AMA 505: Optimization Methods

Subject Lecturer: Ting Kei Pong

Lecture 6
Semidefinite Programming I
Duality Theory

AX is convex function

Semidefinite Programming

affine transformation is conve

Our main focus in this lecture and the next is

$\{X : X \succeq 0\}$ is convex

$$\begin{array}{ll} \underset{X \in \mathcal{S}^n}{\text{Minimize}} & \text{tr}(CX) \\ \text{subject to} & \text{tr}(A_i X) = b_i, \quad i = 1, \ldots, m, \\ & X \succeq 0, \end{array}$$

convex problem: Def

$\begin{cases} \text{tr}(CX) \text{ convex} \\ \text{feasible region is conve} \end{cases}$

Here:

- $\mathcal{S}^n$ is the space of all real symmetric matrices.
- C and A_i are real symmetric matrices.
- For an $Y \in \mathbb{R}^{n \times n}$, $\text{tr}(Y) := \sum_{i=1}^n y_{ii}$.
- The constraint $X \succeq 0$ requires the symmetric matrix X to be positive semidefinite, i.e., all eigenvalues are nonnegative.
- These problems are called semidefinite programming (SDP) problems.
- The feasible region is convex. (CHECK!)
- SDPs are convex problems.

What is $\mathrm{tr}(AX)$?

For $A, B \in \mathcal{S}^n$, we have

$$\mathrm{tr}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}.$$

Note: $\mathrm{tr}(AB)$ is really the vector inner product of the vectors $\mathrm{vec}(A)$ (obtained by stacking columns of A) and $\mathrm{vec}(B)$ (obtained by stacking columns of B).

Example: Let $A = \begin{bmatrix} 2 & 3 \\ 3 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -2 \\ -2 & 6 \end{bmatrix}$. Then $\mathrm{tr}(AB)$ equals

$$2\cdot 1 + 3\cdot(-2) + 3\cdot(-2) + (-1)\cdot 6 = \begin{bmatrix} 2 \\ 3 \\ 3 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1 \\ -2 \\ -2 \\ 6 \end{bmatrix} = [\mathrm{vec}(A)]^T \mathrm{vec}(B).$$

LPs are SDPs

Consider the following linear program.

$$\begin{array}{ll} \underset{x\in\mathbb{R}^2}{\text{Minimize}} & x_1 - x_2 \\ \text{subject to} & 6x_1 + x_2 = 3, \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

We show that this is equivalent to an instance of SDP.

KEY FACT: If $X \in \mathbb{R}^{n\times n}$ is diagonal, then $X \succeq 0$ if and only if $x_{ii} \geq 0$ for all i.

Now, thinking of $X = \begin{bmatrix} x_1 & x_3 \\ x_3 & x_2 \end{bmatrix}$, then the above is equivalent to

$$\begin{array}{ll} \underset{X\in\mathcal{S}^2}{\text{Minimize}} & \mathrm{tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 & x_3 \\ x_3 & x_2 \end{bmatrix}\right) \\ \text{subject to} & \mathrm{tr}\left(\begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 & x_3 \\ x_3 & x_2 \end{bmatrix}\right) = 3, \; \mathrm{tr}\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 & x_3 \\ x_3 & x_2 \end{bmatrix}\right) = 0, \quad X \succeq 0. \end{array}$$

positive definite $\Rightarrow$ symmetric

LPs are SDPs cont.

More generally, consider the following linear program.

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$. We show that this is equivalent to an instance of SDP.

KEY FACT: If $X \in \mathbb{R}^{n \times n}$ is diagonal, then $X \succeq 0$ if and only if $x_{ii} \geq 0$ for all i.

Let $X \in \mathcal{S}^n$ and think of its diagonal to be x. Then

$$c^T x = \text{tr}[\text{Diag}(c)X], \quad \mathbf{a}_j^T x = \text{tr}[\text{Diag}(\mathbf{a}_j)X],$$

where $\mathbf{a}_j^T$ is the jth row of A, and $\text{Diag}(c) \in \mathbb{R}^{n \times n}$ is the diagonal matrix with diagonal being c.

LPs are SDPs cont.

Next, to enforce that X is diagonal, we impose $x_{ij} = 0$ whenever $i \neq j$. These are given by

$$\text{tr}[E_{ij}X] = 0,$$

where E_{ij} is the symmetric matrix that is $\frac{1}{2}$ at the ij and jith entries, and is zero otherwise. Why **two** $\frac{1}{2}$'s?

Thus, (1) is equivalent to the following SDP:

$$\begin{aligned} \underset{X \in \mathcal{S}^n}{\text{Minimize}} \quad & \text{tr}[\text{Diag}(c)X] \\ \text{subject to} \quad & \text{tr}[\text{Diag}(\mathbf{a}_j)X] = b_j, \; j = 1, \ldots, m, \\ & \text{tr}[E_{ij}X] = 0, \; 1 \leq i < j \leq n, \\ & X \succeq 0, \end{aligned}$$

Why SDPs?

- SDPs are generalizations of LPs. They inherit nice properties such as strong duality (extra assumptions needed).
- Many solvers have been developed for SDPs. Solvers based on interior-point methods (IPM) can solve medium-sized problems readily on standard desktops.
- As we shall see later: A large class of problems can be reformulated as SDPs, and a large number of applications can be modeled using SDPs.
- As we shall see later: A software called CVX largely automates the process of transforming problems into standard SDP formats and calling solvers. We will mainly look at its MATLAB interface (which calls free IPM-based solvers SeDuMi or SDPT3). CVX also has interfaces for Python and Julia.

Thm: if $\exists\, \bar{y}$, s.t. $C - \sum \bar{y}_i A_i \succ 0$

Strong duality

$\{(\mathrm{tr}(CX), \mathrm{tr}(A_1X) - \mathrm{tr}(A_mX))^T \mid X \succeq 0\}$ is closed, $\Rightarrow$ primal optimal value is attainable.

Theorem 6.1 (Strong duality for SDPs)
Consider the following primal-dual SDP pairs:

$$\text{Primal}: \begin{cases} \underset{X\in\mathcal{S}^n}{\text{Minimize}} & \mathrm{tr}(CX) \\ \text{subject to} & \mathrm{tr}(A_iX) = b_i, \quad i = 1, \ldots, m, \\ & X \succeq 0, \end{cases}$$

$$\text{Dual}: \begin{cases} \underset{y\in\mathbb{R}^m}{\text{Maximize}} & b^T y \\ \text{subject to} & C - \sum_{i=1}^m y_i A_i \succeq 0, \end{cases}$$

where $C \in \mathcal{S}^n$ and $A_i \in \mathcal{S}^n$ for all i. Let v_p and v_d denote their optimal values respectively. Then the following statements hold.

1. If there exists $\bar{X} \succ 0$ such that $\mathrm{tr}(A_i\bar{X}) = b_i$ for all i, then $v_p = v_d$ and v_d is attained when finite.
2. If there exists $\bar{y} \in \mathbb{R}^m$ such that $C - \sum_{i=1}^m \bar{y}_i A_i \succ 0$, then $v_p = v_d$ and v_p is attained when finite.

if $\bar{X}$ is unique element in feasible region

Def: (slater point): $\bar{X}$ satisfies ($\succ 0$)

Example

Example: Here shows a primal-dual pair of SDP, in standard form.

Primal:

$$\begin{aligned} \underset{X\in\mathcal{S}^2}{\text{Minimize}} \quad & \operatorname{tr}\left(\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} X\right) \\ \text{subject to} \quad & \operatorname{tr}\left(\begin{bmatrix}6 & 0\\ 0 & 1\end{bmatrix} X\right) = 3, \quad \operatorname{tr}\left(\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} X\right) = 0, \\ & X \succeq 0, \end{aligned}$$

Dual:

$$\begin{aligned} \underset{y\in\mathbb{R}^2}{\text{Maximize}} \quad & 3y_1 + 0y_2 \\ \text{subject to} \quad & \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} - y_1\begin{bmatrix}6 & 0\\ 0 & 1\end{bmatrix} - y_2\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \succeq 0, \end{aligned}$$

We next argue that strong duality holds and both primal and dual problems have optimal solutions.

Example cont.

Example cont.:

- **Step 1**: Come up with a primal Slater point. Note that $\bar{X} := \begin{bmatrix}1/6 & 0\\ 0 & 2\end{bmatrix} \succ 0$ and is primal feasible. Thus, by Theorem 6.1, $v_p = v_d$, v_d is attained when finite. Moreover,

$$v_p \le \operatorname{tr}\left(\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\bar{X}\right) = -11/6 < \infty.$$

$\Rightarrow$ any point satisfy ($\succ 0$.)

- **Step 2**: Come up with a dual Slater point. Note that if we take $\bar{y} := \begin{bmatrix}-2 & 0\end{bmatrix}^T$, then

$$\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} - (-2)\begin{bmatrix}6 & 0\\ 0 & 1\end{bmatrix} - 0\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} = \begin{bmatrix}13 & 0\\ 0 & 1\end{bmatrix} \succ 0.$$

  Thus, by Theorem 6.1, $v_p = v_d$, v_p is attained when finite. Moreover, $v_d \ge 3\bar{y}_1 = -6 > -\infty$.

- **Step 3**: Thus, $-11/6 \ge v_p = v_d \ge -6$, showing that $v_p = v_d$ and is finite. Thus, both values are attainable.

$\Rightarrow v_p = v_d$ is finite

we found primal Slater Point & dual Slater —

Nonnegative trace

To understand Theorem 6.1, we need the following.

Theorem 6.2
$\succeq 0$
Let $A \in \mathcal{S}_+^n$ and $C \in \mathcal{S}_+^n$. Then $\text{tr}(AC) \geq 0$.

Proof: Since A is symmetric, there exist an orthogonal matrix U and a diagonal matrix D so that $A = UDU^T$. $tr(UDU^T) = tr(U^TUD) = tr(D) \succeq 0.$

Since all eigenvalues of A are nonnegative, we have $d_{ii} \geq 0$ for all i. Define $W \in \mathbb{R}^{n \times n}$ to be the diagonal matrix so that

$$w_{ij} = \begin{cases} \sqrt{d_{ii}} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Define the square root of A as $A^{\frac{1}{2}} := UWU^T$.

One can show that this definition is independent of the specific eigenvalue decomposition used. Thus, it is really "the" square root.

Then $A^{\frac{1}{2}} \in \mathcal{S}_+^n$ and $(A^{\frac{1}{2}})^2 = A$. Similarly, we can define $C^{\frac{1}{2}}$.

Nonnegative trace cont.

Proof of Theorem 6.2 cont.: Recall that for any two matrices X, $Y \in \mathbb{R}^{n \times n}$, we have $\text{tr}(XY) = \text{tr}(YX)$.

Thus

$$\begin{aligned} \text{tr}(AC) &= \text{tr}(A^{\frac{1}{2}} A^{\frac{1}{2}} C^{\frac{1}{2}} C^{\frac{1}{2}}) = \text{tr}(C^{\frac{1}{2}} A^{\frac{1}{2}} A^{\frac{1}{2}} C^{\frac{1}{2}}) \\ &= \text{tr}([A^{\frac{1}{2}} C^{\frac{1}{2}}]^T A^{\frac{1}{2}} C^{\frac{1}{2}}). \end{aligned}$$

Finally, note that for any matrix $Y \in \mathbb{R}^n$, we have

$$\text{tr}(Y^T Y) = \sum_{i=1}^{n} [Y^T Y]_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} y_{ji} y_{ji} \geq 0.$$

Hence, $\text{tr}(AC) \geq 0$.

Strong duality cont.

Remarks on Theorem 6.1:

- It always holds that $v_p \geq v_d$. Indeed, for any primal feasible X and dual feasible y, we have

$$b^T y = \sum_{i=1}^m b_i y_i = \sum_{i=1}^m \mathrm{tr}(A_i X) y_i = \mathrm{tr}\left(\sum_{i=1}^m y_i A_i X\right)$$
$$= \underbrace{\mathrm{tr}\left(\left[\sum_{i=1}^m y_i A_i - C\right] X\right)}_{\preceq 0 \;\Rightarrow\; \mathrm{tr}(\;) \leq 0} + \mathrm{tr}(CX) \leq \mathrm{tr}(CX),$$

  where the inequality follows from the feasibilities and Theorem 6.2.
  This is known as weak duality.

Strong duality cont.

Remarks on Theorem 6.1 cont.:

- Recall that in the proof of LP duality, another important ingredient there is the closedness of $\left\{ \begin{bmatrix} c^T \\ A \end{bmatrix} x : \; x \geq 0 \right\}$.
- Here, we should look at the object

$$\widehat{\Upsilon} := \{[\mathrm{tr}(CX) \;\; \mathrm{tr}(A_1 X) \;\; \cdots \;\; \mathrm{tr}(A_m X)]^T \in \mathbb{R}^{m+1} : \; X \succeq 0\}.$$

  Unfortunately, this object is not closed in general.
- If there exists $\bar{y} \in \mathbb{R}^m$ such that $C - \sum_{i=1}^m \bar{y}_i A_i \succ 0$, then $\widehat{\Upsilon}$ is closed. Why? See the next slide for a proof.
  One can then similarly use Theorem 4.3 to argue that $v_p = v_d$.
  The attainment of v_p (when finite) also follows from the closedness of $\widehat{\Upsilon}$.

Strong duality cont.

Here we prove the closedness of $\widehat{\Upsilon}$, assuming there exists $\bar{y} \in \mathbb{R}^m$ such that $C - \sum_{i=1}^m \bar{y}_i A_i \succ 0$.

Proposition 6.1
Consider Primal and Dual in Theorem 6.1 and the set $\widehat{\Upsilon}$ on the previous slide. Suppose that there exists $\bar{y} \in \mathbb{R}^m$ such that $C - \sum_{i=1}^m \bar{y}_i A_i \succ 0$. Then $\widehat{\Upsilon}$ is closed.

Proof: Let $\delta > 0$ be such that $C - \sum_{i=1}^m \bar{y}_i A_i \succeq \delta I \succ 0$.

Suppose that $\{u^k\} \subseteq \widehat{\Upsilon}$ and $u^k \to u^*$ for some u^*. We need to show that $u^* \in \widehat{\Upsilon}$.

By definition, there exist $\{X^k\}$ with $X^k \succeq 0$ for all k and

$$u^k = [\mathrm{tr}(CX^k)\ \ \mathrm{tr}(A_1 X^k)\ \cdots\ \mathrm{tr}(A_m X^k)]^T.$$

(ordered)

principal submatrices := $X(I, I)$

Strong duality cont.

I := indicator

Proof of Proposition 6.1 cont.: Now, using Theorem 6.2, we have

$$\delta \mathrm{tr}(X^k) \le \mathrm{tr}\left(\left[C - \sum_{i=1}^m \bar{y}_i A_i\right] X^k\right) = \mathrm{tr}(CX^k) - \sum_{i=1}^m \bar{y}_i \mathrm{tr}(A_i X^k).$$

Since $\mathrm{tr}(CX^k) - \sum_{i=1}^m \bar{y}_i \mathrm{tr}(A_i X^k) \to {u^*}^T \begin{bmatrix} 1 \\ -\bar{y} \end{bmatrix}$, and $X^k \succeq 0$ for all k, the above display shows that $\{X^k\}$ is bounded. $\mathrm{tr}(X^k) = \mathrm{tr}(\Lambda^k) \le L$

Hence, there is a convergent subsequence $X^{k_i} \to X^*$ for some $X^* \succeq 0$. Then $\Lambda^k \geq 0$

$$u^* = \lim_{i\to\infty} u^{k_i} = \lim_{i\to\infty} [\mathrm{tr}(CX^{k_i})\ \ \mathrm{tr}(A_1 X^{k_i})\ \cdots\ \mathrm{tr}(A_m X^{k_i})]^T$$
$$= [\mathrm{tr}(CX^*)\ \ \mathrm{tr}(A_1 X^*)\ \cdots\ \mathrm{tr}(A_m X^*)]^T \in \widehat{\Upsilon},$$

where the last inclusion holds because $X^* \succeq 0$.

$X^{k_i} \to X^* \in$ feasible region, $X^* \succeq 0$.

$f(X^{k_i}) \to f(X^*) = u^* \Rightarrow f(X^*) \in$ set( ). $\square$