An Introduction on SemiDefinite Program

from the viewpoint of computation

Hayato Waki

Institute of Mathematics for Industry, Kyushu University

2015-10-08

Combinatorial Optimization at Work, Berlin, 2015

Institute of Mathematics for Industry
Kyushu University

Contents and Purpose of this lecture

Subject SemiDefinite Program

Contents

Part I Formulations & Strong duality on SDP

Part II Algorithm on SDP – Primal-Dual Interior-Point Methods

Part III Comments of Computation on SDP

Survey M. Todd, "Semidefinite optimization", Acta Numerica 10 (2001), pp. 515–560.

Purpose

- Better understanding for the next lecture (MOSEK on SDP) by Dr. Dahl
- Know the difficulty in solving SDP in Part III

Message : SDP is convex, but also nonlinear

Institute of Mathematics for Industry
Kyushu University

Properties and applications of SDP

$\boxed{\text{Properties}}$: SDP is an extension of LP

- Duality Theorem
- Solvable by primal-dual interior-point methods with up to a given tolerance

$\boxed{\text{Applications}}$

- Combinatorial problems, e.g., Max-Cut by Goemans and Williams
- Control theory, e.g., $\mathbf{H}_\infty$ control problem
- Lift-and-projection approach for nonconvex quadratic problem
- Lasserre's hierarchy for polynomial optimization problems and complexity theory
- Embedding problems, e.g., sensor networks and molecular conformation
- Statistics and machine learning, etc...

Institute of Mathematics for Industry
Kyushu University

From LP To SDP

$\boxed{\text{LP}}$ Primal and Dual

$$\begin{array}{ll|ll} \min_{\mathbf{x}} & \mathbf{c}^T\mathbf{x} & \max_{(\mathbf{y},\mathbf{s})} & \mathbf{b}^T\mathbf{y} \\ \text{s.t.} & \mathbf{a}_j^T\mathbf{x} = \mathbf{b}_j \ (\forall j) & \text{s.t.} & \mathbf{s} = \mathbf{c} - \sum_{j=1}^{m} \mathbf{y}_j\mathbf{a}_j \\ & \mathbf{x} \in \mathbb{R}^n_+ & & \mathbf{s} \in \mathbb{R}^n_+ \end{array}$$

- Minimize/Maximize linear function over the intersection the affine set and $\mathbb{R}^n_+$
- $\mathbb{R}^n_+$ is closed convex cone in $\mathbb{R}^n$

$\boxed{\text{Extension to SDP}}$

- Extension to the space of symmetric matrices $\mathbb{S}^n$

$$\mathbf{c} \in \mathbb{R}^n \to \mathbf{C} \in \mathbb{S}^n, \mathbf{a}_j \in \mathbb{R}^n \to \mathbf{A}_j \in \mathbb{S}^n$$

- Minimize/Maximize linear function over the intersection the affine set and the set of positive semidefinite matrices

Institute of Mathematics for Industry
Kyushu University

LP Primal and Dual

$$
\begin{array}{lll|ll}
\min_x & c^T x & & \max_{(y,s)} & b^T y \\
\text{s.t.} & a_j^T x = b_j \ (\forall j) & & \text{s.t.} & s = c - \sum_{j=1}^m y_j a_j \\
& x \in \mathbb{R}^n_+ & & & s \in \mathbb{R}^n_+
\end{array}
$$

SDP Primal and Dual

$$
\begin{array}{lll|ll}
\min_X & C \bullet X & & \max_{(y,S)} & b^T y \\
\text{s.t.} & A_j \bullet X = b_j \ (\forall j) & & \text{s.t.} & S = C - \sum_{j=1}^m y_j A_j \\
& X \in \mathbb{S}^n_+ & & & S \in \mathbb{S}^n_+
\end{array}
$$

- $\mathbb{S}^n$ is the set of $n \times n$ symmetry matrices,
- $\mathbb{S}^n_+$ is the set of $n \times n$ symmetry positive semidefinite matrices, and
- $A \bullet X := \sum_{k=1}^{n} \sum_{\ell=1}^{n} A_{k\ell} X_{k\ell}$.

Institute of Mathematics for Industry
Kyushu University

1. Definition of positive semidefinite matrices

$\mathbf{X} \in \mathbb{S}^{\mathbf{n}}$ is *positive semidefinite* if for all $\mathbf{z} \in \mathbb{R}^{\mathbf{n}}$, $\mathbf{z}^{\mathbf{T}}\mathbf{X}\mathbf{z} \geq \mathbf{0}$.
Equivalently, all eigenvalues are nonnegative.

Remark

- Eigendecomposition (Spectral decomposition); $\exists \mathbf{Q} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$ (orthogonal) and $\exists \boldsymbol{\lambda}_{\mathbf{i}} \geq \mathbf{0}$ such that

$$\mathbf{X} = \mathbf{Q} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \mathbf{Q}^{\mathbf{T}}$$

- See textbooks of linear algebra for proof
- $\Rightarrow \exists \mathbf{B} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$ such that $\mathbf{X} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$

2. Zero diagonal for positive semidefinite matrices

For $\mathbf{X} \in \mathbb{S}^{\mathbf{n}}_{+}$, each $\mathbf{X}_{\mathbf{ii}}$ is nonnegative. In addition, if $\mathbf{X}_{\mathbf{ii}} = \mathbf{0}$ for some $\mathbf{i}$, then $\mathbf{X}_{\mathbf{ij}} = \mathbf{X}_{\mathbf{ji}} = \mathbf{0}$ for all $\mathbf{j} = \mathbf{1}, \ldots, \mathbf{n}$.

Example of SDP

$$\mathbf{C} = \begin{pmatrix} 2 & \\ & 1 \end{pmatrix}, \mathbf{A}_1 = \begin{pmatrix} 10 & 4 \\ 4 & \end{pmatrix}, \mathbf{A}_2 = \begin{pmatrix} & \\ & -8 \end{pmatrix},$$

$$\mathbf{A}_3 = \begin{pmatrix} & -9 \\ -9 & 2 \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 42 & -8 & 20 \end{pmatrix}^T$$

Primal SDP is formulated as follows:

$$\inf_{\mathbf{X}} \left\{ 2x_{11} + x_{22} : \begin{array}{ll} 10x_{11} + 8x_{12} = 42, & -8x_{22} = -8, \\ -18x_{12} + 2x_{22} = 20, & \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \in \mathbb{S}^2_+ \end{array} \right\}$$

(Fortunately) the primal solution is uniquely fixed:

$$\mathbf{X} = \begin{pmatrix} 5 & -1 \\ -1 & 1 \end{pmatrix} \text{ is positive definite and obj. val.} = \mathbf{11}.$$

Institute of Mathematics for Industry
Kyushu University

Primal SDP is formulated as follows:

$$\inf_{X}\left\{2x_{11}+x_{22} : \begin{array}{ll} 10x_{11}+8x_{12}=42, & -8x_{22}=-8, \\ -18x_{12}+2x_{22}=20, & \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22}\end{pmatrix}\in\mathbb{S}^2_+ \end{array}\right\}$$

Dual SDP is formulated as follows:

$$\sup_{(y,S)}\left\{42y_1-8y_2+20y_3 : \begin{pmatrix} 2-10y_1 & -4y_1+9y_3 \\ -4y_1+9y_3 & 1+8y_2-2y_3\end{pmatrix}\in\mathbb{S}^2_+\right\}$$

A dual solution is $(1/5,-37/360,4/45)$ with the obj. val. $=11$.

Application : Computation of lower bounds of nonconvex QP

$\boxed{\text{QP}}$

$$\theta^* := \inf_{\mathbf{x}} \left\{ \mathbf{x}^T \mathbf{Q} \mathbf{x} + 2\mathbf{c}^T \mathbf{x} : \mathbf{x}^T \mathbf{Q}_j \mathbf{x} + 2\mathbf{c}_j^T \mathbf{x} + \mathbf{r}_j \leq 0 \ (\mathbf{j} = 1, \ldots, \mathbf{m}) \right\}$$

$\boxed{\text{SDP relaxation}}$: Add the following constraint and replace $\mathbf{x}_i \mathbf{x}_j \to \mathbf{X}_{ij}$:

$$\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} (1, \mathbf{x}) \in \mathbb{S}_+^{n+1} \to \mathbf{X} \in \mathbb{S}_+^{n+1}$$

$$\therefore \eta^* := \inf_{\mathbf{x}} \left\{ \begin{pmatrix} 0 & \mathbf{c}^T \\ \mathbf{c} & \mathbf{Q} \end{pmatrix} \bullet \mathbf{X} : \begin{pmatrix} \mathbf{r}_j & \mathbf{c}_j^T \\ \mathbf{c}_j & \mathbf{Q}_j \end{pmatrix} \bullet \mathbf{X} \leq 0, \mathbf{X}_{00} = 1, \mathbf{X} \in \mathbb{S}_+^{n+1} \right\}$$

$\boxed{\text{Remark}}$

- Handle as SDP
- $\eta^* \leq \theta^*$
- binary $\mathbf{x} \in \{0, 1\} \to \mathbf{x}^2 - \mathbf{x} = 0 \Rightarrow$ MIQP with binary variables $=$ QP

Application : Lasserre's SDP relaxation for Polynomial Optimization Problems

$\boxed{\text{POP}}$: $\mathbf{f}, \mathbf{g_j}$ are polynomials on $\mathbf{x} \in \mathbb{R}^{\mathbf{n}}$

$$\theta^* := \inf_{\mathbf{x}} \{\mathbf{f}(\mathbf{x}) : \mathbf{g_j}(\mathbf{x}) \geq \mathbf{0}\ (\mathbf{j} = \mathbf{1}, \ldots, \mathbf{m})\}$$

$\boxed{\text{Lasserre's SDP relaxation}}$

- Generates a sequence of SDP problems : $\{\mathbb{P}_{\mathbf{r}}\}_{\mathbf{r}\geq\mathbf{1}}^{\infty}$
- Optimal value : $\theta_{\mathbf{r}} \leq \theta_{\mathbf{r+1}} \leq \theta^*\ (\forall \mathbf{r})$
- Under assumptions, $\theta_{\mathbf{r}} \to \theta^*\ (\mathbf{r} \to \infty)$
- $\mathbf{r} = \mathbf{2}, \mathbf{3}$, $\theta_{\mathbf{r}} \approx \theta^*$ in practice
- Strongly connected to sum of square polynomials

Institute of Mathematics for Industry
Kyushu University

Compared with LP

Similar points

- Weak and Strong duality holds
- PDIPM also works in SDP

Different points

- SDP may have an irrational optimal solution

$$\text{E.g., } \sup_{\mathbf{y}} \left\{ \mathbf{y} : \begin{pmatrix} 2 & \mathbf{y} \\ \mathbf{y} & 1 \end{pmatrix} \in \mathbb{S}^2_+ \right\}$$

  Optimal solution $\mathbf{y} = \sqrt{2}$, not rational

- Finite optimal value, but $\nexists$ solutions

$$\text{E.g., } \inf_{\mathbf{y}} \left\{ \mathbf{y_1} : \begin{pmatrix} \mathbf{y_1} & 1 \\ 1 & \mathbf{y_2} \end{pmatrix} \in \mathbb{S}^2_+ \right\}$$

Different points (cont'd)

$\exists$ 2 types of infeasibility

$$\text{(LP) } \exists \mathbf{y}; -\mathbf{A}^T\mathbf{y} \in \mathbb{R}^n_+, \mathbf{b}^T\mathbf{y} > 0 \quad \Longleftrightarrow \quad \text{Primal LP is infeasible}$$

$$\text{(SDP) } \exists \mathbf{y}; -\mathbf{A}^T\mathbf{y} \in \mathbb{S}^n_+, \mathbf{b}^T\mathbf{y} > 0 \quad \Rightarrow \quad \text{Primal SDP is infeasible}$$

$\boxed{\text{Remark}}$: Need to consider the following cases

- Finite optimal value, but no optimal solutions for Primal and/or Dual
- Difficult to detect the infeasibility completely

Institute of Mathematics for Industry
Kyushu University

Duality on SDP

$\boxed{\text{Weak duality}}$ for any $\mathbf{X} \in \mathcal{F}_\mathbf{P}$ and $(\mathbf{y}, \mathbf{S}) \in \mathcal{F}_\mathbf{D}$,

$$\mathbf{C} \bullet \mathbf{X} \geq \mathbf{b}^\mathbf{T}\mathbf{y} \quad \therefore \theta^*_\mathbf{P} \geq \theta^*_\mathbf{D}$$

$\boxed{\text{Slater conditon}}$: $\mathbb{S}^\mathbf{n}_{++}$ is the set of positive definite matrices

- Primal satisfies *Slater condition* if $\exists \mathbf{X} \in \mathcal{F}_\mathbf{P}$ such that $\mathbf{X} \in \mathbb{S}^\mathbf{n}_{++}$
- Dual *Slater condition* if $\exists (\mathbf{y}, \mathbf{S}) \in \mathcal{F}_\mathbf{D}$ such that $\mathbf{S} \in \mathbb{S}^\mathbf{n}_{++}$

$\boxed{\text{Strong duality}}$

- Primal satisfies Slater condition and dual is feasible. Then $\theta^*_\mathbf{P} = \theta^*_\mathbf{D}$ and dual has an optimal solution.
- Slater condition are required for both primal and dual for theoretical results on PDIPMs
- See survey on SDP for proof

Institute of Mathematics for Industry
Kyushu University

3. Inner products on positive semidefinite matrices

For all $\mathbf{X}, \mathbf{S} \in \mathbb{S}^{\mathbf{n}}_{+}$, $\mathbf{X} \bullet \mathbf{S} \geq \mathbf{0}$. Moreover, $\mathbf{X} \bullet \mathbf{S} = \mathbf{0}$ iff $\mathbf{XS} = \mathbf{O_n}$

$\boxed{\text{Proof}}$: $\exists \mathbf{B}$ s. t. $\mathbf{X} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$ and $\exists \mathbf{D}$ s.t. $\mathbf{S} = \mathbf{D}\mathbf{D}^{\mathbf{T}}$. Then

$$\begin{aligned} \mathbf{X} \bullet \mathbf{S} &= \mathrm{Trace}(\mathbf{B}\mathbf{B}^{\mathbf{T}}\mathbf{D}\mathbf{D}^{\mathbf{T}}) = \mathrm{Trace}(\mathbf{D}^{\mathbf{T}}\mathbf{B}\mathbf{B}^{\mathbf{T}}\mathbf{D}) \\ &= \mathrm{Trace}((\mathbf{B}^{\mathbf{T}}\mathbf{D})^{\mathbf{T}}(\mathbf{B}^{\mathbf{T}}\mathbf{D})) \geq \mathbf{0} \end{aligned}$$

Moreover, $\mathbf{X} \bullet \mathbf{S} = \mathbf{0} \Rightarrow \mathbf{B}^{\mathbf{T}}\mathbf{D} = \mathbf{O_n} \Rightarrow \mathbf{XS} = \mathbf{O_n}$

$\boxed{\text{Proof of weak duality}}$

In fact, for $\mathbf{X} \in \mathcal{F}_{\mathbf{P}}$ and $(\mathbf{y}, \mathbf{S}) \in \mathcal{F}_{\mathbf{D}}$,

$$\mathbf{C} \bullet \mathbf{X} - \mathbf{b}^{\mathbf{T}}\mathbf{y} = \left(\mathbf{C} - \sum_{\mathbf{j}=1}^{\mathbf{m}} \mathbf{y_j}\mathbf{A_j}\right) \bullet \mathbf{X} = \mathbf{S} \bullet \mathbf{X} \geq \mathbf{0}$$

because both matrices are positive semidefinite.

Institute of Mathematics for Industry
Kyushu University

Remark of 3 (cont'd)

- $\mathbf{X} \in \mathcal{F}_{\mathbf{P}}$: optimal in primal and $(\mathbf{y}, \mathbf{S}) \in \mathcal{F}_{\mathbf{D}}$: optimal in dual
- Then, $\theta_{\mathbf{P}}^* - \theta_{\mathbf{D}}^* = \mathbf{X} \bullet \mathbf{S} = \mathbf{0} \iff \mathbf{XS} = \mathbf{O_n}$
- $\mathbf{XS} = \mathbf{O_n}$ is used in PDIPM

SDP with multiple positive semidefinite cones

SDP

$$\begin{array}{ll} \inf_{X_k} & \sum_{k=1}^{N} C^k \bullet X_k \\ \text{s.t.} & \sum_{k=1}^{N} A_j^k \bullet X_k = b_j \ (j = 1, \ldots, m) \\ & X_k \in \mathbb{S}_+^{n_k} \ (k = 1, \ldots, N) \end{array}$$

where $C^k, A_j^k \in \mathbb{S}^{n_k}$

Example

$$A \bullet X \leq d, X \in \mathbb{S}_+^n \Rightarrow A \bullet X + s = d, X \in \mathbb{S}_+^n \text{ and } s \in \mathbb{S}_+^1 (= \mathbb{R}_+)$$

Dual

$$\sup_{y, S_k} \left\{ b^T y : S_k = A_0^k - \sum_{j=1}^{m} y_j A_j^k \in \mathbb{S}_+^{n_k} \ (k = 1, \ldots, N) \right\}$$

Remark

- SDP with $\mathbb{R}^n_+$, Second order cone $\mathbf{L_n}$ and $\mathbb{S}^n_+$ can be handled as SDP and PDIPM works

$$\mathbf{L_n} := \{(\mathbf{x_0}, \mathbf{x}) \in \mathbb{R}^{\mathbf{n}} : \|\mathbf{x}\|_2 \leq \mathbf{x_0}\}$$

- Free variable can be accepted

$$\begin{aligned} & \mathbf{A} \bullet \mathbf{X} + \mathbf{a}^T\mathbf{x} = \mathbf{d}, \mathbf{X} \in \mathbb{S}^n_+, \mathbf{x} \in \mathbb{R}^n \\ \Rightarrow \quad & \mathbf{A} \bullet \mathbf{X} + \mathbf{a}^T\mathbf{x_1} - \mathbf{a}^T\mathbf{x_2} = \mathbf{d}, \mathbf{X} \in \mathbb{S}^n_+ \text{ and } \mathbf{x_1}, \mathbf{x_2} \in \mathbb{R}^n_+ \end{aligned}$$

Institute of Mathematics for Industry
Kyushu University

Classification of Algorithms for SDP

Algorithms for SDP

- Ellipsoid method
- Interior-point methods
- Bundle method
- first-order methods, etc

Interior-point methods

- Path-following algorithm (= Logarithmic barrier function)
- Potential reduction algorithm
- Self-dual homogeneous embeddings

Path-following algorithm

- Primal
- Dual
- Primal-dual

Institute of Mathematics for Industry
Kyushu University

Path-following method

$\boxed{\text{Optimality conditions}}$: a pair of optimal solutions $(\mathbf{X}, \mathbf{y}, \mathbf{S})$ satisfies

$$\begin{cases} \mathbf{A_j} \bullet \mathbf{X} = \mathbf{b_j}, \mathbf{X} \in \mathbb{S}^n_+, \\ \mathbf{S} = \mathbf{C} - \sum_{j=1}^m \mathbf{y_j A_j}, \mathbf{S} \in \mathbb{S}^n_+, \\ \mathbf{XS} = \mathbf{O_n} (\iff \mathbf{C} \bullet \mathbf{X} - \mathbf{b}^T \mathbf{y} = 0) \end{cases}$$

$\boxed{\text{Perturbed system}}$: for $\mu > 0$,

$$\begin{cases} \mathbf{A_j} \bullet \mathbf{X} = \mathbf{b_j}, \mathbf{X} \in \mathbb{S}^n_{++}, \\ \mathbf{S} = \mathbf{C} - \sum_{j=1}^m \mathbf{y_j A_j}, \mathbf{S} \in \mathbb{S}^n_{++}, \\ \mathbf{XS} = \mu \mathbf{I_n} \end{cases}$$

$\boxed{\text{Remark}}$

- for any $\mu > 0$, $\exists$ unique solution $(\mathbf{X}(\mu), \mathbf{y}(\mu), \mathbf{S}(\mu))$
- Central path $\{(\mathbf{X}(\mu), \mathbf{y}(\mu), \mathbf{S}(\mu)) : \mu > 0\}$ is smooth curve and go to a pair of optimal solutions of primal and dual
- Follows the central path = Path-following method

Institute of Mathematics for Industry
Kyushu University

Algorithm 1: General framework of path-following method

Input: $(\mathbf{X^0}, \mathbf{y^0}, \mathbf{S^0}) \in \mathcal{F}_\mathbf{P} \times \mathcal{F}_\mathbf{D}$ such that $\mathbf{X^0}, \mathbf{S^0} \in \mathbb{S}^\mathbf{n}_{++}$, $\epsilon > \mathbf{0}$, $\mathbf{0} < \theta < \mathbf{1}$ and some parameters

$\mathbf{X} \leftarrow \mathbf{X^0}$, $\mathbf{y} \leftarrow \mathbf{y^0}$ and $\mathbf{S} \leftarrow \mathbf{S^0}$;

while $\mathbf{X} \bullet \mathbf{S} > \epsilon$ **do**

Compute direction $(\mathbf{\Delta X}, \mathbf{\Delta y}, \mathbf{\Delta S})$ from CPE(μ);

Compute step size $\alpha_\mathbf{P}, \alpha_\mathbf{D} > \mathbf{0}$;

$\mathbf{X} \leftarrow \mathbf{X} + \alpha_\mathbf{P} \mathbf{\Delta X}$;

$\mathbf{y} \leftarrow \mathbf{y} + \alpha_\mathbf{D} \mathbf{\Delta y}$; $\mathbf{S} \leftarrow \mathbf{S} + \alpha_\mathbf{D} \mathbf{\Delta S}$;

Compute $\mu \leftarrow \theta\mu$;

end

return $(\mathbf{X}, \mathbf{y}, \mathbf{S})$;

Remark

- Infeasible initial guess is acceptable
- $\#$ of iteration is polynomial in $\mathbf{n}, \mathbf{m}$ and $\mathbf{log}(\epsilon)$
- Computational cost = Computation of direction

Institute of Mathematics for Industry
Kyushu University

$\boxed{\text{Computation of direction}}$: Find $(\mathbf{\Delta X}, \mathbf{\Delta y}, \mathbf{\Delta S})$ such that
$\mathbf{X} + \mathbf{\Delta X} \in \mathcal{F}_{\mathbf{P}}$, $(\mathbf{y} + \mathbf{\Delta y}, \mathbf{S} + \mathbf{\Delta S}) \in \mathcal{F}_{\mathbf{D}}$ and

$$\begin{cases} \mathbf{A_j} \bullet \mathbf{\Delta X} = \mathbf{0}, \\ \mathbf{\Delta S} - \sum_{\mathbf{j=1}}^{\mathbf{m}} \mathbf{\Delta y_j A_j} = \mathbf{O_n}, \\ \mathbf{XS} + \mathbf{\Delta X S} + \mathbf{X \Delta S} = \mu \mathbf{I_n} \end{cases}$$

$\boxed{\text{Remark}}$

- $\mathbf{\Delta X}$ *may not* be symmetry. So, change $\mathbf{XS} = \mu \mathbf{I_n}$ by

$$\frac{\mathbf{1}}{\mathbf{2}} \left(\mathbf{PXSP^{-1}} + \mathbf{P^{-T}SXP^{T}} \right) = \mu \mathbf{I_n},$$

  where $\mathbf{P}$ is nonsingular
- Possible choice of $\mathbf{P}$

$$\begin{aligned} \mathbf{P} &= \mathbf{S^{1/2}} \text{ (HRVW/KSH/M)} \\ \mathbf{P} &= \mathbf{X^{-1/2}} \text{ (dual HRVW/KSH/M)} \\ \mathbf{P} &= \mathbf{W^{1/2}}, \mathbf{W} = \mathbf{X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}} \text{ (NT)} \\ \mathbf{P} &= \ldots \text{ More than 20 types of directions by Todd} \end{aligned}$$

Institute of Mathematics for Industry
Kyushu University

Computational cost in PDIPM

$\boxed{\text{1. Construction of linear system on } \mathbf{\Delta y}}$ for HRVW/KSH/M direction,

$$\mathbf{M\Delta y} = (\text{RHS}), \text{where } \mathbf{M} = (\text{Trace}(\mathbf{A_i X A_j S^{-1}}))_{\mathbf{1 \le i,j \le m}}$$

- Use of sparsity in $\mathbf{A_j}$ is necessary for computation of $\mathbf{M}$
- Almost the same for other search directions

$\boxed{\text{2. Solving the linear system}}$

- $\mathbf{M}$ is dense $\Rightarrow$ takes $\mathbf{O(m^3)}$ computation by Cholesky decomposition
- $\mathbf{M}$ is often sparse in SDP relax for POP $\Rightarrow$ sparse Cholesky decomposition works well

After them , $\mathbf{\Delta S} = \sum_{\mathbf{j=1}}^{\mathbf{m}} \mathbf{\Delta y_j A_j}$ and obtain $\mathbf{\Delta X}$.

Institute of Mathematics for Industry
Kyushu University

Sparsity in SDP

$\boxed{\text{Example}}$ $\mathbf{Q}$ is nonsingular and dense. Then $\mathbb{P}_1$ is equivalent to $\mathbb{P}_2$:

$$\mathbb{P}_1 \quad : \quad \inf_{\mathbf{X}} \left\{ \mathbf{C} \bullet \mathbf{X} : \mathbf{E}_i \bullet \mathbf{X} = 1 \ (i = 1, \ldots, n), \mathbf{X} \in \mathbb{S}^n_+ \right\},$$

$$\mathbb{P}_2 \quad : \quad \inf_{\mathbf{X}} \left\{ (\mathbf{Q}^T \mathbf{C} \mathbf{Q}) \bullet \mathbf{X} : (\mathbf{Q}^T \mathbf{E}_i \mathbf{Q}) \bullet \mathbf{X} = 1 \ (i = 1, \ldots, n), \mathbf{X} \in \mathbb{S}^n_+ \right\}$$

where

$$(\mathbf{E}_i)_{pq} = \begin{cases} 1 & \text{if } p = q = i \\ 0 & \text{o.w.} \end{cases} \quad (p, q = 1, \ldots, n)$$

Institute of Mathematics for Industry
Kyushu University

$\boxed{\text{CPU time}}$: Solved by SeDuMi 1.3 on the MacBook Air (1.7 GHz Intel Core i7)

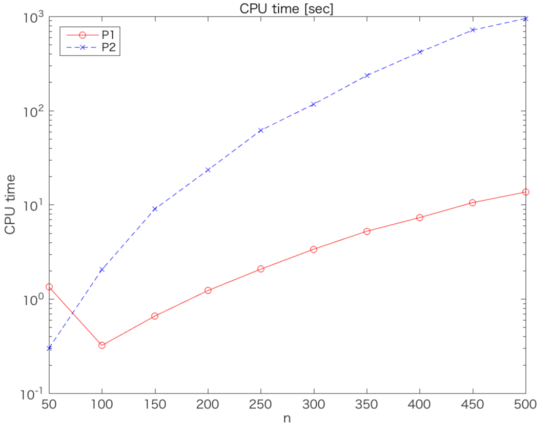

Figure : CPU time on $\mathbb{P}_1$ and $\mathbb{P}_2$

Software

Information from `http://plato.asu.edu/ftp/sparse_sdp.html`

- SeDuMi, SDPT3 (MATLAB)
- SDPA (C++, MATLAB)
- CSDP (C, MATLAB)
- DSDP (C, MATLAB)
- MOSEK

Remark

- Based on PDIPM for almost all software
- Performance depends on SDP problems

Modelling languages on SDP : they can call the above software

- YALMIP
- CVX

Institute of Mathematics for Industry
Kyushu University

Slater conditions

Strong duality

- Require Slater conditions for Primal or Dual
- PDIPM requires Slater conditions for both Primal and Dual
- Sufficient conditions for optimal solutions
- If either Primal or Dual does not satisfy Slater conditions, ...

E.g., Lasserre's SDP relaxation

$$\mathbb{P} : \inf_{x} \left\{ x : x^2 - 1 \geq 0, x \geq 0 \right\}$$

- Gererate SDP relaxation problems $\mathbb{P}_1, \mathbb{P}_2, \ldots,$
- Slater condition fails in all SDP relaxation & all optimal values are $\mathbf{0}$
- SeDuMi and SDPA returns wrong value $\mathbf{1}$
- All SDP relaxation problems are sensitive to numerical errors in the computation of floating points

E.g., Graph Equipartition

- $\mathbf{G(V,E)}$: a weighted undirected graph $\Rightarrow$ Partition the vertex set $\mathbf{V}$ into $\mathbf{L}$ and $\mathbf{R}$
- the minimum total weight of the cut subject to $|\mathbf{L}| = |\mathbf{R}|$
- QOP formulation

$$\inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \sum w_{ij}(1 - x_i x_j) : \sum_{i=1}^{n} x_i = 0, {x_i}^2 = 1 \ (i = 1, \ldots, n) \right\}$$

E.g., Graph Equipartition (cont'd)

- SDP relaxation problem: constant matrices $\mathbf{W}$, $\mathbf{E}$ and $\mathbf{E_i}$

$$\inf_{\mathbf{X} \in \mathbb{S}^n_+} \{\mathbf{W} \bullet \mathbf{X} \mid \mathbf{E} \bullet \mathbf{X} = 0, \mathbf{E_i} \bullet \mathbf{X} = 1\}$$

- Since $\mathbf{E} \in \mathbb{S}^n_+$, $\nexists \mathbf{X} \in \mathbb{S}^n_{++}$ s.t. $\mathbf{E} \bullet \mathbf{X} = \mathbf{0} \Rightarrow$ Slater cond. fails
- Inaccurate value and/or many iterations

Table : SeDuMi 1.3 with ϵ=1.0e-8

SDPLIB	iter	cpusec	duality gap
gpp124-1	30	2.40	-4.63e-05
gpp250-1	29	10.19	-1.60e-04
gpp500-1	34	61.58	-1.90e-04
gpp124-4	40	3.02	-2.14e-08
gpp500-2	40	76.88	-8.26e-06

Institute of Mathematics for Industry
Kyushu University

E.g., Graph Equipartition (cont'd)

$$\inf_{\mathbf{X} \in \mathbb{S}^n_+} \{\mathbf{W} \bullet \mathbf{X} \mid \mathbf{E} \bullet \mathbf{X} = 0, \mathbf{E_i} \bullet \mathbf{X} = 1\}$$

- Transformation of SDP by $\mathbf{V}$:

$$\mathbf{V} = \begin{pmatrix} 1 & & & -1 \\ & 1 & & -1 \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$$

- $\mathbf{X} \to \mathbf{V}^{-T}\mathbf{X}\mathbf{V}^{-1} =: \mathbf{Z}$ and $\mathbf{E} \to \mathbf{V}\mathbf{E}\mathbf{V}^T$
- Then, $\mathbf{X} \in \mathbb{S}^n_+ \iff \mathbf{Z} \in \mathbb{S}^n_+$ and $\mathbf{E} \bullet \mathbf{X} = 0 \iff \mathbf{Z_{nn}} = 0$
- Eliminate $\mathbf{n}$th row and column from transformed SDP $\Rightarrow$ Slater cond. holds

E.g., Graph Equipartition (cont'd)

Table : Numerical Results by SeDuMi 1.3 with ϵ=1.0e-8.

	Slater fails			Slater holds		
Problems	iter	cpusec	d.gap	d.gap	cpusec	iter
gpp100	30	1.78	-2.46e-07	-4.97e-09	0.73	16
gpp124-1	30	2.34	-4.63e-05	-1.75e-08	1.12	19
gpp124-2	26	1.76	-1.41e-06	-1.11e-09	1.03	18
gpp124-3	30	2.56	-4.41e-07	-3.05e-09	1.01	17
gpp124-4	40	2.93	-2.14e-08	-9.52e-11	1.09	17
gpp250-1	29	8.81	-1.60e-04	-1.82e-08	4.71	21
gpp250-2	29	8.61	-1.49e-05	-9.74e-09	4.19	19
gpp250-3	34	9.48	-3.97e-07	-8.12e-10	4.08	18
gpp250-4	35	11.28	-8.80e-07	-7.43e-10	4.37	19
gpp500-1	34	53.45	-1.90e-04	-2.76e-08	31.49	24
gpp500-2	40	68.47	-8.26e-06	-2.20e-09	28.98	22
gpp500-3	28	54.81	-1.00e-05	-2.39e-09	31.35	21
gpp500-4	28	55.06	-1.02e-06	-8.96e-10	32.06	23

Comments : If does not satisfy Slater conditions, ...

- PDIPM computes inaccurate values and/or spends many iter.
- But, reduce the size of SDP

Comments

- A simple (?) transformation generates an SDP in which Slater cond. holds
- More elementary approach :

$$(\text{QOP}) \quad : \quad \inf_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\sum w_{ij}(1 - x_i x_j) : \sum_{i=1}^{n} x_i = 0, {x_i}^2 = 1 \right\}$$

$$(\text{QOP'}) \quad : \quad \text{obtained by substituting } x_1 = -\sum_{i=2}^{n} x_i \text{ in (QOP)}$$

$$\begin{array}{ccc} \text{(QOP)} & \xrightarrow{\text{equiv.}} & \text{(QOP')} \\ \downarrow \text{SDP relax.} & & \text{SDP relax.} \downarrow \\ \text{(SDP)} & \xrightarrow[\text{equiv.}]{\text{V}} & \text{(SDP')} \end{array}$$

- General case : separate $\mathbf{x}$ into basic and nonbasic variables & substitute basic variables $\Rightarrow$ SDP relax

$$\inf_{x} \left\{ x^T Q x + 2c^T x : a_j^T x = b_j \ (j = 1, \ldots, m), x_k \in \{0, 1\} \right\}$$

Institute of Mathematics for Industry Kyushu University

Extension

$\boxed{\text{SDP}}$

$$\inf_{\mathbf{X}} \left\{ \mathbf{C} \bullet \mathbf{X} : \mathbf{A_j} \bullet \mathbf{X} = \mathbf{b_j}, \mathbf{X} \in \mathbb{S}^{\mathbf{n}}_{+} \right\}$$

Slater condition fails in Primal $\iff$ $\exists \mathbf{y} \in \mathbb{R}^{\mathbf{m}} \setminus \{\mathbf{0}\}$ such that

$$\mathbf{b}^{\mathbf{T}}\mathbf{y} \geq \mathbf{0}, -\sum_{\mathbf{j}} \mathbf{y_j A_j} \in \mathbb{S}^{\mathbf{n}}_{+}$$

Moreover, if $\exists \mathbf{y}$ such that $\mathbf{b}^{\mathbf{T}}\mathbf{y} > \mathbf{0}$, then Primal is infeasible

$\boxed{\text{Proof of } (\Leftarrow)}$: Suppose the contrary that Slater condition holds in Primal. $\exists \hat{\mathbf{X}}$ such that $\mathbf{A_j} \bullet \hat{\mathbf{X}} = \mathbf{b_j}$ and $\hat{\mathbf{X}} \in \mathbb{S}^{\mathbf{n}}_{++}$.

$$\mathbf{0} \quad \leq \quad \mathbf{b}^{\mathbf{T}}\mathbf{y} = \sum_{\mathbf{j}} (\mathbf{A_j} \bullet \hat{\mathbf{X}})\mathbf{y_j} = \left(\sum_{\mathbf{j}} \mathbf{A_j y_j} \right) \bullet \hat{\mathbf{X}} < \mathbf{0}(\text{contradiction})$$

Facial Reduction

$\boxed{\text{Idea}}$: Let $\mathbf{W} := -\sum_j \mathbf{A}_j \mathbf{y}_j \in \mathbb{S}^n_+$ and $\mathbf{b}^T\mathbf{y} = 0$

- For any feasible solutions $\mathbf{X}$ in Primal,

$$\mathbf{W} \bullet \mathbf{X} = -\sum_j (\mathbf{A}_j \bullet \mathbf{X})\mathbf{y}_j = -\mathbf{b}^T\mathbf{y} = 0.$$

- Primal is equivalent to

$$\inf_{\mathbf{X}} \left\{ \mathbf{C} \bullet \mathbf{X} : \mathbf{A}_j \bullet \mathbf{X} = \mathbf{b}_j, \mathbf{X} \in \mathbb{S}^n_+ \cap \{\mathbf{W}\}^\perp \right\}$$

where $\{\mathbf{W}\}^\perp := \{\mathbf{X} : \mathbf{X} \bullet \mathbf{W} = 0\}$

- The set $\mathbb{S}^n_+ \cap \{\mathbf{W}\}^\perp$ has nice structure

$$\mathbb{S}^n_+ \cap \{\mathbf{W}\}^\perp = \left\{ \mathbf{X} \in \mathbb{S}^n : \mathbf{X} = \mathbf{Q} \begin{pmatrix} \mathbf{M} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \mathbf{Q}^T, \mathbf{M} \in \mathbb{S}^r_+ \right\}$$

Institute of Mathematics for Industry
Kyushu University

Idea (cont'd)

$$\mathbb{S}^n_+ \cap \{W\}^\perp = \left\{ X \in \mathbb{S}^n : X = Q \begin{pmatrix} M & O \\ O & O \end{pmatrix} Q^T, M \in \mathbb{S}^r_+ \right\}$$

- Assume $Q = I_n$. Then Primal is equivalent to

$$\inf_X \left\{ \tilde{C} \bullet X : \tilde{A}_j \bullet X = b_j, X \in \mathbb{S}^r_+ \right\}$$

  where $\tilde{A}_j$ is $r \times r$ principal matrix
- Compare this SDP with Primal $\Rightarrow$ the size $n \to r$
- May not satisfy Slater cond.
- $\Rightarrow$ Find y and W for the smaller Primal
- This procedure terminates in finitely many iterations
- This procedure is called Facial Reduction Algorithm and acceptable for dual

Institute of Mathematics for Industry
Kyushu University

Histroy of FRA

- Borwein-Wolkowicz in 1980 for general convex optimization
- Ramana, Ramana-Tunçel-Wolkowicz for SDP
- Pataki simplified FRA for the extension
- Apply FRA into SDP relax. for Graph Partition, Quadratic Assignment, Sensor Network by Wolkowicz group
- Apply FRA into SDP relax. for Polynomial Optimization in Waki-Muramatsu
- ...

Summary on Slater condition

- Hope that both Primal and dual satisfy Slater conditions
- Otherwise, may not have any optimal solutions, and wrong value may be obtained
- Obtain inaccurate solutions even if exists optimal solutions, but, one can reduce the size of SDP
- FRA is a general framework to remove the difficulty in Slater cond.

In modeling to SDP...

- Need to be careful in even dual to guarantee the existence of optimal solutions in dual
- A rigorous solution for FRA is necessary

Status of infeasibility

Feasiblity and infeasiblity

$$\inf_{\mathbf{X}} \left\{ \mathbf{C} \bullet \mathbf{X} : \mathbf{A}_j \bullet \mathbf{X} = \mathbf{b}_j, \mathbf{X} \in \mathbb{S}^n_+ \right\}$$

- Strongly feasible if SDP satisfies Slater cond.
- Weakly feasible if SDP is feasible but, does not satisfies Slater cond.
- Strongly infeasible if $\exists$ improving ray **d**, *i.e.*,

$$\mathbf{b}^T\mathbf{d} > 0, -\sum_j \mathbf{d}_j \mathbf{A}_j \in \mathbb{S}^n_+.$$

- Weakly infeasible if SDP is infeasible, but $\nexists$ improving ray

Remark

- Weak infeasibility does not occur in LP
- SOCP and conic optimization also have the four status

Institute of Mathematics for Industry
Kyushu University

$\boxed{\text{Example}}$: Infeasible SDPs

$$\mathbb{P}_1 \qquad \inf_{\mathbf{X}} \left\{ \mathbf{C} \bullet \mathbf{X} : \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \bullet \mathbf{X} = 0, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \bullet \mathbf{X} = 2, \mathbf{X} \in \mathbb{S}^2_+ \right\},$$

$$\mathbb{P}_2 \qquad \inf_{\mathbf{X}} \left\{ \mathbf{C} \bullet \mathbf{X} : \begin{pmatrix} & \\ & 1 \end{pmatrix} \bullet \mathbf{X} = 0, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \bullet \mathbf{X} = 2, \mathbf{X} \in \mathbb{S}^2_+ \right\}$$

$\boxed{\text{Comments}}$

- $\mathbb{P}_1$ is strongly infeasible because $\exists$ certificate $\mathbf{y} = (-1, 1)$
- $\mathbb{P}_2$ is weakly infeasible because $\nexists$ certificate

Characterization of weak infeasibility

- Weakly infeasible SDP; for all $\epsilon >$, $\exists \mathbf{X} \in \mathbb{S}^n_+$

$$|\mathbf{A_j} \bullet \mathbf{X} - \mathbf{b_j}| < \epsilon \ (\mathbf{j} = \mathbf{1}, \ldots, \mathbf{m})$$

- More elementary characterization of Weak infeasibility by recent work by Liu and Pataki

Example $\mathbb{P}_2$ Perturb $\mathbf{b_1} = \mathbf{0} \to \epsilon > \mathbf{0}$

$$\mathbb{P}_2 : \inf_{\mathbf{X}} \left\{ \mathbf{C} \bullet \mathbf{X} : \begin{pmatrix} & \\ & \mathbf{1} \end{pmatrix} \bullet \mathbf{X} = \epsilon, \begin{pmatrix} & \mathbf{1} \\ \mathbf{1} & \end{pmatrix} \bullet \mathbf{X} = \mathbf{2}, \mathbf{X} \in \mathbb{S}^2_+ \right\}$$

Then, perturbed $\mathbb{P}_1$ is feasible:

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}/\epsilon & \mathbf{1} \\ \mathbf{1} & \epsilon \end{pmatrix}$$

Institute of Mathematics for Industry
Kyushu University

Pathological?

$$\text{(POP)} : \inf_{x,y} \{-x - y : xy \leq 1/2, x \geq 1/2, y \geq 1/2\}$$

- Optimal value is -1.5
- Apply Lasserre's SDP hierarchy
- All SDP relaxation is weakly infeasible (in Waki 2012)
- SeDuMi and SDPA returns -1.5 for higher oder SDP relaxation
- Sufficient conditions of (POP) for SDP relaxation to be weakly infeasible (in Waki 2012)

Summary on infeasibility

- Weak infeasibility may occur in SDP, SOCP and conic optimization, but not in LP
- Difficult to detect this type of infeasibility by software
- But, software returns good values for weak infeasible SDP

Summary

- Introduce a part of theoretical and practical aspects in SDP
- Skip applications of SDP, *e.g.*, SDP relaxation for combinatorial problems
- Can read papers on SDP
- Not so easy to handle SDP because it is convex but nonlinear programming

Further Reading I

- M. Anjos and JB Lasserre,
  Handbook of Semidefinite, Conic and Polynomial Optimization: Theory, Algorithms.
  International Series in Operations Research & Management Science, Springer US, 2012.

- E. de Klerk,
  Aspects of semidefinite programming : interior point algorithms and selected applications.
  Applied Optimization, Springer US, 2002.

- B. Gärtner and J. Matoušek
  Approximation Algorithms and Semidefinite Programming.
  Springer, 2012.

Institute of Mathematics for Industry
Kyushu University

Further Reading II

M. Todd,
Semidefinite optimization.
Acta Numerica 10 (2001), pp. 515–560.

L. Tunçel,
Polyhedral and SDP Methods in Combinatorial Optimization.
IFields Institute Monographs, American Mathematical Society, 2012.

H. Wolkowicz, R. Saigal and L. Vandenberghe,
Handbook of Semidefinite Programming.
International Series in Operations Research & Management Science, Springer US, 2000.