Convexity, Gradient Descent, Convergence Guarantees

Avishek Ghosh
Assistant Professor, Dept. of CSE.,IIT Bombay

IIT Bombay

August 3, 2025

Outline

1. Convexity and Smoothness

2. Strong Convexity and Examples

3. Gradient Descent and Decent lemma for Smooth functions

Convex set

Convex set

Definition (Convex Set)

A set $\mathcal{X} \subseteq \mathbb{R}^n$ is *convex* if for any points $x, y \in \mathcal{X}$ and any scalar $\lambda \in [0, 1]$, the combination

$$\lambda x + (1 - \lambda)\, y \ \in \ \mathcal{X}.$$

Convex set

Definition (Convex Set)

A set $\mathcal{X} \subseteq \mathbb{R}^n$ is *convex* if for any points $x, y \in \mathcal{X}$ and any scalar $\lambda \in [0, 1]$, the combination

$$\lambda x + (1 - \lambda) y \in \mathcal{X}.$$

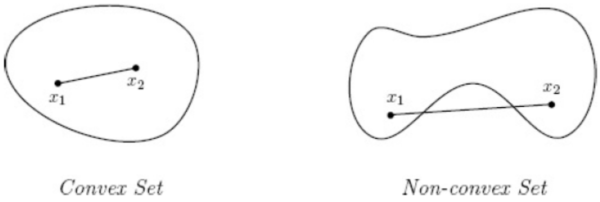


Figure: Convex and Non-convex set.

Convex Function

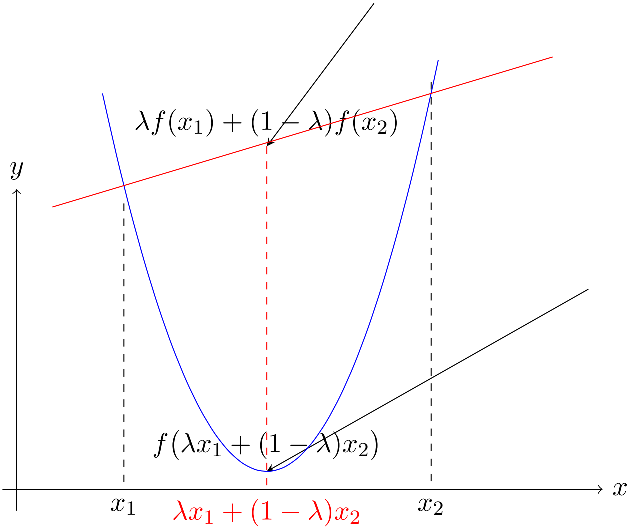
Convex Function

Definition (Convex Function)

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f : \mathcal{X} \to \mathbb{R}$ is called *convex* if, for all $x, y \in \mathcal{X}$ and every $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda) y) \ \leq \ \lambda f(x) + (1 - \lambda) f(y).$$

Illustration of Convex function



First order Characterization of Convex Functions

First order Characterization of Convex Functions

Theorem

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be differentiable. Then f is convex if and only if for all $x, y \in \mathcal{X}$,

First order Characterization of Convex Functions

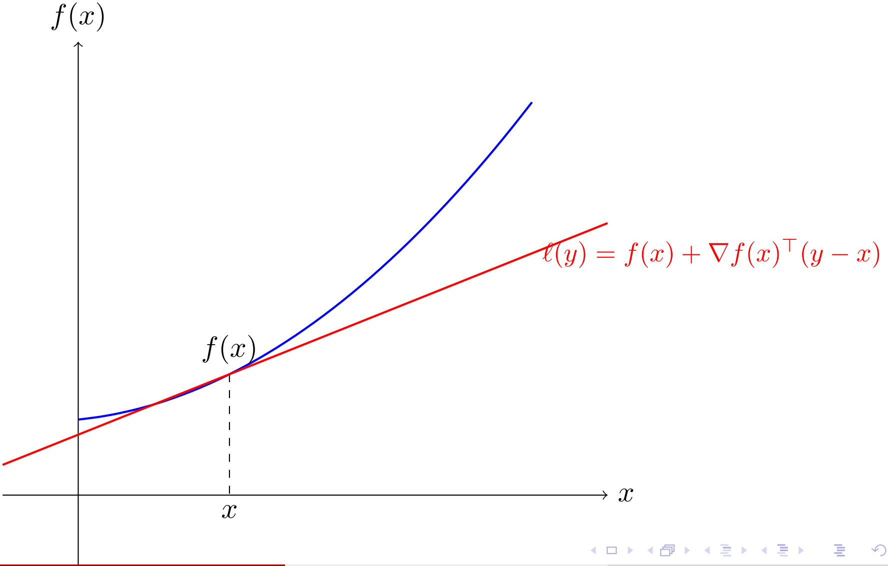
Theorem

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be differentiable. Then f is convex if and only if for all $x, y \in \mathcal{X}$,

$$f(y) \ \geq \ f(x) \ + \ \nabla f(x)^\top (y - x).$$

Proof: Exercise!

Figure of First order charecterization of Convexity



Second order Characterization of Convex Functions

Second order Characterization of Convex Functions

Theorem

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be differentiable. Then f is convex if and only if for all $x, y \in \mathcal{X}$,

Second order Characterization of Convex Functions

Theorem

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be differentiable. Then f is convex if and only if for all $x, y \in \mathcal{X}$,

$$\nabla^2 f(x) \succeq 0$$

This means the Hessian matrix $\nabla^2 f(x)$ is Positive Semi definite (meaning that the eigen values are non-negative) **Examples?**

If $d = 1$, implies $f''(x) \geq 0$ for all x

Second order Characterization of Convex Functions

Theorem

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be differentiable. Then f is convex if and only if for all $x, y \in \mathcal{X}$,

$$\nabla^2 f(x) \succeq 0$$

This means the Hessian matrix $\nabla^2 f(x)$ is Positive Semi definite (meaning that the eigen values are non-negative) **Examples?**

If $d = 1$, implies $f''(x) \geq 0$ for all x

Proof of Theorem: Exercise!

Local and Global Minima of a Function

Definition

Let $f : \mathcal{X} \to \mathbb{R}$. A point $x^* \in \mathcal{X}$ is

- **Local minimum** if there exists an open neighborhood U of x^* such that

$$f(x^*) \leq f(y) \quad \forall\, y \in U.$$

Local and Global Minima of a Function

Definition

Let $f : \mathcal{X} \to \mathbb{R}$. A point $x^* \in \mathcal{X}$ is

- **Local minimum** if there exists an open neighborhood U of x^* such that
$$f(x^*) \leq f(y) \quad \forall\, y \in U.$$
- **Global minimum** if
$$f(x^*) \leq f(y) \quad \forall\, y \in \mathcal{X}.$$

Local and Global Minima of a Function

Definition

Let $f : \mathcal{X} \to \mathbb{R}$. A point $x^* \in \mathcal{X}$ is

- **Local minimum** if there exists an open neighborhood U of x^* such that

$$f(x^*) \leq f(y) \quad \forall\, y \in U.$$

- **Global minimum** if

$$f(x^*) \leq f(y) \quad \forall\, y \in \mathcal{X}.$$

Question: Prove that, for convex functions, all local minima are global minima.

Smooth Functions

Smooth Functions

Definition (Smooth Function)

Suppose $f : \mathbb{R}^d \to \mathbb{R}$ satisfies

$$||\nabla f(x) - \nabla f(y)||_2 \leq M||x - y||_2$$

Where $M > 0$, then we call f as M-smooth/ M-Gradient Lipschitz.

Smooth Functions

Definition (Smooth Function)

Suppose $f : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies

$$||\nabla f(x) - \nabla f(y)||_2 \leq M||x - y||_2$$

Where $M > 0$, then we call f as M-smooth/ M-Gradient Lipschitz.

Example: Quadratic $f(x) = \frac{1}{2}x^T Qx - b^T x$; Q symmetric

Smooth Functions

Definition (Smooth Function)

Suppose $f: \mathbb{R}^d \to \mathbb{R}$ satisfies

$$||\nabla f(x) - \nabla f(y)||_2 \leq M||x - y||_2$$

Where $M > 0$, then we call f as M-smooth/ M-Gradient Lipschitz.

Example: Quadratic $f(x) = \frac{1}{2}x^T Qx - b^T x$; Q symmetric

$$\begin{aligned} \|\nabla f(x) - \nabla f(y)\| &= ||(Qx - b) - (Qy - b)|| \\ &= ||Q(x - y)|| \leq M||x - y|| \end{aligned}$$

Smooth Functions

Definition (Smooth Function)

Suppose $f : \mathbb{R}^d \to \mathbb{R}$ satisfies

$$||\nabla f(x) - \nabla f(y)||_2 \leq M||x - y||_2$$

Where $M > 0$, then we call f as M-smooth/ M-Gradient Lipschitz.

Example: Quadratic $f(x) = \frac{1}{2}x^T Qx - b^T x$; Q symmetric

$$\begin{aligned} \|\nabla f(x) - \nabla f(y)\| &= ||(Qx - b) - (Qy - b)|| \\ &= ||Q(x - y)|| \leq M||x - y|| \end{aligned}$$

Hence Quadratic objective function is M-smooth.

Equivalent Characterizations of Smoothness

Equivalent Characterizations of Smoothness

Theorem

The following statements are equivalent ($\forall x, y$):

1. $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$

Equivalent Characterizations of Smoothness

Theorem

The following statements are equivalent ($\forall x, y$):

1. $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$
2. $\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq M\|x - y\|^2$

Equivalent Characterizations of Smoothness

Theorem

The following statements are equivalent ($\forall x, y$):

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

Equivalent Characterizations of Smoothness

Theorem

The following statements are equivalent ($\forall x, y$):

1. $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$
2. $\langle \nabla f(x) - \nabla f(y), x - y\rangle \leq M\|x - y\|^2$
3. $f(y) - \{f(x) + \langle \nabla f(x), y - x\rangle\} \leqslant \frac{M}{2}\|x - y\|^2$
4. $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{M}\|\nabla f(x) - \nabla f(y)\|^2$

Equivalent Characterizations of Smoothness

Theorem

The following statements are equivalent ($\forall x, y$):

1. $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$
2. $\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq M\|x - y\|^2$
3. $f(y) - \{f(x) + \langle \nabla f(x), y - x \rangle\} \leqslant \frac{M}{2}\|x - y\|^2$
4. $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{M}\|\nabla f(x) - \nabla f(y)\|^2$
5. f *is twice differentiable then:*

$$-M\mathbb{I} \preccurlyeq \nabla^2 f \preccurlyeq M\mathbb{I}$$

where $M\mathbb{I} - \nabla^2 f$ *is P.S.D. matrix.*

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

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

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

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

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

$$\begin{aligned}\langle \nabla f(x) - \nabla f(y), x - y \rangle &\leq \|\nabla f(x) - \nabla f(y)\| \, \|x - y\| \\ &\leq M \, \|x - y\|^2\end{aligned}$$

For $(2) \Rightarrow (3)$, we use the Interpolation method where $G: \; [0,1] \to \mathbb{R}$,

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

$$\begin{aligned} \langle \nabla f(x) - \nabla f(y), x - y \rangle &\leq \| \nabla f(x) - \nabla f(y) \| \, \| x - y \| \\ &\leq M \, \| x - y \|^2 \end{aligned}$$

For $(2) \Rightarrow (3)$, we use the Interpolation method where $G: \ [0,1] \rightarrow \mathbb{R}$,

$$G(t) = f(x + t(y - x)) - f(x) - \langle \nabla f(x), t(y - x) \rangle$$

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

$$\begin{aligned}\langle \nabla f(x) - \nabla f(y), x - y \rangle &\leq \|\nabla f(x) - \nabla f(y)\| \, \|x - y\| \\ &\leq M \, \|x - y\|^2\end{aligned}$$

For $(2) \Rightarrow (3)$, we use the Interpolation method where $G : \; [0,1] \to \mathbb{R}$,

$$G(t) = f(x + t(y - x)) - f(x) - \langle \nabla f(x), t(y - x) \rangle$$

Observe that $G(1) =$ LHS $of(3)$ and $G(0) = 0$.

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

$$\begin{aligned}\langle \nabla f(x) - \nabla f(y), x - y \rangle &\leq \|\nabla f(x) - \nabla f(y)\| \, \|x - y\| \\ &\leq M \, \|x - y\|^2\end{aligned}$$

For $(2) \Rightarrow (3)$, we use the Interpolation method where $G: \ [0,1] \rightarrow \mathbb{R}$,

$$G(t) = f(x + t(y - x)) - f(x) - \langle \nabla f(x), t(y - x) \rangle$$

Observe that $G(1) =$ LHS $of(3)$ and $G(0) = 0$.
By the fundamental theorem of calculus,

Proof of theorem

For $(1) \Rightarrow (2)$, Use Cauchy-Schwartz Ineqality

$$\begin{aligned}\langle \nabla f(x) - \nabla f(y), x - y \rangle &\leq \|\nabla f(x) - \nabla f(y)\| \, \|x - y\| \\ &\leq M \, \|x - y\|^2\end{aligned}$$

For $(2) \Rightarrow (3)$, we use the Interpolation method where $G: \ [0,1] \to \mathbb{R}$,

$$G(t) = f(x + t(y - x)) - f(x) - \langle \nabla f(x), t(y - x) \rangle$$

Observe that $G(1) =$ LHS $of(3)$ and $G(0) = 0$.
By the fundamental theorem of calculus,

$$G(1) - G(0) = \int_0^1 G'(t) dt$$

Proof Contd..

$$\text{LHS of } (3) = \int_0^1 \langle \nabla f(x + t(y-x)), (y-x) \rangle - 0 - \langle \nabla f(x), y - x \rangle \, dt$$

Proof Contd..

$$\begin{aligned}\text{LHS of } (3) &= \int_0^1 \langle \nabla f(x + t(y-x)), (y-x)\rangle - 0 - \langle \nabla f(x), y - x\rangle \, dt \\ &= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x\rangle \, dt\end{aligned}$$

Proof Contd..

$$
\begin{aligned}
\text{LHS of } (3) &= \int_0^1 \langle \nabla f(x + t(y-x)), (y-x) \rangle - 0 - \langle \nabla f(x), y - x \rangle \, dt \\
&= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x \rangle \, dt \\
&= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), t(y - x) \rangle \frac{dt}{t}
\end{aligned}
$$

Proof Contd..

$$
\begin{aligned}
\text{LHS of } (3) &= \int_0^1 \langle \nabla f(x + t(y-x)), (y-x) \rangle - 0 - \langle \nabla f(x), y - x \rangle \; dt \\
&= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x \rangle \; dt \\
&= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), t(y - x) \rangle \frac{dt}{t}
\end{aligned}
$$

Using (2),

Proof Contd..

$$\text{LHS of } (3) \;= \int_0^1 \langle \nabla f(x+t(y-x)), (y-x) \rangle - 0 - \langle \nabla f(x), y-x \rangle \; dt$$

$$= \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x \rangle \; dt$$

$$= \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), t(y-x) \rangle \frac{dt}{t}$$

Using (2),

$$G(1) - G(0) \leq \int_0^1 Mt||x-y||^2 dt$$

Proof Contd..

$$\begin{aligned}
\text{LHS of } (3) &= \int_0^1 \langle \nabla f(x + t(y - x)), (y - x) \rangle - 0 - \langle \nabla f(x), y - x \rangle \, dt \\
&= \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle \, dt \\
&= \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), t(y - x) \rangle \frac{dt}{t}
\end{aligned}$$

Using (2),

$$\begin{aligned}
G(1) - G(0) &\leq \int_0^1 Mt||x - y||^2 dt \\
&= M||x - y||^2 \int_0^1 t \, dt = \frac{M}{2}||x - y||^2
\end{aligned}$$

Proof Contd..

$$\text{LHS of } (3) = \int_0^1 \langle \nabla f(x + t(y-x)), (y-x) \rangle - 0 - \langle \nabla f(x), y-x \rangle \, dt$$
$$= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y-x \rangle \, dt$$
$$= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), t(y-x) \rangle \frac{dt}{t}$$

Using (2),

$$G(1) - G(0) \leq \int_0^1 Mt||x-y||^2 dt$$
$$= M||x-y||^2 \int_0^1 t \, dt = \frac{M}{2}||x-y||^2$$

For $(3) \Rightarrow (5)$, we utilize the Taylor Series Expansion.

Proof Contd..

$$\begin{aligned}
\text{LHS of } (3) &= \int_0^1 \langle \nabla f(x + t(y-x)), (y-x) \rangle - 0 - \langle \nabla f(x), y - x \rangle \, dt \\
&= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x \rangle \, dt \\
&= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), t(y - x) \rangle \frac{dt}{t}
\end{aligned}$$

Using (2),

$$\begin{aligned}
G(1) - G(0) &\leq \int_0^1 Mt||x-y||^2 dt \\
&= M||x-y||^2 \int_0^1 t \, dt = \frac{M}{2}||x-y||^2
\end{aligned}$$

For $(3) \Rightarrow (5)$, we utilize the Taylor Series Expansion.
For $(4) \Rightarrow (1)$, we utilize the Cauchy Schwartz Inequality.

Srong Convexity

Definition (Srong Convexity)

Definition (Srong Convexity)

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set. A function $f : \mathcal{C} \to \mathbb{R}$ is called m-*strongly convex* if there exists $m > 0$ such that for all $x, y \in \mathcal{C}$ and every $\lambda \in [0, 1]$,

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

Remark

Srong Convexity

Definition (Srong Convexity)

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set. A function $f : \mathcal{C} \to \mathbb{R}$ is called m-*strongly convex* if there exists $m > 0$ such that for all $x, y \in \mathcal{C}$ and every $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \;\leq\; \lambda f(x) \;+\; (1-\lambda) f(y) \;-\; \frac{m}{2}\lambda(1-\lambda)\, \|x-y\|^2.$$

Remark

- **Convexity:** Ensures for any $x, y \in \mathcal{C}$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \;\leq\; \lambda f(x) + (1-\lambda) f(y).$$

The graph lies below the chord connecting $(x, f(x))$ and $(y, f(y))$.

Remarks

- m-**Strong Convexity:** Strengthens convexity by quantifying the gap with a quadratic term:

Remarks

- m-**Strong Convexity:** Strengthens convexity by quantifying the gap with a quadratic term:

$$f\big(\lambda x + (1-\lambda)y\big) \;\leq\; \lambda\, f(x) + (1-\lambda)\, f(y) - \tfrac{m}{2}\,\lambda(1-\lambda)\, \|x-y\|^2.$$

Remarks

- m-**Strong Convexity:** Strengthens convexity by quantifying the gap with a quadratic term:

$$f\left(\lambda x+(1-\lambda)y\right) \;\leq\; \lambda\, f(x)+(1-\lambda)\, f(y)-\tfrac{m}{2}\,\lambda(1-\lambda)\,\|x-y\|^2.$$

- **Interpretation:** Strong convexity not only bounds f below its secant, but guarantees a minimum separation (proportional to $\|x-y\|^2$) between the function's graph and that chord (see below Figure).

Figure: Interpretation of m-strongly convex Function

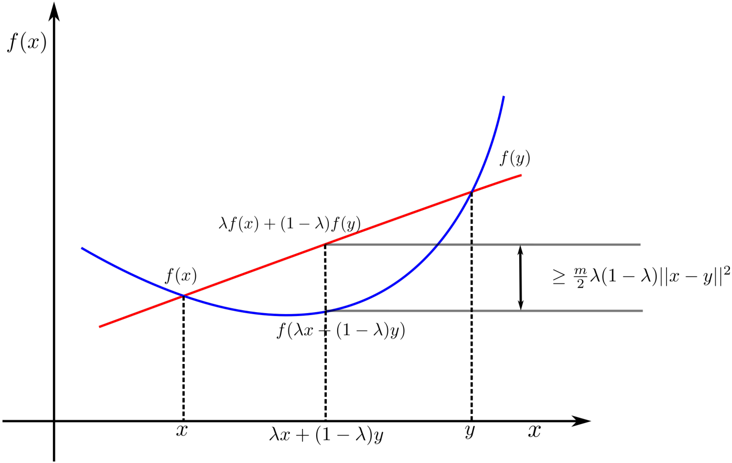


Figure: $m-$strongly convex function f

Equivalent Characterizations of Strong Convexity

Theorem

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set and $f : \mathcal{C} \to \mathbb{R}$ be continuously differentiable. Then the following are equivalent:

a. *f is m-strongly convex.*

Equivalent Characterizations of Strong Convexity

Theorem

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set and $f : \mathcal{C} \to \mathbb{R}$ be continuously differentiable. Then the following are equivalent:

a. f *is m-strongly convex.*
b. $\|\nabla f(x) - \nabla f(y)\| \;\geq\; m\,\|x - y\|$ *for all $x, y \in \mathcal{C}$.*

Theorem

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set and $f : \mathcal{C} \to \mathbb{R}$ be continuously differentiable. Then the following are equivalent:

- (a) f *is* m*-strongly convex.*
- (b) $\|\nabla f(x) - \nabla f(y)\| \;\geq\; m\,\|x - y\|$ *for all* $x, y \in \mathcal{C}$.
- (c) $(\nabla f(x) - \nabla f(y))^\top (x - y) \;\geq\; m\,\|x - y\|^2$ *for all* $x, y \in \mathcal{C}$.

Equivalent Characterizations of Strong Convexity

Theorem

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set and $f : \mathcal{C} \to \mathbb{R}$ be continuously differentiable. Then the following are equivalent:

a. f *is m-strongly convex.*
b. $\|\nabla f(x) - \nabla f(y)\| \ \geq \ m\,\|x - y\|$ *for all* $x, y \in \mathcal{C}$.
c. $(\nabla f(x) - \nabla f(y))^\top (x - y) \ \geq \ m\,\|x - y\|^2$ *for all* $x, y \in \mathcal{C}$.
d. $f(y) - (f(x) + \nabla f(x)^\top (y - x)) \ \geq \ \frac{m}{2}\,\|x - y\|^2$ *for all* $x, y \in \mathcal{C}$.

Equivalent Characterizations of Strong Convexity

Theorem

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set and $f : \mathcal{C} \to \mathbb{R}$ be continuously differentiable. Then the following are equivalent:

- (a) f *is* m*-strongly convex.*
- (b) $\|\nabla f(x) - \nabla f(y)\| \geq m\|x - y\|$ *for all* $x, y \in \mathcal{C}$.
- (c) $(\nabla f(x) - \nabla f(y))^\top (x - y) \geq m\|x - y\|^2$ *for all* $x, y \in \mathcal{C}$.
- (d) $f(y) - (f(x) + \nabla f(x)^\top (y - x)) \geq \frac{m}{2}\|x - y\|^2$ *for all* $x, y \in \mathcal{C}$.
- (e) $(\nabla f(x) - \nabla f(y))^\top (x - y) \leq \frac{1}{m}\|\nabla f(x) - \nabla f(y)\|^2$ *for all* $x, y \in \mathcal{C}$.

Equivalent Characterizations of Strong Convexity

Theorem

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a nonempty convex set and $f : \mathcal{C} \to \mathbb{R}$ be continuously differentiable. Then the following are equivalent:

- (a) f *is* m*-strongly convex.*
- (b) $\|\nabla f(x) - \nabla f(y)\| \;\geq\; m\,\|x - y\|$ *for all* $x, y \in \mathcal{C}$.
- (c) $(\nabla f(x) - \nabla f(y))^\top (x - y) \;\geq\; m\,\|x - y\|^2$ *for all* $x, y \in \mathcal{C}$.
- (d) $f(y) - (f(x) + \nabla f(x)^\top (y - x)) \;\geq\; \frac{m}{2}\,\|x - y\|^2$ *for all* $x, y \in \mathcal{C}$.
- (e) $(\nabla f(x) - \nabla f(y))^\top (x - y) \;\leq\; \frac{1}{m}\,\|\nabla f(x) - \nabla f(y)\|^2$ *for all* $x, y \in \mathcal{C}$.
- (f) *If* f *is twice differentiable, then* $\nabla^2 f(x) \succeq mI$ *for all* $x \in \mathcal{C}$.

Example 1: Quadratic Function

Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2} x^\top Q\, x \;-\; b^\top x,$$

where $Q \in \mathbb{R}^{d \times d}$ is symmetric positive definite and $b \in \mathbb{R}^d$.

Example 1: Quadratic Function

Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2}\, x^\top Q\, x \;-\; b^\top x,$$

where $Q \in \mathbb{R}^{d \times d}$ is symmetric positive definite and $b \in \mathbb{R}^d$.

- $\nabla f(x) - \nabla f(y) = Q\,(x - y)$,

Example 1: Quadratic Function

Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2} x^\top Q\, x \; - \; b^\top x,$$

where $Q \in \mathbb{R}^{d\times d}$ is symmetric positive definite and $b \in \mathbb{R}^d$.

- $\nabla f(x) - \nabla f(y) = Q\,(x - y)$,
- Then

$$\|\nabla f(x) - \nabla f(y)\| = \|Q\,(x-y)\|$$

Example 1: Quadratic Function

Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2} x^\top Q\, x \;-\; b^\top x,$$

where $Q \in \mathbb{R}^{d \times d}$ is symmetric positive definite and $b \in \mathbb{R}^d$.

- $\nabla f(x) - \nabla f(y) = Q\,(x - y)$,
- Then

$$\|\nabla f(x) - \nabla f(y)\| = \|Q\,(x - y)\|$$

- Since $Q \succeq \lambda_{\min}(Q)\, I$, It follows

Example 1: Quadratic Function

Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2} x^\top Q\, x \; - \; b^\top x,$$

where $Q \in \mathbb{R}^{d \times d}$ is symmetric positive definite and $b \in \mathbb{R}^d$.

- $\nabla f(x) - \nabla f(y) = Q\,(x - y)$,
- Then

$$\|\nabla f(x) - \nabla f(y)\| = \|Q\,(x - y)\|$$

- Since $Q \succeq \lambda_{\min}(Q)\, I$, It follows

$$\|\nabla f(x) - \nabla f(y)\| \;\geq\; \lambda_{\min}(Q)\, \|x - y\|.$$

Example 1: Quadratic Function

Let $f \colon \mathbb{R}^d \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2}\, x^{\top} Q\, x \;-\; b^{\top} x,$$

where $Q \in \mathbb{R}^{d \times d}$ is symmetric positive definite and $b \in \mathbb{R}^d$.

- $\nabla f(x) - \nabla f(y) = Q\,(x - y)$,
- Then

$$\|\nabla f(x) - \nabla f(y)\| = \|Q\,(x - y)\|$$

- Since $Q \succeq \lambda_{\min}(Q)\, I$, It follows

$$\|\nabla f(x) - \nabla f(y)\| \;\geq\; \lambda_{\min}(Q)\, \|x - y\|.$$

Thus By Above Theorem , f is $\lambda_{\min}(Q)$-strongly convex.

Example 2: Log–Sum–Exp Function

Consider

$$f : [0, \infty) \to \mathbb{R}, \quad f(x) = \log(1 + e^x).$$

Then:

- **Gradient:**

Example 2: Log–Sum–Exp Function

Consider

$$f : [0, \infty) \to \mathbb{R}, \quad f(x) = \log(1 + e^x).$$

Then:

- **Gradient:**

$$\nabla f(x) = \frac{e^x}{1 + e^x}.$$

- **Hessian:**

Example 2: Log–Sum–Exp Function

Consider

$$f\colon [0,\infty) \to \mathbb{R}, \quad f(x) = \log(1 + e^x).$$

Then:

- **Gradient:**

$$\nabla f(x) = \frac{e^x}{1 + e^x}.$$

- **Hessian:**

$$\nabla^2 f(x) = \frac{e^x}{(1 + e^x)^2} \ \geq 0,$$

so f is convex on $[0, \infty)$.

Example 2: Log–Sum–Exp Function

Consider

$$f\colon [0,\infty) \to \mathbb{R}, \quad f(x) = \log(1+e^x).$$

Then:

- **Gradient:**
$$\nabla f(x) = \frac{e^x}{1+e^x}.$$
- **Hessian:**
$$\nabla^2 f(x) = \frac{e^x}{(1+e^x)^2} \geq 0,$$
so f is convex on $[0,\infty)$.
- **Convex but not strongly convex on $[0,\infty)$:**

Example 2: Log–Sum–Exp Function

Consider

$$f\colon [0,\infty) \to \mathbb{R}, \quad f(x) = \log(1 + e^x).$$

Then:

- **Gradient:**

$$\nabla f(x) = \frac{e^x}{1 + e^x}.$$

- **Hessian:**

$$\nabla^2 f(x) = \frac{e^x}{(1 + e^x)^2} \geq 0,$$

 so f is convex on $[0, \infty)$.

- **Convex but not strongly convex on $[0,\infty)$:** $\inf_{x \geq 0} \nabla^2 f(x) = 0$, so no $m > 0$ exists with $\nabla^2 f(x) \geq m$ for all $x \geq 0$.

Example 2: Contd..

- **Strong convexity on** $[0, \beta]$**:**

Example 2: Contd..

- **Strong convexity on $[0, \beta]$:** For any $\beta > 0$, set

$$m = \min_{x \in [0,\beta]} \nabla^2 f(x) = \frac{e^{\beta}}{(1 + e^{\beta})^2} > 0.$$

Example 2: Contd..

- **Strong convexity on $[0, \beta]$:** For any $\beta > 0$, set

$$m = \min_{x \in [0,\beta]} \nabla^2 f(x) = \frac{e^\beta}{(1+e^\beta)^2} > 0.$$

  Then $\nabla^2 f(x) \geq m$ on $[0, \beta]$, so f is m-strongly convex there.
- **Smoothness:**

Example 2: Contd..

- **Strong convexity on** $[0, \beta]$**:** For any $\beta > 0$, set

$$m = \min_{x \in [0,\beta]} \nabla^2 f(x) = \frac{e^\beta}{(1+e^\beta)^2} > 0.$$

  Then $\nabla^2 f(x) \geq m$ on $[0, \beta]$, so f is m-strongly convex there.
- **Smoothness:** $\nabla^2 f(x) \leq \nabla^2 f(0) = \frac{1}{4}$ for all $x \geq 0$, hence f is M-smooth with

$$M = \frac{1}{4}.$$

Example 3: Quartic Function on $[-1, 1]$

- Let $f : [-1, 1] \to \mathbb{R}$ be defined by $f(x) = x^4$.

Example 3: Quartic Function on $[-1, 1]$

- Let $f\colon [-1,1] \to \mathbb{R}$ be defined by $f(x) = x^4$.
- Its second derivative is
$$f''(x) = 12x^2,$$

Example 3: Quartic Function on $[-1, 1]$

- Let $f\colon [-1,1] \to \mathbb{R}$ be defined by $f(x) = x^4$.
- Its second derivative is
$$f''(x) = 12x^2,$$
so $0 \le f''(x) \le 12$ for all $x \in [-1,1]$.
- Hence f is convex and M-smooth with $M = 12$.

Example 3: Quartic Function on $[-1, 1]$

- Let $f : [-1, 1] \to \mathbb{R}$ be defined by $f(x) = x^4$.
- Its second derivative is
$$f''(x) = 12x^2,$$
so $0 \leq f''(x) \leq 12$ for all $x \in [-1, 1]$.
- Hence f is convex and M-smooth with $M = 12$.
- However, since $f''(0) = 0$, there is no $m > 0$ such that $f''(x) \geq m$ on $[-1, 1]$; thus f is not strongly convex.

Example 4: Invex Function (Non-Convex but Smooth)

- Define

$$f : [-\tfrac{1}{2}, \infty) \times [1, \infty) \to \mathbb{R}, \qquad f(x, y) = y\,(x^2 - 1)^2.$$

Example 4: Invex Function (Non-Convex but Smooth)

- Define

$$f\colon [-\tfrac{1}{2}, \infty) \times [1, \infty) \to \mathbb{R}, \qquad f(x, y) = y\,(x^2 - 1)^2.$$

- Hessian matrix

Example 4: Invex Function (Non-Convex but Smooth)

- Define

$$f\colon [-\tfrac{1}{2},\infty)\times[1,\infty)\to\mathbb{R}, \qquad f(x,y)=y\,(x^2-1)^2.$$

- Hessian matrix

$$\nabla^2 f(x,y)=\begin{pmatrix}4y\,(3x^2-1) & 4x\,(x^2-1)\\ 4x\,(x^2-1) & 0\end{pmatrix}.$$

Example 4: Invex Function (Non-Convex but Smooth)

- Define

$$f\colon [-\tfrac{1}{2}, \infty) \times [1, \infty) \to \mathbb{R}, \qquad f(x, y) = y\,(x^2 - 1)^2.$$

- Hessian matrix

$$\nabla^2 f(x, y) = \begin{pmatrix} 4y\,(3x^2 - 1) & 4x\,(x^2 - 1) \\ 4x\,(x^2 - 1) & 0 \end{pmatrix}.$$

- Since $\nabla^2 f(x, y)$ is not positive definite on the entire domain, f is *not convex*.

Example 4: Invex Function (Non-Convex but Smooth)

- Define

$$f\colon [-\tfrac{1}{2},\infty) \times [1,\infty) \to \mathbb{R}, \qquad f(x,y) = y\,(x^2-1)^2.$$

- Hessian matrix

$$\nabla^2 f(x,y) = \begin{pmatrix} 4y\,(3x^2-1) & 4x\,(x^2-1) \\ 4x\,(x^2-1) & 0 \end{pmatrix}.$$

- Since $\nabla^2 f(x,y)$ is not positive definite on the entire domain, f is *not convex*.
- $f \in C^2$ with continuous Hessian $\Rightarrow$ on any bounded subset of $[-\frac{1}{2},\infty) \times [1,\infty)$, f is M-smooth.

Gradient Descent(GD) Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} f(x)$$

.

Gradient Descent(GD) Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} f(x)$$

.

Gradient Descent Update:

Gradient Descent(GD) Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} f(x)$$

.

Gradient Descent Update:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad \alpha_k > 0 \text{ (step size)}.$$

Gradient Descent(GD) Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} f(x)$$

.

Gradient Descent Update:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad \alpha_k > 0 \text{ (step size)}.$$

- Choose α_k appropriately(for example, through line search methods)

Gradient Descent(GD) Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} f(x)$$

.

Gradient Descent Update:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad \alpha_k > 0 \text{ (step size)}.$$

- Choose α_k appropriately(for example, through line search methods)

Now from next slide ,we'll see about GD convergence results in the following cases:

- Smooth functions

Gradient Descent(GD) Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} f(x)$$

.

Gradient Descent Update:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad \alpha_k > 0 \text{ (step size)}.$$

- Choose α_k appropriately(for example, through line search methods)

Now from next slide ,we'll see about GD convergence results in the following cases:

- Smooth functions
- Smooth + Convex functions

Gradient Descent(GD) Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} f(x)$$

.

Gradient Descent Update:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad \alpha_k > 0 \text{ (step size)}.$$

- Choose α_k appropriately(for example, through line search methods)

Now from next slide ,we'll see about GD convergence results in the following cases:

- Smooth functions
- Smooth + Convex functions
- Smooth + Strongly Convex functions

References I

[1] Benjamin Recht and Stephen J. Wright.
Optimization for Data Analysis. 2022.

[2] Stephen Boyd and Lieven Vandenberghe.
Convex Optimization.
Cambridge University Press, 2004.

[3] Jorge Nocedal and Stephen J. Wright.
Numerical Optimization, 2nd Edition.
Springer, 2006.

Thank You!