Recent advances in structure exploitation for semidefinite programming

Xin Jiang

Lehigh University
PhD from University of California, Los Angeles

Harvard ENG-SCI 257: Semidefinite optimization and relaxation
April 17, 2024

sparsity vs. **low rank**

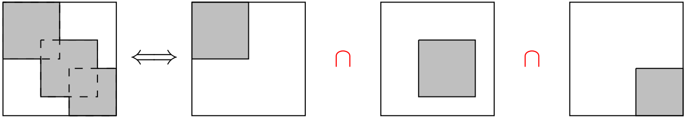
Outline

Sparse SDP
- Chordal sparsity
- SDP with chordal sparsity pattern
- Exploiting sparsity in solving SDPs

Low-rank SDP

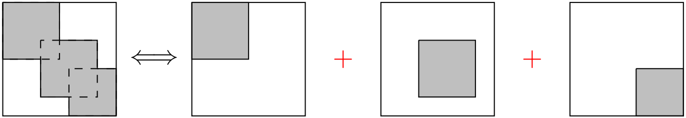
Clique decomposition of chordal sparsity pattern

decomposition for PSD completable matrices with chordal pattern



$X \in \mathbb{S}^n_+(\mathcal{E}, ?)$ $\iff$ $X_1 \succeq 0$ $\cap$ $X_2 \succeq 0$ $\cap$ $X_3 \succeq 0$

decomposition for PSD matrices with chordal pattern



$Z \in \mathbb{S}^n_+(\mathcal{E}, 0)$ $\iff$ $Z_1 \succeq 0$ $+$ $Z_2 \succeq 0$ $+$ $Z_3 \succeq 0$

Grone et al. (1988); Griewank & Toint (1984), Agler et al. (1984)

Sparse Cholesky factorization

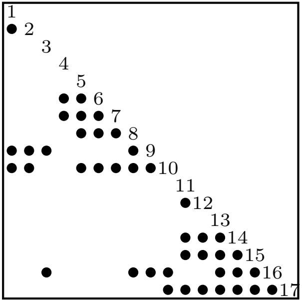
$$PAP^T = LDL^T$$

- A is positive definite, and P is a permutation matrix
- L is unit lower triangular, D positive diagonal

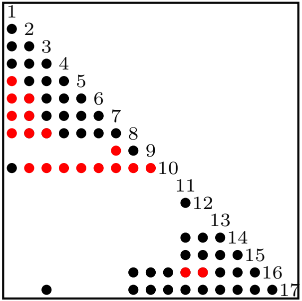
Sparsity pattern

$$P^T(L + L^T)P \in \mathbb{S}^n_{\mathcal{E}'}$$

- fill-in $\mathcal{E}' \setminus \mathcal{E}$ determines positions of added nonzeros



Pattern of A and L where $A = LL^T$



Pattern of $\widetilde{L}$ where $PAP^T = \widetilde{L}\widetilde{L}^T$

Cholesky factorization and chordal sparsity

Chordal pattern

if $A \in \mathbb{S}^n_+(\mathcal{E}, 0)$, then there is a permutation P such that

$$P^T(L + L^T)P \in \mathbb{S}^n_{\mathcal{E}}$$

A has a "zero fill" Cholesky factorization

Non-chordal pattern

If $\mathcal{E}$ is not chordal, then for every P there exists $A \in \mathbb{S}^n_+(\mathcal{E}, 0)$ such that

$$P^T(L + L^T)P \notin \mathbb{S}^n_{\mathcal{E}}$$

Rose (1970)

Logarithmic barriers for sparse matrix cones

Definition: the function $\phi_*\colon \mathbb{S}^n_{\mathcal{E}} \to \mathbb{R}$ with

$$\phi_*(Z) = -\log\det Z, \qquad \operatorname{dom}\phi_* = \operatorname{int}(\mathbb{S}^n_+(\mathcal{E},0)) \equiv \mathbb{S}^n_{++}(\mathcal{E},0)$$

the log-barrier for $\mathbb{S}^n_+(\mathcal{E},?)$ is the negative conjugate of ϕ_*:

$$\phi(X) = \sup_{Z\in\operatorname{int}(\mathbb{S}^n_+(\mathcal{E},0))} \big(-\langle X,Z\rangle + \log\det Z\big)$$

Value: efficiently computed from Cholesky factorization $Z = LDL^T$

Gradient: the negative of the projected inverse

$$\nabla\phi_*(Z) = -\Pi_{\mathcal{E}}(Z^{-1})$$

Hessian: for arbitrary $Y \in \mathbb{S}^n_{\mathcal{E}}$:

$$\nabla^2\phi_*(Z)[Y] = \frac{d}{dt}\nabla\phi(S+tY)\Big|_{t=0} = \Pi_{\mathcal{E}}(Z^{-1}YZ^{-1})$$

Andersen, Dahl, and Vandenberghe (2013)

Various formulations of SDP

Standard SDP

$$\begin{array}{ll} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \in \mathbb{S}^n_+ \end{array} \qquad \begin{array}{ll} \text{maximize} & \langle b, y \rangle \\ \text{subject to} & C - \mathcal{A}^*(y) = Z \\ & Z \in \mathbb{S}^n_+ \end{array}$$

Sparse SDP

$$\begin{array}{ll} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \in \mathbb{S}^n_+(\mathcal{E}, ?) \end{array} \qquad \begin{array}{ll} \text{maximize} & \langle b, y \rangle \\ \text{subject to} & C - \mathcal{A}^*(y) = Z \\ & Z \in \mathbb{S}^n_+(\mathcal{E}, 0) \end{array}$$

Various formulations of SDP: decomposed SDP

Primal decomposed SDP

$$
\begin{array}{ll}
\text{minimize} & \sum_{j=1}^{p} \langle C_j, X_j \rangle \\
\text{subject to} & \sum_{j=1}^{p} \langle A_{i,j}, X_j \rangle = b,\ i \in [m] \\
& E_{\mathcal{C}_j \cap \mathcal{C}_\ell} \left(E_{\mathcal{C}_j}^T X_j E_{\mathcal{C}_j} - E_{\mathcal{C}_\ell}^T X_\ell E_{\mathcal{C}_\ell} \right) E_{\mathcal{C}_j \cap \mathcal{C}_\ell}^T = 0, \quad \forall j \neq \ell, \mathcal{C}_j \cap \mathcal{C}_\ell \neq \emptyset \\
& X_j \succeq 0, \quad j \in [p]
\end{array}
$$

primal variables: $X_j \in \mathbb{S}^{|\mathcal{C}_j|}$, $j \in [p]$

Dual decomposed SDP

$$
\begin{array}{ll}
\text{maximize} & \langle b, y \rangle \\
\text{subject to} & C - \mathcal{A}^*(y) = \sum_{j=1}^{p} E_{\mathcal{C}_j}^T Z_k E_{\mathcal{C}_j} \\
& Z_j \succeq 0, \quad j \in [p]
\end{array}
$$

dual variables: $Z_j \in \mathbb{S}^{|\mathcal{C}_j|}$, $j \in [p]$

Exploiting sparsity in solving SDPs

	sparse SDP	decomposed SDP
first-order methods		
interior-point methods		

Interior-point methods for SDP

- (Symmetric) IPMs for the standard SDP
  exploit sparsity when forming "Schur complement" equations
  Fukuda et al. (2000), Benson & Ye (2008), Gao et al. (2022)

- (Non-symmetric) IPMs for the sparse SDP
  Fukuda et al. (2000), Srijuntongsiri et al. (2004), Andersen et al. (2010), ...

- standard IPMs for the decomposed SDP
  Nakata et al. (2003), Andersen et al. (2010), Zhang & Lavaei (2021), ...

ADMM for decomposed SDP

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax + By = c \end{array}$$

Alternating direction method of multipliers (ADMM)

$$\begin{aligned} x^{(k+1)} &= \operatorname*{argmin}_x \mathcal{L}_\rho(x, y^{(k)}, z^{(k)}) \\ y^{(k+1)} &= \operatorname*{argmin}_y \mathcal{L}_\rho(x^{(k+1)}, y^{(k)}, z^{(k)}) \\ z^{(k+1)} &= z^{(k)} + \rho(Ax^{(k+1)} + By^{(k+1)} - c) \end{aligned}$$

where the augmented Lagrangian is defined as

$$\mathcal{L}_\rho(x, y, z) = f(x) + g(y) + \langle z, Ax + By - c \rangle + \frac{\rho}{2} \|Ax + By - c\|_2^2$$

ADMM applied to decomposed SDP

- different splitting yields different algorithms
- scalability comes at a price of accuracy
- computational bottleneck: eigen-decomposition needed for PSD projection

Zheng et al. (2020)

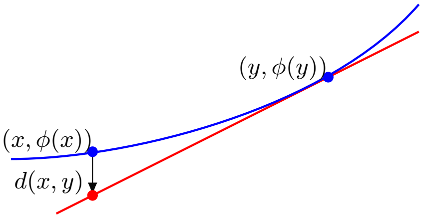
First-order methods for sparse SDP

Question: How to remove the bottleneck of eigen-decomposition in ADMM?

Attempt: Use *generalized* projection rather than Euclidean projection

Bregman divergence (generalized distance)

$$d(x, y) = \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle$$



- *kernel function* ϕ is convex and continuously differentiable on $\mathrm{int}(\mathrm{dom}\, \phi)$

Bregman (1967), Censor and Zenios (1997)

(Euclidean) proximal operator

Proximal operator (or **proximal mapping**) for closed convex function f

$$\mathrm{prox}_f(x) = \underset{y}{\mathrm{argmin}} \left(f(y) + \frac{1}{2}\|x - y\|_2^2 \right)$$

it exists and is unique for all $x \in \mathbb{R}^n$

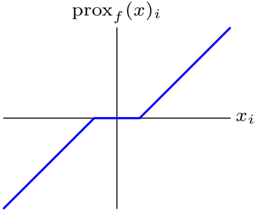
Example

- $f(x)$ is indicator function of closed convex set C: prox_f is projection on C

$$\mathrm{prox}_f(x) = \underset{u \in C}{\mathrm{argmin}} \|u - x\|_2^2 = \Pi_C(x)$$

- $f(x) = \|x\|_1$: prox_f is the "soft-threshold" (shrinkage) operation

$$\mathrm{prox}_f(x)_i = \begin{cases} x_i - 1 & x_i > 1 \\ 0 & |x_i| \le 1 \\ x_i + 1 & x_i < -1 \end{cases}$$



Moreau (1965)

Generalized proximal operator

- proximal operator of f with Bregman distance d generated by ϕ:

$$\operatorname{prox}_f^{\phi}(y,a) = \operatorname*{argmin}_x \big(f(x) + \langle a, x\rangle + d(x,y) \big)$$

- for $d(x,y) = \frac{1}{2}\|x-y\|_2^2$, this is the standard proximal operator

$$\begin{aligned} \operatorname{prox}_f^{\phi}(y,a) &= \operatorname*{argmin}_x \big(f(x) + \langle a, x\rangle + \tfrac{1}{2}\|x-y\|_2^2 \big) \\ &= \operatorname*{argmin}_x \big(f(x) + \tfrac{1}{2}\|x-y+a\|_2^2 \big) \\ &= \operatorname{prox}_f(y-a) \end{aligned}$$

Requirements

- minimizer $\hat{x}$ exists and is unique for all $y \in \operatorname{int}(\operatorname{dom}\phi)$ and all a
- minimizer $\hat{x}$ is inexpensive to compute

Example: relative entropy

$$d(x,y) = \sum_{i=1}^{n} (x_i \log(x_i/y_i) - x_i + y_i), \qquad \operatorname{dom} d = \mathbb{R}^n_+ \times \mathbb{R}^n_{++}$$

- the kernel function is

$$\phi(x) = \sum_{i=1}^{n} x_i \log x_i, \qquad \operatorname{dom} \phi = \mathbb{R}^n_+$$

- generalized projection (prox-operator for $f = \delta_{\mathcal{H}}$) on $\mathcal{H} = \{x \mid \mathbf{1}^T x = 1\}$

$$\operatorname*{argmin}_{\mathbf{1}^T x = 1} \big(\langle a, x\rangle + d(x,y)\big) = \frac{1}{\sum\limits_{j=1}^{n} y_j e^{-a_j}} \begin{bmatrix} y_1 e^{-a_1} \\ \vdots \\ y_n e^{-a_n} \end{bmatrix}$$

Generalized proximal operator: applications

- signal processing [Chao & Vandenberghe, 2018]
- optimal transport [Chambolle & Contreras, 2022]
- matrix optimization problem [Dhillon & Tropp, 2008]
- nonnegative matrix approximation [Dhillon & Sra, 2006; Li et al., 2012]
- statistical estimation [Taskar et al., 2006]
- machine learning [Kulis et al., 2009; Roman & d'Aspremont, 2020]
- *etc.*

Proximal splitting algorithms

$$\text{minimize} \quad f(x) + g(Ax) + h(x)$$

- h is convex, differentiable, and L-smooth
- f and g are convex and have simple proximal operators
- A is large and structured

Algorithms

- $g = 0$: proximal gradient method $x^{(k+1)} = \mathrm{prox}_{\tau f}(x^{(k)} - \tau \nabla h(x^{(k)}))$
- $h = 0$: ADMM, Douglas–Rachford (DRS), PDHG (Chambolle–Pock)
- $f = 0$: Loris–Verhoeven (a.k.a. PDFP^{2}O, PAPC)
- $A = I$: Davis–Yin
- three-operator splitting algorithms: Condat–Vũ, PD3O, PDDY

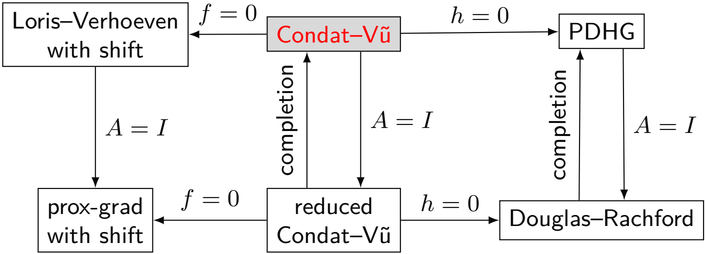
Boyd et al. (2010), Chambolle and Pock (2011, 2016)
Loris and Verhoeven (2011), Chen et al. (2013), Drori et al. (2015), Davis and Yin (2015)
Condat (2013), Vũ (2013), Yan (2018), Salim et al. (2020)

Condat–Vũ three-operator splitting algorithm

Algorithm

$$x^{(k+1)} = \mathrm{prox}_{\tau f}\left(x^{(k)} - \tau(A^T z^{(k)} + \nabla h(x^{(k)}))\right)$$
$$z^{(k+1)} = \mathrm{prox}_{\sigma g^*}\left(z^{(k)} + \sigma A(2x^{(k+1)} - x^{(k)})\right)$$

Relations with other proximal methods



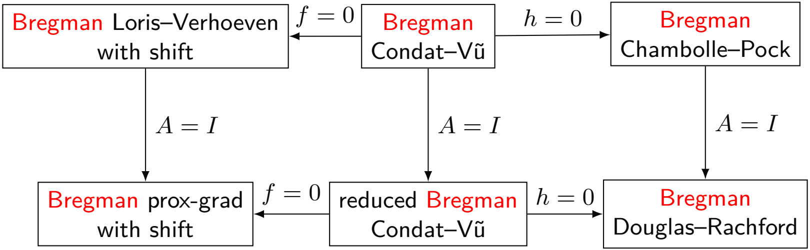
similar diagrams also exist for PD3O and PDDY

“completion” trick: O’Connor and Vandenberghe (2020)

Bregman Condat–Vũ algorithm

$$x^{(k+1)} = \mathrm{prox}_{\tau f}^{\phi_\mathrm{p}}\left(x^{(k)}, \tau(A^T z^{(k)} + \nabla h(x^{(k)}))\right)$$
$$z^{(k+1)} = \mathrm{prox}_{\sigma g^*}^{\phi_\mathrm{d}}\left(z^{(k)}, -\sigma A(2x^{(k+1)} - x^{(k)})\right)$$

Relations with other Bregman proximal methods



- "completion" trick may not be applicable in Bregman case
- similar diagram also exists for Bregman PD3O
- it is still unclear how to extend PDDY to Bregman distance

First-order proximal methods for SDP

Algorithmic toolbox

- Euclidean and Bregman proximal operators (and projections)
- Euclidean and Bregman proximal splitting methods

Optimization problem: (sparse) SDP

$$\begin{array}{ll} \text{minimize} & \langle C, X\rangle \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \in \mathcal{K} \end{array} \qquad \begin{array}{ll} \text{maximize} & \langle b, y\rangle \\ \text{subject to} & C - \mathcal{A}^*(y) = Z \\ & Z \in \mathcal{K}^* \end{array}$$

- $(\mathcal{K}, \mathcal{K}^*) = (\mathbb{S}^n_+, \mathbb{S}^n_+)$ or $(\mathcal{K}, \mathcal{K}^*) = \big(\mathbb{S}^n_+(\mathcal{E}, ?), \mathbb{S}^n_+(\mathcal{E}, 0)\big)$
- prior work considers $(\mathcal{K}, \mathcal{K}^*) = (\mathbb{S}^n_+, \mathbb{S}^n_+)$ and (Dhillon & Tropp (2008))

$$\tilde{\phi}(X) = -\log\det X, \quad \operatorname{dom}\tilde{\phi} = \mathbb{S}^n_{++}$$

Logarithmic barrier for sparse SDPs

- $\phi_*(Z) = -\log\det Z$ is the log-det barrier for $\mathcal{K}^* = \mathbb{S}^n_+(\mathcal{E}, 0)$
- the primal barrier ϕ for $\mathcal{K} = \mathbb{S}^n_+(\mathcal{E}, ?)$ is the negative conjugative of $\phi_*(Z)$
- for chordal $\mathcal{E}$: efficient algorithms for computing ϕ, ϕ_*, $\nabla\phi$, $\nabla\phi_*$, *etc.*

Bregman proximal operator in sparse SDP

$$\begin{array}{ll} \text{minimize} & \langle C, X\rangle \\ \text{subject to} & \mathcal{A}(X) = b, \quad \langle I, X\rangle = 1, \quad X \in \mathbb{S}^n_+(\mathcal{E}, ?) \end{array}$$

- objective restricted to $\langle I, X\rangle = \operatorname{tr} X = 1$ and $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$

$$f(X) = \langle C, X\rangle + \delta_{\mathcal{H}}(X) + \delta_{\mathbb{S}^n_+(\mathcal{E},?)}(X), \qquad \text{where } \mathcal{H} = \{X \mid \operatorname{tr} X = 1\}$$

- prox-operator $\widehat{X} = \operatorname{prox}^{\phi}_{f}(Y, A)$, using Bregman distance generated by ϕ:

$$\begin{aligned} \widehat{X} &= \operatorname*{argmin}_X \left(f(X) + \langle A, X\rangle + \tfrac{1}{\tau} d(X, Y)\right) \\ &= \operatorname*{argmin}_X \{\langle B, X\rangle + \phi(X) \mid \operatorname{tr} X = 1\}, \end{aligned}$$

  where $B \in \mathbb{S}^n_{\mathcal{E}}$ depends on Y, A, C, and τ

- owing to the definition of ϕ, the constraint $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$ is always satisfied

- dual problem (scalar λ is the multiplier for $\operatorname{tr} X = 1$)

$$\text{maximize} \quad -\lambda + \log\det(B + \lambda I)$$

Jiang and Vandenberghe (2022)

Newton's method for Bregman proximal operator

- use Newton's method to find unique solution $\hat{\lambda}$ of the nonlinear equation

$$\operatorname{tr}((B + \lambda I)^{-1}) = 1 \qquad \text{(with } B + \lambda I \succ 0\text{)}$$

- for chordal sparsity patterns $\mathcal{E}$, efficient algorithms exist for computing

$$g(\lambda) = \operatorname{tr}((B + \lambda I)^{-1}), \qquad g'(\lambda) = -\operatorname{tr}((B + \lambda I)^{-2})$$

  from sparse Cholesky factorization of $B + \lambda I$

- from λ, the primal solution $\widehat{X}$ is computed as an projection on $\mathbb{S}^n_{\mathcal{E}}$:

$$\widehat{X} = \Pi_{\mathcal{E}}\big((B + \lambda I)^{-1}\big)$$

complexity $\approx$ # Newton iterations $\times$ cost of sparse Cholesky factorization

Jiang and Vandenberghe (2022)

SDP relaxation of graph partitioning

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(P^T L P X) \\ \text{subject to} & \mathrm{diag}(P X P^T) = \mathbf{1} \\ & X \succeq 0 \end{array}$$

- columns of P are sparse basis of $\{x \mid \mathbf{1}^T x = 0\}$
- four problems from SDPLIB, four graphs from SuiteSparse collection

	n	PDHG iterations	time per Cholesky factorization	Newton steps per prox-evaluation	time per PDHG iteration
gpp100	100	305	0.01	2.43	0.02
gpp124-1	124	392	0.01	2.00	0.02
gpp250-1	250	365	0.01	2.65	0.03
gpp500-1	500	394	0.02	3.01	0.07
delaunay_n10	1024	403	0.37	4.36	1.76
delaunay_n11	2048	420	0.48	4.70	2.54
delaunay_n12	4096	367	0.60	4.43	3.05
delaunay_n13	8192	375	1.02	4.42	4.98

Summary: sparse SDP

Chordal sparsity

- clique decomposition of two sparse matrix cones
- chordal sparsity offers zero fill-in sparse Cholesky factorization

Algorithms for SDPs with chordal sparsity pattern

	sparse SDP	decomposed SDP
interior-point methods		
first-order methods		

- interior-point methods are extensively studied and well developed
- first-order methods are relatively new
  - proximal methods for decomposed SDP

    consistency constraints and expensive eigen-decomposition
  - Bregman proximal methods for sparse SDP

    suitable Bregman distance $\Longrightarrow$ sparse Cholesky factorization

Outline

Sparse SDP

Low-rank SDP

Burer–Monteiro factorization

Standard primal SDP

$$\begin{array}{ll} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & \langle A_i, X \rangle = b_i, \quad i \in [m] \\ & X \succeq 0 \end{array}$$

- denote $r^\star = \min\{\text{rank}(X^\star) \mid X^\star \text{ is optimal for the primal SDP}\}$
- the rank r of any extreme point of the feasible set satisfies

$$r(r+1)/2 \leq m;$$

  there is an optimal solution $X^\star$ whose rank satisfies the above inequality

Burer–Monteiro low-rank SDP

$$\begin{array}{ll} \text{minimize} & \langle C, YY^T \rangle \\ \text{subject to} & \langle A_i, YY^T \rangle = b_i, \quad i \in [m] \end{array}$$

with variable $Y \in \mathbb{R}^{n \times r}$

- if $r \geq r^\star$, the two problems share the same *global* minimum

Barvinok (1995), Pataki (1998); Burer and Monteiro (2003, 2005)

Can we solve BM-SDP to global optimum?

When does BM-SDP has no bad local minima?

No bad local minima

Problem formulation

$$\begin{array}{ll} \text{minimize} & f(X) \\ \text{subject to} & X \succeq 0 \end{array} \qquad\qquad \text{minimize} \quad g(Y) := f(YY^T)$$

- the decision variables are $X \in \mathbb{S}^n$ (left) and $Y \in \mathbb{R}^{n \times r}$ (right)
- $r^\star = \mathrm{rank}(X^\star)$ is the optimal rank, and r is the search rank
- f is μ-strongly convex and L-smooth
- denote its condition number by $\kappa_f = L/\mu \in [1, \infty)$

No bad local minima: every second order critical point is a global minimizer

$$\nabla \psi(x) = 0, \;\; \nabla^2 \psi(x) \succeq 0 \qquad \iff \qquad \psi(x) = \psi^\star$$

Guarantees for no bad local minima

Condition-based guarantees

- statistical learning: $\kappa_f < \infty$ with "enough" samples
- if $\kappa_f < 3/2$, then g has no bad local minima (Bhojanapalli et al. (2016))
- if $\kappa_f \geq 3$, then counter-example exists (Zhang et al. (2018))

empirical evidence: well-conditioned problems have no bad local min
worse case: need $\kappa \approx 1$ to eliminate bad local min

Overparameterization guarantees

- if $r \geq n$, then g has no bad local minima (Boumal et al. (2020))
  but it obviates the computational advantage of BM
- if $r = \Omega(n)$, then counter-example exists (Waldspurger & Waters (2020))

empirical evidence: $r \ll n$ escapes all bad local minima
worst case: need $r \approx n$ to eliminate bad local minima

Question: Can these two results talk to each other?

Guarantees for no bad local minima

Condition-based guarantees

- statistical learning: $\kappa_f < \infty$ with "enough" samples
- if $\kappa_f < 3/2$, then g has no bad local minima (Bhojanapalli et al. (2016))
- if $\kappa_f \geq 3$, then counter-example exists (Zhang et al. (2018))

Overparameterization guarantees

- if $r \geq n$, then g has no bad local minima (Boumal et al. (2020))
- if $r = \Omega(n)$, then counter-example exists (Waldspurger & Waters (2020))

Combined guarantees involving κ and overparameterization

- if $r > r^\star + \gamma$, then g has no bad local minima
- if $r \leq (1 + \gamma) r^\star$, then counter-example exists

where $\gamma = \frac{1}{4}(\kappa_f - 1)^2 - 1$ (Zhang (2022))

Algorithms for BM-SDP

$$\begin{array}{ll} \text{minimize} & f(X) \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \succeq 0 \end{array} \qquad \begin{array}{ll} \text{minimize} & g(Y) := f(YY^T) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

- second-order methods?
- gradient-based methods
  (Bhojanapalli et al. (2016), Zhang et al. (2022), Yang et al. (2022), . . .)
- augmented Lagrangian methods
  (Burer & Monteiro (2003, 2005), Lee et al. (2022), Monteiro et al. (2024), . . .)
- proximal gradient methods (Bai, Duchi, & Mei (2019))
- ADMM (Chen & Goulart (2023))
- block coordinate descent methods (Erdogdu et al. (2022))
- manifold optimization (Wang & Hu (2023))
- more, and more to come

sparsity vs. **low rank**

sparsity and **low rank**

Motivation

Standard SDP

$$\begin{array}{ll} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \in \mathbb{S}^n_+ \end{array} \qquad \begin{array}{ll} \text{maximize} & \langle b, y \rangle \\ \text{subject to} & C - \mathcal{A}^*(y) = Z \\ & Z \in \mathbb{S}^n_+ \end{array}$$

- the data matrices $C, A_1, \ldots, A_m$ are sparse
- the optimal solution is often very dense
- in many applications, we want a low-rank solution

Question: Can we exploit data sparsity and low rank of solution?

Sparse SDP

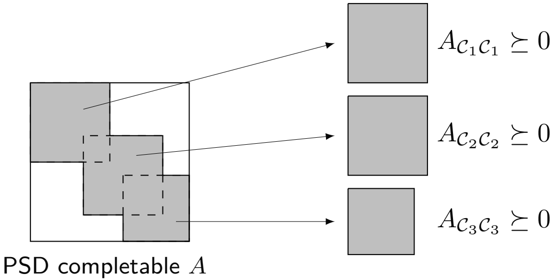
$$\begin{array}{ll} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \in \mathbb{S}^n_+(\mathcal{E}, ?) \end{array} \qquad \begin{array}{ll} \text{maximize} & \langle b, y \rangle \\ \text{subject to} & C - \mathcal{A}^*(y) = Z \\ & Z \in \mathbb{S}^n_+(\mathcal{E}, 0) \end{array}$$

- primal optimum $X^\circ \in \mathbb{S}^n_+(\mathcal{E}, ?)$ is sparse (or incomplete)
- any completion $X^\bullet$ is a solution for the original SDP

Minimum rank PSD completion with chordal sparsity

recall that $A \in \mathbb{S}^n_{\mathcal{E}}$ has a positive semidefinite completion if and only if

$$A_{\mathcal{C}_i\mathcal{C}_i} \succeq 0 \quad \text{for all cliques } \mathcal{C}_i$$



Minimum rank PSD completion

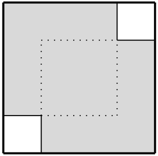
if $\mathcal{E}$ is chordal, then there is a PSD completion $X \in \mathbb{S}^n_+$ with rank

$$\operatorname{rank}(X) = \max_{\text{cliques } \mathcal{C}_i} \operatorname{rank}(A_{\mathcal{C}_i\mathcal{C}_i})$$

Dancis (1992)

Two-block completion problem

find the minimum rank positive semidefinite completion of



$$A = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & A_{23} \\ 0 & A_{32} & A_{33} \end{bmatrix}$$

- a PSD completion exists if and only if

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \succeq 0, \qquad \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} \succeq 0$$

- define $r = \max\{r_1, r_2\}$ where

$$r_1 = \mathrm{rank} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad r_2 = \mathrm{rank} \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix}$$

Two-block completion algorithm

- compute matrices U, V, $\widetilde{V}$, W of column dimension r such that

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix}^T, \qquad \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} \widetilde{V} \\ W \end{bmatrix} \begin{bmatrix} \widetilde{V} \\ W \end{bmatrix}^T$$

- since $VV^T = \widetilde{V}\widetilde{V}^T$, the matrices V and $\widetilde{V}$ have SVDs

$$V = P\Sigma Q_1^T, \qquad \widetilde{V} = P\Sigma Q_2^T;$$

  hence $V = \widetilde{V}Q$, where $Q = Q_2 Q_1^T$ is an orthogonal $r \times r$ matrix

- a completion of rank r is given by

$$\begin{bmatrix} UQ^T \\ \widetilde{V} \\ W \end{bmatrix} \begin{bmatrix} UQ^T \\ \widetilde{V} \\ W \end{bmatrix}^T = \begin{bmatrix} A_{11} & A_{12} & UQ^T W^T \\ A_{21} & A_{22} & A_{23} \\ WQU^T & A_{32} & A_{33} \end{bmatrix}$$

SDP relaxation of optimal power flow problem

	n	max. clique	MOSEK 8		SeDuMi v1.05		SDPT3 v4.0	
			$\mathrm{rank}(X^\circ)$	$\mathrm{rank}(X^\bullet)$	$\mathrm{rank}(X^\circ)$	$\mathrm{rank}(X^\bullet)$	$\mathrm{rank}(X^\circ)$	$\mathrm{rank}(X^\bullet)$
IEEE-118	118	20	1	1	1	1	1	1
IEEE-300	300	17	5	1	5	1	5	1
2383wp	2383	31	17	1	17	1	17	1
2736sp	2736	30	1	1	1	1	1	1
2737sop	2737	29	44	1	43	1	43	1
2746wop	2746	30	32	1	32	1	32	1
2746wp	2746	31	1	1	1	1	1	1
3012wp	3012	32	346	13	346	13	337	17
3120sp	3120	32	514	27	572	32	519	27
3375wp	3375	33	451	19	451	19	454	21

- benchmark problems from MATPOWER package
- rank is number of eigenvalues greater than $10^{-5}\sqrt{n}\lambda_{\max}$
- X° is the (Hermitian) solution of sparse SDP relaxation
- $X^\bullet$ is minimum rank PSD completion of $\Pi_{\mathcal{E}}(X^\circ)$

Jiang, Sun, Andersen, and Vandenberghe (2023)

Summary

Exploiting chordal sparsity

- key properties: clique decomposition and zero fill-in Cholesky factorization
- interior-point methods are well studied
- development of first-order methods is far from mature

Exploiting low rank via Burer–Monteiro factorization

- various theoretical guarantees for no bad local minima
- first-order methods are prevalent

Minimum low rank PSD completion with chordal sparsity

- minimum rank PSD completion is easy when $\mathcal{E}$ is chordal
- NP-hard when $\mathcal{E}$ is not chordal
- can be used as a *post-processing* technique after solving sparse SDP