and \( \mathcal{G}_n \) have matching first and second moments, we get 3rd derivatives, hence \( \|\mathcal{O}^{(3)}\|_1 \)

- Since \( \mathcal{O} \) is smooth, all derivatives are “small” so \( \|\mathcal{O}^{(3)}\|_1 \) is also “small”
Establishing the invariance principle

Main result of HKM’13 Invariance principle for $\tau$-regular polytopes

$$\left| \mathbb{E}_{x \sim \mathcal{U}_n} [Wx \leq \theta] - \mathbb{E}_{g \sim \mathcal{G}_n} [Wg \leq \theta] \right| \leq \tau \text{polylog } k \quad (2)$$

How to prove this?

1. **Smooth invariance.** Establish (2) for smooth mollifiers $\mathcal{O} : \mathbb{R}^n \to \mathbb{R}$, i.e.,

$$\left| \mathbb{E}_{x \sim \mathcal{U}_n} [\mathcal{O}(Wx)] - \mathbb{E}_{g \sim \mathcal{G}_n} [\mathcal{O}(Wg)] \right| \leq \tau \log k \cdot \|\mathcal{O}(3)\|_1 \quad (3)$$

2. **Bentkus mollifier.** Care about $[Wx \leq \theta]$ not $\mathcal{O}(Wx)$. Bentkus’90 established a mollifier $B : \mathbb{R}^k \to \mathbb{R}$ that approximates the orthant function, i.e.,

$$B(z_1, \ldots, z_k) \approx \left[ \max_i z_i \leq \theta \right] \quad \text{and} \quad \|B^\ell\|_1 \leq \log^\ell/2k$$

3. **Anti-concentration.** From above $B$ “approximately agrees” with $\mathcal{O}$.

- Around the “boundary” of the polytope is where $B$ and $\mathcal{O}$ disagree
- If probability of $x \in \mathcal{G}^n$ lying in boundary is “small”, maybe it is ok? YES
- Gaussian surface area! Nazarov’03 showed GSA of $k$-facet polytopes is $\sqrt{\log k}$

Putting everything together. All dependence are logarithmic factors, so (3) $\implies$ (2).
Let’s quantize everything!

A halfspace $F(x) = [\sum_i w_i x_i \leq \theta]$. Spectrahedron is $F(x) = [x_1 A^1 + \cdots + x_n A^n \preceq B]$

1. **Hybrid method?**
   - **1.** Spectrahedron naturally deals with eigenvalues of matrices
   - **2.** Unknown if Lindeberg-type argument works for spectral mollifiers (i.e., smooth functions acting on the eigenspectrum of matrices)

An invariance principle for the Bentkus mollifier of arbitrary regular spectrahedra

2. **Anti-concentration?** Even if GSA of spectrahedra are small, they are “funky” geometric objects, not clear how to go from mollifier-closeness to CDF closeness

Prove a Littlewood-Offord theorem for positive regular spectrahedra
Invariance principle: two new definitions

1. **Spectral function** $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ acts on the eigenvalues of matrices:

$$f(M) = g \circ \lambda(M) = g(\lambda_1(M), \ldots, \lambda_n(M))$$

for some $g : \mathbb{R}^n \to \mathbb{R}$. Examples include determinants, trace, matrix norms.

2. **Derivatives of matrix-valued functions** $f : \mathbb{R}^{n \times n} \to \mathbb{R}^{m \times m}$.

   **Taylor series.** Let $h : \mathbb{R} \to \mathbb{R}$, then Taylor series of $h$ is

   $$h(x) = h(a) + \frac{h'(a)}{1!} (x - a) + \frac{h''(a)}{2!} (x - a)^2 + \frac{h'''(a)}{3!} (x - a)^3 + \cdots,$$

   where

   $$h'(a) = \lim_{s \to 0} \frac{1}{s} \cdot (h(a + s) - h(a))$$

   **Fréchet derivatives.** Derivatives in Banach spaces. “Similar” to standard calculus. For $A, B \in \mathbb{R}^{n \times n}$, we have

   $$Df(A)[B] = \lim_{s \to 0} \frac{1}{s} \cdot (f(A + sB) - f(A)),$$

   $$D^t f(A)[B] = \lim_{s \to 0} \frac{1}{s} \cdot (D^{t-1}f(A + sB)[B] - D^{t-1}f(A)[B])$$

   Fréchet derivatives are hard to compute. Poorly understood: basic properties as continuity, Lipschitz continuity, differentiability proven in last 2 decades.
Invariance principle: Part I

**Goal:** Invariance principle for Bentkus mollifier $B$

$$\left| \Pr_{x \sim U_n} [(B \circ \lambda)(\sum_i x_i A^i - B)] - \Pr_{g \sim G_n} [(B \circ \lambda)(\sum_i g_i A^i - B)] \right| \leq \tau \text{ polylog } k$$

1. **Hybrid method.** Hash the sum over $[n]$ into $t$ blocks: let $Q_x = \sum_{i=1}^{n/t} x_i A^i$,

$$\left| \mathbb{E}_{x,g} [(B \circ \lambda)(Q_x + P_{x,g})] - \mathbb{E}_{x,g} [(B \circ \lambda)(Q_g + P_{x,g})] \right|$$  \hspace{1cm} (4)

2. **Taylor expansion.** Write out Fréchet series for both these terms.

$$B_\lambda(Q_x + P_{x,g}) = B_\lambda(P_{x,g}) + DB_\lambda(P_{x,g})[Q_x] + \frac{1}{2} D^2 B_\lambda(P_{x,g})[Q_x, Q_x] + \frac{1}{6} D^3 B_\lambda(P_{x,g})[Q_x, Q_x, Q_x]$$

$$B_\lambda(Q_g + P_{x,g}) = B_\lambda(P_{x,g}) + DB_\lambda(P_{x,g})[Q_g] + \frac{1}{2} D^2 B_\lambda(P_{x,g})[Q_g, Q_g] + \frac{1}{6} D^3 B_\lambda(R_{x,g})[Q_g, Q_g, Q_g]$$

Same colour terms are equal in expectation

So bounding Eq (4) amounts to proving. **Goal:** upper bound

$$\left| \mathbb{E}_{x,g} \left[ D^3 B_\lambda(P_{x,g})'[Q_x, Q_x, Q_x] - D^3 B_\lambda(R_{x,g})'[Q_g, Q_g, Q_g] \right] \right| \leq \tau \text{ polylog } k$$

3. **Sendov to the rescue.** For us, Sendov provided a tensorial representation of Fréchet series for spectral functions
Invariance principle: Part II

Recall: Goal is to upper bound

\[
\left\| \mathbb{E}_{x,g} \left[ D^3 B_{\lambda}(P'_{x,g})(Q_x, Q_x, Q_x) - D^3 B_{\lambda}(R'_{x,g})(Q_g, Q_g, Q_g) \right] \right\|
\]

Hope: use Sendov’s tensor-result. BUT, if you write it out, we get:

\[
H = VQVT. \text{ Then } D^3 F(P)[Q, Q, Q] \text{ is the summation of the following terms.}
\]

1. \( \sum_{i_1} \nabla^3_{i_1,i_1,i_1} f(x) H^3_{i_1,i_1} \)
2. \( \sum_{i_1 \neq i_2} \nabla^3_{i_1,i_2,i_1} f(x) H^2_{i_1,i_1} H_{i_2,i_2} \)
3. \( \sum_{i_1 \neq i_2 \neq i_3} (\nabla^3_{i_1,i_2,i_2} f(x)) \cdot H_{i_1,i_1} H_{i_2,i_2} H_{i_3,i_3} \)
4. \( \sum_{i_1 \neq i_2} \left( \frac{\nabla^2_{i_2,i_2} - \nabla^2_{i_1,i_1}}{x_{i_2} - x_{i_1}} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} \right) f(x) H_{i_2,i_2} H^2_{i_2,i_1} \)
5. \( \sum_{i_1 \neq i_2 \neq i_3} \frac{\nabla^2_{i_2,i_3} - \nabla^2_{i_1,i_1}}{x_{i_2} - x_{i_1}} f(x) H^2_{i_1,i_2} H_{i_3,i_3} \)
6. \( \sum_{i_1 \neq i_2 \neq i_3} \left( \frac{\nabla_{i_3} - \nabla_{i_1}}{(x_{i_3} - x_{i_2})(x_{i_3} - x_{i_1})} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_3} - x_{i_2})(x_{i_2} - x_{i_1})} \right) f(x) H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} \)
7. \( \sum_{i_1 \neq i_2 \neq i_3} \left( \frac{\nabla_{i_2} - \nabla_{i_3}}{(x_{i_3} - x_{i_1})(x_{i_2} - x_{i_3})} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_3} - x_{i_1})(x_{i_2} - x_{i_1})} \right) f(x) H_{i_1,i_3} H_{i_2,i_1} H_{i_3,i_2} \)
Recall: Goal is to upper bound
\[ \left| \mathbb{E}_{x, g} \left[ D^3 B_\lambda(P'_{x,g})[Q_x, Q_x, Q_x] - D^3 B_\lambda(R'_{x,g})[Q_g, Q_g, Q_g] \right] \right| \]

**Technical contribution.**
- Bound each of the 7 terms by polylog \( k \) times norms of \( Q_g, Q_x, P'_{x,g}, R'_{x,g} \).
- Completely open up the Bentkus mollifier (prior works used it as a blackbox)

4. **Final step.** Understand \( \mathbb{E}_x[\| Q_x \|_4^4] \) and similar quantities.
- We use matrix Rosenthal’s inequality (proved “recently”) gives good concentration for Schatten norms of \( \| \sum_i x_i A^i \|_p^p \)
- Also matrix Rosenthal is true when \( (x_1, \ldots, x_n) \) is \( p \)-wise independent

Putting everything together.
\[ \left| \Pr_{x \sim \mathcal{U}_n} [(B \circ \lambda)(\sum_i x_i A^i - B)] - \Pr_{g \sim \mathcal{G}_n} [(B \circ \lambda)(\sum_i g_i A^i - B)] \right| \leq \tau \cdot \text{polylog } k \]
Recall: positive spectrahedron $S = \{ x : x_1 A^1 + \cdots + x_n A^n \preceq B \}$ where $A^i, B \succeq 0$

So far.

$$\left| \Pr_{x \sim U_n} \left[ (B \circ \lambda) \left( \sum_i x_i A^i - B \right) \right] - \Pr_{g \sim G^n} \left[ (B \circ \lambda) \left( \sum_i g_i A^i - B \right) \right] \right| \leq \text{polylog } k \quad (5)$$

$$\approx \left[ \sum_i x_i A^i \preceq B \right] \quad \approx \left[ \sum_i g_i A^i \preceq B \right]$$

But we care about CDF distance

$$\left| \Pr_{x \sim U_n} \left[ \sum_i x_i A^i \preceq B \right] - \Pr_{g \sim G^n} \left[ \sum_i g_i A^i \preceq B \right] \right| \leq \text{polylog } k \quad (6)$$

Intuition for this approximation: What $\approx$ means?

If $\lambda_1, \ldots, \lambda_k \in [-1/100, 1/100]$, then $(B \circ \lambda) \left( \sum_i x_i A^i - B \right) \not\approx \left[ \sum_i x_i A^i \preceq B \right]$

Else $\lambda_1, \ldots, \lambda_k \not\in [-1/100, 1/100]$, then $(B \circ \lambda) \left( \sum_i x_i A^i - B \right) \approx \left[ \sum_i x_i A^i \preceq B \right]$

For uniform $x$, $\lambda_{\text{max}} \left( \sum_i x_i A^i - B \right) \in [-\frac{1}{100}, \frac{1}{100}]$ with tiny probability
Recall: positive spectrahedron $S = \{x : x_1 A^1 + \cdots + x_n A^n \preceq B\}$ where $A^i, B \succeq 0$

So far.

$$\left| \Pr_{x \sim \mathcal{U}_n} \left[ (B \circ \lambda)(\sum_i x_i A^i - B) \right] - \Pr_{g \sim \mathcal{G}_n} \left[ (B \circ \lambda)(\sum_i g_i A^i - B) \right] \right| \leq \text{polylog } k \quad (7)$$

$$\approx \left[ \sum_i x_i A^i \preceq B \right] \quad \approx \left[ \sum_i g_i A^i \preceq B \right]$$

But we care about CDF distance

$$\left| \Pr_{x \sim \mathcal{U}_n} \left[ \sum_i x_i A^i \preceq B \right] - \Pr_{g \sim \mathcal{G}_n} \left[ \sum_i g_i A^i \preceq B \right] \right| \leq \text{polylog } k \quad (8)$$

Our result: Littlewood-Offord for spectrahedra

Let $A^1, \ldots, A^n$ be positive matrices s.t. $\sum_i \|A^i\|^2 \geq 1$. For every $\Lambda$

$$\Pr_{x \sim \mathcal{U}_n} \left[ \lambda_{\max} \left( \sum_i x_i A^i - B \right) \in [-\Lambda, \Lambda] \right] \leq O(\Lambda).$$

Prior: Littlewood-Offord’43, Erdős’45 proved it for halfspaces, OST’19 for polytopes

Hence (7) implies (8) except tiny probability. Done!
Recall: positive spectrahedron \( F(x) = [x_1 A^1 + \cdots + x_n A^n \preceq B] \) where \( A^i, B \succeq 0 \)

A PRG that \( \delta \)-fools the class of positive width-\( M \) spectrahedra with seed length \( \text{poly}(\log n, \log k, M, 1/\delta) \)

Open questions:

1. Remove regularity?
2. Remove positivity?
3. What is the Gaussian surface area of spectrahedron?
4. Improve the \( 1/\delta \) dependence?
5. A general invariance principle for spectral functions?

THANK YOU