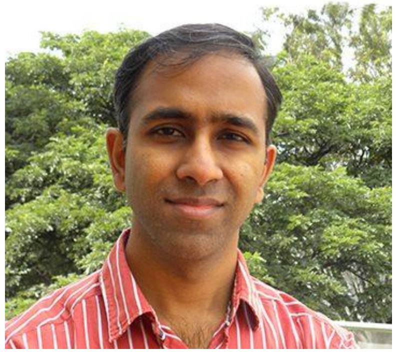
Smoothed analysis for low-rank solutions to SDPs

Praneeth Netrapalli
Microsoft Research India

Joint work with:





Srinadh Bhojanapalli
TTI Chicago

Nicolas Boumal
Princeton

Prateek Jain
MSR India

Semi-definite programs (SDPs)

$$\min_{X \in R^{n \times n}} \langle C, X \rangle \quad s.t. \quad \langle A_i, X \rangle = b_i, 1 \leq i \leq m$$
$$X \succcurlyeq 0$$

- Several applications
  - Clustering (max-cut)
  - Control
  - Sum-of-squares
  - …
- Polynomial time solutions exist but can be slow
  - Interior-point methods
  - Multiplicative weight update

Burer-Monteiro 2003

- Much faster
- Empirically works well
- No proof of correctness

Low rank solutions always exist!

- (Barvinok'95, Pataki'98): For any feasible SDP, at least one solution exists with rank $k^* \leq \sqrt{2m}$
- In several applications $m \sim n$. So $k^* \ll n$.

Burer-Monteiro: Optimize in low rank space; iterations are fast!

Burer-Monteiro factorization

$$\min_{X \in R^{n\times n}} \langle C, X\rangle$$
$$s.t. \quad \langle A_i, X\rangle = b_i;\ i = 1, \cdots, m$$
$$X \succcurlyeq 0$$

$$\min_{U \in R^{n\times k}} \langle C, UU^T\rangle$$
$$s.t. \quad \langle A_i, UU^T\rangle = b_i$$

$$k \sim \sqrt{m}$$

n^2 dimensional problem

nk dimensional problem

Burer-Monteiro factorization

$$\min_{X \in R^{n \times n}} \langle C, X \rangle$$
$$s.t.\quad \langle A_i, X \rangle = b_i;\ i = 1, \cdots, m$$
$$X \succcurlyeq 0$$

$$\min_{U \in R^{n \times k}} \langle C, UU^T \rangle$$
$$s.t.\quad \langle A_i, UU^T \rangle = b_i$$

$$k \sim \sqrt{m}$$

Penalty Version

Penalty parameter

$$\min_{U \in R^{n \times k}} f(U) = \langle C, UU^T \rangle + \mu \sum_i \left(\langle A_i, UU^T \rangle - b_i \right)^2$$

Nonconvex problem!

What can be done for nonconvex problems?

- First order stationary points (FOSP)

$$\|\nabla f(x)\| \leq \epsilon$$

- Second order stationary points (SOSP)

$$\|\nabla f(x)\| \leq \epsilon \text{ and } \nabla^2 f(x) \geq -\epsilon \mathbb{I}$$

- Lot of recent work on how to find SOSPs efficiently

Low rank SDP

$$\min_{U \in R^{n \times k}} f(U) = \langle C, UU^T \rangle + \mu \sum_i \left(\langle A_i, UU^T \rangle - b_i \right)^2$$

Boumal et al. 2016: if $k \geq \sqrt{2m}$, for almost all C, SOSP $=$ global optimum

Open questions

- Are there C for which SOSP $\neq$ global optimum?
- Are approximate SOSP $=$ approximate global optima?

Our results

- Yes, there are C for which SOSP $\neq$ global optimum
- Yes, for perturbed SDPs,

  approximate SOSP $=$ approximate global optima

Smoothed analysis

$$\min_{U \in R^{n\times k}} f(U) = \langle C, UU^T \rangle + \mu \sum_i \left(\langle A_i, UU^T \rangle - b_i \right)^2$$

$$\min_{U \in R^{n\times k}} f(U) = \langle C + G, UU^T \rangle + \mu \sum_i \left(\langle A_i, UU^T \rangle - b_i \right)^2$$

- G: symmetric Gaussian matrix with $G_{ij} \sim N(0, \sigma_G^2)$
  - $\sigma_G \approx \Omega(\epsilon)$

- If $k = \Omega(\sqrt{m \log 1/\epsilon})$ then with high probability

every ϵ SOSP = ϵ global optimum

Main Ideas of the Proof

Two key steps

1. SOSP that is rank deficient is global optimum [Burer-Monteiro 2003]

U SOSP and $\sigma_k(U) = 0 \Rightarrow U$ is a global optimum

↳ k^{th} largest singular value of U

2. For perturbed SDPs, with probability 1, if $k \geq \sqrt{2m}$, then

all FOSPs have $\sigma_k(U) = 0$. [Boumal et al. 2016]

Two key steps

$\min_{U \in \mathbb{R}^{n \times k}} f(UU^\top)$ $f(\cdot)$ convex

1. approximate U SOSP and $\sigma_k(U)$ small $\Rightarrow$ U is a approximate global optimum

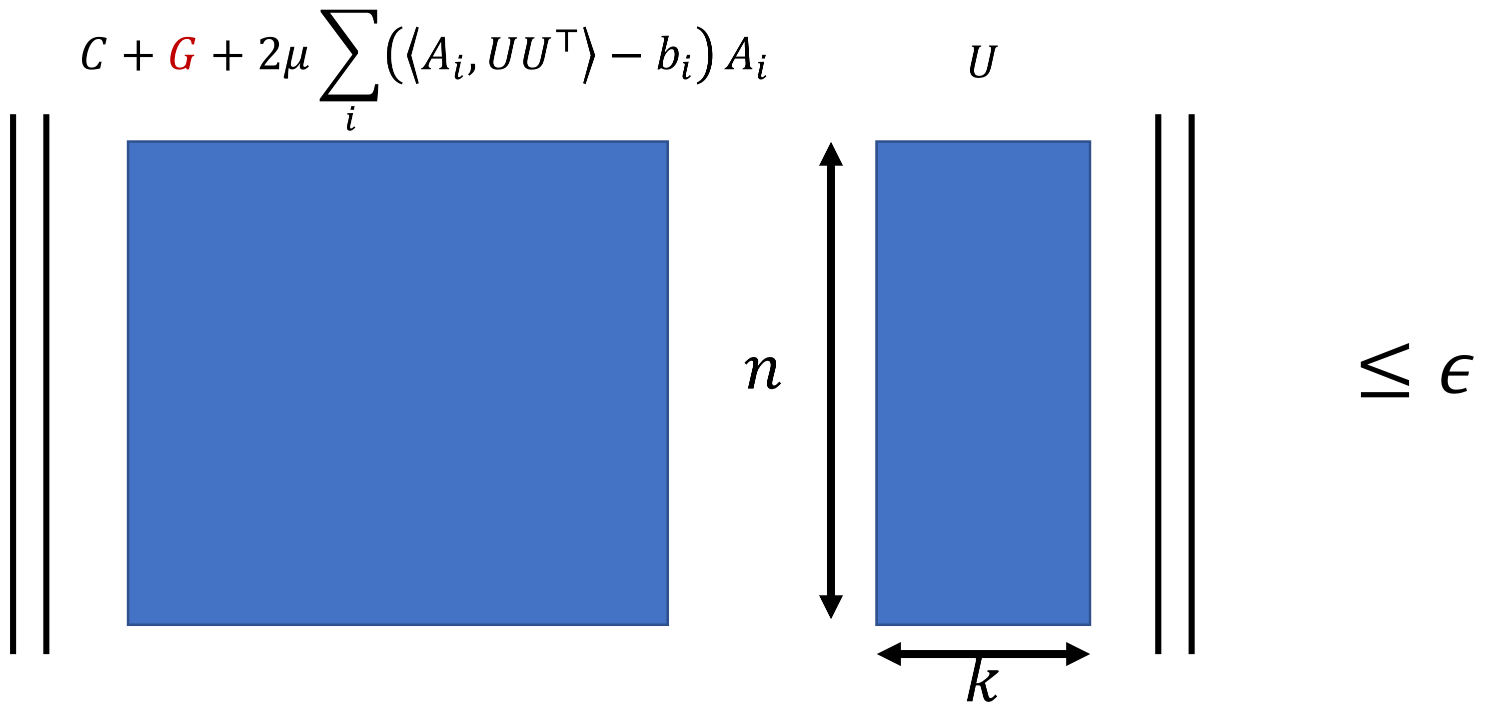
2. For perturbed SDPs, with ~~probability 1~~ high probability, if ~~$k \geq \sqrt{2m}$~~ $k \geq \sqrt{m \log 1/\epsilon}$, then

approximate all FOSPs have small $\sigma_k(U)$.

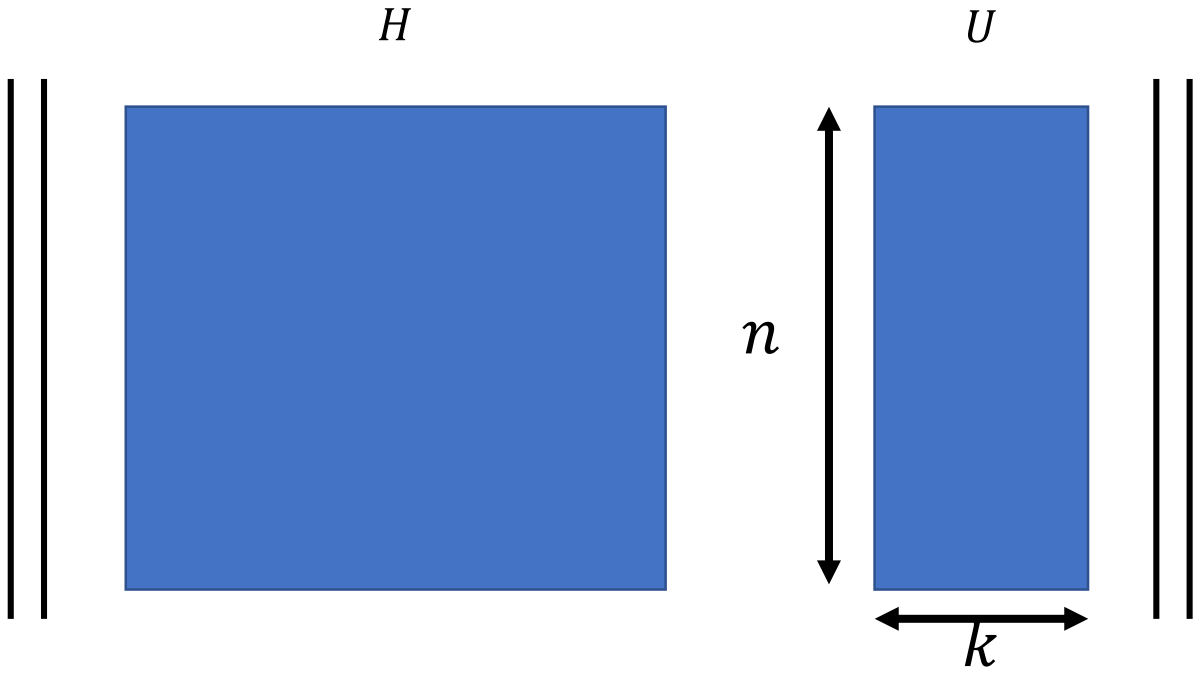
FOSP $\Rightarrow \sigma_k(U)$ is small

$$\min_{U \in R^{n\times k}} f(U) = \langle C + G, UU^T \rangle + \mu \sum_i \left(\langle A_i, UU^T \rangle - b_i \right)^2$$

- Approximate FOSP: $\left\| \left(C + G + 2\mu \sum_i \left(\langle A_i, UU^\top \rangle - b_i\right) A_i \right) U \right\| \le \epsilon$



Aside: Lower bound on product of matrices

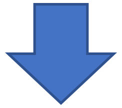


$$\geq \sigma_{n-k+1}(H) \cdot \sigma_k(U)$$

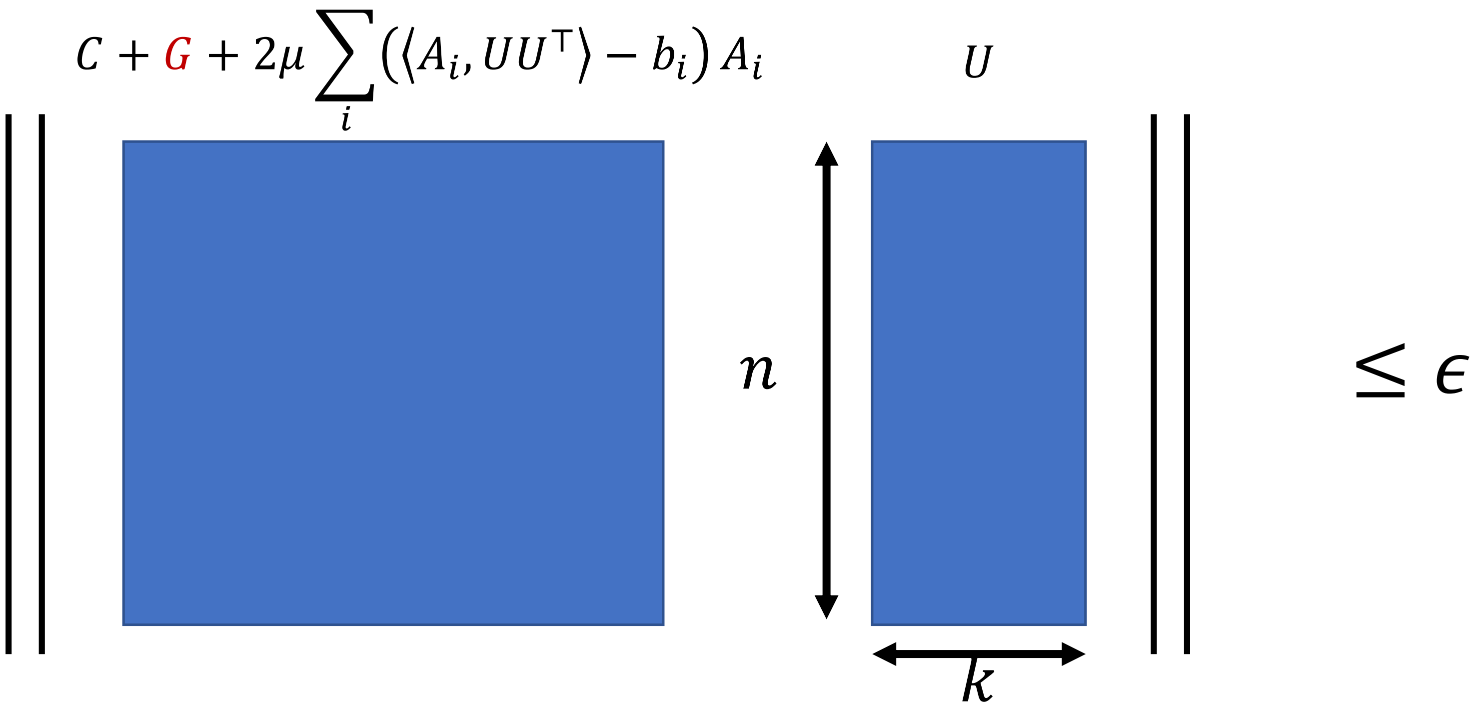
$$\sigma_k(U) \leq \frac{\|HU\|}{\sigma_{n-k+1}(H)}$$

FOSP $\Rightarrow \sigma_k(U)$ is small

$\sigma_{n-k+1}\left(C + G + 2\mu \sum_i \left(\langle A_i, UU^\top \rangle - b_i\right) A_i\right)$ large



$\sigma_k(U)$ is small



Smallest singular values of Gaussian matrices

- $\sigma_i(G)$ denotes the i^{th} singular value of G.

$$\mathbb{P}[\sigma_n(G) = 0] = 0$$

- In general, $\sigma_{n-k}(G) \sim \frac{k}{\sqrt{n}}$.
- Can obtain large deviation bounds [Nguyen 2017]

$$\mathbb{P}\left[\sigma_{n-k}(G) < c\,\frac{k}{\sqrt{n}}\right] < \exp\!\left(-Ck^2 + k\log n\right)$$

- Can extend the above to $G + A$ for any fixed matrix A

Need $k^2 \geq m \log 1/\epsilon$

Coming back to SDPs

Unknown quantity

- Interested in bounding

$$\sigma_{n-k}\left(G + C + 2\mu \sum_{i=1}^{m} \overbrace{\left(\langle A_i, UU^\top \rangle - b_i\right)}A_i\right)$$

- Do an ϵ-net of $(\lambda_1, \cdots, \lambda_m) \in \mathbb{R}^m$ and apply large deviation bound for

$$\mathbb{P}\left[\sigma_{n-k}\left(G + C + \sum_{i=1}^{m} \lambda_i A_i\right) < c\frac{k}{\sqrt{n}}\right] < \exp\left(-Ck^2 + k\log n\right)$$

- Taking union bound over ϵ-net gives additional $\left(\frac{1}{\epsilon}\right)^m$ factor

Technical issues

- Can do ϵ-net only over a finite size ball
- Need to show $\langle A_i, UU^\top \rangle - b_i$ does not become unbounded at SOSPs
- Requires us to show that all SOSPs are uniformly bounded
- Can show this for compact SDPs i.e., feasible set is compact
- Not obvious – SOSPs in nonconvex world may be infeasible

Approx low rank SOSP ⇒ approx. global opt

- $f(\cdot)$ convex: $UU^\top$ suboptimal ⇒ there exists descent direction
- In fact, ∃ descent direction increasing the rank by at most 1
- If U was rank deficient, this direction exists in factorized space
- Since U is approx. rank deficient, can construct a direction that does not increase the rank

Summary

- Low rank solutions to SDPs useful from both application and algorithmic perspectives
- Burer-Monteiro approach tries to leverage this idea
- May not work in the worst case
- This work: Burer-Monteiro works in the smoothed analysis sense

Open directions

- We believe the results are not tight

- Extension of these results to augmented Lagrangian methods (ALM)
  - The one actually used in practice
  - Significantly better than penalty methods

- Preliminary results on exact-ALMs but no results for inexact ALMs

- Obtain solutions of rank $\ll \sqrt{m}$ for special problems

Open directions – random matrix theory

- Main bottleneck in our results – ϵ-net argument
- Distance of a random matrix from a subspace Well understood!
- Distance of a random matrix from low rank matrices Well understood!
- Distance of a random matrix from a subspace + low rank matrices

  Not understood!
- Leads to interesting Mathematical questions + applications (in SDPs)

Thank you!

Questions?