IE 539: Convex Optimization **KAIST, Fall 2022**
Lecture #3: Convex functions, first-order and second-order optimality conditions
September 6, 2022
Lecturer: Dabeen Lee

1 Outline

In this lecture, we study

- Convex functions,
- First-order optimality conditions, and
- Second-order optimality condition

2 Convex functions

2.1 Definition

Definition 3.1. A function $f : \mathbb{R}^d \to \mathbb{R}$ is *convex* if the domain $\mathrm{dom}(f)$ is convex and for all $x, y \in \mathrm{dom}(f)$, we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \text{for } 0 \leq \lambda \leq 1.$$

In words, f when evaluated at a point between x and y lies below the line segment joining $f(x)$ and $f(y)$.

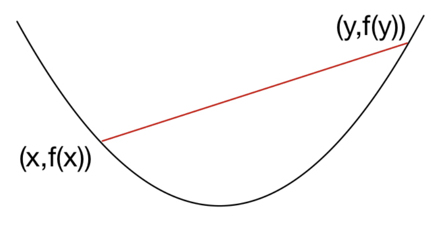

Figure 3.1: Illustration of a convex function in $\mathbb{R}^2$

Definition 3.2. We say that $f : \mathbb{R}^d \to \mathbb{R}$ is *concave* if $-f$ is convex.

Definition 3.3. A function $f : \mathbb{R}^d \to \mathbb{R}$ is

- *strictly convex* if $\mathrm{dom}(f)$ is convex and for any distinct $x, y \in \mathrm{dom}(f)$, we have

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \quad \text{for } 0 < \lambda < 1.$$

- *strongly convex* if $f(x) - \alpha\|x\|^2$ is convex for some $\alpha > 0$ and norm $\|\cdot\|$.

Note that strong convexity implies strict convexity, and strict convexity implies convexity.

2.2 Examples

Univariate functions (on $\mathbb{R}$)

- Exponential function: e^{ax} for any $a \in \mathbb{R}$.
- Power function: x^a for $a \geq 1$ over $\mathbb{R}$ and x^a for $a < 0$ over $\mathbb{R}_{++}$.
  x^a for $0 \leq a < 1$ over $\mathbb{R}_+$ is concave.
- Logarithm: $\log x$ is concave on $\mathbb{R}_{++}$.
- Negative entropy: $x \log x$ on $\mathbb{R}_{++}$.

Multivariate functions (on $\mathbb{R}^d$)

- Linear function: $a^\top x + b$ where $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ is both convex and concave.
- Quadratic function: $\frac{1}{2}x^\top Ax + b^\top x + c$ where $A \succeq 0$, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$.
- Least squares loss: $\|b - Ax\|_2^2$ for any A.
- Norm: Any norm $\|\cdot\|$ is concex, because a norm is subadditive and homogeneous.
- Maximum eigenvalue of a symmetric matrix.
- Indicator function: When C is convex, its indicator function, given by,

  $$I_C(x) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases}$$

  is convex.
- Support function: Given a convex set C, its support function is defined as

  $$I_C^*(x) = \sup_{y \in C} \left\{ y^\top x \right\}.$$

- Conjugate function: Given an arbitrary function $f : \mathbb{R}^d \to \mathbb{R}$, the conjugate function f^* is defined as

  $$f^*(x) = \sup_{y \in \mathbb{R}^d} \left\{ y^\top x - f(y) \right\}.$$

2.3 Properties of convex functions

Definition 3.4. The *epigraph* of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\mathrm{epi}(f) = \{(x, t) \in \mathrm{dom}(f) \times \mathbb{R} : \ f(x) \leq t\}.$$

The following is another definition of convex functions with respect to the epigraph.

Exercise 3.5. Prove that f is a convex function if and only if the epigraph is a convex set.

Example 3.6. Recall that the norm cone $\{(x, t) \in \mathbb{R}^d \times \mathbb{R} : \|x\| \leq t\}$ is a convex cone. This implies that any norm $f(x) = \|x\|$ is a convex function.

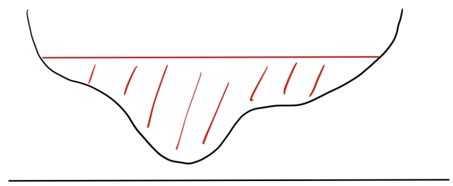

Figure 3.2: Convex level sets from a nonconvex function

Remark 3.7. A level set of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\{x \in \mathrm{dom}(f) : \ f(x) \le \alpha\}$$

for any $\alpha \in \mathbb{R}$. If f is convex, then all level sets are covex. However, the converse does not hold as Figure 3.2 demonstrates.

The following results provides a first-order characterization of convex functions.

Theorem 3.8. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. Then f is convex if and only if dom(f) is convex and*

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

for all $x, y \in$ dom(f).

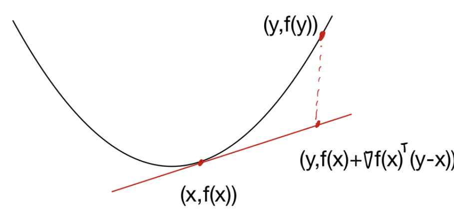

Figure 3.3: Illustration of the first-order characterization

Proof. ($\Rightarrow$) We first consider the $d = 1$ case. If f is convex, then for any $x, y \in \mathrm{dom}(f)$ and $\lambda \in [0, 1]$,

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

Moving the $(1 - \lambda)f(x)$ term to the other side and dividing each side by λ, we obtain

$$f(y) \ge f(x) + \frac{f(x + \lambda(y - x)) - f(x)}{\lambda}.$$

Then

$$f(y) \ge f(x) + \lim_{\lambda \to 0^+} \frac{f(x + \lambda(y - x)) - f(x)}{\lambda} = f(x) + (y - x)f'(x)$$

as f is differentiable and thus the limit exists.

Now we consider the general case. We take a function $g(\lambda) := f(x + \lambda(y - x))$ for $\lambda \in [0, 1]$. Then g is differentiable as $g'(\lambda) = (y - x)^\top \nabla f(x + \lambda(y - x))$. If f is convex, then g is convex. By the $d = 1$ case, $g(1) \ge g(0) + g'(0)$, which implies that $f(y) \ge f(x) + \nabla f(x)^\top (y - x)$.

($\Leftarrow$) Let $x, y \in \mathrm{dom}(f)$ and $\lambda \in [0, 1]$. Take $z = \lambda x + (1 - \lambda)y$. Then

$$f(x) \ge f(z) + \nabla f(z)^\top (x - z), \quad f(y) \ge f(z) + \nabla f(z)^\top (y - x).$$

Multiplying the first and second by λ and $(1-\lambda)$, respectively, and adding the resulting inequalities, it follows that

$$\lambda f(x) + (1 - \lambda)y) \ge f(z) + \nabla f(z)^\top (\lambda x + (1 - \lambda)y - z) = f(\lambda x + (1 - \lambda)y),$$

so f is convex. $\square$

What follows is another first-order characterization.

Theorem 3.9. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. Then f is convex if and only if dom(f) is convex and*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$$

for all $x, y \in$ dom(f).

Proof. ($\Rightarrow$) By Theorem 3.8, we have

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

Add these two to obtain $(\nabla f(x) - \nabla f(y))^\top (x - y) \geq 0$.

($\Leftarrow$) By the fundamental theorem of calculus, we obtain

$$\begin{aligned}\int_0^1 \nabla f(x + \lambda(y - x))^\top (y - x) d\lambda &= \int_0^1 \left(\frac{d}{d\lambda} f(x + \lambda(y - x)) \right) d\lambda \\ &= f(x + \lambda(y - x)) \Big|_{\lambda=0}^1 \\ &= f(y) - f(x).\end{aligned}$$

Moreover, for any $\lambda > 0$, we have

$$\nabla f(x + \lambda(y - x))^\top (y - x) - \nabla f(x)^\top (y - x) = \frac{1}{\lambda} \langle \nabla f(x + \lambda(y - x)) - \nabla f(x), \lambda(y - x) \rangle \geq 0,$$

implying in turn that

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

for any $\lambda > 0$. Note that this inequality trivially holds when $\lambda = 0$. Therefore,

$$f(y) - f(x) = \int_0^1 \nabla f(x + \lambda(y - x))^\top (y - x) d\lambda \geq \nabla f(x)^\top (y - x).$$

Then f is convex by Theorem 3.8. □

Next, we consider the second-order characterization.

Theorem 3.10. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function*[1]*. Then f is convex if and only if dom(f) is convex and*

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

for all $x \in$ dom(f).

Proof. ($\Rightarrow$) We first consider the $d = 1$ case. By Theorem 3.8, we have $f(x) \geq f(y) + f'(y)(x - y)$ and $f(y) \geq f(x) + f'(x)(y - x)$. Adding these up and dividing each side by $(y - x)^2$, we obtain

$$\frac{f'(y) - f'(x)}{y - x} \geq 0.$$

Taking the limit as $y \to x$, we obtain $f''(x) \geq 0$.

Next, let us consider the general case. Let $x \in \mathrm{dom}(f)$ and $v \in \mathbb{R}^d$. As $\mathrm{dom}(f)$ is open, we have a sufficiently small $\epsilon > 0$ such that $x + \lambda v \in \mathrm{dom}(f)$ for any $\lambda \in (-\epsilon, \epsilon)$. Let $g(\lambda) = f(x + \lambda v)$

[1] $\nabla^2 f$ exists at any point in $\mathrm{dom}(f)$, and $\mathrm{dom}(f)$ is open.

for $\lambda \in (-\epsilon, \epsilon)$. Since f is convex, g is also convex. Note that $g'(\lambda) = v^\top \nabla f(x + \lambda v)$ and that $g''(\lambda) = v^\top \nabla^2 f(x + \lambda v)v$. By the $d = 1$ case,

$$g''(0) = v^\top \nabla^2 f(x)v \geq 0.$$

Therefore, we have proved that $\nabla^2 f(x)$ is positive semidefinite.

($\Leftarrow$) By the fundamental theorem of calculus, we obtain

$$\begin{aligned}
\int_0^1 (y-x)^\top \nabla^2 f(x + \lambda(y-x)) d\lambda &= \int_0^1 \left(\frac{d}{d\lambda} \nabla f(x + \lambda(y-x)) \right) d\lambda \\
&= \nabla f(x + \lambda(y-x)) \Big|_{\lambda=0}^1 \\
&= \nabla f(y) - \nabla f(x).
\end{aligned}$$

Then

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle = \int_0^1 (y-x)^\top \nabla^2 f(x + \lambda(y-x))(y-x) d\lambda \geq 0$$

where the inequality follows because $\nabla^2 f$ is positive semidefinite. Then f is convex by Theorem 3.9.

$\square$