# Lecture 3. Function Properties

Advanced Optimization (Fall 2022)

**Peng Zhao**

zhaop@lamda.nju.edu.cn

Nanjing University

# Outline

- Smoothness

- Strong Convexity

# Lipschitz Continuity

**Definition 1** (Continuity). A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\mathbf{x} \in \operatorname{dom} f$ if for all $\epsilon > 0$ there exists a $\delta > 0$ with $\mathbf{y} \in \operatorname{dom} f$, such that

$$\|\mathbf{y} - \mathbf{x}\|_2 \leq \delta \Rightarrow \|f(\mathbf{y}) - f(\mathbf{x})\|_2 \leq \epsilon.$$

**Definition 2** (Lipschitz Continuity). A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is $G$-Lipschitz-continuous if for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$,

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq G \|\mathbf{x} - \mathbf{y}\|.$$

# Lipschitzness and Subgradient

- Relationship between ***Lipschitzness*** and ***bounded subgradient***

**Theorem 1.** *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Suppose that $\mathcal{X} \subseteq \mathrm{int}(\mathrm{dom}\, f)$. Consider the following two claims:*

*(i) Lipschitzness:* $|f(\mathbf{x}) - f(\mathbf{y})| \leq G\|\mathbf{x} - \mathbf{y}\|$ *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

*(ii) Bounded subgradient:* $\|\mathbf{g}\| \leq G$ *for any* $\mathbf{g} \in \partial f(\mathbf{x}), \mathbf{x} \in \mathcal{X}$.

*Then*

*(a) (ii) $\Rightarrow$ (i).*

*(b) if $\mathcal{X}$ is open, then (i) $\Leftrightarrow$ (ii).*

# Smoothness

**Definition 3** (Smoothness)**.** A function $f$ is $L$-smooth if, for any $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|.$$

Smoothness is also called ***gradient Lipschitz***.

The class of $L$-smooth functions over domain $\mathcal{X}$ is denoted by $C_L^{1,1}(\mathcal{X})$.

# Smoothness

**Definition 4.** Let $\mathcal{X} \subseteq \mathbb{R}^d$. We denote by $C_L^{a,b}(\mathcal{X})$ the class of functions with the following properties:

- any $f \in C_L^{a,b}(\mathcal{X})$ is $a$ times continuously differentiable on $\mathcal{X}$.
- $f$'s $b$-th derivative is Lipschitz continuous on $\mathcal{X}$ with constant $L$:

$$\left\|\nabla^b f(\mathbf{x}) - \nabla^b f(\mathbf{y})\right\| \leq L\|\mathbf{x} - \mathbf{y}\|, \ \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}.$$

*Ref: Lectures on Convex Optimization, Yurii Nesterov. Page 23-24.*

- Lipschitz continuous functions belong to $C_L^{0,0}(\mathcal{X})$.
- Smoothness is in fact the *Lipshitzness of gradients*.

# Smoothness

**Example 1.** Linear function $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + c$ is 0-smooth.

**Example 2.** Quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + \mathbf{w}^\top \mathbf{x} + c$ is $\|A\|_{\mathrm{op},p}$-smooth w.r.t. $\|\cdot\|_p$ norm.

***Proof.*** The proof is direct by the definition of smoothness and the operator norm:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_p = \|A\mathbf{x} - A\mathbf{y}\|_p \leq \|A\|_{\mathrm{op},p}\|\mathbf{x} - \mathbf{y}\|_p.$$

**Definition 6** (Matrix Operator Norm). The operator norm (or called induced norm) of a matrix $A \in \mathbb{R}^{m\times n}$ is defined by

$$\|A\|_{\mathrm{op},p} \triangleq \max\left\{\frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \;\middle|\; \mathbf{x} \in \mathbb{R}^d, \mathbf{x} \neq \mathbf{0}\right\}.$$

# Smoothness

**Example 3.** Log-sum-exp function $f(\mathbf{x}) = \log\left(e^{x_1} + e^{x_2} + \cdots + e^{x_n}\right)$ is 1-smooth w.r.t. $\ell_2$-norm and $\ell_\infty$-norm.

**Example 4.** Function $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_p^2$ is $(p-1)$-smooth w.r.t. $\ell_p$-norm.

**Example 5.** Function $f(\mathbf{x}) = \sqrt{1 + \|\mathbf{x}\|_2^2}$ is 1-smooth w.r.t. $\ell_2$-norm.

**Example 6.** Function $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \Pi_{\mathcal{X}}(\mathbf{x})\|^2$ is 1-smooth, where $\Pi_{\mathcal{X}}(\mathbf{x})$ denotes the Euclidean projection of $\mathbf{x}$ onto a *convex* domain $\mathcal{X}$.

# Smoothness

**Example 6.** Function $f(\mathbf{x}) = \sqrt{1 + \|\mathbf{x}\|_2^2}$ is 1-smooth w.r.t. $\ell_2$-norm.

***Proof:***

$$\nabla f(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{\|\mathbf{x}\|_2^2 + 1}}$$

$$\Longrightarrow \nabla^2 f(\mathbf{x}) = \frac{1}{\sqrt{\|\mathbf{x}\|_2^2 + 1}}\left(I - \frac{\mathbf{x}\mathbf{x}^\top}{\|\mathbf{x}\|_2^2 + 1}\right) \preceq \frac{1}{\sqrt{\|\mathbf{x}\|_2^2 + 1}} I \preceq I \qquad \square$$

**Example 7.** Function $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \Pi_{\mathcal{X}}(\mathbf{x})\|^2$ is 1-smooth, where $\Pi_{\mathcal{X}}(\mathbf{x})$ denotes the Euclidean projection of $\mathbf{x}$ onto a *convex* domain $\mathcal{X}$.

# Smoothness

The next lemma is an ***equivalent*** condition of smoothness.

**Lemma 1** (Descent Lemma). *Let $f$ be an $L$-smooth function over a given convex set $\mathcal{X}$. Then for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$*

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

***Proof:***

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \quad \text{(calculus)}$$

$$\Longrightarrow f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle = \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t$$

$$\text{(Cauchy-Schwarz)} \leq \int_0^1 \|\nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x})\| \, \|\mathbf{y} - \mathbf{x}\| \, \mathrm{d}t$$

$$\text{(smoothness)} \leq L\|\mathbf{y} - \mathbf{x}\|^2 \int_0^1 t \mathrm{d}t \leq \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2 \qquad \square$$

# Smoothness

**Theorem 2** (*First-order* Characterizations of $L$-smoothness). *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function, differentiable over $\mathcal{X}$. Then the following claims are equivalent:*

*(i)* $f$ *is L-smooth.*

*(ii)* $f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2$ *for all* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

*(iii)* $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2$ *for all* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

*(iv)* $\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \frac{1}{L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2$ *for all* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

*(v)* $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) - \frac{L}{2}\lambda(1-\lambda)\|\mathbf{x} - \mathbf{y}\|^2$ *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ *and* $\lambda \in [0, 1]$.

# Smoothness

**Theorem 3** (*Second-order* Characterization of $L$-smoothness). *Let $f$ be a twice continuously differentiable function over $\mathbb{R}^d$. Then for a given $L \geq 0$, $L$-smoothness w.r.t. the $\ell_p$-norm ($p \in [1, \infty]$) is equivalent to*

$$\left\|\nabla^2 f(\mathbf{x})\right\|_{op,p} \leq L,$$

*for any* $\mathbf{x} \in \mathbb{R}^d$.

# Strong Convexity

**Definition 5** (Strong Convexity)**.** A function $f$ is $\sigma$-strongly convex if, for any $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$ and $\lambda \in [0, 1]$,

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|^2.$$

- Clearly, for generally convex functions, $\sigma = 0$.

***Examples:***

- $f(\mathbf{x}) = \|\mathbf{x}\|_p^2$ is 2-strongly-convex with respect to norm $\|\cdot\|_p$.
- Negative entropy $f(\mathbf{x}) = \sum_{i=1}^d x_i \ln x_i$ over probability distribution (i.e., $x_i \in [0,1]$ and $\sum_{i=1}^d x_i = 1$) is 1-strongly-convex.

# Strong Convexity

**Theorem 3** (*First-order* Characterizations of Strong Convexity)**.** *Let $f$ be a proper closed and convex function. Then for a given $\sigma > 0$, the followings equal:*

*(i)* *$f$ is $\sigma$-strongly convex.*

*(ii)* *For any $\mathbf{x} \in \mathrm{dom}(\partial f), \mathbf{y} \in \mathrm{dom}(f)$ and $\mathbf{g} \in \partial f(\mathbf{x})$,*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle + \frac{\sigma}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$ ***most common***

*(iii)* *For any $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(\partial f)$, and $\mathbf{g}_\mathbf{x} \in \partial f(\mathbf{x}), \mathbf{g}_\mathbf{y} \in \partial f(\mathbf{y})$,*

$$\langle \mathbf{g}_\mathbf{x} - \mathbf{g}_\mathbf{y}, \mathbf{x} - \mathbf{y} \rangle \geq \sigma\|\mathbf{x} - \mathbf{y}\|^2.$$

*(iv)* *Function $f(\cdot) - \frac{\sigma}{2}\|\cdot\|^2$ is convex.*

# Strong Convexity

***Proof: (i)→(ii)***

$$f(\lambda \mathbf{y} + (1-\lambda)\mathbf{x}) \leq \lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{x}) - \frac{\sigma}{2}\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|^2$$

$$\Rightarrow \frac{f(\mathbf{x} + \lambda(\mathbf{y}-\mathbf{x})) - f(\mathbf{x})}{\lambda} \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma}{2}(1-\lambda)\|\mathbf{x}-\mathbf{y}\|^2 \quad \text{(rearrange)}$$

$$\Rightarrow f'(\mathbf{x};\mathbf{y}-\mathbf{x}) \triangleq \lim_{\lambda\to 1} \frac{f(\mathbf{x} + \lambda(\mathbf{y}-\mathbf{x})) - f(\mathbf{x})}{\lambda} \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma}{2}\|\mathbf{x}-\mathbf{y}\|^2$$

$f'(\mathbf{x};\mathbf{y}-\mathbf{x})$: the ***directional derivative*** of $f$ at point $\mathbf{x}$ along direction $\mathbf{y}-\mathbf{x}$

$$\forall \mathbf{g} \in \partial f(\mathbf{x}), \quad \langle \mathbf{g}, \mathbf{y}-\mathbf{x}\rangle \leq f'(\mathbf{x};\mathbf{y}-\mathbf{x})$$

Plugging $\mathbf{g} = \nabla f(\mathbf{x})$ finishes the proof. □

# Strong Convexity

**Theorem 4.** *Let $\mathcal{X}$ be a Euclidean space. Then $f$ is $\sigma$-strongly convex if and only if the function $f(\cdot) - \frac{\sigma}{2}\|\cdot\|^2$ is convex.*

***$f$ is "as least as convex" as a quadratic function.***

**Example 8.** $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + \mathbf{w}^\top\mathbf{x} + c$ is $\sigma$-strongly convex if and only if $A \succeq \sigma I$.

***Proof:***

$f$ is $\sigma$-strongly convex if and only if $h(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top(A - \sigma I)\mathbf{x} + \mathbf{w}^\top\mathbf{x} + c$ is convex

$\Rightarrow$ $\nabla^2 h(\mathbf{x}) = A - \sigma I \succeq 0$ $\square$

# Strong Convexity

**Theorem 5** (*Second-order* Characterization of Strong Convexity). *Let $\mathcal{X}$ be a Euclidean space. Then $f$ is $\sigma$-strongly convex if and only if for any $\mathbf{x}, \mathbf{w} \in \mathcal{X}$,*

$$\mathbf{w}^\top \nabla^2 f(\mathbf{x}) \mathbf{w} \geq \sigma \|\mathbf{w}\|^2 .$$

*Furthermore, when using $\ell_2$-norm, it is equivalent to $\nabla^2 f(\mathbf{x}) \succeq \sigma I$.*

**Theorem 6.** *Let $f$ be a proper closed and $\sigma$-strongly convex function. Then*

- *$f$ has a unique minimizer, denoted by $\mathbf{x}^\star$.*
- *$f(\mathbf{x}) - f(\mathbf{x}^\star) \geq \frac{\sigma}{2}\|\mathbf{x} - \mathbf{x}^\star\|^2$ for all $\mathbf{x} \in \mathrm{dom}(f)$.*

# Relationship

If function $f$ is both $\sigma$-strongly convex and $L$-smooth, then

- $\sigma I \preccurlyeq \nabla^2 f(\mathbf{x}) \preccurlyeq LI$
- $f$ is *$\gamma$-well-conditioned* where $\gamma$ is called the condition number of $\gamma = \sigma/L \leq 1$.
- *Conjugate Correspondence:*

  (a) If $f$ is a $\frac{1}{\sigma}$-smooth and convex, then $f^*$ is $\sigma$-strongly convex w.r.t. the dual norm $\|\cdot\|_*$.

  (b) If $f$ is proper closed $\sigma$-strongly convex, then $f^*$ is $\frac{1}{\sigma}$-smooth.

Conjugate function: $f^*(\mathbf{y}) = \max_{\mathbf{x}\in\mathcal{X}}\{\langle \mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x})\}$