Mathematical methods in communication	December 27th, 2011
Lecture 7	
Lecturer: Haim Permuter	*Scribe: Moti Teitel*

I. Convex Optimization

In convex optimization, our goal is to minimize an objective a convex function subject to convex inequality and affine equality constraints. The problem can be mathematically written as:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq b_i \quad i = 1, \cdots, k \\ & g_j(x) = 0 \quad j = 1, \cdots, l \end{aligned} \tag{1}$$

where $f_0(x)$ and $\{f_i(x)\}_{i=1}^{k}$ are convex functions, and $\{g_j(x)\}_{j=1}^{l}$ are affine.

Throughout the following lectures, the following notations and definitions will be used:

A. Notations

1) $\mathbb{R}$ - the set of all real numbers
2) $\mathbb{R}_+$ - the set of all non-negative real numbers
3) $\mathbb{R}_{++}$ - the set of all positive real numbers
4) $\mathbb{R}^n$ - the set of all real n-dimensional vectors
5) $\mathbb{R}^{m\times n}$ - the set of real $m \times n$ matrices
6) (a, b, c) - a column vector whose elements (by order) are a, b and c.
7) $(\cdot)^T$ - the transpose operator.
8) $\mathbf{1}$ - a column vector composed of one's: $(1, 1, \ldots, 1)$
9) $\mathbf{x}$ - a vector
10) $\mathbf{x}_i$ - the i^{th} element of $\mathbf{x}$.
11) S^k - the set of all (real) symmetric $k \times k$ matrices: $S^k = \left\{\mathbf{A} \in \mathbb{R}^{k\times k} : A^T = A\right\}$
12) S_+^k - the set of all (real) symmetric positive semi-definite (PSD) $k \times k$ matrices: $S_+^k = \left\{\mathbf{A} \in S^{k\times k} : S \succeq 0\right\}$
13) S_{++}^k - the set of all (real) symmetric positive semi-definite (PSD) $k \times k$ matrices: $S_{++}^k = \left\{\mathbf{A} \in S^{k\times k} : S \succ 0\right\}$
14) $\mathbf{x} \succeq \mathbf{y}$ (vectors) - element-wise inequality: $\forall i \quad \mathbf{x}_i \geq \mathbf{y}$
15) $\mathbf{A} \succeq \mathbf{B}$ (matrices) - $(\mathbf{A} - \mathbf{B}) \in S_+^k$
16) $\mathrm{dom}(f)$ - the domain on which f is defined. For example: $\log : \mathbb{R} \to \mathbb{R}$, and $\mathrm{dom}(f) = \mathbb{R}_{++}$
17) $\bar{\theta} = 1 - \theta$

Note the difference between the interpretations of inequalities in $\mathbb{R}^n$ when $n = 1$ and $n > 1$. As $\mathbb{R}$ is well-ordered, the elements $a, b \in \mathbb{R}$ satisfy $a < b, a = b$ or $a > b$. However. for $\mathbb{R}^2$ the vectors $(0, 1)$ and $(1, 0)$ do not satisfy any of the relations $<, =, >$.

B. *Definitions*

Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ be two arbitrary vectors. Then:

1) The **line** that goes through $\mathbf{x}_1$ and $\mathbf{x}_2$ is the set of all points $\theta\mathbf{x}_1 + \bar{\theta}\mathbf{x}_2$, where $\theta \in \mathbb{R}$.

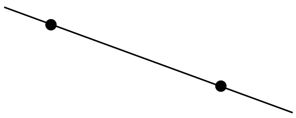

Fig. 1. A line

2) The **line segment** that goes through $\mathbf{x}_1$ and $\mathbf{x}_2$ is the set of all points $\theta\mathbf{x}_1 + \bar{\theta}\mathbf{x}_2$, where $\theta \in [0, 1]$.

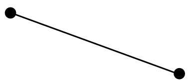

Fig. 2. A line segment

3) A set $C \subseteq \mathbb{R}^n$ is an **affine set** if any line that goes through any two points in C is contained in C:
$\forall\theta \in \mathbb{R}: \quad \mathbf{x}_1, \mathbf{x}_2 \in C \Rightarrow \theta\mathbf{x}_1 + \bar{\theta}\mathbf{x}_2 \in C$.
4) A set $C \subseteq \mathbb{R}^n$ is a **convex set** if the line segment that goes through any two points in C is contained in C:
$\forall\theta \in [0, 1]: \quad \mathbf{x}_1, \mathbf{x}_2 \in C \Rightarrow \theta\mathbf{x}_1 + \bar{\theta}\mathbf{x}_2 \in C$. For example, see Figure 3.

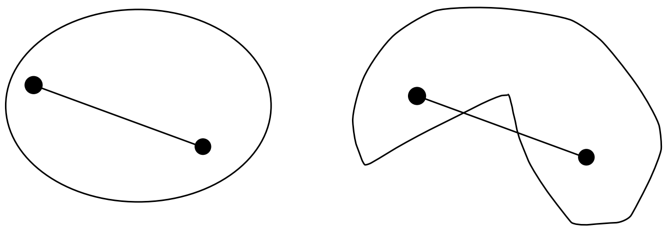

Fig. 3. To the left: a convex set. To the right: a non-convex set.

5) A **convex combination** of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_x)$ is any linear combination $\sum_{i=1}^{k} \theta_i \mathbf{x}_i$, where $\theta_i \geq 0$ and $\sum_{i=1}^{k} \theta_i = 1$.
6) a **convex hull**, denoted $\mathbf{Conv}(C)$, is the set of all convex combinations of points in C. A convex hull is the smallest convex set that contains C. For example, in Figure 4, the convex hull of the non-convex set is obtained by adding the area inside the dashed line.

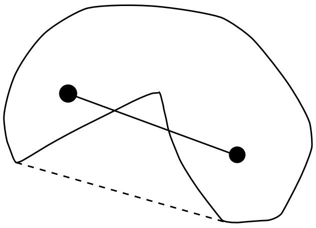

Fig. 4. Convex hull

7) A set C is called a **cone** if $\forall \mathbf{x} \in C$ and $\theta \geq 0$ we have $\theta \mathbf{x} \in C$.

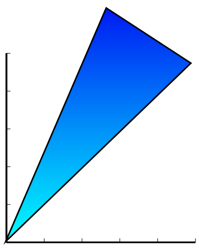

Fig. 5. A Cone

8) A cone C is called a **convex cone** if it is convex:
$\forall \theta_1, \theta_2 \geq 0 : \quad \mathbf{x}_1, \mathbf{x}_2 \in C \Rightarrow \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 \in C$.
9) A **conic combination** of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_x)$ is any linear combination $\sum_{i=1}^{k} \theta_i \mathbf{x}_i$, where $\theta_i \geq 0$.
10) A **conic hull** is the set of all conic combinations of points in C.
11) A cone C is called **pointed** if $\mathbf{x} \in C$ and $-\mathbf{x} \in C$ implies that $\mathbf{x} = 0$. This means that the cone C contains no line.
12) A cone $C \subset \mathbb{R}^n$ is called a **proper cone** if the following requirements hold:
    a) C is a convex cone.
    b) C is a closed set.
    c) C has a non-empty interior.
    d) C is pointed.
13) A **ray** is a set of points $\{\mathbf{x} : \mathbf{x} = \mathbf{x}_0 + \theta \mathbf{y} | \theta \geq 0\}$.
14) A **Hyperplane** is a set $\left\{\mathbf{x} : \mathbf{a}^T \mathbf{x} = b\right\}$, where $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$, $\mathbf{a} \neq 0$ and $b \in \mathbb{R}$.
Another possible notation is $\left\{\mathbf{x} : \mathbf{a}^T (\mathbf{x} - \mathbf{x}_0) = b\right\}$, where $\mathbf{x}_0$ is any point on the hyperplane (so that $\mathbf{a}^T \mathbf{x}_0 = b$).

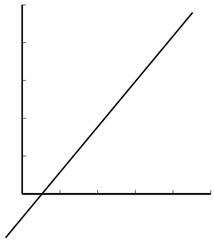

Fig. 6. A Hyperplane

$\mathbf{a}$ is called the **normal** to the hyperplane.

15) A **Half-space** is the set of points above (or below) a hyperplane: $\left\{\mathbf{x} : \mathbf{a}^T\left(\mathbf{x} - \mathbf{x}_0\right) > b\right\}$

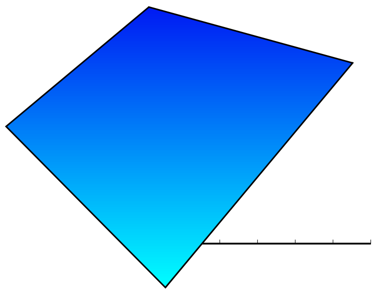

Fig. 7. Half space

16) A **Polyhedra** is the set of points that satisfy a set of linear equalities & inequalities. For instance:

$$P = \{\mathbf{x} \in \mathbb{R}^n\} \tag{2}$$
$$s.t. \begin{cases} \mathbf{a}_i^T\mathbf{x} \leq b_i & i = 1, 2, \ldots, K \\ \mathbf{c}_j^T\mathbf{x} = d_j & j = 1, 2, \ldots, J \end{cases}$$

A compact notation is

$$P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{C}\mathbf{x} = \mathbf{d}\} \tag{3}$$

where $\mathbf{A} = \left(\mathbf{a}_1^T, \ldots, \mathbf{a}_K^T\right)$ and $\mathbf{C} = \left(\mathbf{c}_1^T, \ldots, \mathbf{c}_J^T\right)$.
A Polyhedra is the outcome of an intersection of half-spaces and hyperplanes.

C. *Examples*

1) The empty set $\emptyset$, a single point in $\mathbf{x}_0 \in \mathbb{R}^n$ and the entire space are affine.
2) Any line is affine.
3) Any line that passes through the origin is a cone.

4) A line segment is convex, but not affine.
5) A ray is convex but not affine. It is a cone iff $\mathbf{x}_0 = 0$.
6) A subspace is convex, affine & a cone.

We start with the following lemma regarding convex sets:

Lemma 1 Convexity is preserved under intersections: Let S_1, S_2 be convex sets. Then $S_1 \cap S_2$ is convex.

Excercise: Prove Lemma 1.

Later we will see that if $f_1(x)$ and $f_2(x)$ are convex, then $f(x) = \max(f_1(x), f_2(x))$ is convex, and we will see that this is equivalent to Lemma 1.

II. Generalized Inequalities

A. *Definition*

Let $K \subset \mathbb{R}^n$ be a proper cone. We define the generalized inequality with respect to K as following:

$$\mathbf{x} \preceq_K \mathbf{y} \Leftrightarrow \mathbf{x} - \mathbf{y} \in K \tag{4}$$

B. *Properties of the Generalized Inequality*

1) The G.I. is preserved under addition: $\mathbf{x}_1 \preceq_K \mathbf{y}_1, \quad \mathbf{x}_2 \preceq_K \mathbf{y}_2 \Rightarrow \mathbf{x}_1 + \mathbf{x}_2 \preceq_K \mathbf{y}_1 + \mathbf{y}_2$
2) The G.I. is transitive: $\mathbf{a} \preceq_K \mathbf{b}, \quad \mathbf{b} \preceq_K \mathbf{c} \Rightarrow \mathbf{a} \preceq_K \mathbf{c}$
3) The G.I. is preserved under non-negative scaling: $\mathbf{x} \preceq_K \mathbf{y}, \quad a \geq 0 \Rightarrow a\mathbf{x} \preceq_K a\mathbf{y}$
4) The G.I. is reflexive: $\forall \mathbf{x}: \quad \mathbf{x} \preceq_K \mathbf{x}$
5) The G.I. is preserved under limits: if $\forall i \quad \mathbf{x}_i \preceq_K \mathbf{y}_i$, and $\mathbf{x} = \lim_{i\to\infty} \mathbf{x}_i$ and $\mathbf{y} = \lim_{i\to\infty} \mathbf{y}_i$ exist, then $\mathbf{x} \preceq_K \mathbf{y}$

Excercise: Prove properties 1-5.

Theorem 1 Separation Theorem
Let C, D be two convex sets in $\mathbb{R}^n$, such that $C \cap D = \emptyset$. Then, there exists a hyperplane $\mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) = 0$ such that:

- $\forall \mathbf{x} \in C: \quad \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) \leq 0$, and
- $\forall \mathbf{x} \in D: \quad \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) \geq 0$.

Proof: Assume that C, D are closed sets. Define the following distance measure :

$$\operatorname{dist}(C, D) \triangleq \min_{\mathbf{u}\in C, \mathbf{v}\in D} \|\mathbf{u} - \mathbf{v}\|^2 \tag{5}$$

Let $\mathbf{c} \in C, \mathbf{d}\ D$ be the elements that achieve the minimum:

$$\|\mathbf{c} - \mathbf{d}\|^2 = \operatorname{dist}(C, D) = \min_{\mathbf{u}\in C, \mathbf{v}\in D} \|\mathbf{u} - \mathbf{v}\|^2 \tag{6}$$

and let $\mathbf{a}$ be the line connecting $\mathbf{c}, \mathbf{d}$ ($\mathbf{a} = \mathbf{d} - \mathbf{c}$), and $\mathbf{x}_0 = \frac{\mathbf{c}+\mathbf{d}}{2}$. Define the following functional:

$$f(\mathbf{x}) = \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) = \mathbf{a}^T\left(\mathbf{x} - \frac{\mathbf{c}+\mathbf{d}}{2}\right) = (\mathbf{d} - \mathbf{c})^T\left(\mathbf{x} - \frac{\mathbf{c}+\mathbf{d}}{2}\right) \tag{7}$$

$$= (\mathbf{d}-\mathbf{c})^T \left(\mathbf{x}-\mathbf{d}+\frac{\mathbf{d}-\mathbf{c}}{2}\right) = (\mathbf{d}-\mathbf{c})^T (\mathbf{x}-\mathbf{d}) + \frac{1}{2}\|\mathbf{d}-\mathbf{c}\|^2$$

We show that for every $\mathbf{u} \in D$ we have $f(\mathbf{u}) \geq 0$:
Assume that there exists a vector $\mathbf{u} \in D$ s.t. $f(\mathbf{u}) < 0$. Then we may write

$$0 > f(\mathbf{u}) = (\mathbf{d}-\mathbf{c})^T (\mathbf{u}-\mathbf{d}) + \frac{1}{2}\|\mathbf{d}-\mathbf{c}\|^2 > (\mathbf{d}-\mathbf{c})^T (\mathbf{u}-\mathbf{d}) \tag{8}$$

Define the following function:

$$\begin{aligned} g(t) &= \|\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c}\|^2 = \|(1-t)\,\mathbf{d}+t\mathbf{u}-\mathbf{c}\|^2 \\ &= ((1-t)\,\mathbf{d}+t\mathbf{u}-\mathbf{c})^T((1-t)\,\mathbf{d}+t\mathbf{u}-\mathbf{c}) \\ &= (\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c})^T(\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c}) \end{aligned} \tag{9}$$

$$\begin{aligned} \frac{d}{dt}g(t) = \frac{d}{dt}&\left[(\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c})^T\right](\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c}) \\ &+ (\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c})^T\frac{d}{dt}(\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c}) \\ =& (\mathbf{u}-\mathbf{d})^T(\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c}) + (\mathbf{d}+t(\mathbf{u}-\mathbf{d})-\mathbf{c})^T(\mathbf{u}-\mathbf{d}) \end{aligned} \tag{10}$$

and inspect

$$\frac{d}{dt}g(t)|_{t=0} = 2\,(\mathbf{d}-\mathbf{c})^T(\mathbf{u}-\mathbf{d}) \tag{11}$$

By (8) we have that

$$\frac{d}{dt}g(t)|_{t=0} < 0 \tag{12}$$

so $g(t)$ is decreasing at $t=0$. If so, there exists a (small) t_0 s.t. $0 < t_0 < 1$ for which $g(t_0) < g(0)$, so

$$\|(1-t_0)\,\mathbf{d}+t_0\mathbf{u}-\mathbf{c}\|^2 < \|\mathbf{d}-\mathbf{c}\|^2 = \mathrm{dist}\,(C,D) \tag{13}$$

But since D is convex, and by assumption we have $\mathbf{u},\mathbf{d} \in D$, we also have $\mathbf{u}' = (1-t_0)\mathbf{d}+t_0\mathbf{u} \in D$. So, we may write

$$\|\mathbf{u}'-\mathbf{c}\|^2 < \|\mathbf{d}-\mathbf{c}\|^2 = \mathrm{dist}\,(C,D) \tag{14}$$

which contradicts the assumption that $\mathbf{c},\mathbf{d}$ satisfy the minimum in (6). Therefore we conclude that $f(\mathbf{u}) \geq 0$ for all $\mathbf{u} \in D$. The proof of $f(\mathbf{v}) \leq 0$ for all $\mathbf{v} \in C$ follows from similar steps. ■