SYS 6003: Optimization **Fall 2016**

Lecture 4

Instructor: Quanquan Gu *Date: Sep 5^{th}*

Recall the general optimization problem as follows

$$\min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}) \tag{1}$$

For a convex optimization problem, the objective function $f(\mathbf{x})$ needs to be convex, and the constraint set $\mathcal{D}$ needs to be convex. Therefore, it is important to know the definitions of convex functions and convex sets. In this lecture, we study convex sets.

Definition 1 (Affine Set) *A set $\mathcal{S}$ is called an affine set if for any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ and any real number θ, $\theta\,\mathbf{x} + (1-\theta)\,\mathbf{y} \in \mathcal{S}$.*

It can be easily verified that $\mathbb{R}$ and $\mathbb{R}^d$ are affine sets. Also, the set of solutions to the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$, i.e., $\{\mathbf{x}|\mathbf{A}\mathbf{x} = \mathbf{b}\}$ is an affine set.

Definition 2 (Convex Set) *A set $\mathcal{S}$ is called a convex set if for any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ and any real number $\theta \in [0, 1]$, $\theta\,\mathbf{x} + (1-\theta)\,\mathbf{y} \in \mathcal{S}$.*

Note that an affine set is also a convex set, however, the converse is not true.

Example 1 (Hyperplane) *A hyperplane in $\mathbb{R}^d$ is defined as $\mathcal{H} \equiv \{\mathbf{x} \in \mathbb{R}^d \,|\, \mathbf{a}^\top\mathbf{x} = b, \mathbf{a} \in \mathbb{R}^d, b \in \mathbb{R}\}$*

Hyperplane is an affine set, and therefore it is also a convex set.

Example 2 (Half Space) *A upper (lower, respectively) half space is the set $\{\mathbf{x} \in \mathbb{R}^d \,|\, \mathbf{a}^\top\mathbf{x} \geq b\}$ ($\{\mathbf{x} \in \mathbb{R}^d \,|\, \mathbf{a}^\top\mathbf{x} \leq b\}$, respectively).*

It can be easily verified that a half space is convex but not affine.

We will now define various combination methods.

Definition 3 *An affine combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^d$ is any $\mathbf{y} \in \mathbb{R}^d$ that can be written as $\mathbf{y} = \sum_{i=1}^k \alpha_i\mathbf{x}_i$, where $\alpha_i \in \mathbb{R}$ for $1 \leq i \leq k$ and $\sum_{i=1}^k \alpha_i = 1$.*

Definition 4 *A convex combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^d$ is any $\mathbf{y} \in \mathbb{R}^d$ that can be written as $\mathbf{y} = \sum_{i=1}^k \alpha_i\mathbf{x}_i$, where $\alpha_i \geq 0$ for $1 \leq i \leq k$ and $\sum_{i=1}^k \alpha_i = 1$.*

If a set is not convex, we can find a convex set which contains this set. It can be achieved by finding its convex hull.

Definition 5 *The convex hull $\mathcal{C}$ of a set $\mathcal{S}$ is*

$$\mathcal{C} = \left\{\mathbf{y} : \mathbf{y} = \sum_{i=1}^n \alpha_i\mathbf{x}_i \text{ for some } n \text{ and } \{\mathbf{x}_i\}_{i=1}^n \subseteq \mathcal{S} \text{ and } \alpha_i \geq 0 \text{ for } 1 \leq i \leq n \text{ s.t. } \sum_{i=1}^n \alpha_i = 1\right\}$$

This set is denoted by $\mathbf{conv}(\mathcal{S})$.

The smallest convex set that contains this set is called its convex hull. In other words, convex hull is the tightest convex relaxation of a nonconvex set. In particular, the convex hull of a convex set is itself.

In the following, we show more examples of convex sets.

Example 3 *A set* $\mathcal{C} \in \mathbb{R}^d$ *is a cone if for any* $\mathbf{x} \in \mathcal{C}$*,* $\{\lambda \mathbf{x} | \lambda \geq 0\} \subseteq \mathcal{C}$*.*

Example 4 *A set* $\mathcal{C} \in \mathbb{R}^d$ *is a convex cone if for any* $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ *and any* $\alpha_1 \geq 0, \alpha_2 \geq 0$*,* $\alpha_1 \mathbf{x} + \alpha_2 \mathbf{y} \in \mathcal{C}$*.*

Example 5 *Define a norm cone as*

$$\mathcal{C} \equiv \{(\mathbf{x}, t) : \mathbf{x} \in \mathbb{R}^d, t \geq 0, \|\mathbf{x}\| \leq t\} \subseteq \mathbb{R}^{d+1}$$

We will prove the norm cone is convex by using the definition of convex sets.
Proof: Let $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2) \in \mathcal{C}$. By definition $t_1, t_2 \geq 0$, thus $\alpha t_1 + (1-\alpha) t_2 \geq 0$ for any $\alpha \in [0, 1]$. Next, we need to show the convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$ is contained in the norm cone. By the triangle inequality, for any $\alpha \in [0, 1]$, we have

$$\begin{aligned} \|\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 &\leq \|\alpha \mathbf{x}_1\| + \|(1-\alpha)\mathbf{x}_2\| \\ &\leq \alpha \cdot \|\mathbf{x}_1\| + (1-\alpha) \cdot \|\mathbf{x}_2\| \\ &\leq \alpha t_1 + (1-\alpha) t_2, \end{aligned}$$

where the last inequality follows from the assumption that $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2) \in \mathcal{C}$. Therefore, $\alpha(\mathbf{x}_1, t_1) + (1-\alpha)(\mathbf{x}_2, t_2) \in \mathcal{C}$ ■

Two more examples of convex sets include Euclidean balls centered at a point $\mathbf{x}$, as well as ellipsoids centered at a point $\mathbf{x}$. Since Euclidean balls are a special case of ellipsoids, we only show that an ellipsoid is convex.

Example 6 *An ellipsoid around a point* $\mathbf{x} \in \mathbb{R}^d$ *is the set* $\mathcal{E}(\mathbf{x})$*, where*

$$\mathcal{E}(\mathbf{x}) = \{\mathbf{y} : (\mathbf{y} - \mathbf{x})^\top \mathbf{P} (\mathbf{y} - \mathbf{x}) \leq 1, \mathbf{P} \succ 0, \mathbf{y} \in \mathbb{R}^d\}$$

Before proving that ellipsoids centered about $\mathbf{x}$ are convex, we introduce a new norm. We will show that this norm is a valid norm, in that it conforms to the three conditions for a norm.

Definition 6 *The Mahalanobis norm is defined as* $\|\mathbf{x}\|_\mathbf{P} = \sqrt{\mathbf{x}^\top \mathbf{P} \mathbf{x}}$*, where* $\mathbf{x} \in \mathbb{R}^d$*, and* $\mathbf{P} \succ 0$*.*

Next, we will prove that $\| \cdot \|_\mathbf{P}$ is indeed a norm.

Proof: First, note that

$$\|\alpha \mathbf{x}\|_\mathbf{P} = \sqrt{(\alpha \mathbf{x})^\top \mathbf{P} (\alpha \mathbf{x})} = |\alpha| \cdot \sqrt{\mathbf{x}^\top \mathbf{P} \mathbf{x}} = |\alpha| \cdot \|x\|_\mathbf{P}.$$

Next, we observe that $\|\mathbf{x}\|_{\mathbf{P}} \geq 0$ and $\|\mathbf{x}\|_{\mathbf{P}} = 0$ iff $\mathbf{x} = 0$ by the definition of $P \succ 0$.
The third component of proving $\|\cdot\|_{\mathbf{P}}$ is a norm is to show the triangle inequality holds. By the definition of the Mahalanobis norm, we have

$$\|\mathbf{x}+\mathbf{y}\|_{\mathbf{P}}^2 = (\mathbf{x}+\mathbf{y})^\top \mathbf{P}(\mathbf{x}+\mathbf{y}) = \mathbf{x}^\top \mathbf{P}\mathbf{x} + \mathbf{y}^\top \mathbf{P}\mathbf{y} + 2\mathbf{x}^\top \mathbf{P}\mathbf{y}. \tag{2}$$

Since $\mathbf{P} \succ 0$, $\mathbf{P}$ has the eigendecompostion $\mathbf{P} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U}$ is an orthogonal matrix, $\mathbf{\Lambda}$ is a diagonal matrix with all diagonal entries being positive. Hence, $\mathbf{\Lambda}^{1/2}$ is well defined, so is $\mathbf{P}^{1/2}$ (defined as $\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top$). From (2) and the definition of $\|\cdot\|_{\mathbf{P}}$, it then follows that

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|_{\mathbf{P}}^2 &= \|\mathbf{x}\|_{\mathbf{P}}^2 + \|\mathbf{y}\|_{\mathbf{P}}^2 + 2\mathbf{x}^\top \mathbf{P}^{\frac{1}{2}}\mathbf{P}^{\frac{1}{2}}\mathbf{y} \\
&\leq \|\mathbf{x}\|_{\mathbf{P}}^2 + \|\mathbf{y}\|_{\mathbf{P}}^2 + 2\|\mathbf{P}^{\frac{1}{2}}\mathbf{x}\|_2 \cdot \|\mathbf{P}^{\frac{1}{2}}\mathbf{y}\|_2 \\
&= \|\mathbf{x}\|_{\mathbf{P}}^2 + \|\mathbf{y}\|_{\mathbf{P}}^2 + 2\|\mathbf{x}\|_{\mathbf{P}} \cdot \|\mathbf{y}\|_{\mathbf{P}},
\end{aligned} \tag{3}$$

where the inequality follows from the Cauchy-Schwarz inequality, and the last equality holds since $\|\mathbf{P}^{1/2}\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\top \mathbf{P}\mathbf{x}} = \|\mathbf{x}\|_{\mathbf{P}}$. Note that (3) can be rewritten as

$$\|\mathbf{x}+\mathbf{y}\|_{\mathbf{P}}^2 \leq (\|\mathbf{x}\|_{\mathbf{P}} + \|\mathbf{y}\|_{\mathbf{P}})^2,$$

which is equivalent to the triangle inequality. Therefore, $\|\cdot\|_{\mathbf{P}}$ is a norm. ■

Given that the Mahalanobis norm is indeed a norm, we can now show that an ellipsoid centered at $\mathbf{x}$ is a convex set.
Proof: Since $(\mathbf{y}-\mathbf{x})^\top \mathbf{P}(\mathbf{y}-\mathbf{x}) = \|\mathbf{y}-\mathbf{x}\|_{\mathbf{P}}^2$, we can redefine ellipsoid as

$$\mathcal{E}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}-\mathbf{x}\|_{\mathbf{P}}^2 \leq 1,\ \mathbf{P} \succ 0,\ \mathbf{x} \in \mathbb{R}^d\},$$

or, equivalently,

$$\mathcal{E}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}-\mathbf{x}\|_{\mathbf{P}} \leq 1,\ P \succ 0,\ \mathbf{x} \in \mathbb{R}^d\}.$$

To show $\mathcal{E}(\mathbf{x})$ is convex, we need to show that for any $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{E}(\mathbf{x})$ and any $\alpha \in [0,1]$, $\alpha\mathbf{y}_1 + (1-\alpha)\mathbf{y}_2 \in \mathcal{E}(\mathbf{x})$, i.e. $\|\alpha\mathbf{y}_1 + (1-\alpha)\mathbf{y}_2 - \mathbf{x}\|_{\mathbf{P}} \leq 1$ holds. This is equivalent to showing

$$\|\alpha\mathbf{y}_1 - \alpha\mathbf{x} + (1-\alpha)\mathbf{y}_2 - (1-\alpha)\mathbf{x}\|_{\mathbf{P}} \leq 1. \tag{4}$$

Applying the triangle inequality gives

$$\begin{aligned}
\|\alpha\mathbf{y}_1 - \alpha\mathbf{x} + (1-\alpha)\mathbf{y}_2 - (1-\alpha)\mathbf{x}\|_{\mathbf{P}} &\leq \|\alpha\mathbf{y}_1 - \alpha\mathbf{x}\|_{\mathbf{P}} + \|(1-\alpha)\mathbf{y}_2 - (1-\alpha)\mathbf{x}\|_{\mathbf{P}} \\
&= \alpha \cdot \|\mathbf{y}_1 - x\|_{\mathbf{P}} + (1-\alpha) \cdot \|\mathbf{y}_2 - x\|_{\mathbf{P}} \\
&\leq \alpha + (1-\alpha) = 1,
\end{aligned}$$

where the last inequality follows from the assumption that $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{E}(\mathbf{x})$. Hence, inequality (4) holds and $\mathcal{E}(\mathbf{x})$ is convex. ■

To show a set is convex, we can prove it by definition. However, proving the convexity of a set by definition is not the only way to prove that a set is convex.

In general, there are two major proof techniques for convex sets:

1. By definition (for any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$, and any $\alpha \in [0, 1]$, show that $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in \mathcal{S}$)

2. Start with a known convex set $\mathcal{S}$, and show that another set $\mathcal{C}$ can be obtained using a convexity preserving operation based on $\mathcal{S}$

For the first proof technique, we have already seen several examples. We therefore will focus on the second technique in the sequel. To use the second proof technique, the key is to know what are typical convexity preserving operations. In the following, we will show two operations. More operations will be introduced next time.

1. Intersection - the intersection of convex sets is also convex. Specifically, suppose $\mathcal{S}_i$ is convex for all $i \in \mathcal{A}$ where $\mathcal{A}$ is a set. Then $\cap_{i \in \mathcal{A}} \mathcal{S}_i$ is also convex.

2. Affine Function: $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{x} \in \mathbb{R}^d, \mathbf{A} \in \mathbb{R}^{m \times d}$, and $\mathbf{b} \in \mathbb{R}^m$. We define the image of $\mathcal{S}$ under f, denoted by $f(\mathcal{S})$, as

   $$f(\mathcal{S}) = \{\mathbf{y} : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \text{ for all } \mathbf{x} \in \mathcal{S}\}.$$

   It can be shown that if $\mathcal{S}$ is convex and f is affine, then $f(\mathcal{S})$ is convex. Next, we define the inverse image of $\mathcal{S}$ under f, denoted by $f^{-1}(\mathcal{S})$, as

   $$f^{-1}(\mathcal{S}) = \{\mathbf{x} : \text{ there exists } \mathbf{y} \in \mathcal{S} \text{ such that } \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}\}.$$

   It can also be verified that if $\mathcal{S}$ is convex and f is affine, then $f^{-1}(\mathcal{S})$ is convex.