Fundamentals of Mathematical and Computing Sciences: Applied Mathematical Sciences

Tokyo Institute of Technology
Department Mathematical and Computing Sciences
Mitsuhiro Fukuda

Spring Semester of 2013

Outline of the Lecture

This course introduces several basic concepts of mathematical optimization, probability and statistics, and is intended to provide key knowledge necessary for advanced study in Mathematical and Computing Sciences.

References

[Ben-Tal-Nemirovski] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (SIAM, Philadelphia, PA, 2001).

[Renegar] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization* (SIAM, Philadelphia, PA, 2001).

Evaluation

Final exam and/or reports.

1 Preliminaries

We assume that convexity and closeness (openness) of sets are familiar concepts for the readers.

Let $\langle\ ,\ \rangle$ be an arbitrary inner product on $\mathbb{R}^n$. Given a linear operator $\mathcal{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, its *adjoint* operator $\mathcal{A}^* : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is such that

$$\langle \mathcal{A}(\boldsymbol{x}), \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \mathcal{A}^*(\boldsymbol{y}) \rangle, \quad \forall \boldsymbol{x} \in \mathbb{R}^n,\ \forall \boldsymbol{y} \in \mathbb{R}^m.$$

Definition 1.1 A set $\mathcal{K} \subseteq \mathbb{R}^n$ is called *cone* if for any positive scalar $\alpha > 0$ and an arbitrary element $\boldsymbol{x}$ of $\mathcal{K}$, $\alpha \boldsymbol{x} \in \mathcal{K}$.

Definition 1.2 A cone is said to be *pointed* if $\mathcal{K} \cap -\mathcal{K} = \{\mathbf{0}\}$.

Definition 1.3 Given a cone $\mathcal{K} \subseteq \mathbb{R}^n$, its *dual* cone is defined as $\mathcal{K}^* := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0,\ \forall \boldsymbol{y} \in \mathcal{K}\}$.

Definition 1.4 If a cone is such that $\mathcal{K}^* = \mathcal{K}$, it is called **self-dual**.

Theorem 1.5 (Separation theorem for convex sets [Ben-Tal-Nemirovski]) Let A, B nonempty non-intersecting convex subsets of $\mathbb{R}^n$. Then, $\exists \boldsymbol{s} \in \mathbb{R}^n$, $\boldsymbol{s} \neq 0$ such that

$$\sup_{\boldsymbol{a} \in A} \langle \boldsymbol{a}, \boldsymbol{s} \rangle \leq \inf_{\boldsymbol{b} \in B} \langle \boldsymbol{b}, \boldsymbol{s} \rangle.$$

2 Conic Linear Program

The Linear Program (LP) is the most basic mathematical optimization problem. We will start defining a generalization of the LP.

The *Conic Linear Program* (CLP) is defined as follows[1]:

$$(\text{CLP}) \left\{ \begin{array}{ll} \text{minimize} & \langle \boldsymbol{c}, \boldsymbol{x} \rangle \\ \text{subject to} & \mathcal{A}(\boldsymbol{x}) = \boldsymbol{b}, \\ & \boldsymbol{x} \in \mathcal{K}, \end{array} \right.$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$, $\mathcal{A}(\cdot)$ is an linear operator, and $\mathcal{K}$ is a <u>closed convex cone</u> in $\mathbb{R}^n$.

The dual problem of (CLP) is defined as[2]:

$$(\text{DCLP}) \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{b}, \boldsymbol{y} \rangle \\ \text{subject to} & \mathcal{A}^*(\boldsymbol{y}) + \boldsymbol{s} = \boldsymbol{c}, \\ & \boldsymbol{s} \in \mathcal{K}^*, \end{array} \right.$$

where the <u>inner product is defined on $\mathbb{R}^m$</u> now. Notice that $\mathcal{K}^*$ is a <u>closed convex cone</u>, too.

Example 2.1 If we chose $\mathcal{K} = \mathbb{R}^n_+$, $\mathcal{A} := \boldsymbol{A} \in \mathbb{R}^{m \times n}$, and $\langle \boldsymbol{c}, \boldsymbol{x} \rangle = \boldsymbol{c}^T \boldsymbol{x}$, (CLP) becomes an LP. Likewise, taking $\mathcal{K} = \mathbb{S}^n_+$, the cone of positive semidefinite symmetric matrices, and the inner product which defines the Frobenius norm, we have a *Semidefinite Program* (SDP); $\mathcal{K} = \mathbb{Q}^n_+ := \{\boldsymbol{x} \in \mathbb{R}^n \mid x_1^2 \geq \sum_{i=2}^{n-1} x_i^2\}$, the second-order cone, we have a *Second-Order Cone Program* (SOCP).

The following result known as *weak duality* is a simple consequence of above facts.

Lemma 2.2 (Weak Duality) Let $\boldsymbol{x}$ be feasible for (CLP) and $(\boldsymbol{y}, \boldsymbol{s})$ feasible for (DCLP). Then $\langle \boldsymbol{b}, \boldsymbol{y} \rangle \leq \langle \boldsymbol{c}, \boldsymbol{x} \rangle$.

[1] strictly speaking, the term "minimize" should be replaced by "infimum"

[2] strictly speaking, the term "maximize" should be replaced by "supremum"

Proof:
$\langle \boldsymbol{c}, \boldsymbol{x} \rangle - \langle \boldsymbol{b}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{c} \rangle - \langle \mathcal{A}(\boldsymbol{x}), \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{c} - \mathcal{A}^*(\boldsymbol{y}) \rangle = \langle \boldsymbol{x}, \boldsymbol{s} \rangle \geq 0$ since $\boldsymbol{x} \in \mathcal{K}$ and $\boldsymbol{s} \in \mathcal{K}^*$. ∎

The following example shows that **strong duality** does not hold in general.

$$\left\{\begin{array}{ll} \text{minimize} & \left\langle \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \right\rangle \\ \text{subject to} & \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ & \boldsymbol{x} \in \mathcal{K} = \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1^2 + x_2^2 \leq x_3^2,\ x_3, x_4 \geq 0\} \end{array}\right.$$

$$\left\{\begin{array}{ll} \text{maximize} & \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle \\ \text{subject to} & \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} y_1 \\ 0 \\ 0 \\ y_1 \end{pmatrix} - \begin{pmatrix} 0 \\ y_2 \\ y_2 \\ 0 \end{pmatrix} \in \mathcal{K}^* = \mathcal{K}. \end{array}\right.$$

Both problems are feasible, but the optimal value of the primal is 0 while for the dual is -1.

Theorem 2.3 (Strong Duality) If (CLP) is bounded from below and it is strictly feasible (*i.e*, $\exists \boldsymbol{x} \in \operatorname{int}(\mathcal{K})$ and $\mathcal{A}(\boldsymbol{x}) = \boldsymbol{b}$), then (DCLP) is solvable and its optimal value coincides with the one of (CLP). The result is valid if the roles of (CLP) and (DCLP) are exchanged.

Proof:
Let c_{val} be the optimal value of (CLP) which exists by the assumption. We need to show that (DCLP) is solvable and have the same optimal value.

For $\boldsymbol{c} = \boldsymbol{0}$, $c_{\text{val}} = 0$ and the existence of the feasible solution $\boldsymbol{y} = \boldsymbol{0}$, $\boldsymbol{s} = \boldsymbol{0}$ for (DCLP) is evident.

Now let $\boldsymbol{c} \neq \boldsymbol{0}$. Consider the set

$$M = \{\boldsymbol{x} \in \mathbb{R}^n \mid \mathcal{A}(\boldsymbol{x}) = \boldsymbol{b},\ \langle \boldsymbol{c}, \boldsymbol{x} \rangle \leq c_{\text{val}}\}.$$

It is clear that $M \neq \emptyset$. Also $M \cap \operatorname{int}(\mathcal{K}) = \emptyset$. In fact, if we assume on the contrary that $\exists \bar{\boldsymbol{x}} \in M \cap \operatorname{int}(\mathcal{K})$, since $\boldsymbol{c} \neq \boldsymbol{0}$ and $\bar{\boldsymbol{x}}$ is an interior point, we can always construct a $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ feasible for (CLP) which $\langle \boldsymbol{c}, \hat{\boldsymbol{x}} \rangle < c_{\text{val}}$ with contradicts the optimality. If we are in the case where this is not possible, $\bar{\boldsymbol{x}}$ will be the optimal solution of (CLP) and this will be treated in the next theorem.

From Theorem 1.5, $\exists \bar{\boldsymbol{s}} \in \mathbb{R}^n$ such that $\bar{\boldsymbol{s}} \neq \boldsymbol{0}$ and

$$\sup_{\boldsymbol{x} \in M} \langle \boldsymbol{x}, \bar{\boldsymbol{s}} \rangle \leq \inf_{\boldsymbol{x} \in \operatorname{int}(\mathcal{K})} \langle \boldsymbol{x}, \bar{\boldsymbol{s}} \rangle.$$

Since M is nonempty and $\mathcal{K}$ is a cone, we have in fact that

$$\sup_{\boldsymbol{x} \in M} \langle \boldsymbol{x}, \bar{\boldsymbol{s}} \rangle \leq 0 = \inf_{\boldsymbol{x} \in \mathcal{K}} \langle \boldsymbol{x}, \bar{\boldsymbol{s}} \rangle. \tag{1}$$

Therefore, due to this fact, $\bar{\boldsymbol{s}} \in \mathcal{K}^*$. From the definition of M, we can conclude that in fact $\exists \bar{\alpha}$, $\exists \bar{\beta} \geq 0$, and $\exists \bar{\boldsymbol{y}} \in \mathbb{R}^m$ such that $\bar{\boldsymbol{s}} = \bar{\alpha} \mathcal{A}^*(\bar{\boldsymbol{y}}) + \bar{\beta} \boldsymbol{c}$. This can be seen since $\forall \boldsymbol{x} \in M$,

$$\begin{aligned} \langle \boldsymbol{x}, \bar{\boldsymbol{s}} \rangle &= \langle \mathcal{A}(\boldsymbol{x}), \bar{\alpha} \bar{\boldsymbol{y}} \rangle + \langle \boldsymbol{x}, \bar{\beta} \boldsymbol{c} \rangle \\ &= \langle \boldsymbol{b}, \bar{\alpha} \bar{\boldsymbol{y}} \rangle + \bar{\beta} \langle \boldsymbol{x}, \boldsymbol{c} \rangle \leq \text{constant} + \bar{\beta} c_{\text{val}}. \end{aligned}$$