Large-scale optimization with positive semidefinite matrices

Hamza Fawzi

University of Cambridge

CCIMI Conference 2020

Matrix optimization

- Many optimization problems involve *matrix variables*
- One often encounters *positive semidefinite* constraint on matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$:
  - $X_{ij} = X_{ji}$ (i.e., $\mathbf{X}$ is symmetric)
  - All eigenvalues of $\mathbf{X}$ are nonnegative

  Notation: $\mathbf{X} \succeq 0$
- Applications:
  - Learning covariance matrix from data
  - Graph partitioning
  - Clustering using ellipsoids
  - Robustness of neural networks
  - ...

Covariance estimation

- Estimate covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ from n samples of a Gaussian vector $\mathbf{x} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$

- Sample covariance matrix

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$$

- Enforcing sparsity of $\boldsymbol{\Sigma}^{-1}$ reveals conditional independence structure in $\mathbf{x}$.

- Optimization formulation

$$\max_{\mathbf{K} \succ 0} \quad \log\det \mathbf{K} - \mathrm{Tr}\left[\mathbf{KS}\right] - \lambda \|\mathbf{K}\|_1$$

Term $\|\mathbf{K}\|_1$ promotes sparsity of $\mathbf{K}$.
Approach due to [Yuan-Lin, d'Aspremont et al, Friedman et al.]

Graph partitioning

Cluster graph G nodes into two communities A and B, so that

- Connections within communities $\gg$ connections across communities
- A and B have similar size

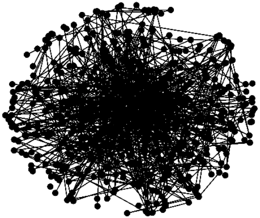


Figure from "Exact recovery in the stochastic block model", by Abbe, Bandeira, Hall

Graph partitioning

Cluster graph G nodes into two communities A and B, so that

- Connections within communities $\gg$ connections across communities
- A and B have similar size

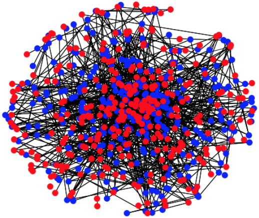


Figure from "Exact recovery in the stochastic block model", by Abbe, Bandeira, Hall

Graph partitioning

Cluster graph G nodes into two communities A and B, so that

- Connections within communities $\gg$ connections across communities
- A and B have similar size

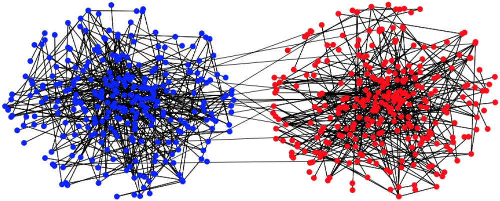


Figure from "Exact recovery in the stochastic block model", by Abbe, Bandeira, Hall

Graph partitioning

Cluster graph G nodes into two communities A and B, so that

- Connections within communities $\gg$ connections across communities
- A and B have similar size

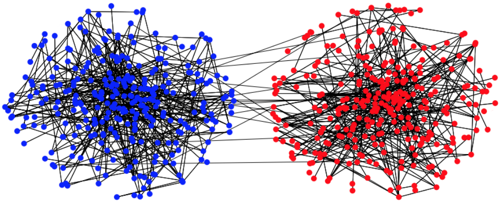


Figure from "Exact recovery in the stochastic block model", by Abbe, Bandeira, Hall

Popular optimization-based solution:

$$\max_{\mathbf{X} \in \mathbb{R}^{n \times n}} \sum_{1 \le i,j \le n} B_{ij} X_{ij} \quad \text{s.t.} \quad \mathbf{X} \succeq 0, \quad X_{ii} = 1 \; \forall 1 \le i \le n.$$

where $B_{ij} = \begin{cases} +1 & \text{if } i,j \text{ connected} \\ -1 & \text{otherwise} \end{cases}$

Many other applications

- Estimating robustness of neural networks
  [Raghunatan et al., Hassani et al., Dathatri et al., ...]
- Quantum information [Gross et al., Navascués et al.,]
- Control and dynamical systems [Boyd et al., ...]
- Polynomial optimization [Parrilo, Lasserre, ...]
- ...

General structure of the problem

$$\begin{array}{ll} \min_{\mathbf{X} \in \mathbb{R}^{n \times n}} & f(\mathbf{X}) \\ \text{s.t.} & \mathbf{X} \succeq 0 \end{array}$$

- Positive semidefinite constraint $\mathbf{X} \succeq 0$ is convex
- However these are difficult convex problems
- Generic algorithms scale poorly with n. Challenging for $n \gtrsim 500$

General structure of the problem

$$\begin{array}{ll} \min_{\mathbf{X}\in\mathbb{R}^{n\times n}} & f(\mathbf{X}) \\ \text{s.t.} & \mathbf{X} \succeq 0 \end{array}$$

- Positive semidefinite constraint $\mathbf{X} \succeq 0$ is convex
- However these are difficult convex problems
- Generic algorithms scale poorly with n. Challenging for $n \gtrsim 500$
- **This talk:**

  Can we replace, or approximate,
  the constraint $\mathbf{X} \succeq 0$ by a simpler one?

  We seek a geometric point of view on why such problems are hard

Approximating the positive semidefinite constraint

Simplest type of constraint in optimization is a linear one:

$$\sum_{1 \leq i,j \leq n} \mathbf{A}_{ij}\mathbf{X}_{ij} + b \geq 0.$$

Approximating the positive semidefinite constraint

Simplest type of constraint in optimization is a linear one:

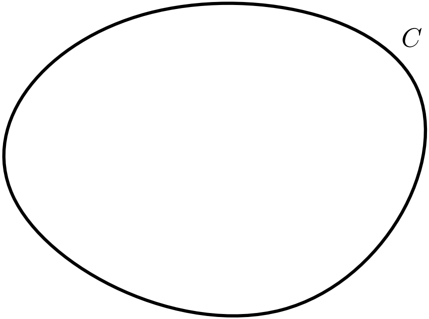
$$\sum_{1 \le i,j \le n} \mathbf{A}_{ij}\mathbf{X}_{ij} + b \ge 0.$$

$$\begin{array}{ll} \min & f(\mathbf{X}) \\ \text{s.t.} & \mathbf{X} \succeq 0 \end{array} \quad \overset{?}{\iff} \quad \min\ f(\mathbf{X})\ \text{ s.t. } \begin{cases} \langle \mathbf{A}^{(1)}, \mathbf{X} \rangle + b^{(1)} \ge 0 \\ \quad \vdots \\ \langle \mathbf{A}^{(N)}, \mathbf{X} \rangle + b^{(N)} \ge 0 \end{cases}$$

To have exact equivalence need $N = +\infty$!

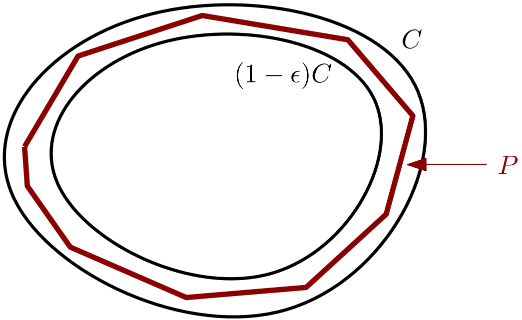
A question in convex geometry

Let $C = \{\mathbf{X} \in \mathbb{R}^{n \times n} : \mathbf{X} \succeq 0,\ \mathrm{Tr}(\mathbf{X}) = 1\}$



A question in convex geometry

Let $C = \{\mathbf{X} \in \mathbb{R}^{n\times n} : \mathbf{X} \succeq 0,\ \mathrm{Tr}(\mathbf{X}) = 1\}$



How "complex" does P have to be?

Complexity of polytopes

How to measure complexity of polytope?

$$P = \{\mathbf{X} : \langle \mathbf{A}^{(i)}, \mathbf{X} \rangle + b^{(i)} \geq 0 \ \ \forall i = 1, \dots, N\}$$

# of linear inequalities?

Complexity of polytopes

How to measure complexity of polytope?

$$P = \{\mathbf{X} : \langle \mathbf{A}^{(i)}, \mathbf{X} \rangle + b^{(i)} \geq 0 \ \ \forall i = 1, \ldots, N\}$$

# of linear inequalities?

The extension complexity $\mathrm{xc}(P)$ is the smallest integer K such that

- $P = \pi(Q)$ where π linear projection map
- Q is a polytope defined using K linear inequalities.

Complexity of polytopes

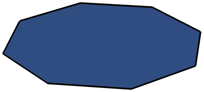
How to measure complexity of polytope?

$$P = \{\mathbf{X} : \langle \mathbf{A}^{(i)}, \mathbf{X} \rangle + b^{(i)} \geq 0 \ \ \forall i = 1, \ldots, N\}$$

$\#$ of linear inequalities?

The extension complexity $\mathrm{xc}(P)$ is the smallest integer K such that

- $P = \pi(Q)$ where π linear projection map
- Q is a polytope defined using K linear inequalities.



Complexity of polytopes

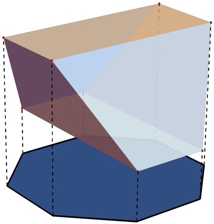
How to measure complexity of polytope?

$$P = \{\mathbf{X} : \langle \mathbf{A}^{(i)}, \mathbf{X} \rangle + b^{(i)} \geq 0 \;\; \forall i = 1, \ldots, N\}$$

$\#$ of linear inequalities?

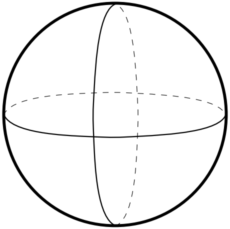
The extension complexity $\mathrm{xc}(P)$ is the smallest integer K such that

- $P = \pi(Q)$ where π linear projection map
- Q is a polytope defined using K linear inequalities.



Lifting matters!

Consider $B = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$



- Any approximation B must have at least $\approx \exp(n)$ linear inequalities
- However can write such an approximation as $P = \pi(Q)$ where Q has $\approx n$ inequalities! [Ben-Tal and Nemirovski]

Main theorem

Question: Can we approximate

$$C = \{\mathbf{X} \in \mathbb{R}^{n \times n} : \mathbf{X} \succeq 0 \text{ and } \mathrm{Tr}\,(X) = 1\}$$

with a polytope of small extension complexity?

Main theorem

Question: Can we approximate

$$C = \{\mathbf{X} \in \mathbb{R}^{n \times n} : \mathbf{X} \succeq 0 \text{ and } \mathrm{Tr}\,(X) = 1\}$$

with a polytope of small extension complexity?

Theorem (Main - F.)

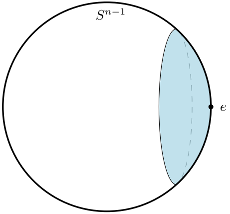
Assume P is a polytope such that

$$(1-\epsilon) \bullet C \subset P \subset C$$

Then the extension complexity of P is at least $\exp(c\sqrt{n})$ where $c > 0$ is a constant depending only on ϵ.

In words, it is not possible to approximate
$\{\mathbf{X} \in \mathbb{R}^{n \times n} : \mathbf{X} \succeq 0\}$
with a "simple" linear program

Proof of theorem. Geometry in high dimensions



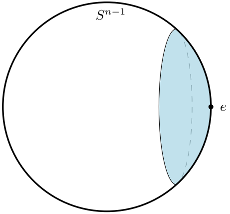
$S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$

- For any $e \in S^{n-1}$, the surface area of

$$\{x \in S^{n-1} : \langle x, e \rangle \geq 2/3\}$$

  is exponentially small in n!
- Need exponentially many spherical caps to cover the sphere

Proof of theorem. Geometry in high dimensions



$S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$

- For any $e \in S^{n-1}$, the surface area of

$$\{x \in S^{n-1} : \langle x, e \rangle \geq 2/3\}$$

  is exponentially small in n!
- Need exponentially many spherical caps to cover the sphere
- This shows that approximating C requires at least $\approx \exp(cn)$ inequalities
- Getting the main theorem on extension complexity requires more advanced tools.
- Key technical tool is a certain hypercontractive inequality

Discussion

- No generic way to approximate p.s.d. constraint with a small # of linear inequalities

- Ways out:
  - Adaptive approximation using a sequence of linear programs (see e.g., [Ahmadi and Majumdar])
  - Exploiting specific structures of problems
  - Very cheap methods with low accuracy

H. Fawzi, *On polyhedral approximations of the positive semidefinite cone*, Mathematics of Operations Research

Discussion

- No generic way to approximate p.s.d. constraint with a small # of linear inequalities

- Ways out:
  - Adaptive approximation using a sequence of linear programs (see e.g., [Ahmadi and Majumdar])
  - Exploiting specific structures of problems
  - Very cheap methods with low accuracy

Thank you!

H. Fawzi, *On polyhedral approximations of the positive semidefinite cone*, Mathematics of Operations Research