Convex Function and Strong Convexity

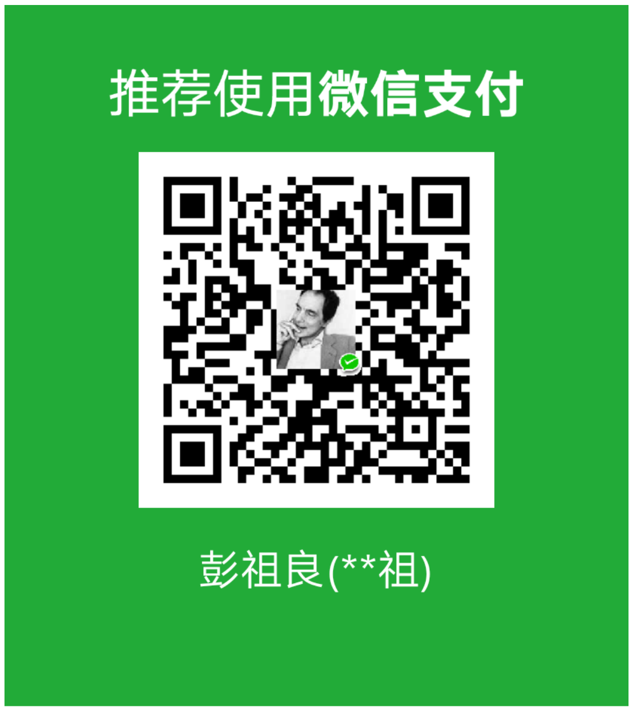
Liangzu Peng
School of Information Science and Technology
ShanghaiTech University
penglz@shanghaitech.edu.cn

April 25, 2018

Preface

What documented here is an incomplete summary, as well as proofs absent in many mainstream textbooks, of the properties of strong convexity. Feel free to use, and be willing to pay (`no pay no gay`).

A final disclaimer: due to the ignorance and inability of the author, this note is by no means rigorous, and errors occur in different places and in various ways. That said, it would not trouble the interested readers with curious mind.



微信支付

Contents

Convex Function and Strong Convexity

For the sake of simplicity, we assume

1. that it is always the case that $\theta_1 \geq 0, \theta_2 \geq 0$, and $\theta_1 + \theta_2 = 1$, and
2. that dom f is convex.

1.1 Convex Function

Definition 1.1.1 (convex function)**.** A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be convex if for all $x_1, x_2 \in \operatorname{dom} f$, we have

$$f(\theta_1 x_1 + \theta_2 x_2) \leq \theta_1 f(x_1) + \theta_2 f(x_2). \tag{1.1}$$

Definition 1.1.2 (functions restricted to a line)**.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function, $x \in \operatorname{dom} f$, and $v \in \mathbb{R}^d$. Then the function

$$g_{x,v}(\lambda) = f(x + \lambda v),$$

where the domain of $g_{x,v}$ is

$$\operatorname{dom} g_{x,v} = \{\lambda \in \mathbb{R} : x + \lambda v \in \operatorname{dom} f\},$$

is said to be the restriction of f on the line $l = \{x + \lambda v : \lambda \in \mathbb{R}\}$.

Proposition 1.1.3 (convexity of a function and its restrictions)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if for any $x \in \operatorname{dom} f$ and $v \in \mathbb{R}^d$, the restriction $g_{x,v}$ is convex.*

Proof. Suppose f is convex. We want to prove that for any $x \in \operatorname{dom} f$ and $v \in \mathbb{R}^d$ we have

$$g_{x,v}(\theta_1 \lambda_1 + \theta_2 \lambda_2) \leq \theta_1 g_{x,v}(\lambda_1) + \theta_2 g_{x,v}(\lambda_2).$$

This is equivalent to

$$f(\theta_1(x + \lambda_1 v) + \theta_2(x + \lambda_2 v)) \leq \theta_1 f(x + \lambda_1 v) + \theta_2 f(x + \lambda_2 v),$$

which follows from the convexity of f.

On the other hand, suppose

$$g_{x,v}(\theta_1 \lambda_1 + \theta_2 \lambda_2) \leq \theta_1 g_{x,v}(\lambda_1) + \theta_2 g_{x,v}(\lambda_2)$$

for any $x \in \operatorname{dom} f$, $v \in \mathbb{R}^d$, and $\lambda_1, \lambda_2 \in \operatorname{dom} g_{x,v}$. That is,

$$f(\theta_1(x + \lambda_1 v) + \theta_2(x + \lambda_2 v)) \leq \theta_1 f(x + \lambda_1 v) + \theta_2 f(x + \lambda_2 v)$$

for any $x \in \operatorname{dom} f$ and $v \in \mathbb{R}^d$. For any $x_1, x_2 \in \operatorname{dom} f$, let $x = x_1, v = x_2 - x_1$ and let $\lambda_1 = 0, \lambda_2 = 1$ (note that $\lambda_1, \lambda_2 \in \operatorname{dom} g_{x,v}$). Then we have

$$f(\theta_1 x_1 + \theta_2 x_2) \leq \theta_1 f(x_1) + \theta_2 f(x_2).$$

This implies f is convex. $\square$

Proposition 1.1.4 (first order conditions). *A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if*

$$f(x_1) - f(x_2) \geq \langle \nabla f(x_2), x_1 - x_2 \rangle \tag{1.2}$$

holds for all $x_1, x_2 \in \operatorname{dom} f$.

Proof. Suppose f is convex and $\theta_1 \neq 0$. Then we have for all $x_1, x_2 \in \operatorname{dom} f$,

$$f(x_2 + \theta_1(x_1 - x_2)) \leq f(x_2) + \theta_1(f(x_1) - f(x_2)).$$

This implies

$$f(x_1) - f(x_2) \geq \frac{f(x_2 + \theta_1(x_1 - x_2)) - f(x_2)}{\theta_1}.$$

By taking the limit of above equation, we have

$$f(x_1) - f(x_2) \geq \lim_{\theta_1 \to 0} \frac{f(x_2 + \theta_1(x_1 - x_2)) - f(x_2)}{\theta_1} = \langle \nabla f(x_2), x_1 - x_2 \rangle,$$

where the last equality holds because of the definition of directional derivative.

On the other hand, Suppose Eq. 1.2 holds for all $x_1, x_2 \in \operatorname{dom} f$. Let $c = \theta_1 x_1 + \theta_2 x_2$, then we have

$$f(x_1) - f(c) \geq \langle \nabla f(c), x_1 - c \rangle \Rightarrow \theta_1 f(x_1) - \theta_1 f(c) \geq \langle \nabla f(c), \theta_1(x_1 - c) \rangle$$

and

$$f(x_2) - f(c) \geq \langle \nabla f(c), x_2 - c \rangle \Rightarrow \theta_2 f(x_2) - \theta_2 f(c) \geq \langle \nabla f(c), \theta_2(x_2 - c) \rangle.$$

This implies $\theta_1 f(x_1) + \theta_2 f(x_2) \geq f(\theta_1 x_1 + \theta_2 x_2)$.

$\square$

Lemma 1.1.5 (convexity conditions for single-variable functions). *A function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if one of the following holds:*

1. $(f'(x) - f'(x_2))(x_1 - x_2) \geq 0$.

2. $f''(x) \geq 0$.

Proof. Note that conditions (1) and (2) are equivalent. It is enough to show that the convexity of f implies (1) and follows from (2).

Suppose f is convex and thus Eq. 1.2 holds. Then for any $x_1, x_2 \in \operatorname{dom} f$ with $x_1 \neq x_2$, we have

$$\begin{aligned}
&\begin{cases} f(x_1) - f(x_2) \geq f'(x_2)(x_1 - x_2) \\ f(x_2) - f(x_1) \geq f'(x_1)(x_2 - x_1) \end{cases} \\
\Rightarrow &(f'(x_1) - f'(x_2))(x_1 - x_2) \geq 0 \\
\Rightarrow &\frac{f'(x_1) - f'(x_2)}{x_1 - x_2} \geq 0.
\end{aligned} \tag{1.3}$$

This also means that $f''(x) \geq 0$ for each $x \in \operatorname{dom} f$.

On the other hand, suppose Eq. 1.6 holds. By Taylor's Theorem, for each $x_1, x_2 \in \operatorname{dom} f \subset \mathbb{R}$, we have

$$f(x_2) = f(x_1) + f'(x_1)(x_2 - x_1) + \frac{1}{2} f''(c)(x_2 - x_1)^2$$

for some c between x_1 and x_2. This implies Eq. 1.2 holds, and thus f is convex. □

Proposition 1.1.6 (monotone gradient condition)**.** *A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if*

$$\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \geq 0 \tag{1.4}$$

holds for all $x_1, x_2 \in \operatorname{dom} f$.

Proof.

1. ($d = 1$). It has been proved in Lemma 1.1.5.

2. ($d > 1$). For any $x \in \operatorname{dom} f$ and $v \in \mathbb{R}^d$, let $g_{x,v}(\lambda) = f(x + \lambda v)$ be a restriction of f on the line. Then it is enough to show that $g_{x,v}$ is convex if and only if $g_{x,v}$ is convex if and only if

$$\begin{aligned}
(g'_{x,v}(\lambda_1) - g'_{x,v}(\lambda_2))(\lambda_1 - \lambda_2) &= \langle \nabla f(x + \lambda_1 v) - \nabla f(x + \lambda_2 v), v \rangle (\lambda_1 - \lambda_2) \\
&= \langle \nabla f(x + \lambda_1 v) - \nabla f(x + \lambda_2 v), (\lambda_1 - \lambda_2) v \rangle \\
&= \langle \nabla f(x + \lambda_1 v) - \nabla f(x + \lambda_2 v), (x + \lambda_1 v) - (x + \lambda_2 v) \rangle \\
&\geq 0
\end{aligned} \tag{1.5}$$

for any $\lambda_1, \lambda_2 \in \operatorname{dom} g$, which follows from the case $d = 1$.

□

Proposition 1.1.7 (second order conditions)**.** *A twice differentiable function* $f : \mathbb{R}^d \to \mathbb{R}$ *is convex if and only if*

$$\nabla^2 f(x) \succeq 0 \tag{1.6}$$

for all $x \in \operatorname{dom} f$.

Proof.

1. ($d = 1$). It has been proved in Lemma 1.1.5.

2. ($d > 1$). For any $x \in \operatorname{dom} f$ and $v \in \mathbb{R}^d$, let $g_{x,v}(\lambda) = f(x + \lambda v)$ be a restriction of f on the line. Then it is enough to show that $g_{x,v}$ is convex if and only if

$$g_v''(\lambda) = v^T \nabla^2 f(x + \lambda v) v \geq 0$$

for any $\lambda \in \operatorname{dom} g$, which follows from the case $d = 1$.

□

Remark 1.1.8. For a twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, the following are equivalent.

1. f is μ-strongly convex.

2. $f(\theta_1 x_1 + \theta_2 x_2) \leq \theta_1 f(x_1) + \theta_2 f(x_2)$ for all $x_1, x_2 \in \operatorname{dom} f$ (Definition 1.1.1).

3. $f(x_1) - f(x_2) \geq \langle \nabla f(x_2), x_1 - x_2 \rangle$ for all $x_1, x_2 \in \operatorname{dom} f$ (Proposition 1.1.4).

4. $\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \geq 0$ for all $x_1, x_2 \in \operatorname{dom} f$ (Proposition 1.1.6).

5. $\nabla^2 f(x) \succeq 0$ for all $x_1, x_2 \in \operatorname{dom} f$ (Proposition 1.1.7).

1.2 Strong Convexity

Definition 1.2.1 (strongly convex function)**.** A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be strongly convex with constant $\mu > 0$, or μ-strongly convex, if the function $f(x) - \frac{\mu}{2} \|x\|_2^2$ is convex.

Remark 1.2.2. The definition of strong convexity immediately implies that $f : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex if and only if

$$\begin{aligned} & f(\theta_1 x_1 + \theta_2 x_2) - \frac{\mu}{2} \|\theta_1 x_1 + \theta_2 x_2\|_2^2 \le \theta_1 (f(x_1) - \frac{\mu}{2} \|x_1\|_2^2) + (\theta_2 f(x_2) - \frac{\mu}{2} \|x_2\|_2^2) \\ \iff & f(\theta_1 x_1 + \theta_2 x_2) \le \theta_1 f(x_1) + \theta_2 f(x_2) - \frac{\theta_1 \theta_2 \mu}{2} \|x_1 - x_2\|_2^2 \end{aligned} \tag{1.7}$$

for any $x_1, x_2 \in \operatorname{dom} f$.

Remark 1.2.3 (first order condition). Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable. Then by Proposition 1.1.4, we have that $f : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex if and only if

$$\begin{aligned} & (f(x_1) - \frac{\mu}{2} \|x_1\|_2^2) - (f(x_2) - \frac{\mu}{2} \|x_2\|_2^2) \ge \langle \nabla f(x_2 - \frac{\mu}{2} \|x_2\|_2^2), x_1 - x_2 \rangle \\ \iff & f(x_1) - f(x_2) \ge \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\mu}{2} \|x_1 - x_2\|_2^2, \end{aligned} \tag{1.8}$$

for all $x_1, x_2 \in \operatorname{dom} f$.

Remark 1.2.4 (monotone gradient condition). Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable. Then by Proposition 1.1.6, we have that $f : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex if and only if

$$\begin{aligned} & \langle \nabla (f(x_1) - \frac{\mu}{2} \|x_1\|_2^2) - \nabla (f(x_2) - \frac{\mu}{2} \|x_2\|_2^2)), x_1 - x_2 \rangle \ge 0 \\ \iff & \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \ge \mu \|x_1 - x_2\|_2^2 \end{aligned} \tag{1.9}$$

for all $x_1, x_2 \in \operatorname{dom} f$.

Remark 1.2.5 (second order condition). Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable. Then by Proposition 1.1.7, we have that $f : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex if and only if

$$\begin{aligned} & \nabla^2 (f(x) - \frac{\mu}{2} \|x\|_2^2) \succeq 0 \\ \iff & \nabla^2 f(x) \succeq \mu I, \end{aligned} \tag{1.10}$$

where $I \in \mathbb{R}^{d \times d}$ is the identity matrix, for all $x \in \operatorname{dom} f$.

Remark 1.2.6. For a twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, the following are equivalent.

1. f is μstrongly convex.
2. $f(\theta_1 x_1 + \theta_2 x_2) \le \theta_1 f(x_1) + \theta_2 f(x_2) - \frac{\theta_1 \theta_2 \mu}{2} \|x_1 - x_2\|_2^2$ (Remark 1.2.2).
3. $f(x_1) - f(x_2) \ge \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\mu}{2} \|x_1 - x_2\|_2^2$ (Remark 1.2.3).
4. $\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \ge \mu \|x_1 - x_2\|_2^2$ (Remark 1.2.4).
5. $\nabla^2 f(x) \succeq \mu I$ (Remark 1.2.5).

1.3 Consequences of Strong Convexity

The main reference for this subsection is this blog post.

Proposition 1.3.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be μ-strongly convex. Then the following conditions hold.*

1. $\|\nabla f(x_1) - \nabla f(x_2)\|_2 \geq \mu \|x_1 - x_2\|_2$ *for any* $x_1, x_2 \in \operatorname{dom} f$.
2. $f(x_1) - f(x_2) \leq \frac{1}{2\mu} \|\nabla f(x_1)\|_2^2$, *for any* $x_1, x_2 \in \operatorname{dom} f$. *Specifically,*
$$f(x) - f(x^*) \leq \frac{1}{2\mu} \|\nabla f(x)\|_2^2,$$
where x^* is a minimizer of f, for any $x \in \operatorname{dom} f$.
3. $f(x_1) - f(x_2) \leq \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{1}{2\mu} \|\nabla f(x_1) - \nabla f(x_2)\|_2^2$ *for all* $x_1, x_2 \in \operatorname{dom} f$.
4. $\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \leq \frac{1}{\mu} \|\nabla f(x_1) - \nabla f(x_2)\|_2^2$.

Proof.

1. For any $x_1, x_2 \in \operatorname{dom} f$ with $x_1 \neq x_2$, we have
$$\|\nabla f(x_1) - \nabla f(x_2)\|_2 \|x_1 - x_2\|_2 \geq \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \overset{\text{Eq. 1.9}}{\geq} \mu \|x_1 - x_2\|_2^2.$$
2. For any $x_1, x_2 \in \operatorname{dom} f$, we have
$$\begin{aligned} f(x_1) - f(x_2) &\overset{\text{Eq. 1.8}}{\leq} \langle \nabla f(x_1), x_1 - x_2 \rangle - \frac{\mu}{2} \|x_1 - x_2\|_2^2 \\ &= \frac{1}{2\mu} \|\nabla f(x_1)\|_2^2 - \frac{1}{2\mu} \|\nabla f(x_1) - \mu(x_1 - x_2)\|_2^2 \\ &\leq \frac{1}{2\mu} \|\nabla f(x_1)\|_2^2. \end{aligned} \tag{1.11}$$
3. For any $x_2 \in \operatorname{dom} f$, the function $h(x_1) = f(x_1) - \langle \nabla f(x_2), x_1 \rangle$ is μ-strongly convex (proved by applying Eq. 1.9). Then apply the condition 2 will result the condition 3.
4. For any $x_1, x_2 \in \operatorname{dom} f$, we have
$$\begin{aligned} \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle &\leq \|\nabla f(x_1) - \nabla f(x_2)\|_2 \|x_1 - x_2\|_2 \\ &\overset{1.}{\leq} \frac{1}{\mu} \|\nabla f(x_1) - \nabla f(x_2)\|_2^2 \end{aligned} \tag{1.12}$$

□