IE 539: Convex Optimization **KAIST, Fall 2024**
Lecture #3: convex functions September 9, 2024
Lecturer: Dabeen Lee

1 Outline

In this lecture, we study

- Convex functions and properties
- Epigraphs.
- First-order and second-order characterizations of convex functions.
- Operations preserving convexity

2 Convex functions

2.1 Definition

Definition 3.1. A function $f : \mathbb{R}^d \to \mathbb{R}$ is *convex* if the domain, denoted $\mathrm{dom}(f)$, 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, function f evaluated at a point between x and y lies below the line segment joining $f(x)$ and $f(y)$.

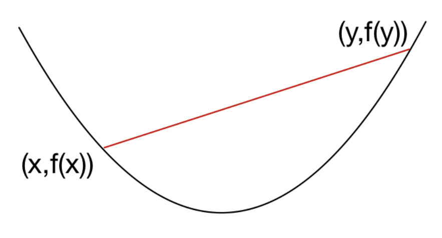

Figure 3.1: Illustration of a convex function in $\mathbb{R}^2$

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 $b \in \mathbb{R}$ are 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

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

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.

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

$$\{x \in \text{dom}(f) : \ f(x) \leq \alpha\}$$

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.

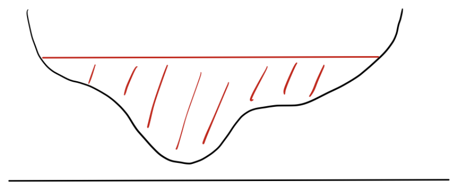

Figure 3.2: Convex level sets from a nonconvex function

3 First-order and second-order characterizations of convex functions

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*

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

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

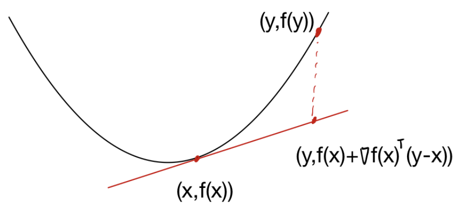

Figure 3.3: Illustration of the first-order characterization

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

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

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

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

Then

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

as f is differentiable and thus the limit exists.

Now we consider the general case. We define a function g over $\lambda \in [0, 1]$ as follows.

$$g(\lambda) := f(x + \lambda(y - x)).$$

Here, we can argue that if f is convex, then g is convex. More precisely, we have for $\alpha \in [0, 1]$ and $\lambda_1, \lambda_2 \in [0, 1]$,

$$\begin{aligned} g(\alpha\lambda_1 + (1 - \alpha)\lambda_2) &= f(x + (\alpha\lambda_1 + (1 - \alpha)\lambda_2)(y - x)) \\ &= f(\alpha(x + \lambda_1(y - x)) + (1 - \alpha)(x + \lambda_2(y - x))) \\ &\leq \alpha f(x + \lambda_1(y - x)) + (1 - \alpha)f(x + \lambda_2(y - x)). \end{aligned}$$

Moreover, g is differentiable as

$$g'(\lambda) = (y - x)^\top \nabla f(x + \lambda(y - x)).$$

By the $d = 1$ case, $g(1) \geq g(0) + g'(0)$, which implies that $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$.

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

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

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

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

so f is convex. □

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 us define g over $\lambda \in (-\epsilon, \epsilon)$ as follows.

$$g(\lambda) = f(x + \lambda v).$$

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. □

4 Operations preserving convexity

For many problems, it is important to recognize underlying convex structures. We can determine whether certain sets and functions are convex by understanding basic rules. Moreover, based on these rules, we can build complex convex sets and functions from simpler ones.

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

4.1 Set operations

We first consider set operations that preserve convexity.

- Intersection: The intersection of any (possibly infinite) collection of covex sets is convex.
- Scaling: Given a convex set C and $\alpha \in \mathbb{R}$,
$$\alpha C = \{\alpha x : \ x \in C\}.$$
- Minkowski sum: Given convex sets $C_i \subseteq \mathbb{R}^d$ for $i = 1, \ldots, k$, the Minkowski sum of them, defined by
$$C_1 + \cdots + C_k = \{x^1 + \cdots + x^k : \ x^i \in C_i \text{ for } i = 1, \ldots, k\}$$
is convex.
- Cartesian Product: Given convex sets $C_i \subseteq \mathbb{R}^{d_i}$ for $i = 1, \ldots, k$, the Cartesian product of them, defined by
$$C_1 \times \cdots \times C_k = \{(x^1, \ldots, x^k) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} : \ x^i \in C_i \text{ for } i = 1, \ldots, k\}$$
is convex.
- Affine image: Given a convex set C and matrices $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$, we define an affine mapping $f(x) = Ax + b : \mathbb{R}^d \to \mathbb{R}^p$. Then
$$f(C) = \{Ax + b : \ x \in C\}.$$
- Inverse affine image: Given a convex set C and matrices $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$, we define an affine mapping $f(x) = Ax + b : \mathbb{R}^d \to \mathbb{R}^p$. Then
$$f^{-1}(C) = \{x : \ Ax + b \in C\}.$$

4.2 Function operations

We next consider function operations preserving convexity.

- Nonnegative weighted sum: Let $f_1, \ldots, f_k : \mathbb{R}^d \to \mathbb{R}$ be convex functions. Then for any $\alpha_1, \ldots, \alpha_k \geq 0$,
$$\alpha_1 f_1 + \cdots + \alpha_k f_k$$
is convex.
- Maximum of arbitrary collection of convex functions: Let $\{f_\gamma\}_{\gamma \in \Gamma}$ be a collection of convex functions. Then $\max_{\gamma \in \Gamma} f_\gamma$ is also convex. Here, Γ may be infinite.
- Minimizing out variables: Let $g(x, y)$ be convex function in (x, y). Define f by $f(x) = \inf_{y \in C} g(x, y)$ for some convex set C. Then f is convex.
- Perspective function: Let $g(x)$ be a convex function. Then $f(x, t) = tg(x/t)$ is a convex function in $(x, t) \in \mathbb{R}^d \times \mathbb{R}_{++}$. Here, f is called the perspective of g.
- Affine composition: Let $g : \mathbb{R}^p \to \mathbb{R}$ be a convex function, and take matrices $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$. Then $f : \mathbb{R}^d \to \mathbb{R}$ defined by $f(x) = g(Ax + b)$ is convex.

- Compositions: Let $h : \mathbb{R} \to \mathbb{R}$ be a univariate non-decreasing convex function, and let $g : \mathbb{R}^d \to \mathbb{R}$ be convex. Then $f = h \circ g$ is convex.

Example 3.11. Let C be an arbitrary set of locations. Note that

$$f_1(x) = \max_{y \in C} ||x - y||$$

measures the longest distance from x to a location in C, and

$$f_2(x) = \min_{y \in C} \|x - y\|$$

measures the shortest distance from x to a location in C. Let us show that both f_1 and f_2 are convex. Observe first that

$$g(x, y) = \|x - y\|$$

is convex in x and y. Then f_1 is convex as it is the pointwise maximum of some convex functions. Furthermore, if C is convex, then f_2 is convex because it is a partial minimization of a convex function. In summary, f_1 is convex regardless of whether C is convex or not, while f_2 is convex if the set C is convex.