# Lecture 2: Convex Analysis

*Lecturer: Jacob Abernethy* *Scribes: Ning Jiang*



## 2.1 A few concepts

For a differentiable function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$, the gradient of $f$ at a point $\mathbf{x} \in \text{dom}f$ is the vector containing the partial derivatives of the function at that point, namely, $\nabla f(\mathbf{x}) = (\frac{\partial f}{\partial x_1}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x}))$.

For a twice differentiable function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$, the Hessian of $f$ at a point $\mathbf{x} \in \text{dom}f$ is the matrix containing the second derivatives of the function at that point, namely, $\nabla^2 f(\mathbf{x})$ is the matrix with elements given by

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}), 1 \leq i, j \leq n$$

We say that a function $f$ is $c$-Lipschiz with respect to a norm $\|\cdot\|$ for $c \in \mathbb{R}^+$ if

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq c\|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \text{dom}(f).$$

**Claim 2.1.** *Let $f$ be a real-valued differentiable function. Then, $\|\nabla f(\mathbf{x})\| \leq c$ if and only if $f$ is c-Lipschitz.*

**Proof: the "$\Rightarrow$" direction**: Assume $\forall \mathbf{x} \in \text{dom}(f), \|\nabla f(\mathbf{x})\| \leq c$. Then for $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$, there exists $t \in [0, 1]$ such that

$$|f(\mathbf{x}) - f(\mathbf{y})| = |\nabla f(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))^T(\mathbf{x} - \mathbf{y})|$$

By the Schwarz's inequality, the equation gives the estimate:

$$\begin{aligned} |f(\mathbf{x}) - f(\mathbf{y})| &\leq \|\nabla f(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))\| \|(\mathbf{x} - \mathbf{y})\| \\ &\leq c\|(\mathbf{x} - \mathbf{y})\| \end{aligned}$$

**the "$\Leftarrow$" direction**: Assume $\forall \mathbf{x}, \mathbf{y} \in \text{dom}(f), f(\mathbf{x}) - f(\mathbf{y}) \leq c\|\mathbf{x} - \mathbf{y}\|$. Then the directional derivative of $f$ along $u$ is:

$$\nabla f(\mathbf{x})^T \mathbf{u} = \lim_{\delta \to 0} \frac{f(\mathbf{x} + \delta \mathbf{u}) - f(\mathbf{x})}{\delta} \leq \lim_{\delta \to 0} \frac{c\|\mathbf{x} + \delta \mathbf{u} - \mathbf{x}\|}{\delta} = c\|\mathbf{u}\|$$

Set $\mathbf{u} = \frac{(\nabla f(\mathbf{x}))^T}{\|\nabla f(\mathbf{x})\|}$, then we have $\|\nabla f(\mathbf{x})\| \leq c$. ■

## 2.2 Convexity

**Definition 2.2** (convex set)**.** *A set $U \subseteq \mathbb{R}^n$ is convex if for all $\mathbf{x}, \mathbf{y} \in U$ and all $\alpha$ in the interval $[0, 1]$, the point $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}$ also belongs to $U$.*

**Definition 2.3** (convex function)**.** *Let $X$ be a convex set in $\mathbb{R}^n$ and let $f : X \to \mathbb{R}$ be a function. We say that $f$ is* ***convex*** *if $\forall \mathbf{x}, \mathbf{y} \in X, \forall \alpha \in [0, 1] : f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$. We say that $f$ is* ***strictly convex*** *if $\forall \mathbf{x} \neq \mathbf{y} \in X, \forall \alpha \in (0, 1) : f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$.*

Here are some alternative characterizations of convexity:

- A function $f$ is convex if and only if it satisfies the Jensen's inequality everywhere: $\forall \mathbf{x} \in \mathrm{dom}(f), \mathbb{E}(f(\mathbf{x})) \geq f(\mathbb{E}(\mathbf{x}))$.
- A differentiable function $f$ is convex if and only if $f(\mathbf{x}+\mathbf{u}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T \mathbf{u}$.
- A twice differentiable function $f$ is convex if and only if $\forall \mathbf{x} \in \mathrm{dom}(f), \nabla^2 f(x) \succeq 0$.

Here are some examples of convex functions:

- $f(\mathbf{x}) = \|\mathbf{x}\|^2$,
- $f(\mathbf{x}) = \mathbf{x}^T M \mathbf{x}$, when $M$ is positive semidefinite,
- If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ (e.g. $f(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2$), then $g_1(x) = \mathbb{E}_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$ and $g_2(x) = \sup_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$ are convex.

**Definition 2.4** (strongly convex)**.** *A differentiable function $f$ is $c$-strongly convex with respect to a norm $\|\cdot\|$ if for all $\mathbf{x}, \mathbf{u}$ such that $\mathbf{x}, \mathbf{x}+\mathbf{u} \in \mathrm{dom} f$, the following inequality holds:*

$$f(\mathbf{x}+\mathbf{u}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T \mathbf{u} + \frac{c}{2}\|\mathbf{u}\|^2.$$

**Definition 2.5** (strongly smooth)**.** *A differentiable function $f$ is $c$-strongly smooth with respect to a norm $\|\cdot\|$ if for all $\mathbf{x}, \mathbf{u}$ such that $\mathbf{x}, \mathbf{x}+\mathbf{u} \in \mathrm{dom} f$, the following inequality holds:*

$$f(\mathbf{x}+\mathbf{u}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T \mathbf{u} + \frac{c}{2}\|\mathbf{u}\|^2.$$

For example, $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$ is both 1-strongly convex and 1-strongly smooth.

**Fact 2.6.** *When $f$ is twice differentiable, $f$ is $c$-strongly convex with respect to $\|\cdot\|_2$ if and only if $\nabla^2 f(x) \succeq cI$, and $f$ is $c$-strongly smooth with respect to $\|\cdot\|_2$ if and only if $cI \succeq \nabla^2 f(x)$.*

**Theorem 2.7.** *To generalize the above notion, a twice-differentiable function $f$ is $c$-strongly convex with respect to a norm $\|\cdot\|$ if and only if $\inf_{\mathbf{x}:\|\mathbf{x}\|=1} \mathbf{x}^\top \nabla^2 f(x) \mathbf{x} \geq c$.*

*Similarly, a twice-differentiable function $f$ is $c$-strongly smooth with respect to a norm $\|\cdot\|$ if and only if $\sup_{\mathbf{x}:\|\mathbf{x}\|=1} \mathbf{x}^\top \nabla^2 f(x) \mathbf{x} \leq c$.*

**Proof:** Left as exercise. ∎

## 2.3 Bregman divergence

**Definition 2.8** (Bregman divergence)**.** *The Bregman divergence associated with $f$ is a function $D_f : \mathrm{dom}(f) \times \mathrm{dom}(f) \to \mathbb{R}$ defined by $D_f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})^T(\mathbf{x}-\mathbf{y})$.*

Here are some examples:

- $f(\mathbf{x}) = \|\mathbf{x}\|^2, D_f(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|^2$,
- $f(\mathbf{p}) = \sum_{i=1}^n p_i \log p_i, D_f(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i}$, which is the Kullback-Leibler divergence.

Here are some properties of Bregman Divergence:

- If $f$ is convex, $D_f(\mathbf{x}, \mathbf{y}) \geq 0$.
- $\forall \mathbf{x} \in \mathrm{dom}(f), D_f(\mathbf{x}, \mathbf{x}) = 0$.
- In general, $D_f(\mathbf{x}, \mathbf{y}) \neq D_f(\mathbf{y}, \mathbf{x})$.

**Fact 2.9.** *If $f$ is $c$-strongly convex, $D_f(\mathbf{x}, \mathbf{y}) \geq \frac{c}{2}\|\mathbf{x}-\mathbf{y}\|^2$.*

## 2.4 convex conjugate

**Definition 2.10** (Fenchel conjugate)**.** *For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, its **Fenchel conjugate** is*

$$f^*(\theta) = \sup_{\mathbf{x}\in\mathbb{R}^n} \mathbf{x}^T\theta - f(\mathbf{x}).$$

For example, we have

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2, f^*(\theta) = \frac{1}{2}\|\theta\|^2$.
- $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{M}\mathbf{x}$ and $\mathbf{M}$ is positive semidefinite, then $f^*(\mathbf{x}) = \frac{1}{2}\theta^T\mathbf{M}^{-1}\theta$.

**Fact 2.11** (biconjugate)**.** *Under a weak condition*[1]*, $f = f^{**}$.*

**Fact 2.12.** *If $f$ is differentiable and strongly convex, $\forall \mathbf{x} \in \mathrm{dom}(f), \theta \in \mathrm{dom}(f^*)$ we have $\nabla f^*(\nabla f(\mathbf{x})) = \mathbf{x}$ and $\nabla f(\nabla f^*(\theta)) = \theta$.*

**Fact 2.13.** *If $f$ is strictly convex and differentiable, $D_f(\mathbf{x}, \mathbf{y}) = D_{f^*}(\nabla f(\mathbf{y}), \nabla f(\mathbf{x}))$.*

[1] $f$ is closed convex