A GRADUATE COURSE

IN

OPTIMIZATION

A GRADUATE COURSE

IN

OPTIMIZATION

CIE6010 Notebook

Prof. Yin Zhang

The Chinese University of Hongkong, Shenzhen

香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

Contents

Acknowledgments

This book is from the CIE6010 in fall semester, 2018.

CUHK(SZ)

Notations and Conventions

X	Set
$\inf X \subseteq \mathbb{R}$	Infimum over the set X
$\mathbb{R}^{m\times n}$	set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m\times n}$	set of all $m \times n$ complex-valued matrices
x_i	ith entry of column vector $\boldsymbol{x}$
a_{ij}	(i,j)th entry of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i$	ith column of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i^{\mathrm{T}}$	ith row of matrix $\boldsymbol{A}$
$\mathbb{S}^n$	set of all $n \times n$ real symmetric matrices, i.e., $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $a_{ij} = a_{ji}$ for all i,j
$\mathbb{H}^n$	set of all $n \times n$ complex Hermitian matrices, i.e., $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ and $\bar{a}_{ij} = a_{ji}$ for all i,j
$\boldsymbol{A}^{\mathrm{T}}$	transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}$ means $b_{ji} = a_{ij}$ for all i,j
$\boldsymbol{A}^{\mathrm{H}}$	Hermitian transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{H}}$ means $b_{ji} = \bar{a}_{ij}$ for all i,j
$\mathrm{trace}(\boldsymbol{A})$	sum of diagonal entries of square matrix $\boldsymbol{A}$
$\mathbf{1}$	A vector with all 1 entries
$\mathbf{0}$	either a vector of all zeros, or a matrix of all zeros
$\boldsymbol{e}_i$	a unit vector with the nonzero element at the ith entry
$\mathcal{C}(\boldsymbol{A})$	the column space of $\boldsymbol{A}$
$\mathcal{R}(\boldsymbol{A})$	the row space of $\boldsymbol{A}$
$\mathcal{N}(\boldsymbol{A})$	the null space of $\boldsymbol{A}$
$\mathrm{Proj}_{\mathcal{M}}(\boldsymbol{A})$	the projection of $\boldsymbol{A}$ onto the set $\mathcal{M}$

Chapter 14

Week14

14.1. Monday

Notation. The inner product $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle$ is defined as:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \sum_{i,j=1}^{n} x_{ij} y_{ij} = \operatorname{trace}(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{Y})$$

The conic programming aims to optimize

$$\begin{aligned} \min \quad & \langle \boldsymbol{C}, \boldsymbol{X} \rangle \\ & \boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \\ & \boldsymbol{X} \in \mathcal{K} \end{aligned} \tag{14.1}$$

where $\mathcal{K}$ is a **convex closed pointed cone**. The pointed cone is a cone that contains no line (from $-\infty$ to ∞ is called a line).

1. For the linear programming, $\mathcal{K} = \mathbb{R}^n_+ = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \geq 0\}$
2. For the SOCP, $\mathcal{K} = \mathcal{S}_2^{n+1} = \{(\xi, \boldsymbol{X}) \mid \xi \geq \|\boldsymbol{X}\|\}$
3. For the semidefinite programming, $\mathcal{K} = \{\boldsymbol{X} \in \mathbb{R}^{n \times n} \mid \boldsymbol{X} = \boldsymbol{X}^{\mathrm{T}} \succeq 0\}$

Definition 14.1 [Dual cone] The dual cone is defined as

$$\mathcal{K}^* = \{\boldsymbol{X} \mid \langle \boldsymbol{X}, \boldsymbol{Y} \rangle \geq 0, \forall Y \in \mathcal{K}\}$$

Proposition 14.1 For LP, SOCP, SDP, the dual cone $\mathcal{K}^* = \mathcal{K}$.

Proof. We pick the SDP case as an example.

1. We show $\mathcal{K} \subseteq \mathcal{K}^*$ first.

   For $\boldsymbol{S} \in \mathcal{K}$, we have

$$\begin{aligned}\boldsymbol{S} = \boldsymbol{U}\Lambda\boldsymbol{U}^{\mathrm{T}} &:= \boldsymbol{U}\Lambda^{1/2}\Lambda^{1/2}\boldsymbol{U}^{\mathrm{T}} \\ &= (\boldsymbol{U}\Lambda^{1/2}\boldsymbol{U}^{\mathrm{T}})(\boldsymbol{U}\Lambda^{1/2}\boldsymbol{U}^{\mathrm{T}}) \\ &:= \boldsymbol{S}^{1/2}\boldsymbol{S}^{1/2}\end{aligned}$$

   It suffices to show that the inner product between $\boldsymbol{S}$ and any other elements in $\mathcal{K}$ is non-negative:

$$\langle \boldsymbol{S}, \boldsymbol{X} \rangle = \mathrm{trace}(\boldsymbol{S}\boldsymbol{X}) = \mathrm{trace}(\boldsymbol{S}^{1/2}\boldsymbol{S}^{1/2}\boldsymbol{X}) = \mathrm{trace}(\boldsymbol{S}^{1/2}\boldsymbol{X}\boldsymbol{S}^{1/2}) \geq 0,$$

   the last inequality is because $\boldsymbol{S}^{1/2}\boldsymbol{X}\boldsymbol{S}^{1/2}$ is psd and the operator $\mathrm{trace}(\cdot)$ outputs the sum of eigenvalues.

2. Let $\boldsymbol{Y} \in \mathcal{K}^*$, and any element $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} \in \mathcal{K}$, we have

$$\langle \boldsymbol{Y}, \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} \rangle = \mathrm{trace}(\boldsymbol{Y}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}) = \mathrm{trace}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{V}) = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{V} \geq 0,$$

   where the last inequality is because $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{V}$ is a scalar. Therefore, $\boldsymbol{Y} \in \mathcal{K}$, and $\mathcal{K}^* = \mathcal{K}$.

■

The self-dual properites allows us to apply the primal-dual interior method to solver the conic programming.

Dual programming.

$$\begin{aligned}\max \quad & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} \\ & \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} + \boldsymbol{S} = \boldsymbol{C} \\ & \mathcal{S} \in \mathcal{K}^*\end{aligned} \tag{14.2}$$

For SDP,

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b},$$

where $\boldsymbol{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ and $\boldsymbol{A}^{\mathrm{H}} : \mathbb{R}^m \to \mathbb{R}^{n\times n}$ are linear operators. The pair $(\boldsymbol{A}, \boldsymbol{A}^{\mathrm{H}})$ admits the property

$$\langle \boldsymbol{A}\boldsymbol{X}, \boldsymbol{y}\rangle = \langle \boldsymbol{X}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y}\rangle$$

■ **Example 14.1** [Primal-SDP] Suppose we want to optimize

$$\begin{aligned} \min \quad & \langle \boldsymbol{C}, \boldsymbol{X}\rangle \\ & \boldsymbol{A}_i X = \boldsymbol{b}_i, \quad i = 1,\dots,m \\ & \boldsymbol{X} \succeq 0 \end{aligned} \tag{14.3}$$

for

$$\boldsymbol{C} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{A}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{A}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

This problem can be transformed into

$$\begin{aligned} \min \quad & x_{33} \\ & x_{11} = 0 \\ & 2x_{12} + 2x_{33} = 2 \\ & X \succeq 0 \end{aligned} \tag{14.4}$$

After computation, we obtain the optimal solution

$$\begin{pmatrix} 0 & 0 & \times \\ 0 & x_{22} & \times \\ \times & \times & 1 \end{pmatrix} \succeq 0$$

with the optimal value 1.

Next we try to solve the dual SDP:

$$\begin{aligned}\max \quad & 2y_2 \\ & y_1\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} + y_2\begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 2\end{pmatrix} \preceq \begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\end{aligned}$$

Or equivalently, the constraint should be

$$\begin{pmatrix}-y_1 & -y_2 & 0\\ -y_2 & 0 & 0\\ 0 & 0 & 1-2y_2\end{pmatrix} \succeq 0$$

After computation, we obtain the optimal value $y_2 = 0$. ■

R The strong duality does not necessarily hold for conic programming.