# Polynomial time guarantees for the Burer-Monteiro method

Diego Cifuentes

Department of Mathematics
Massachusetts Institute of Technology

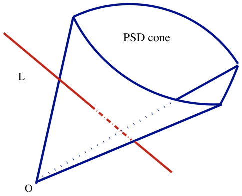
Joint work with **Ankur Moitra** (MIT)
arXiv:1912.01745

Applied Math Seminar - UMass Lowell - 2020

# Semidefinite programming

A *semidefinite program* is

$$(\text{SDP}) \qquad \min_{X \in \mathbb{S}^n} \quad C \bullet X \quad \text{s.t.} \quad A_i \bullet X = b_i \text{ for } i \in [m], \quad X \succeq 0$$



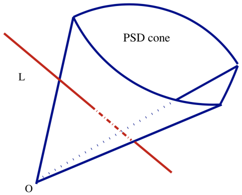

where $C, A_1, \ldots, A_m \in \mathbb{S}^n$, $b \in \mathbb{R}^m$.

- This is a *convex* problem.
- Polynomial time solvable with interior point methods (Nesterov-Nemirovski'87).
- Many different applications.

# Semidefinite programming

A *semidefinite program* is

$$\text{(SDP)} \qquad \begin{aligned} \min_{X \in \mathbb{S}^n} \quad & C \bullet X \\ \text{s.t.} \quad & A_i \bullet X = b_i \text{ for } i \in [m] \\ & X \succeq 0 \end{aligned}$$

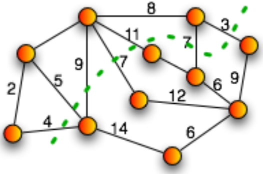



where $C, A_1, \ldots, A_m \in \mathbb{S}^n,\ b \in \mathbb{R}^m$.

- This is a *convex* problem.
- Polynomial time solvable with interior point methods (Nesterov-Nemirovski'87).
- Many different applications.

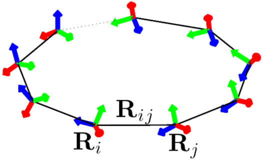
**Goal:** Show that SDPs can be solved in polynomial time with a more recent class of methods. The proof is remarkably geometric.

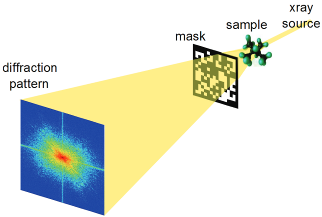
# SDP applications

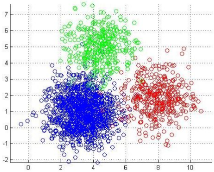



| | | | | |
|---|---|---|---|---|
| | | -1 | | |
| | | | 1 | |
| 1 | 1 | -1 | 1 | -1 |
| 1 | | | | -1 |
| | | -1 | | |

| | | | | |
|---|---|---|---|---|
| 1 | 1 | -1 | 1 | -1 |
| 1 | 1 | -1 | 1 | -1 |
| 1 | 1 | -1 | 1 | -1 |
| 1 | 1 | -1 | 1 | -1 |
| 1 | 1 | -1 | 1 | -1 |



# Solving large scale SDPs

Interior point methods intractable for large SDPs (too much memory)

**Practical methods**

- Low rank factorization (Burer, Monteiro)
- Frank-Wolfe / CGM (Hazan, Jaggi, Freund, Grigas, Mazumder)
- Sketching (Yurtsever, Ding, Udell, Tropp, Cevher)
- Bundle (Helmberg, Rendl, Oustry)
- Subgradients (Nesterov, Yurtsever, Tran Dinh, Cevher)

The Burer-Monteiro method is one of the most widely used in practice.

# Burer-Monteiro method

$$\text{(SDP)} \qquad \min_{X \in \mathbb{S}^n} \quad C \bullet X \quad \text{s.t.} \quad A_i \bullet X = b_i \text{ for } i \in [m], \quad X \succeq 0$$

Assume the optimal solution has rank $\leq p$, and factorize

$$X = YY^T \quad \text{where } Y \in \mathbb{R}^{n \times p}$$

Use local optimization to solve the *nonconvex* problem

$$\text{(BM)} \qquad \min_{Y \in \mathbb{S}^n} \quad C \bullet YY^T \quad \text{s.t.} \quad A_i \bullet YY^T = b_i \text{ for } i \in [m]$$

How large should we choose $p$?

# Burer-Monteiro method

(SDP) $$\min_{X \in \mathbb{S}^n} \quad C \bullet X \quad \text{s.t.} \quad A_i \bullet X = b_i \text{ for } i \in [m], \quad X \succeq 0$$

Assume the optimal solution has rank $\leq p$, and factorize

$$X = YY^T \quad \text{where } Y \in \mathbb{R}^{n \times p}$$

Use local optimization to solve the *nonconvex* problem

(BM) $$\min_{Y \in \mathbb{S}^n} \quad C \bullet YY^T \quad \text{s.t.} \quad A_i \bullet YY^T = b_i \text{ for } i \in [m]$$

How large should we choose $p$?

Theorem (Barvinok-Pataki bound)

*BM is equivalent to SDP when* $\binom{p+1}{2} \geq m$. *So we need* $p \gtrsim \sqrt{2m}$.

# Known results about BM

**Below Barvinok-Pataki bound**

There might be spurious local minima [Waldspurger, Waters].

# Known results about BM

**Below Barvinok-Pataki bound**
There might be spurious local minima [Waldspurger, Waters].

**Above Barvinok-Pataki bound**
Assume $\binom{p+1}{2} > m$.

- BM *should* work [Burer-Monteiro, Journée et al.]

# Known results about BM

**Below Barvinok-Pataki bound**

There might be spurious local minima [Waldspurger, Waters].

**Above Barvinok-Pataki bound**

Assume $\binom{p+1}{2} > m$.

- BM *should* work [Burer-Monteiro, Journée et al.]

Assume the feasible set is smooth.

- For generic $C$, no spurious local minima [Boumal-Voroninski-Bandeira]

# Known results about BM

**Below Barvinok-Pataki bound**
There might be spurious local minima [Waldspurger, Waters].

**Above Barvinok-Pataki bound**
Assume $\binom{p+1}{2} > m$.

- BM *should* work [Burer-Monteiro, Journée et al.]

Assume the feasible set is smooth.

- For generic $C$, no spurious local minima [Boumal-Voroninski-Bandeira]

Assume compactness and $\binom{p+1}{2} > \frac{9}{2} m \log(\sigma^{-1})$ for some $\sigma > 0$.

- Randomly perturb $C$ (magnitude $\sigma$). No spurious approximate local minima w.h.p. [Pumir-Jelassi-Boumal]

# Known results about BM

**Below Barvinok-Pataki bound**
There might be spurious local minima [Waldspurger, Waters].

**Above Barvinok-Pataki bound**
Assume $\binom{p+1}{2} > m$.

- BM *should* work [Burer-Monteiro, Journée et al.]

Assume the feasible set is smooth.

- For generic $C$, no spurious local minima [Boumal-Voroninski-Bandeira]

Assume compactness and $\binom{p+1}{2} > \frac{9}{2} m \log(\sigma^{-1})$ for some $\sigma > 0$.

- Randomly perturb $C$ (magnitude $\sigma$). No spurious approximate local minima w.h.p. [Pumir-Jelassi-Boumal]

No polynomial time guarantees known!
*Obstacle:* How to find a local minimum that is exactly feasible?

# Our results

*Assumptions:*

- $\binom{p+1}{2} > (1+\eta)m$ for a fixed $\eta > 0$.
- Constraint set is smooth and compact.
- Solve BM using local method with 2nd order guarantees.

**Theorem (Polytime optimality)**

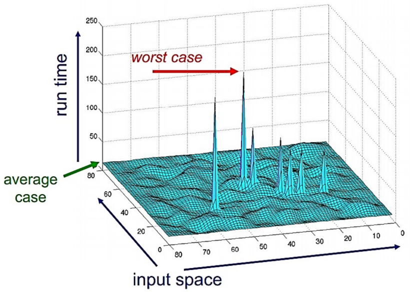
*Randomly perturb C (magnitude $\sigma$). Given an initial approx feasible point, then BM computes a point that is approx feasible & approx optimal w.h.p. in $\text{poly}(n, \sigma^{-1})$ iterations.*

# Smoothed analysis [Spielman-Teng '01]

**Theorem (Polytime optimality)**

*Randomly perturb C (magnitude $\sigma$). Given an initial approx feasible point, then BM computes a point that is approx feasible & approx optimal w.h.p. in* $\mathrm{poly}(n, \sigma^{-1})$ *iterations.*

worst-case complexity

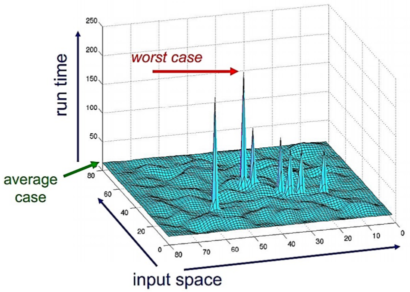


# Smoothed analysis [Spielman-Teng '01]

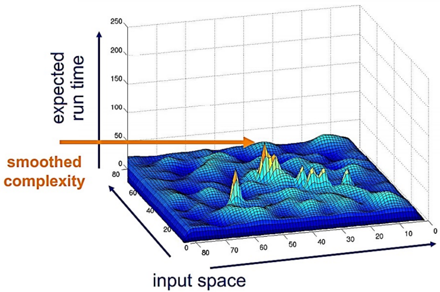
**Theorem (Polytime optimality)**

*Randomly perturb C (magnitude $\sigma$). Given an initial approx feasible point, then BM computes a point that is approx feasible & approx optimal w.h.p. in* $\text{poly}(n, \sigma^{-1})$ *iterations.*

worst-case complexity




smoothed complexity



# BM for SDP feasibility

Goal:

$$\text{find } X \quad \text{s.t.} \quad A_i \bullet X \approx b_i \text{ for } i \in [m], \quad X \succeq 0$$

Consider the least squares problem:

$$(\text{SDP}_{ls}) \qquad \min_{X \in \mathbb{S}^n} \sum_i (A_i \bullet X - b_i)^2 \quad \text{s.t.} \quad X \succeq 0$$

The associated Burer-Monteiro problem is

$$(\text{BM}_{ls}) \qquad \min_{Y \in \mathbb{S}^n} \sum_i (A_i \bullet YY^T - b_i)^2$$

Previous work on $\text{BM}_{ls}$ relies on RIP [Bhojanapalli-Neyshabur-Srebro].

# Our results

*Assumptions:*

- $\binom{p+1}{2} > (1+\eta)m$ for a fixed $\eta > 0$
- Constraint set is smooth and compact
- Solve BM using local method with 2nd order guarantees.

**Theorem (Polytime optimality)**

*Randomly perturb $C$ (magnitude $\sigma$). Given an initial approx feasible point, then BM computes a point that is approx feasible & approx optimal w.h.p. in* $\text{poly}(n, \sigma^{-1})$ *iterations.*

**Theorem (Polytime feasibility)**

*Randomly perturb $A_1, \ldots, A_m$ (magnitude $\sigma$). Then $BM_{ls}$ computes an approx feasible point w.h.p. in* $\text{poly}(n, \sigma^{-1})$ *iterations.*

# Optimality/Criticality conditions

(SDP) $\quad \min \quad C \bullet X \quad \text{s.t.} \quad A_i \bullet X = b_i \text{ for } i \in [m], \quad X \succeq 0$

**Lemma**

*Assuming strong duality, $X$ is optimal for SDP iff there exists $\lambda$ such that*

$$A_i \bullet X = b_i, \quad X \succeq 0, \quad S(\lambda)X = 0, \quad S(\lambda) \succeq 0,$$

*where $S(\lambda) := C - \sum_i \lambda_i A_i$.*

# Optimality/Criticality conditions

(SDP) $\min \quad C \bullet X \quad \text{s.t.} \quad A_i \bullet X = b_i \text{ for } i \in [m], \quad X \succeq 0$

**Lemma**

*Assuming strong duality, $X$ is optimal for SDP iff there exists $\lambda$ such that*

$$A_i \bullet X = b_i, \quad X \succeq 0, \quad S(\lambda)X = 0, \quad S(\lambda) \succeq 0,$$

*where $S(\lambda) := C - \sum_i \lambda_i A_i$.*

(BM) $\min \quad C \bullet YY^T \quad \text{s.t} \quad A_i \bullet YY^T = b_i \text{ for } i \in [m]$

**Lemma**

*$Y$ is a 2-critical point of BM iff there exists $\lambda$ such that*

$$A_i \bullet YY^T = b_i, \quad S(\lambda)Y = 0, \quad S(\lambda) \bullet UU^T \geq 0 \;\; \forall U : A_i \bullet UY^T = 0$$

A critical point $Y$ of BM is spurious if $YY^T$ is not optimal for SDP.

# Spurious critical points

A critical point $Y$ of BM is spurious if $YY^T$ is not optimal for SDP.

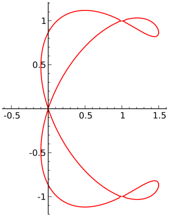
**Theorem (Boumal-Voroninski-Bandeira)**

*Spurious critical points may only exist if*

$$C \in \mathcal{M}_{n-p} + \mathcal{L} \subset \mathbb{S}^n$$

*where* $\mathcal{M}_{n-p} = \{S : \mathrm{rank} S \leq n-p\}$ *and* $\mathcal{L} = \mathrm{span}\{A_1, \ldots, A_m\}$

- Assume that $\binom{p+1}{2} > m$.
- $\mathcal{M}_{n-p} + \mathcal{L}$ is an *algebraic variety* of codimension $\binom{p+1}{2} - m$.
- Generically, no spurious points.




# Approximate optimality/criticality

**(SDP)** $X$ is $\delta$-optimal if there exists $\lambda$ such that

$$\|A_i \bullet X - b_i\| \leq \delta, \qquad \|S(\lambda)X\| \leq \delta, \qquad X \succeq 0, \qquad S(\lambda) \succeq -\delta I_n.$$

**(BM)** $Y$ is $\epsilon$-critical for BM if there exists $\lambda$ such that

$$\|A_i \bullet YY^T - b_i\| \leq \epsilon, \qquad \|S(\lambda)Y\| \leq \epsilon^2,$$
$$S(\lambda) \bullet UU^T \geq -\epsilon \quad \forall U : \|U\| \leq 1,\ \|A_i \bullet UY^T\| \leq \epsilon.$$

Theorem (C.-Moitra)

*Assume domain is smooth and compact. Given an approx feasible point, local optimization can produce an approx critical point in* $\mathrm{poly}(\epsilon^{-1})$*.*

# Spurious approximate critical points

An $\epsilon$-critical point $Y$ of BM is spurious if $YY^T$ is not $\delta$-optimal for SDP, with $\delta = O(\epsilon)$.

# Spurious approximate critical points

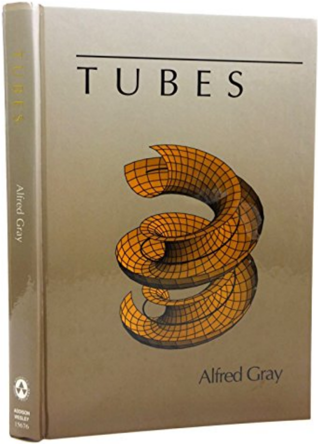
An $\epsilon$-critical point $Y$ of BM is spurious if $YY^T$ is not $\delta$-optimal for SDP, with $\delta = O(\epsilon)$.

**Theorem (C.-Moitra)**

*Spurious $\epsilon$-critical points may only exist if*

$$C \in \text{tube}_\epsilon(\mathcal{M}_{n-p} + \mathcal{L}) \subset \mathbb{S}^n$$

*where* $\text{tube}_\epsilon(\mathcal{W}) := \{X : \text{dist}(X, \mathcal{W}) \leq \epsilon\}$



# Volumes of tubes

**Theorem (Weyl 1939)**

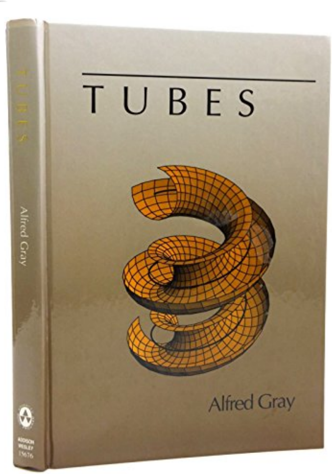
*Let $V \subset \mathbb{R}^n$ manifold of codimension $c$. There are curvature constants $k_i(V)$ such that*

$$\mathrm{Vol}[\mathrm{tube}_\epsilon(V)] = \sum_{i=c}^{n} k_i(V)\, \epsilon^i$$

**Theorem (Lotz 2015)**

*Let $V \subset \mathbb{R}^n$ variety of codimension $c$ defined by polynomials of degree $D$. Let $x$ uniformly distributed on a ball of radius $\sigma$. Then*

$$\Pr[x \in \mathrm{tube}_\epsilon(V)] \leq O(nD\epsilon/\sigma)^c$$



# Polytime optimality

**Theorem (C.-Moitra)**

Assumptions: $\binom{p+1}{2} > (1+\eta)m$, *constraint set smooth and compact, solve BM with local method with 2nd order guarantees.*

*Randomly perturb C (magnitude $\sigma$). Given an approx feasible point, BM computes an approx optimal solution w.h.p. in* $\text{poly}(n, \sigma^{-1})$ *iterations.*

# Polytime optimality

**Theorem (C.-Moitra)**

Assumptions: $\binom{p+1}{2} > (1+\eta)m$, *constraint set smooth and compact, solve BM with local method with 2nd order guarantees.*

*Randomly perturb C (magnitude $\sigma$). Given an approx feasible point, BM computes an approx optimal solution w.h.p. in* $\text{poly}(n, \sigma^{-1})$ *iterations.*

**Proof.**

Method converges to an $\epsilon$-critical point in $\text{poly}(\epsilon^{-1})$. Using tubes,

$$\Pr[\text{spurious}] \;\leq\; \Pr[C \in \text{tube}_\epsilon(\mathcal{M}_{n-p} + \mathcal{L})] \;\leq\; \epsilon^{\binom{p+1}{2}-m} \cdot O\left(n^3/\sigma\right)^{\binom{p+1}{2}}$$

For $\binom{p+1}{2} > (1+\eta)m$ and $\epsilon = O(\sigma/n^3)^{1+1/\eta}$ the probability is tiny. □

# Polytime feasibility

$$(\mathrm{BM}_{ls}) \qquad \min_{Y\in\mathbb{S}^n} \quad \sum_i (A_i \bullet YY^T - b_i)^2$$

**Theorem (C.-Moitra)**

*Spurious $\epsilon$-critical points may only exist if*

$$\mathrm{tube}_\epsilon(\mathcal{M}_{n-p}) \cap \mathcal{L} \text{ nontrivial}$$

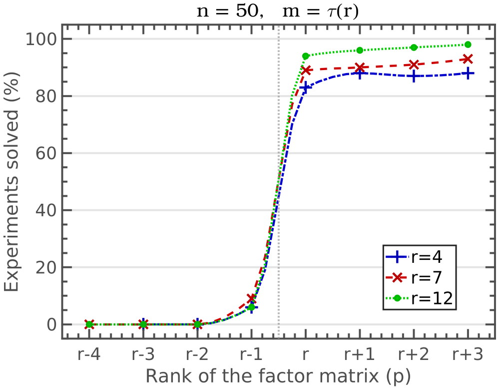
**Theorem (C.-Moitra)**

Assumptions: $\binom{p+1}{2} > (1+\eta)m$, *constraint set smooth and compact, solve $\mathrm{BM}_{ls}$ with local method with 2nd order guarantees.*

*Randomly perturb $A_1, \ldots, A_m$ (magnitude $\sigma$). Then $\mathrm{BM}_{ls}$ computes an approximately feasible solution w.h.p. in $\mathrm{poly}(n, \sigma^{-1})$ iterations.*

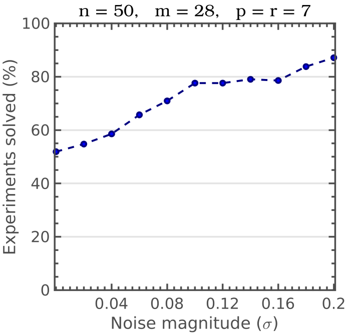
# Experiments

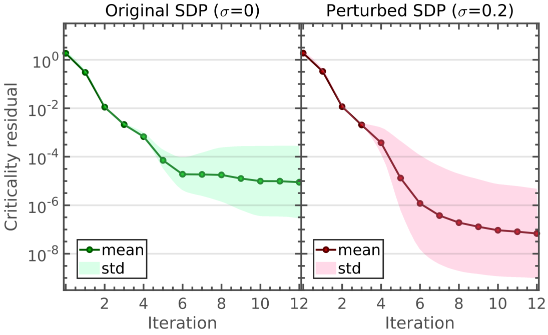
- Generate planted matrix of rank $r$.
- Generate SDP for which planted marix is optimal.
- Solve BM using augmented Lagrangians.



# Experiments

- Fix an SDP instance for which BM behaves badly.
- Perturb the problem with small noise.
- Solve BM using augmented Lagrangians.



# Summary

- We proved the first polynomial time guarantees for the Burer-Monteiro method.
- Guarantees work arbitrarily close to the Barvinok-Pataki bound.
- Proof relies on geometric ideas (varieties, tubes).

# Summary

- We proved the first polynomial time guarantees for the Burer-Monteiro method.
- Guarantees work arbitrarily close to the Barvinok-Pataki bound.
- Proof relies on geometric ideas (varieties, tubes).

References:


- D. Cifuentes, A. Moitra, *Polynomial time guarantees for the Burer-Monteiro method*, `arXiv:1912.01745`.
- D. Cifuentes *Burer-Monteiro guarantees for general semidefinite programs*, `arXiv:1904.07147`.

# Thanks for your attention