Iterative Inner/outer Approximations for Scalable Semidefinite Programs

Feng-Yi Liao, Yang Zheng

ECE department, UC San Diego
Scalable Optimization and Control (SOC) Lab

Feb 16, 2023

Semidefinite Prgramms

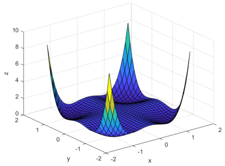
Primal SDP

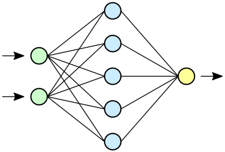
$$
\begin{aligned}
\min_{X} \quad & \langle C, X \rangle \\
\text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \\
& X \in \mathbb{S}^n_+.
\end{aligned}
$$

Dual SDP

$$
\begin{aligned}
\max_{y, Z} \quad & b^{\mathsf{T}} y \\
\text{subject to} \quad & Z + \sum_{i=1}^{m} A_i y_i = C, \\
& Z \in \mathbb{S}^n_+.
\end{aligned}
$$

- **SDPs are powerful tools in broad areas.**
- **Application:** Control theory, combinatorial problem, polynomial optimization, neural network verification, etc.





Primal SDP

$$
\begin{aligned}
p^\star := \min_X \quad & \langle C, X \rangle \\
\text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \\
& X \in \mathbb{S}_+^n.
\end{aligned}
$$

General purpose solver: Interior-point method

- Standard complexity $\mathcal{O}(n^3 m + n^2 m^2 + m^3)$ per iteration.
- Cannot efficiently handle large-scale SDPs ($n \approx 1000$, and m : a few thousands).

Active research directions

- Explore problem sparsity and structures[1].

[1] Yang Zheng, Giovanni Fantuzzi, and Antonis Papachristodoulou (2021). "Chordal and factor-width decompositions for scalable semidefinite and polynomial optimization". In: *Annual Reviews in Control* 52, pp. 243–279.

Something simpler: inner/outer approximations

Inner approximation

- Restrict the feasible region to a simpler cone $\mathcal{K} \subset \mathbb{S}^n_+$.

$$\begin{aligned} \min_X \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \\ & X \in \mathcal{K}. \end{aligned}$$

- Gives us an upper bound on $p^\star$.

Outer approximation

- Relax the feasible region by a simpler cone $\hat{\mathcal{K}} \supset \mathbb{S}^n_+$.

$$\begin{aligned} \min_X \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \\ & X \in \hat{\mathcal{K}}. \end{aligned}$$

- Gives us a lower bound on $p^\star$.

Which cone to choose?

- **Diagonally dominant:**
  A symmetric matrix $X \in \mathbb{S}^n$ is diagonally dominant if and only if

$$X_{ii} \geq \sum_{j \neq i} |X_{ij}|, i = 1, 2, \ldots, n.$$

- Let $\mathcal{DD}_n = \{X \in \mathbb{S}^n \mid X \text{ is diagonally dominant}\} \subset \mathbb{S}^n_+$.

Gershgorin's circle theorem
Given an $n \times n$ matrix X, every eigenvalue of X lies in at least one of the discs D_i in the complex plane, where

$$D_i = |\lambda - X_{ii}| \leq \sum_{j \neq i} |X_{ij}|$$

- Diagonally dominant

$$X_{ii} \geq \sum_{j \neq i} |X_{ij}| \Longrightarrow |\lambda - X_{ii}| \leq X_{ii} \Longrightarrow \lambda \geq 0.$$

Diagonally dominant

- Optimizing over $\mathcal{DD}_n$ leads to LP.
- For each $|X_{ij}|$, Introduce variable T_{ij} such that

$$-T_{ij} \le X_{ij} \le T_{ij}, \quad \sum_{j \ne i} T_{ij} \le X_{ii}, \; i = 1, 2, \ldots, n.$$

- Replace $\mathbb{S}_+^n$ by $\mathcal{DD}_n$

$$\begin{aligned} \min_{X} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \\ & X \in \mathcal{DD}_n. \end{aligned}$$

- This is equivalent to

$$\begin{aligned} \min_{X, T_{ij}} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_k, X \rangle = b_k, \quad k = 1, \ldots, m, \\ & -T_{ij} \le X_{ij} \le T_{ij}, \quad \sum_{j \ne i} T_{ij} \le X_{ii}, \; i = 1, 2, \ldots, n. \end{aligned}$$

Which cone to choose?

- **Scaled-diagonally dominant:**
  A symmetric matrix $X \in \mathbb{S}^n$ is scaled-diagonally dominant if and only if there exists a diagonal matrix D with nonnegative elements such that

  $$DXD \text{ is diagonally dominant.}$$

Another Interpretation of $\mathcal{SDD}_n$: A symmetric X belongs to $\mathcal{SDD}_n$ if and only if there exist $Z_{ij} \in \mathbb{S}^2_+$ such that

$$X = \sum_{1 \le i < j \le n} E_{ij}^\top Z_{ij} E_{ij},$$

where $E_{ij} = \begin{bmatrix} E_i \\ E_j \end{bmatrix}$, and $E_i \in \mathbb{R}^{1 \times n}$ is zero everywhere except the i-th component being 1. $E_i = \begin{bmatrix} 0 \dots 1 \dots 0 \end{bmatrix} \in \mathbb{R}^{1 \times n}$.

- Let $\mathcal{SDD}_n = \{X \in \mathbb{S}^n \mid X \text{ is scaled-diagonally dominant}\} \subset \mathbb{S}^n_+$.

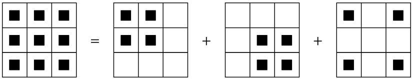


Figure: Illustration of $\mathcal{FW}_2^n$ (or $\mathcal{SDD}$) matrices.

Scaled-diagonally dominant

- A 2×2 semidefinite constraint is equivalent to a (rotated) second-order cone constraint.

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0 \iff a \geq 0, c \geq 0, ac \geq \|b\|_2^2 \iff \left(b, a, \frac{1}{2}c\right) \in \mathcal{L}_{\mathrm{rot}}^{n+2},$$

where $\mathcal{L}_{\mathrm{rot}}^{n+2} = \{(x, y, z) \in \mathbb{R}^{n+2} | 2yz \geq \|x\|_2^2, y \geq 0, z \geq 0\}$.

- Optimizing over $\mathcal{SDD}_n$ leads to SOCP.

$$\begin{aligned} \min_{X} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \\ & X \in \mathcal{SDD}_n. \end{aligned}$$

- This is equivalent to

$$\begin{aligned} \min_{X, Z_{ij}} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_k, X \rangle = b_i, \quad k = 1, \ldots, m, \\ & X = \sum_{1 \leq i < j \leq n} E_{ij}^{\mathsf{T}} Z_{ij} E_{ij}, \\ & Z_{ij} \succeq 0 \iff \left((Z_{ij})_{12}, (Z_{ij})_{11}, \frac{1}{2}(Z_{ij})_{22}\right) \in \mathcal{L}_{\mathrm{rot}}^{n+2}. \end{aligned}$$

Approximation quality

$$\begin{aligned} \min_{X} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \\ & X \in \mathcal{SDD}_n \text{ (or } \mathcal{DD}_n). \end{aligned}$$

- The approximation quality might be conservative

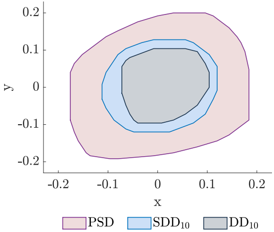


Figure: Feasible region of PSD, SDD_{10}, or DD_{10} over a 10×10 LMI

- $\mathcal{DD}_n$ requires $\mathcal{O}(n^2)$ linear constraints.
- $\mathcal{SDD}_n$ requires $\mathcal{O}(n^2)$ small SOCP constraints.
- $\mathcal{DD}_n$ and $\mathcal{SDD}_n$ encounter problems for large n.

Comparison of computational time

2N (# states)	4	6	8	10	12	14	16	18	20	22
DSOS	< 1	0.44	2.04	3.08	9.67	25.1	74.2	200.5	492.0	823.2
SDSOS	< 1	0.72	6.72	7.78	25.9	92.4	189.0	424.74	846.9	1275.6
SOS (SeDuMi)	< 1	3.97	156.9	1697.5	23676.5	∞	∞	∞	∞	∞
SOS (MOSEK)	< 1	0.84	16.2	149.1	1526.5	∞	∞	∞	∞	∞

Figure: Time consumption of using LP and SOCP approximation.

- The table is taken from Amir Ali Ahmadi's paper[2]

[2] Amir Ali Ahmadi and Anirudha Majumdar (2019). "DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization". In: *SIAM Journal on Applied Algebra and Geometry* 3.2, pp. 193–230.

Factor-width-two matrices

Another Interpretation of $\mathcal{SDD}_n$: A symmetric X belongs to $\mathcal{SDD}_n$ if and only if there exist $Z_{ij} \in \mathbb{S}^2_+$ such that

$$X = \sum_{1 \le i < j \le n} E_{ij}^\mathsf{T} Z_{ij} E_{ij},$$

where $E_{ij} = \begin{bmatrix} E_i \\ E_j \end{bmatrix}$, and $E_i \in \mathbb{R}^{1\times n}$ is zero everywhere except the i-th component being 1. $E_i = \begin{bmatrix} 0 \ldots 1 \ldots 0 \end{bmatrix} \in \mathbb{R}^{1\times n}$.

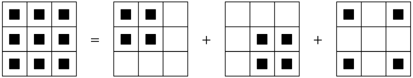


Figure: Illustration of $\mathcal{FW}_2^n$ (or $\mathcal{SDD}$) matrices.

Let $\mathcal{SDD}_n = \mathcal{FW}_2^n$. Optimizing over $\mathcal{FW}_2^n$ is equivalent to an SDP over the cone product

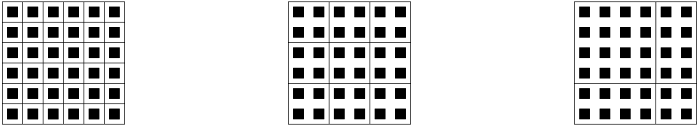
$$\mathbb{S}^2_+ \times \ldots \times \mathbb{S}^2_+.$$

Block factor-width-two matrices

Given a set of integers $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ with $\sum_{i=1}^{p} \alpha_i = n$, we say a matrix $A \in \mathbb{R}^{n \times n}$ is block-partitioned by α if we can write A as

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ A_{21} & A_{22} & \dots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \dots & A_{pp} \end{bmatrix},$$

where $A_{ij} \in \mathbb{R}^{\alpha_i \times \alpha_j}, \forall i, j = 1, 2, \dots, p$.



(a) $\alpha = \{1, 1, 1, 1, 1, 1\}$ (b) $\beta = \{2, 2, 2\}$ (c) $\gamma = \{4, 2\}$

Figure: Different partitions for a 6×6 matrix

Definition (Zheng et al. 2022)
A symmetric matrix X with partition $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_p\}$ belongs to *block-factor-width-two* matrices, denoted as $\mathcal{FW}^n_{\alpha,2}$, if there exist X_{ij} such that

$$X = \sum_{1 \le i < j \le p}^{p} (E^\alpha_{ij})^\mathsf{T} Z_{ij} E^\alpha_{ij}, \tag{1}$$

with $Z_{ij} \in \mathbb{S}_+^{\alpha_i+\alpha_j}$, $E^\alpha_{ij} = \begin{bmatrix} E^\alpha_i \\ E^\alpha_j \end{bmatrix} \in \mathbb{R}^{(\alpha_i+\alpha_j)\times n}$, for $i \neq j$ and $E^\alpha_i = \begin{bmatrix} 0 & \dots & I_{\alpha_i} & \dots & 0 \end{bmatrix} \in \mathbb{R}^{\alpha_i \times n}$.

- We denote

$$\mathcal{FW}^n_{\alpha,2} = \{X \in \mathbb{S}^n \mid X \text{ is } \alpha\text{-block-factor-width-two}\} \subset \mathbb{S}^n_+.$$

- $\mathcal{SDD}_n$ is a special case of $\mathcal{FW}^n_{\alpha,2}$ with partition $\alpha = \{1, \dots, 1\}$.

Block-factor-width-two matrices

Optimizing over $\mathcal{FW}^n_{\alpha,2}$ is equivalent to an SDP over the cone product

$$\mathbb{S}_+^{\alpha_1+\alpha_2} \times \ldots \times \mathbb{S}_+^{\alpha_{p-1}+\alpha_p}.$$

- $\mathcal{FW}^2_{\alpha,2}$ allows different size of submatrices

$$X = \sum_{1\leq i<j\leq p} E_{ij}^{\mathsf{T}} Z_{ij} E_{ij}, \text{ with } Z_{ij} \in \mathbb{S}_+^{\alpha_i+\alpha_j}.$$

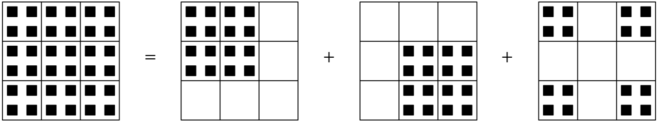


Figure: Illustration of $\mathcal{FW}^n_{\alpha,2}$ matrices.

- The flexibility of $\mathcal{FW}^n_{\alpha,2}$ improves the approximation quality and numerical efficiency.
- Number of PSD constraints has been reduced $\binom{n}{2} \Longrightarrow \binom{p}{2}$.

A hierarchy of inner/outer approximations

- We say a partition α is a *finer* partition of β, denoted as $\alpha \sqsubseteq \beta$, if α can be formed by breaking down some blocks in β.

Theorem (Zheng et al. 2022)
Given $\{1,1,\ldots,1\} \sqsubseteq \alpha \sqsubseteq \beta \sqsubseteq \gamma = \{\gamma_1, \gamma_2\}$ with $\gamma_1 + \gamma_2 = n$, we have a converging hierarchy of inner and outer approximations

$$\begin{aligned} &\mathcal{DD}_n \subseteq \mathcal{SDD}_n \subseteq \mathcal{FW}^n_{\alpha,2} \subseteq \mathcal{FW}^n_{\beta,2} \subseteq \mathcal{FW}^n_{\gamma,2} = \mathbb{S}^n_+ \\ &= (\mathcal{FW}^n_{\gamma,2})^* \subseteq (\mathcal{FW}^n_{\beta,2})^* \subseteq (\mathcal{FW}^n_{\alpha,2})^* \subseteq (\mathcal{SDD}_n)^* \subseteq (\mathcal{DD}_n)^*, \end{aligned} \tag{2}$$

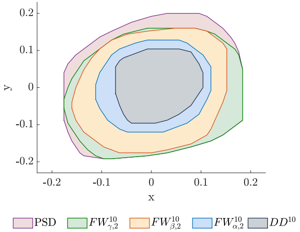


Figure: Feasible region of $\mathcal{FW}^{10}_{\alpha,2}$, $\mathcal{FW}^{10}_{\beta,2}$, $\mathcal{FW}^{10}_{\gamma,2}$, and $\mathcal{DD}^{10}$ over a 10×10 LMI, where $\alpha = \{1,1,\ldots,1\}$, $\beta = \{2,2,2,2,2\}$, $\gamma = \{4,4,2\}$.

$$\mathcal{FW}^n_{\alpha,2} = \left\{ X \in \mathbb{S}^n_+ | X = \sum_{1 \le i < j \le p}^{p} (E^\alpha_{ij})^\mathsf{T} Z_{ij} E^\alpha_{ij}, Z_{ij} \succeq 0 \right\}$$

$$\begin{aligned}
\left(\mathcal{FW}^n_{\alpha,2}\right)^* &= \left\{ Y \in \mathbb{S}^n | \langle Y, X \rangle \ge 0, \forall X \in \mathcal{FW}^n_{\alpha,2} \right\} \\
&= \left\{ Y \in \mathbb{S}^n | \left\langle Y, \sum_{1 \le i < j \le p}^{p} (E^\alpha_{ij})^\mathsf{T} Z_{ij} E^\alpha_{ij} \right\rangle \ge 0, \forall Z_{ij} \succeq 0 \right\} \\
&= \left\{ Y \in \mathbb{S}^n | \sum_{1 \le i < j \le p}^{p} \left\langle E^\alpha_{ij} Y (E^\alpha_{ij})^\mathsf{T}, Z_{ij} \right\rangle \ge 0, \forall Z_{ij} \succeq 0 \right\} \\
&= \left\{ Y \in \mathbb{S}^n | E^\alpha_{ij} Y (E^\alpha_{ij})^\mathsf{T} \succeq 0, \forall 1 \le i < j \le p \right\}
\end{aligned}$$

Primal

$$\begin{aligned}
\min_{X} \quad & \langle C, X \rangle \\
\text{subject to} \quad & \langle A_k, X \rangle = b_k, \quad k = 1, \ldots, m, \\
& X = \sum_{1 \le i < j \le p}^{p} (E^\alpha_{ij})^\mathsf{T} Z_{ij} E^\alpha_{ij}, \\
& Z_{ij} \succeq 0.
\end{aligned}$$

Dual

$$\begin{aligned}
\max_{y,Z} \quad & b^\mathsf{T} y \\
\text{subject to} \quad & Z + \sum_{k=1}^{m} A_k y_k = C, \\
& E^\alpha_{ij} Z (E^\alpha_{ij})^\mathsf{T} \succeq 0, \\
& \forall 1 \le i < j \le p.
\end{aligned}$$

Iterative inner approximations

$$\begin{aligned} \min_{X} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \\ & X \in \mathcal{FW}^n_{\alpha,2}. \end{aligned}$$

- A coarser partition naturally provides a tighter upper bound on $p^\star$.
- However, a coarser partition leads to a larger PSD constraint.
- **Key idea:** we keep an acceptable partition size and iteratively tighten the upper bound by **basis pursuit**.

Ahmadi and Hall[3] introduces an iterative method over $\mathcal{DD}_n$ and $\mathcal{SDD}_n$. It can be naturally extended to $\mathcal{FW}^n_{\alpha,2}$.

- **Basis pursuit:**

$$\begin{aligned} \mathrm{U}^t_\alpha := \min_X \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \ldots, m, \\ & X \in \mathcal{FW}^n_{\alpha,2}(V_t), \end{aligned}$$

where $\mathcal{FW}^n_{\alpha,2}(V) := \{M \in \mathbb{S}^n \mid M = V^{\mathsf{T}} Q V,\ Q \in \mathcal{FW}^n_{\alpha,2}\}$.

- We choose the sequence of matrices $\{V_t\}$ as

$$\begin{aligned} V_1 &= I \\ V_{t+1} &= \mathrm{chol}(X^\star_t). \end{aligned}$$

[3] Amir Ali Ahmadi and Georgina Hall (2017). "Sum of squares basis pursuit with linear and second order cone programming". In: *Algebraic and geometric methods in discrete mathematics* 685, pp. 27–53.

Iterative inner approximations

$$V_1 = I$$
$$V_{t+1} = \mathrm{chol}(X_t^\star).$$

- **Key idea:** the optimal solution $X_t^\star$ at iteration t is contained in the feasible set $\mathcal{FW}^n_{\alpha,2}(V_{t+1})$.

$$\begin{aligned} X_t^\star &= V_{t+1}^\star V_{t+1} \\ &= V_{t+1}^\star \times I \times V_{t+1} \end{aligned}$$

- Note that $I \in \mathcal{FW}^n_{\alpha,2} \implies X_t^\star \in \mathcal{FW}^n_{\alpha,2}(V_{t+1}) \implies \mathrm{U}_\alpha^t \geq \mathrm{U}_\alpha^{t+1}$.
- Instead of Cholesky factorization, other decompositions such as spectral decomposition also work.

Proposition (Monotonic decreasing upper bounds)

Given any partition α, inner approximations with matrices $\{V_t\}$ lead to

$$\mathrm{U}_\alpha^1 \geq \mathrm{U}_\alpha^2 \geq \ldots \geq \mathrm{U}_\alpha^t \geq \mathrm{U}_\alpha^{t+1} \geq p^\star.$$

$$\min_{x,y} \quad -x-y$$
$$\text{subject to} \quad I + xA + yB \succeq 0.$$

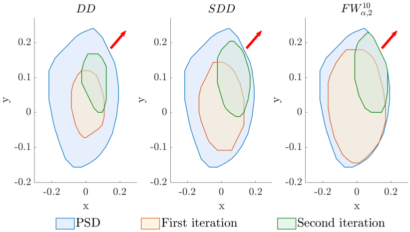


Figure: Feasible regions of inner approximations using $\mathcal{DD}_n$, $\mathcal{SDD}_n$, and $\mathcal{FW}^n_{\alpha,2}$ with $\alpha = \{2,2,2,2,2\}$. The red arrows denote the decreasing direction of the cost value.

Iterative outer approximations

- The dual cone of $\mathcal{FW}^n_{\alpha,2}$ naturally gives us an outer approximation

$$\mathcal{FW}^n_{\alpha,2} \subseteq \mathbb{S}^n_+ \subseteq (\mathcal{FW}^n_{\alpha,2})^*.$$

- Similar to inner approximation, we have

$$\begin{aligned} \mathsf{L}^t_\alpha := \min_X \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \qquad i = 1, \ldots, m, \\ & X \in (\mathcal{FW}^n_{\alpha,2}(V_t))^*. \end{aligned}$$

- We choose the sequence of matrices $\{V_t\}$ as

$$\begin{aligned} V_1 &= I \\ V_{t+1} &= \text{chol}\left(C - \sum_{i=1}^m y_i^{t,\star} A_i\right). \end{aligned}$$

Proposition (Monotonic increasing lower bounds)
Given any partition α, inner approximations with matrices $\{V_t\}$ lead to

$$\mathsf{L}_\alpha^1 \leq \mathsf{L}_\alpha^2 \leq \ldots \leq \mathsf{L}_\alpha^t \leq \mathsf{L}_\alpha^{t+1} \leq p^\star.$$

$$\begin{aligned} \min_{x,y} \quad & -x-y \\ \text{subject to} \quad & I + xA + yB \succeq 0. \end{aligned}$$

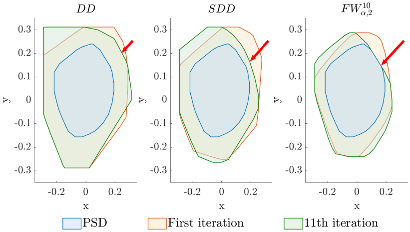


Figure: Feasible regions of outer approximations using $\mathcal{DD}_n$, $\mathcal{SDD}_n$, and $\mathcal{FW}^n_{\alpha,2}$ with $\alpha = \{2,2,2,2,2\}$. The red arrows denote the decreasing direction of the cost value.

$$\min_{x,y} \quad -x-y$$
$$\text{subject to} \quad I + xA + yB \succeq 0.$$

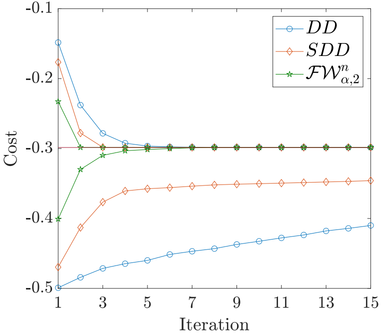


Figure: The evaluation of the cost value by different inner/outer approximations.

Table: Computational results of 7 different large-scale SDPs using inner approximation with $\alpha = \{10, \ldots, 10\}$ and $\beta = \{20, \ldots, 20\}$. f_1 denotes the cost value of the first iteration. f_{30} denotes the cost value after 30 minutes. The time consumption (in seconds) for solving the original SDP is listed in the last column.

	$\mathcal{FW}^n_{\alpha,2}$			$\mathcal{FW}^n_{\beta,2}$			PSD
n	f_1	f_{30}	Gap	f_1	f_{30}	Gap	Time
1500	5.63e6	4.76e6	0.03	5.20e6	4.76e6	0.03	603
2000	3.33e6	2.86e6	0.10	3.09e6	2.86e6	0.05	1 201
2500	6.11e6	5.29e6	0.07	5.70e6	5.29e6	0.05	2 893
3000	1.81e7	1.32e7	0.79	1.57e7	1.32e7	0.79	5 508
3500	8.96e6	7.08e6	0.10	8.02e6	7.07e6	0.08	7 369
4000	9.52e6	6.89e6	0.15	8.21e6	6.89e6	0.11	10 689
4500	2.05e7	1.70e7	0.08	1.88e7	1.69e7	0.06	16 989

- Different cones

$$\mathcal{DD}_n \subset \mathcal{SDD}_n \subset \mathcal{FW}^n_{\alpha,2} \subset \mathbb{S}^n_+$$

$$\text{LP} \implies \text{SOCP} \implies \text{Small SDP} \implies \text{SDP}$$

- Block-factor-width-two matrices

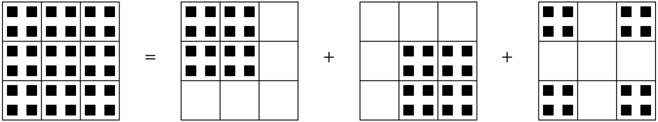


Figure: Illustration of block-factor-width-two matrices matrices ($\mathcal{FW}^n_{\alpha,2}$).

- A tight approximation quality with iterative inner/outer approximations.

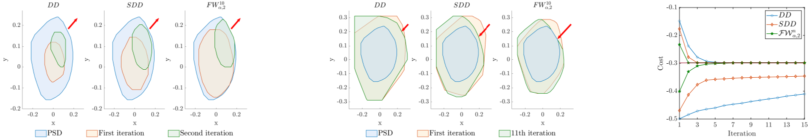


Figure: Iterative inner/outer approximation.

Thank you for your attention!

Q & A

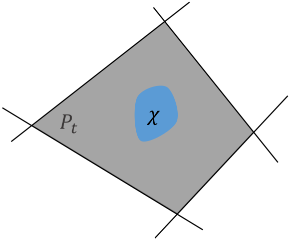
Cutting plane method

$$\min_x \quad f(x)$$
$$\text{subject to} \quad f_i(x) \leq 0.$$

- $f_0, f_1, \ldots, f_m$ are convex
- Suppose f is differentiable, f is convex if and only if

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \forall y \in \mathbb{R}^n.$$

- $\{f_i\}$ forms the feasible region $\mathcal{X}$
- $\mathcal{X}$ is complex and hard to optimize over
- Consider a bigger but simpler feasible region



Cutting plane method

- At iteration t, we consider

$$x_t^\star =: \min_x \quad f(x)$$
$$\text{subject to} \quad x \in P_t.$$

- If $x_t^\star \in \mathcal{X}, X_t^\star$ is the optimal solution.
- If $x_t^\star \notin \mathcal{X}$, there exists j such that

$$f_j(x_t^\star) > 0.$$

By first-order condition for convex functions

$$f_j(x) \geq f_j(x_t^\star) + \langle \nabla f_j(x_t^\star), x - x_t^\star \rangle, \forall x \in \mathbb{R}^n.$$

If $f_j(x_t^\star) + \langle \nabla f_j(x_t^\star), x - x_t^\star \rangle > 0$, then $f(x) > 0$ violates the constraint.

- Therefore, we need to impose

$$f_j(x_t^\star) + \langle \nabla f_j(x_t^\star), x - x_t^\star \rangle \leq 0.$$

Algorithm of cutting plane method

1. Given a simple set P_0 that contains the feasible region $\mathcal{X}$.
2. **(Initialization)** Initialize $x_0 \in \mathbb{R}^n$.
3. For $t \leq t_{\max}$
4. Solve
$$x_t^\star =: \min_x \quad f(x)$$
$$\text{subject to} \quad x \in P_t.$$
5. If $x_t \in \mathcal{X}$, quit.
6. $P_{t+1} = P_t \cap \{x \in \mathbb{R}^n | f_j(x_t^\star) + \langle \nabla f_j(x_t^\star), x - x_t^\star \rangle \leq 0.\}$
7. End For loop

Cutting plane method

- How to use it in SDP?
- Equivalent SDPs

$$\begin{aligned} \min_X \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \\ & X \in \mathbb{S}_+^n. \end{aligned} \qquad \begin{aligned} \min_X \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \\ & \lambda_{\min}(X) \geq 0. \end{aligned}$$

- $\lambda_{\min}(X) \geq 0 \iff \lambda_{\max}(-X) \leq 0$
- Consider

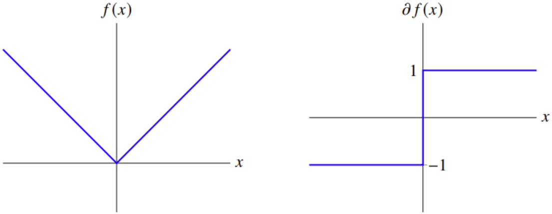
$$\begin{aligned} \min_X \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \\ & \lambda_{\max}(-X) \leq 0. \end{aligned}$$

- $g(X) = \lambda_{\max}(-X)$ is not differentiable. Fortunately, a subgradient exists!

Given a convex function $f : \mathbb{R}^n \to \mathbb{R}$, $z \in \mathbb{R}^n$ is a subgradient of f at $x \in dom(f)$ if

$$f(y) \geq f(x) + \langle z, y - x \rangle, \forall y \in dom(f)$$

- **Subdifferential example**
  $f(x) = |x|$



the picture is taken from Prof. L. Vandenberghe's lecture note.

Let $f(X) = \lambda_{\max}(-X)$. A subgradient of f at X can be computed as

$$-vv^{\mathsf{T}},$$

where v is the unit eigenvector of $\lambda_{\max}(-X)$.

- Suppose $X_t \notin \mathbb{S}^n_+$, $\lambda_{max}(-X_t) > 0$.
- From the subgradient inequality,

$$f(X) \geq f(X_t) + \left\langle -vv^{\mathsf{T}}, X - X_t \right\rangle$$

- We need to impose

$$\begin{aligned} & f(X_t) + \left\langle -vv^{\mathsf{T}}, X - X_t \right\rangle \leq 0 \\ \iff \quad & \lambda_{\max}(-X_t) + \left\langle -vv^{\mathsf{T}}, X - X_t \right\rangle \leq 0 \\ \iff \quad & \lambda_{\max}(-X_t) - \langle vv^{\mathsf{T}}, X \rangle - \lambda_{\max}(-X_t) \leq 0 \\ \iff \quad & \langle vv^{\mathsf{T}}, X \rangle \geq 0 \end{aligned}$$

Algorithm of cutting plane method for SDPs

1. Given a simple set P_0 that contains the feasible region $\mathcal{X}$.
2. **(Initialization)** Initialize $X_0 \in \mathbb{S}^n$
3. For $t \leq t_{\max}$
4. Solve

$$\begin{aligned} X_t^\star =: \min_X \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, i = 1, \ldots, m, \\ & X \in P_t. \end{aligned}$$

5. If $X_t \succeq 0$, quit.
6. Compute the eigenvector(v) of $\lambda_{max}(-X_t)$.
7. Set $P_{t+1} = P_t \cap \{x \in \mathbb{R}^n | \langle vv^\mathsf{T}, X \rangle \geq 0\}$.
8. End For loop