DATA8004: Optimization for Statistical Learning

Subject Lecturer: Man-Chung YUE

Lecture 8
Semidefinite Programming
Duality Theory and Reformulations

Semidefinite Programming

Semidefinite programming (SDP) problems:

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

Here:

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

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

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

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

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).

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.

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

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

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.

LPs are SDPs

Consider the following linear program.

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

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

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

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.

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.

LPs are SDPs cont.

More generally, consider the following linear program.

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

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

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

where $\mathbf{a}_j^T$ is the jth row of A, and $\mathrm{Diag}(c) \in \mathbb{R}^{n \times n}$ 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

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

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?

LPs are SDPs cont.

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

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

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:

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

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.
- A large class of problems can be reformulated as SDPs, and a large number of applications can be modeled using SDPs.
- (If we have time) 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.

Strong duality

Theorem 8.1 (Strong duality for SDPs)
Consider the following primal-dual SDP pairs:

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

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

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

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:

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

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 8.1, $v_p = v_d$, v_d is attained when finite. Moreover,

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

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 8.1, $v_p = v_d$, v_d is attained when finite. Moreover,

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

- **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 8.1, $v_p = v_d$, v_p is attained when finite. Moreover, $v_d \geq 3\bar{y}_1 = -6 > -\infty$.

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 8.1, $v_p = v_d$, v_d is attained when finite. Moreover,

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

- **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 8.1, $v_p = v_d$, v_p is attained when finite. Moreover, $v_d \geq 3\bar{y}_1 = -6 > -\infty$.

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

Nonnegative trace

To understand Theorem 8.1, we need the following.

Theorem 8.2
Let $A \in \mathcal{S}_+^n$ and $C \in \mathcal{S}_+^n$. Then $\mathrm{tr}(AC) \geq 0$.

Proof: Since A is symmetric, there exist an orthogonal matrix U and a diagonal matrix D so that $A = UDU^T$.

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

Thus

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

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

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

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

Strong duality cont.

Remarks on Theorem 8.1:

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

$$b^T y = \sum_{i=1}^m b_i y_i = \sum_{i=1}^m \operatorname{tr}(A_i X) y_i = \operatorname{tr}\left(\sum_{i=1}^m y_i A_i X\right)$$
$$= \operatorname{tr}\left(\left[\sum_{i=1}^m y_i A_i - C\right] X\right) + \operatorname{tr}(CX) \leq \operatorname{tr}(CX),$$

  where the inequality follows from the feasibility of y and Theorem 8.2. This is known as weak duality.

Strong duality cont.

Remarks on Theorem 8.1 cont.:

- The proof of the strong duality requires the closedness of the set

  $$\widehat{\Upsilon} := \{[\mathrm{tr}(CX)\ \ \mathrm{tr}(A_1X)\ \cdots\ \mathrm{tr}(A_mX)]^T \in \mathbb{R}^{m+1} :\ X \succeq 0\}.$$

  Unfortunately, this set is not closed in general. To see this, consider the example where $m = 1$, $n = 2$, and

  $$C = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Strong duality cont.

Remarks on Theorem 8.1 cont.:

- The proof of the strong duality requires the closedness of the set

$$\widehat{\Upsilon} := \{[\mathrm{tr}(CX)\ \ \mathrm{tr}(A_1X)\ \cdots\ \mathrm{tr}(A_mX)]^T \in \mathbb{R}^{m+1} :\ X \succeq 0\}.$$

  Unfortunately, this set is not closed in general. To see this, consider the example where $m = 1$, $n = 2$, and

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

- 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 the existence of $\overline{y} \in \mathbb{R}^m$ such that $C - \sum_{i=1}^m \overline{y}_i A_i \succ 0$.

Proposition 8.1

Consider Primal and Dual in Theorem 8.1 and the set $\widehat{\Upsilon}$ on the previous slide. Suppose that there exists $\overline{y} \in \mathbb{R}^m$ such that $C - \sum_{i=1}^m \overline{y}_i A_i \succ 0$. Then $\widehat{\Upsilon}$ is closed.

Strong duality cont.

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

Proposition 8.1
Consider Primal and Dual in Theorem 8.1 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

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

Strong duality cont.

Proof of Proposition 8.1 cont.: Now, using Theorem 8.2, 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.

Strong duality cont.

Proof of Proposition 8.1 cont.: Now, using Theorem 8.2, we have

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

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.

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

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

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

Schur complement

The following result is crucial in reformulating problems into SDPs.

Theorem 8.3

Let $A \in \mathcal{S}^m$, $C \in \mathcal{S}^n$, $B \in \mathbb{R}^{m \times n}$, and $A \succ 0$. Then

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \iff C - B^T A^{-1} B \succeq 0.$$

Note: We call $C - B^T A^{-1} B$ the Schur complement of A in $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$.

Schur complement

The following result is crucial in reformulating problems into SDPs.

Theorem 8.3

Let $A \in \mathcal{S}^m$, $C \in \mathcal{S}^n$, $B \in \mathbb{R}^{m \times n}$, and $A \succ 0$. Then

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \iff C - B^T A^{-1} B \succeq 0.$$

Note: We call $C - B^T A^{-1} B$ the Schur complement of A in $\left[\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right]$.

Proof: Note that

$$\begin{bmatrix} I & 0 \\ (-A^{-1}B)^T & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix}.$$

Schur complement cont.

Proof of Theorem 8.3 cont.: First, suppose that $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$. Then

$$x^T(C - B^TA^{-1}B)x = \begin{bmatrix} 0^T & x^T \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^TA^{-1}B \end{bmatrix} \begin{bmatrix} 0 \\ x \end{bmatrix}$$

$$= \begin{bmatrix} 0^T & x^T \end{bmatrix} \begin{bmatrix} I & 0 \\ (-A^{-1}B)^T & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 \\ x \end{bmatrix} \geq 0$$

for any $x \in \mathbb{R}^n$. Thus, $C - B^TA^{-1}B \succeq 0$.

Schur complement cont.

Proof of Theorem 8.3 cont.: First, suppose that $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$. Then

$$x^T(C - B^T A^{-1} B)x = \begin{bmatrix} 0^T & x^T \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix} \begin{bmatrix} 0 \\ x \end{bmatrix}$$

$$= \begin{bmatrix} 0^T & x^T \end{bmatrix} \begin{bmatrix} I & 0 \\ (-A^{-1}B)^T & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 \\ x \end{bmatrix} \geq 0$$

for any $x \in \mathbb{R}^n$. Thus, $C - B^T A^{-1} B \succeq 0$.

Conversely, suppose that $C - B^T A^{-1} B \succeq 0$.

Fix any $x \in \mathbb{R}^{m+n}$ and let y be such that

$$\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} y = x.$$

Why does such a y exist?

Schur complement cont.

Proof of Theorem 8.3 cont.: Then,

$$x^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} x = y^T \begin{bmatrix} I & 0 \\ (-A^{-1}B)^T & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} y$$
$$= y^T \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix} y.$$

Now, write $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, where $y_1 \in \mathbb{R}^m$, $y_2 \in \mathbb{R}^n$. Then

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}^T \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = {y_1}^T A y_1 + {y_2}^T \underbrace{(C - B^T A^{-1} B)}_{\succeq\ 0 \text{ by assump.}} y_2 \geq 0.$$

Hence, $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0.$

Convex QPs are SDPs

Consider the following convex quadratic program.

$$\begin{array}{rl} \underset{x \in \mathbb{R}^2}{\text{Minimize}} & x_1^2 + 2x_1 x_2 + 2x_2^2 - x_1 - x_2 \\ \text{subject to} & 6x_1 + x_2 \leq 3. \end{array}$$

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

Convex QPs are SDPs

Consider the following convex quadratic program.

$$\begin{array}{ll} \underset{x \in \mathbb{R}^2}{\text{Minimize}} & x_1^2 + 2x_1x_2 + 2x_2^2 - x_1 - x_2 \\ \text{subject to} & 6x_1 + x_2 \leq 3. \end{array}$$

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

Notice that

$$x_1^2 + 2x_1x_2 + 2x_2^2 = x^T \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} x = \operatorname{tr}\left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} xx^T \right).$$

Since $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \succeq 0$, from Theorem 8.2, we know that the above problem is equivalent to

$$\begin{array}{ll} \underset{x \in \mathbb{R}^2, Y \in \mathcal{S}^2}{\text{Minimize}} & \operatorname{tr}\left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} Y \right) - x_1 - x_2 \\ \text{subject to} & 6x_1 + x_2 \leq 3, \ Y \succeq xx^T. \end{array}$$

Why?

Convex QPs are SDPs cont.

Next, we apply Theorem 8.3 to deduce that

$$\begin{bmatrix} 1 & x^T \\ x & Y \end{bmatrix} \succeq 0 \iff Y - xx^T \succeq 0.$$

Thus, the above problem is further equivalent to

$$\begin{array}{ll} \underset{x\in\mathbb{R}^2, u\in\mathbb{R}, Y\in\mathcal{S}^2}{\text{Minimize}} & \operatorname{tr}\left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} Y\right) - x_1 - x_2 \\ \text{subject to} & 6x_1 + x_2 + u = 3, \\ & \begin{bmatrix} 1 & x^T \\ x & Y \end{bmatrix} \succeq 0, \quad u \geq 0. \end{array}$$

To put this into standard form, we need a matrix variable of the form

$$\begin{bmatrix} u & 0 & 0 \\ 0 & 1 & x^T \\ 0 & x & Y \end{bmatrix}$$

Convex QPs are SDPs cont.

Then the above problem is further equivalent to

$$
\begin{array}{rl}
\underset{x\in\mathbb{R}^2, u\in\mathbb{R}, Y\in\mathcal{S}^2}{\text{Minimize}} & \operatorname{tr}\left(\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix} Y\right) - x_1 - x_2\\
\text{subject to} & 6x_1 + x_2 + u = 3,\\
& \begin{bmatrix} u & 0 & 0\\ 0 & 1 & x^T\\ 0 & x & Y\end{bmatrix} \succeq 0.
\end{array}
$$

Convex QPs are SDPs cont.

And further:

$$\underset{x\in\mathbb{R}^2,u\in\mathbb{R},Y\in\mathcal{S}^2}{\text{Minimize}} \quad \operatorname{tr}\left(\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & -0.5 & -0.5\\ 0 & -0.5 & 1 & 1\\ 0 & -0.5 & 1 & 2\end{bmatrix}\begin{bmatrix}u & 0 & 0 & 0\\ 0 & 1 & x_1 & x_2\\ 0 & x_1 & y_{11} & y_{12}\\ 0 & x_2 & y_{12} & y_{22}\end{bmatrix}\right)$$

$$\text{subject to} \quad \operatorname{tr}\left(\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 3 & 0.5\\ 0 & 3 & 0 & 0\\ 0 & 0.5 & 0 & 0\end{bmatrix}\begin{bmatrix}u & 0 & 0 & 0\\ 0 & 1 & x_1 & x_2\\ 0 & x_1 & y_{11} & y_{12}\\ 0 & x_2 & y_{12} & y_{22}\end{bmatrix}\right) = 3,$$

$$\begin{bmatrix}u & 0 & 0\\ 0 & 1 & x^T\\ 0 & x & Y\end{bmatrix} \succeq 0.$$

To bring this into standard primal form, replace the matrix variable by X and impose constraints such as $x_{22} = 1$, etc.

Convex QPs are SDPs cont.

And further:

$$\underset{X \in \mathcal{S}^4}{\text{Minimize}} \quad \operatorname{tr}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & -0.5 \\ 0 & -0.5 & 1 & 1 \\ 0 & -0.5 & 1 & 2 \end{bmatrix} X\right)$$

$$\text{subject to} \quad \operatorname{tr}\left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0.5 \\ 0 & 3 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \end{bmatrix} X\right) = 3,$$

$$x_{12} = x_{13} = x_{14} = 0, x_{22} = 1,$$

$$X \succeq 0.$$

This can be brought to standard primal form by using E_{ij}'s on Slide 5.

Convex QPs are SDPs cont.

More generally, consider the following convex quadratic program.

$$\begin{array}{rl} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & x^T Q x + 2c^T x \\ \text{subject to} & Ax \le b. \end{array}$$

where $Q \in \mathcal{S}^n_+$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$.

Convex QPs are SDPs cont.

More generally, consider the following convex quadratic program.

$$\begin{array}{rl} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & x^T Q x + 2c^T x \\ \text{subject to} & Ax \leq b. \end{array}$$

where $Q \in \mathcal{S}_+^n$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$.

Invoking Theorem 8.2 and Theorem 8.3, the above problem can be equivalently transformed into

$$\begin{array}{rl} \underset{Y \in \mathcal{S}^n, x \in \mathbb{R}^n, u \in \mathbb{R}^m}{\text{Minimize}} & \text{tr}(QY) + 2c^T x \\ \text{subject to} & Ax + u = b, u \geq 0, \\ & \begin{bmatrix} 1 & x^T \\ x & Y \end{bmatrix} \succeq 0. \end{array}$$

This can be brought to standard primal form similarly as before.