Optimization Theory and Algorithm	**Lecture 8 - 05/21/2021**

Lecture 8

Lecturer:Xiangyu Chang — *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

1 Convex Set

Before define what convex set is, let us consider what a line is. A line is determined by two distinct points, namely

$$\{\mathbf{y}|\mathbf{y} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 = \mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)\}$$

is a line. Obviously, if $\theta = 0$, $\mathbf{y} = \mathbf{x}_2$ and $\theta = 1$, $\mathbf{y} = \mathbf{x}_1$. Thus, this line is through the points $\mathbf{x}_1$ and $\mathbf{x}_2$ with respect to the direction $\mathbf{x}_1 - \mathbf{x}_2$.

Then the *line segment* could be denoted as

$$\{\mathbf{y}|\mathbf{y} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 = \mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2), 0 \le \theta \le 1\}. \quad (1)$$

Definition 1 *A set C is convex if the line segment between any tow points in C lies in C. Mathematical formulation: for any $\mathbf{x}_1, \mathbf{x}_2 \in C$, it has $\mathbf{y} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C$.*

Definition 2 *Convex combination of $\{\mathbf{x}_i\}_{i=1}^m$ is $\mathbf{y} = \sum_{i=1}^m \theta_i\mathbf{x}_i$ and $\theta_i \ge 0, \sum_{i=1}^m \theta_i = 1$.*

Definition 3 *Convex hull of set C is a set which contains all convex combination of points in C. Denoted as $Conv(C) = \{\mathbf{y} = \sum_{i=1}^m \theta_i\mathbf{x}_i, \theta_i \ge 0, \sum_{i=1}^m \theta_i = 1, m \ge 1\}$*

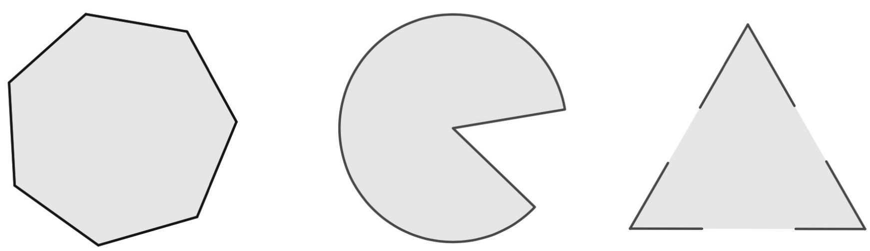

Figure 1: Examples of Convex and Nonconvex Sets

Example 1 *Give examples of convex set:*

- *See Figure 1.*
- *Hyperplane: $C = \{\mathbf{x}|\mathbf{a}^\top\mathbf{x} = \mathbf{b}\}$. Suppose that $\mathbf{x}_0$ is on the hyperplane and $\mathbf{a}$ is perpendicular to C, then for any $\mathbf{x} \in C$, it has $\langle a, \mathbf{x} - \mathbf{x}_0\rangle = 0$. Thus, $\mathbf{a}^\top\mathbf{x} = \mathbf{a}^\top\mathbf{x}_0$. Denote $\mathbf{a}^\top\mathbf{x}_0 = \mathbf{b}$, so a hyperplane is denoted as $\mathbf{a}^\top\mathbf{x} = \mathbf{b}$.*
- *Halfspace: $C = \{\mathbf{x}|\mathbf{a}^\top\mathbf{x} \le \mathbf{b}\}$. See Figure 2.*

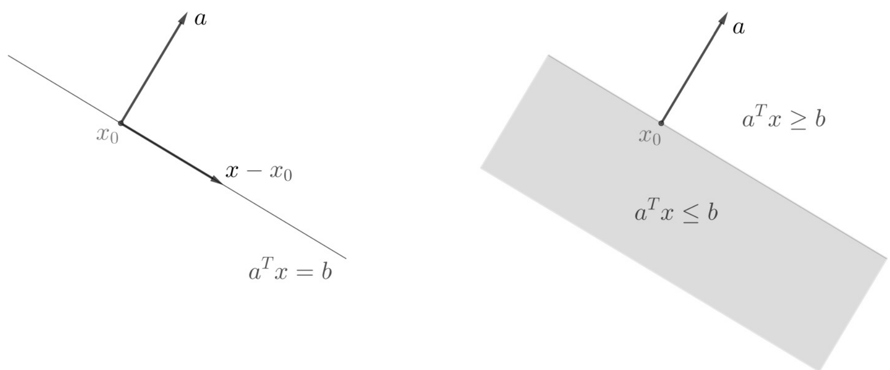

Figure 2: Hyperplane and Halfspace.

- *Norm Ball:* $B(\mathbf{x}_c, r) = \{\mathbf{x}|\ \|\mathbf{x} - \mathbf{x}_c\| \leq r\} = \{\mathbf{x}|\mathbf{x} = \mathbf{x}_c + r\mathbf{v}, \|\mathbf{v}\| \leq 1\}$.
- *Ellipsoid:* $E(\mathbf{x}_c) = \{\mathbf{x}|(\mathbf{x} - \mathbf{x}_c)^\top A(\mathbf{x} - \mathbf{x}_c) \leq 1, A$ *is a positive and definite matrix.*$\} = \{\mathbf{x}|\mathbf{x} = \mathbf{x}_c + A^{-1/2}\mathbf{v}, \|\mathbf{v}\| \leq 1\}$.**Q:** *How to define* $A^{-1/2}$.
- *Cone:* $\{(\mathbf{x}, t)|\ \|\mathbf{x}\| \leq t\}$.
- *Polyhedron:* $\mathcal{P} = \{\mathbf{x}|\ \mathbf{a}_i^\top \mathbf{x} \leq b_j, j = 1, \ldots, m,$ *and* $\mathbf{c}_j^\top \mathbf{x} = d_j, j = 1, \ldots, l\} = \{\mathbf{x}|\ A\mathbf{x} \leq \mathbf{b}, C\mathbf{x} = \mathbf{d}\}$. *Polyhedron is the intersection of a finite numbers of halfspace and hyperplane.*

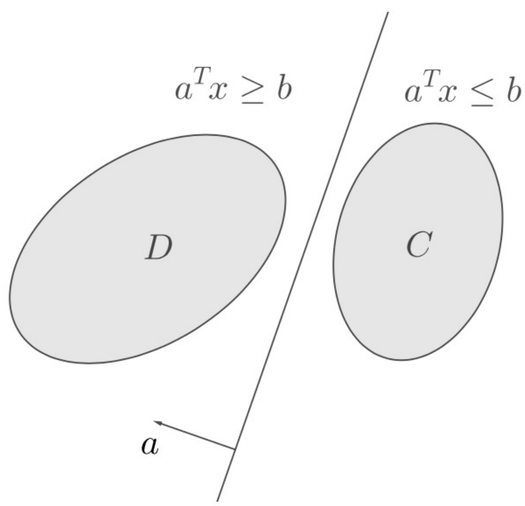

Figure 3: Separating Hyperplane Theorem

Theorem 1 *(Separating Hyperplane Theorem) Suppose that there are two convex sets* C *and* D *satisfies* $C \cap D = \emptyset$. *Then there exists* $\mathbf{a} \neq 0$ *and* $\mathbf{b}$ *such that*

$$\mathbf{a}^\top \mathbf{x} \leq \mathbf{b} \text{ for any } \mathbf{x} \in C, \text{ and } \mathbf{a}^\top \mathbf{x} \geq \mathbf{b} \text{ for any } \mathbf{x} \in D. \tag{2}$$

Proof 1 *See Figure 3.*

Theorem 2 *(Supporting Hyperplan Theory) Suppose that C is a convex set and $\mathbf{x}_0$ is a point on the boundary of C. Then there exists a vector $\mathbf{a}$ such that*

$$\mathbf{a}^\top \mathbf{x} \leq \mathbf{a}^\top \mathbf{x}_0 \textit{ for any } \mathbf{x} \in C, \tag{3}$$

where $\{\mathbf{x}| \ \mathbf{a}^\top \mathbf{x} = \mathbf{a}^\top \mathbf{x}_0\}$ is called a supporting hyperplan of C at $\mathbf{x}_0$.

Operations preserve the convexity:

- If $C_i, i = 1, \dots, \infty$ are convex sets, then $\cap_i C_i$ is convex. This results can be extended as $\cap_{i \in \mathcal{I}} C_i$ is convex if the indicator set $\mathcal{I}$ is convex.
- If C is convex, the $f(C) = \{\mathbf{y}|\mathbf{y} = f(\mathbf{x}) = A\mathbf{x} + b, \mathbf{x} \in C\}$ is convex.

1.0.1 Convex Function

Definition 4 *We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if $dom(f)$ is convex and for any $\mathbf{x}, \mathbf{y} \in dom(f)$*

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}). \tag{4}$$

Example 2 *Let us give some examples of convex function:*

- $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ *or* $f(\mathbf{x}) = A\mathbf{x}$. *Is* $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ *convex?*
- $f(\mathbf{x}) = \|\mathbf{x}\|$.
- $f(x) = \exp(ax)$ *for* $a, x \in \mathbb{R}$.
- $f(x) = x\log(x), x > 0$.
- $f(A) = -\log(\det(A))$ *for any* $A \in \mathcal{S}^n_{++}$.

At here, we briefly introduce a very important theorem about the smooth convex function. It helps us to understand what is convex function in different expressions.

Theorem 3 *Suppose $f \in C^{1,1}_L$. Then the following are equivalent:*

1. *f is convex.*
2. $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle$
3. $\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \geq 0$ *(monotonicity)*
4. *Additionally, if $f \in C^{2,1}_L$, then $\nabla^2 f \succeq 0$ everywhere ($\nabla^2 f$ is positive semi-definite).*

Proof 2

- (1) $\Rightarrow$ (2)*:*

  Write $f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))$ two ways:

  $$f(\mathbf{x} + t(\mathbf{y}-\mathbf{x})) = f(\mathbf{x}) + t\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + o(t)$$
  $$f(\mathbf{x} + t(\mathbf{y}-\mathbf{x})) = f(t\mathbf{y} + (1-t)\mathbf{x}) \leq t f(\mathbf{y}) + (1-t) f(\mathbf{x})$$

  Therefore:

  $$t\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + o(t) \leq t(f(\mathbf{y}) - f(\mathbf{x}))$$
  $$\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{o(t)}{t} \leq f(\mathbf{y}) - f(\mathbf{x})$$

  Taking the limit as $t \to 0$,

  $$\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \leq f(\mathbf{y}) - f(\mathbf{x}).$$

- (2) $\Rightarrow$ (3):
  If we exchange the roles in the inequality in (2), we could get, for any $\mathbf{x}, \mathbf{y} \in dom(f)$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$
$$f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

  And if we sum those two inequalities we could obtain (3).

- (3) $\Rightarrow$ (2):
  Define $\mathbf{x}_t = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$ *and* $\phi(t) = f(\mathbf{x}_t)$. *Observe that*

$$\phi'(s) = \langle \nabla f(\mathbf{x}_s), \mathbf{y} - \mathbf{x} \rangle, \quad \phi(0) = f(\mathbf{x}), \quad \phi(1) = f(\mathbf{y}).$$

  Suppose $t > s$. *Then*

$$\begin{aligned} \phi'(t) - \phi'(s) &= \langle \nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}_s), \mathbf{y} - \mathbf{x} \rangle \\ &= \frac{1}{t-s} \langle \nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}_s), \mathbf{x}_t - \mathbf{x}_s \rangle \geq 0, \end{aligned}$$

  so ϕ' *is nondecreasing.*

$$\begin{aligned} & f(y) = \phi(0) + \int_0^1 \phi'(\tau) d\tau \geq \phi(0) + \phi'(0) \\ \Rightarrow \quad & f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \end{aligned}$$

- (2) $\Rightarrow$ (1):
  Let's define $l_{\mathbf{x}}(\mathbf{y}) := f(\mathbf{x}) + t\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$, *and from (2) we know that, for any* $\mathbf{y} \in dom(f)$,

$$f(\mathbf{y}) = \max_{\mathbf{x} \in dom(f)} l_{\mathbf{x}}(\mathbf{y})$$

  Notice that the reason why we could put $=$ *there is because,* $f(\mathbf{y}) = l_{\mathbf{y}}(\mathbf{y})$. *And for each* $\mathbf{x}$ *we know that* $l_{\mathbf{x}}(\mathbf{y})$ *is a affine function and the point-wise maximum of arbitrary convex function is still convex. Then we know that* f *is convex.*

And for smooth convex function we have a really nice to judge if a point is optimal.

Theorem 4 (Optimality Conditions) *The following are equivalent for a convex* C^1 *function:*

1. $\mathbf{x}^*$ *is a global minimum.*
2. $\mathbf{x}^*$ *is a local minimum.*
3. $\mathbf{x}^*$ *is a stationary points (*$\nabla f(\mathbf{x}^*) = 0$*).*

Proof 3
- *(1)* $\Rightarrow$ *(2)*
  This direction is trivial. If x^* *is global minimum then it definitely is a local minimum.*

- *(2)* $\Rightarrow$ *(3)*
  Assume that $\mathbf{x}^*$ *is the local minimum, then there exists a ball* $B(\mathbf{x}^*, \epsilon) = \{bx | \; \|\mathbf{x} - \mathbf{x}^*\| \leq \epsilon\}$ *such that* $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ *for any* $\mathbf{x} \in B(\mathbf{x}^*, \epsilon)$. *Based on Taylor expansion, it has*

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \langle \nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle + o(\|\mathbf{x} - \mathbf{x}^*\|).$$

  Let $\mathbf{x} - \mathbf{x}^* = -s\nabla f(\mathbf{x}^*)$, *and the* s *makes* $\mathbf{x} \in B(\mathbf{x}^*, \epsilon)$. *Thus,* $f(\mathbf{x}) = f(\mathbf{x}^*) - s\|\nabla f(\mathbf{x}^*)\|^2 + o(s)$. *So,*

$$0 \leq \frac{f(\mathbf{x}) - f(\mathbf{x}^*)}{s} = -s\|\nabla f(\mathbf{x}^*)\|^2 + o(s) \leq 0. \tag{5}$$

  We obtain that $\nabla f(\mathbf{x}^*) = 0$.

- *(3) ⇒ (1)*
  Assume x^* *is a critical point,* $\nabla f(x^*) = 0$*, and from convexity we have*

  $$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*)$$

  for every x*, then* x^* *is a global minimizer.*

Theorem 5 $f(\mathbf{x})$ *is a convex function if and only if for any* $\mathbf{x} \in (f), \mathbf{d} \in \mathbb{R}^n$*, function* $\phi : \mathbb{R} \to \mathbb{R}$*,*

$$\phi(t) := f(\mathbf{x} + t\mathbf{d}), (\phi) = \{t | \ \mathbf{x} + t\mathbf{d} \in dom(f)\},$$

is convex.

Proof 4 *See Theorem 2.8 on Page 48.*

Definition 5 *Denote that the epigraph of a function* f *as the set* $epi(f) = \{(\mathbf{x}, t) | \ f(\mathbf{x}) \leq t\}$*.*

Theorem 6 *function* f *is convex if and only if* $epi(f)$ *is a convex set.*

Proposition 1
- *If* $f_1, f_2, \ldots, f_m$ *are convex, then* $g(\mathbf{x}) = \max(f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ *is convex.*
- $f(\mathbf{x}, \mathbf{y})$ *is convex with respect to* $\mathbf{x}$*, then* $g(\mathbf{x}) = sup_{\mathbf{y} \in \mathcal{Y}} f(\mathbf{x}, \mathbf{y})$ *is convex.*

References