# Smoothness and Strong Convexity

**Instructor: Jin Zhang**

Department of Mathematics
Southern University of Science and Technology

*Fall 2023*

# Contents


1. $L$-Smooth Functions
2. Strong Convexity
3. Smoothness and Strong Convexity Correspondence

# 1. $L$-Smoothness

## Recall

- **(vector $l_p$ norm and its dual).** For a given $1 \leq p < \infty$, the $l_p$-norm on $\mathbb{R}^n$ is given by

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \text{ for all } x = (x_1, \cdots, x_n)^T \in \mathbb{R}^n.$$

The $l_\infty$-norm on $\mathbb{R}^n$ is defined by

$$\|x\|_\infty = \max_{i=1,\cdots,n} |x_i|.$$

$l_q$-norm is the dual of $l_p$-norm $(1 \leq p \leq \infty)$, where $q$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$.

## Recall

- **(induced matrix norm).** Given $l_a$-norm and $l_b$-norm on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, the induced matrix norm on $\mathbb{R}^{m\times n}$ is given by

$$\|A\|_{a,b} = \max_{x\in\mathbb{R}^n} \left\{ \|Ax\|_b \;\middle|\; \|x\|_a \le 1 \right\} \text{ for all } A \in \mathbb{R}^{m\times n}.$$

It can be easily shown that for any $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$,

$$\|Ax\|_b \le \|Ax\|_{a,b} \|x\|_a .$$

## Definition: ($L-$smoothness)

Let $L \geq 0$. Then $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be $L$-smooth over a set $D \in \mathbb{R}^n$ if it is differentiable over $D$ and satisfies

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L \|x - y\| \text{ for all } x, y \in D.$$

The constant $L$ is called the smoothness parameter.

## Remark:

1. $C_L^{1,1}(D) = \left\{f : \mathbb{R}^n \to \overline{\mathbb{R}} \,|\, f \text{ is } L\text{-smooth over } D \subset \mathbb{R}^n\right\}$.
2. Let $0 \leq L_1 \leq L_2$, then $C_{L_1}^{1,1}(D) \subset C_{L_2}^{1,1}(D)$.
3. We say $f : \mathbb{R}^n \to \mathbb{R}$ is $L$-smooth if $f$ is $L$-smooth over $\mathbb{R}^n$.

**Example ($0$-smoothness of affine functions).**

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \langle b, x \rangle + c$, where $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then $f$ is $0$-smooth.

Recall: Let $C \subset \mathbb{R}^n$ be nonempty, then $P_C$ is the orthogonal projection mapping associated with $C$ defined by

$$P_C(x) = \arg\min_{y \in C} \|y - x\| .$$

In addition, if $C$ is closed and convex, $P_C$ is well-defined and the function $\varphi_C(x) : \mathbb{R}^n \to \mathbb{R}$ given by

$$\varphi_C(x) \equiv \frac{1}{2} d_C^2(x) = \frac{1}{2} \|x - P_C(x)\|^2$$

is convex. Under the $l_2$-norm setting, $\varphi_C$ is differentiable with

$$\nabla \varphi_C(x) \equiv x - P_C(x).$$

**Theorem:**

Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Then

(a). **(firm nonexpansiveness)** For any $v, w \in \mathbb{R}^n$,

$$\langle P_C(v) - P_C(w), v - w \rangle \geq \|P_C(v) - P_C(w)\|^2 .$$

(b). **(nonexpansiveness)** For any $v, w \in \mathbb{R}^n$,

$$\|P_C(v) - P_C(w)\| \leq \|v - w\| .$$

## Example (1-smoothness of $\frac{1}{2}d_C^2$).

Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Consider the function $\varphi_C(x) = \frac{1}{2}d_C^2$, then $\varphi_C$ is 1-smooth with respect to (w.r.t.) the $l_2$-norm (over $\mathbb{R}^n$).

## Example (1-smoothness of $\frac{1}{2}\|\cdot\|_2^2 - \frac{1}{2}d_C^2$).

Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Consider the function $\psi_C(x) = \frac{1}{2}\|x\|_2^2 - \frac{1}{2}d_C^2(x)$. Then $\psi_C$ is 1-smooth with respect to (w.r.t.) the $l_2$-norm.

**Lemma: (descent lemma).**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be $L$-smooth over a given convex set $\Omega$. Then for any $x, y \in \Omega$,

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|x - y\|^2 .$$

**Theorem: (characterizations of $L$-smoothness).**

Let $f : \mathbb{R}^n \to \mathbb{R}$ be real-valued convex and differentiable over $\mathbb{R}^n$, and let $L > 0$. Then the following claims are equivalent:

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

(ii). $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|x - y\|^2$ for all $x, y \in \mathbb{R}^n$.

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

(iv). $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|_*^2$ for all $x, y \in \mathbb{R}^n$.

(v). $f(\lambda x + (1 - \lambda) y) \geq \lambda f(x) + (1 - \lambda) f(y) - \frac{L}{2} \lambda (1 - \lambda) \|x - y\|^2$ for any $x, y \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$.

Recall: Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable over $\mathbb{R}^n$. Then for any $x, y \in \mathbb{R}^n$,

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

## Remark: necessity of convexity for $L$-smoothness characterization.

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = -\frac{1}{2} \|x\|_2^2$. Then $f$ is 1-smooth w.r.t. the $l_2$-norm but is not $L$-smooth w.r.t. the $l_2$-norm for $L < 1$. However, $-f$ is convex, so for any $x, y \in \mathbb{R}^n$,

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

or equivalently,

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle,$$

which implies that (ii) of Theorem (characterization of $L$-smoothness) is satisfied with $L = 0$.

**Theorem: ($L$-smoothness and boundedness of the Hessian).**

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Then for a given $L \geq 0$, the following two claims are equivalent:

(i). $f$ is $L$-smooth w.r.t. the $l_p$-norm $(1 \leq p \leq \infty)$.

(ii). $\left\|\nabla^2 f(x)\right\|_{p,q} \leq L$ for any $x \in \mathbb{R}^n$, where $q$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$.

Recall:

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable convex over $\mathbb{R}^n$. Then $\nabla f(x) \geq 0$ for any $x \in \mathbb{R}^n$.
- If $A \in \mathbb{R}^{n \times n}$ is nonnegative definite, $\|A\|_{2,2} = \|A\|_2 = \lambda_{\max}(A)$.

**Example: (1-smoothness of $\sqrt{1 + \|\cdot\|_2^2}$).**

The function $f : \mathbb{R}^n \to \mathbb{R}$, given by

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

is 1-smooth w.r.t. the $l_2$-norm.

### Definition: (strong convexity).

$f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called $\sigma$-strongly convex for a given $\sigma > 0$ if $\mathrm{dom}(f)$ is convex and the following inequality holds for any $x, y \in \mathrm{dom}(f)$ and $0 \leq \lambda \leq 1$:

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

### Theorem

$f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is $\sigma$-strongly convex w.r.t the $l_2$-norm if and only if $f(\cdot) - \frac{\sigma}{2}\left\|\cdot\right\|_2^2$ is convex.

**Example: (strong convexity of quadratic functions).**

Consider the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c,$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.
Then $f$ is strongly convex w.r.t. the $l_2$-norm if and only if $A > 0$, and in that case, $\lambda_{\text{min}}(A)$ is its largest possible strong convexity parameter.

**Lemma:**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be $\sigma$-strongly convex and let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. Then $f + g$ is $\sigma$-strongly convex.

**Example:** (strong convexity of $\frac{1}{2}\left\|\cdot\right\|_2^2 + \delta_C$).

Let $C \subset \mathbb{R}^n$ be nonempty convex. Then the function $\frac{1}{2}\left\|\cdot\right\|_2^2 + \delta_C$ is 1-strongly convex.

**Lemma:**

Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be closed convex and let $[a, b] \subseteq \operatorname{dom} f (a \leq b)$. Then

$$f(b) - f(a) = \int_a^b h(t)dt,$$

where $h : (a, b) \to \mathbb{R}$ satisfies $h(t) \in \partial f(t)$ for any $t \in (a, b)$

**Lemma: (line segment principle).**

Let $C \subset \mathbb{R}^n$ be nonempty convex. Suppose that $x \in \mathrm{ri}(C)$, $y \in \mathrm{cl}(C)$, and let $\lambda \in (0, 1]$. Then $\lambda x + (1 - \lambda)y \in \mathrm{ri}(C)$.

## Theorem: (first-order characterizations of strong convexity).

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex. Then for a given $\sigma > 0$, the following three claims are equivalent:

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

(ii). For any $x \in \mathrm{dom}\partial f, y \in \mathrm{dom} f$ and $g \in \partial f(x)$ we have

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

(iii). For any $x, y \in \mathrm{dom}\partial f$, and $g_x \in \partial f(x), g_y \in \partial f(y)$ we have

$$\langle g_x - g_y, x - y \rangle \geq \sigma \|x - y\|^2 .$$

**Theorem: (existence and uniqueness of a minimizer of closed strongly convex functions)**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and $\sigma$-convex. Then

(a). $f$ has a unique minimizer;

(b). $f(x) - f(x^*) \geq \frac{\sigma}{2} \|x - x^*\|^2$ for all $x \in \mathrm{dom}(f)$, where $x^*$ is the unique minimizer of $f$.

**Theorem: (conjugate correspondence theorem).**

Let $\sigma > 0$. Then

(a). If $f : \mathbb{R}^n \to \mathbb{R}$ is $\frac{1}{\sigma}$-smooth and convex, then $f^*$ is $\sigma$-strongly convex with respect to the dual norm $\| \ \|_*$.

(b). If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper closed and $\sigma$-strongly convex, then $f^* : \mathbb{R}^n \to \mathbb{R}$ is $\frac{1}{\sigma}$-smooth w.r.t the dual norm $\| \ \|_*$.

**Theorem: (smoothness of the infimal convolution).**

Let $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let real-valued function $w:\mathbb{R}^n \to \mathbb{R}$ be $L$-smooth and convex. Assume that $f\Box w$ is real-valued. Then the following hold:

(a). $f\Box w$ is $L$-smooth.

(b). Let $x \in \mathbb{R}^n$, and assume that $u(x)$ is a minimizer of

$$\min_u \{f(u) + w(x-u)\}.$$

Then $\nabla(f\Box w)(x) = \nabla w\,(x - u(x))$.