**EE3001 Machine Learning** **Fall 2021**

Lecture 07. Convex Functions

Lecturer: Jie Wang Date: Oct 25, 2021

## 1 Introduction

An optimization problem is convex if both its objective function and problem domain are convex. We have seen convex sets last lecture. In this lecture, we will focus on convex functions. The major reference of this lecture is [1, 2, 3].

## 2 Definitions

**Definition 1.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if **dom** $f$ is a convex set, and if for all $\mathbf{x}, \mathbf{y} \in$ **dom** $f$, and $\theta \in [0, 1]$, we have

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}). \tag{1}$$

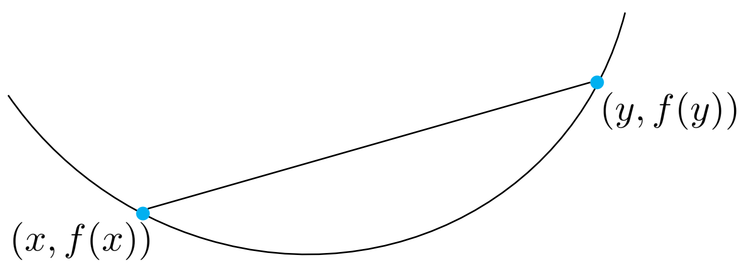


Figure 1: Convex function.

**Question 1.** What can we say about the continuity and differentiability of convex functions in view of Definition 1?

**Definition 2.** We have several variants of convexity.

- A function $f$ is **strictly convex** if strict inequality in Eq. (1) holds whenever $\mathbf{x} \neq \mathbf{y}$ and $\theta \in (0, 1)$.
- A function $f$ is **strongly convex** with parameter $\mu > 0$ if $f - \frac{\mu}{2}\|\mathbf{x}\|_2^2$ is convex.
- A function $f$ is **concave** if $-f$ is convex, **strictly concave** if $-f$ is strictly convex, and **strongly concave** if $-f$ is strongly convex.

**Example 1.** We give a few commonly seen examples of convex functions.

1. Affine function: $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} + b$, where $\mathbf{a} \neq 0$ and $b \in \mathbb{R}$.
2. Norms. Every norm on $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$.
3. Negative entropy: $f(\mathbf{x}) = x \log x$ is convex on $\mathbb{R}_{++}$.

**Definition 3 (Sublevel sets).** The **$\alpha$-sublevel set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$C_\alpha = \{x \in \textbf{dom } f : f(\mathbf{x}) \leq \alpha\}.$$

**Proposition 1.** *Sublevel sets of a convex function are convex, for any value of $\alpha$.*

We next provide another definition of the convexity of functions, which bridges the convexity of functions and that of sets.

**Definition 4.** The **epigraph** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\textbf{epi } f = \{(\mathbf{x}, t) : \mathbf{x} \in \textbf{dom } f, f(\mathbf{x}) \leq t\},$$

which is a subset of $\mathbb{R}^{n+1}$.

Epi means above, and thus epigraph means above the graph.

**Theorem 1.** *A function is convex if and only of its epigraph is a convex set.*

*Proof.* $\Rightarrow$ Suppose that $f$ is convex, and $(\mathbf{x}, t)$ and $(\mathbf{y}, s)$ belong to **epi** $f$ (of course, $\mathbf{x}, \mathbf{y} \in \textbf{dom } f$). To show that **epi** $f$ is convex, it suffices to show that the line segment joining $(\mathbf{x}, t)$ and $(\mathbf{y}, s)$ belongs to **epi** $f$, which is equivalent to

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta t + (1-\theta)s, \, \forall\, \theta \in [0, 1].$$

This can be seen easily from the convexity of $f$:

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) \leq \theta t + (1-\theta)s,$$

as $f(\mathbf{x}) \leq t$ and $f(\mathbf{y}) \leq s$ by the definition of epigraph.

$\Leftarrow$ Suppose that **epi** $f$ is convex. Consider $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$. Clearly, we have $(\mathbf{x}, f(\mathbf{x})), (\mathbf{y}, f(\mathbf{y})) \in$ **epi** $f$. As **epi** $f$ is convex, the line segment joining $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ belongs to **epi** $f$, i.e.,

$$(\theta\mathbf{x} + (1-\theta)\mathbf{y}, \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})) \in \textbf{epi } f.$$

The convexity of $f$ follows immediately by the definition of **epi** $f$. □

Theorem 1 is useful to determine the convexity of functions for some seemingly complicated cases.

**Lemma 1.** *If $f$ is a convex function, then for all $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ and all nonnegative $\alpha_i$, $i = 1, 2, \ldots, m$, such that $\sum_{i=1}^m \alpha_m = 1$, we have*

$$f\left(\sum_{i=1}^m \alpha_i \mathbf{x}_m\right) \leq \sum_{i=1}^m \alpha_i f(\mathbf{x}_i).$$

*Proof.* We can see that, the points

$$\begin{pmatrix} \mathbf{x}_i \\ f(\mathbf{x}_i) \end{pmatrix}, \, i = 1, 2, \ldots, m,$$

belong to the epigraph of $f$. As $f$ is a convex function, its epigraph **epi** $f$ is convex. Thus, any convex combination of the points $(\mathbf{x}_i, f(\mathbf{x}_i))^\top$, $i = 1, 2, \ldots, m$, belong to **epi** $f$, which leads to the claim immediately. □

**Theorem 2.** *A function $f : \mathbb{R}^n$ is convex if and only if* **dom** $f$ *is convex and its restriction to any line intersecting its domain is convex. By restriction to a line we mean that, for any $\mathbf{x}_0 \in$* **dom** $f$ *and $\mathbf{v} \in \mathbb{R}^n$, the function*

$$g(t) = f(\mathbf{x}_0 + t\mathbf{v}),$$

*is convex over its domain* **dom** $g = \{t : \mathbf{x}_0 + t\mathbf{v} \in \textbf{dom } f\}$.

# 3 Smooth Convex Functions

## 3.1 First-order conditions

**Theorem 3.** *Suppose that $f$ is continuously differentiable. Then, $f$ is convex if and only if* **dom** $f$ *is convex and*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \forall\, \mathbf{x}, \mathbf{y} \in \mathbf{dom}\ f.$$

*Proof.* $\Rightarrow$ The convexity of $f$ implies that, $\forall\, \theta \in (0, 1)$, we have

$$f(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x})) \leq f(\mathbf{x}) + \theta(f(\mathbf{y}) - f(\mathbf{x})).$$

This leads to

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \lim_{\theta \downarrow 0} \frac{f(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\theta} = \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle.$$

$\Leftarrow$ Let $\mathbf{z} = \theta \mathbf{x} + (1 - \theta)\mathbf{y}$. Then,

$$f(\mathbf{x}) \geq f(\mathbf{z}) + \langle \nabla f(\mathbf{z}), \mathbf{x} - \mathbf{z} \rangle, \quad f(\mathbf{y}) \geq f(\mathbf{z}) + \langle \nabla f(\mathbf{z}), \mathbf{y} - \mathbf{z} \rangle.$$

Multiplying the first inequality by $\theta$, the second by $1 - \theta$, and adding them together lead to the desired result. □

**Theorem 4.** *Suppose that $f$ is continuously differentiable. Then, $f$ is convex if and only if* **dom** $f$ *is convex and*

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq 0.$$

*Proof.* $\Rightarrow$ The convexity of $f$ implies that

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle.$$

Adding them together leads to desired result.

$\Leftarrow$ Let $\mathbf{x}_t = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$. Then,

$$\begin{aligned} f(\mathbf{y}) =& f(\mathbf{x}) + \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt \\ =& f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \frac{1}{t} \langle \nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}), \mathbf{x}_t - \mathbf{x} \rangle dt \\ \geq& f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle. \end{aligned}$$

□

**Example 2.** Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ defined as the quadratic form

$$f(\mathbf{x}) = \langle \mathbf{x}, A\mathbf{x} \rangle,$$

where $A \in \mathbb{S}^n$ is a symmetric matrix. Then, $f$ is convex if and only if $A$ is a positive semidefinite matrix, and strictly convex if and only if $A$ is a positive definite matrix.

Indeed, as

$$\nabla f(\mathbf{x}) = 2A\mathbf{x},$$

we have

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle =& \langle \mathbf{y}, A\mathbf{y} \rangle - \langle \mathbf{x}, A\mathbf{x} \rangle - 2\langle A\mathbf{x}, \mathbf{y} - \mathbf{x} \rangle \\ =& \langle \mathbf{y}, A\mathbf{y} \rangle + \langle \mathbf{x}, A\mathbf{x} \rangle - 2\langle A\mathbf{x}, \mathbf{y} \rangle \\ =& \langle \mathbf{y} - \mathbf{x}, A(\mathbf{y} - \mathbf{x}) \rangle. \end{aligned}$$

## 3.2 Second-order conditions

**Theorem 5.** *Suppose that $f$ is twice continuously differentiable. Then, $f$ is convex if and only if* **dom** $f$ *is convex and* $\nabla^2 f(\mathbf{x}) \succeq 0$.

*Proof.* $\Rightarrow$ Let $\mathbf{x}_t = \mathbf{x} + t\mathbf{s}$, $t > 0$. Then,

$$\begin{aligned} 0 \leq & \frac{1}{t^2}\langle \nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}), \mathbf{x}_t - \mathbf{x}\rangle = \frac{1}{t}\langle \nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}), \mathbf{s}\rangle \\ &= \frac{1}{t}\int_0^t \langle \nabla^2 f(\mathbf{x} + \tau\mathbf{s})\mathbf{s}, \mathbf{s}\rangle d\tau. \end{aligned} \tag{2}$$

By the mean value theorem, we can find an $\alpha \in (0, t)$ such that

$$\int_0^t \langle \nabla^2 f(\mathbf{x} + \tau\mathbf{s})\mathbf{s}, \mathbf{s}\rangle d\tau = t\langle \nabla^2 f(\mathbf{x} + \alpha\mathbf{s})\mathbf{s}, \mathbf{s}\rangle. \tag{3}$$

Plugging Eq. (3) into the inequality in (2) leads to

$$0 \leq \langle \nabla^2 f(\mathbf{x} + \alpha\mathbf{s})\mathbf{s}, \mathbf{s}\rangle.$$

As the above inequality holds for any $t > 0$ and $\alpha \in (0, t)$, we have

$$0 \leq \langle \nabla^2 f(\mathbf{x})\mathbf{s}, \mathbf{s}\rangle$$

by letting $t \downarrow 0$. We further note that $s$ is an arbitrary vector. Thus, the Hessian $\nabla^2 f(\mathbf{x})$ must be positive semi-definite, i.e., $\nabla^2 f(\mathbf{x}) \succeq 0$.

$\Leftarrow$ Let $g(t) = f(\mathbf{x} + t\mathbf{s})$. Then, $g'(0) = \langle \nabla f(\mathbf{x}), \mathbf{s}\rangle$ and $g''(0) = \langle \nabla^2 f(\mathbf{x})\mathbf{s}, \mathbf{s}\rangle$.

$$\begin{aligned} g(1) =& g(0) + \int_0^1 g'(t)dt = g(0) + \int_0^1 \left[g'(0) + \int_0^t g''(\tau)d\tau\right] dt \\ =& g(0) + g'(0) + \int_0^1 \left[\int_0^t g''(\tau)d\tau\right] dt \\ \geq& g(0) + g'(0) \end{aligned}$$

$\square$

**Example 3.** The log-determinant function

$$f(X) = -\log\det X$$

is convex with **dom** $f = \mathbb{S}^n_{++}$.

To see this, let $X_0 \in \mathbb{S}^n_{++}$ and $V\mathbb{S}^n$. We define

$$g(t) = f(X_0 + tV)$$

with **dom** $g = \{t : X_0 + tV \in \mathbb{S}^n_{++}\}$. Thus

$$\begin{aligned} g(t) &= -\log\det(X_0 + tV) \\ &= -\log\det(X_0^{1/2}(I + tX_0^{-1/2}VX_0^{-1/2})X_0^{1/2}) \\ &= -\sum_{i=1}^n \log(1 + t\lambda_i) - \log\det X_0, \end{aligned}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $X_0^{-1/2} V X_0^{-1/2}$. Therefore, we have

$$g'(t) = -\sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i}, \quad g''(t) = \sum_{i=1}^{n} \frac{\lambda_i^2}{(1 + t\lambda_i)^2}.$$

As $g''(t) \geq 0$, we conclude that $f$ is convex.

### 3.3 Extended-value extensions

**Definition 5.** If $f$ is convex, we define its *extended-value extension* $\tilde{f} : \mathbf{R}^n \to \mathbb{R} \cup \{\infty\}$ by

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \mathbf{x} \in \mathbf{dom}\ f, \\ \infty, & \mathbf{x} \notin \mathbf{dom}\ f. \end{cases}$$

**Example 4.** Let $C \subseteq \mathbb{R}^n$ be a convex set. Its *indicator function* $I_C : C \to \mathbb{R}$ is zero for all $\mathbf{x} \in C$. The extended-value extension of $I_C$ is

$$\tilde{I}_C(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in C, \\ \infty, & \mathbf{x} \notin C. \end{cases}$$

***Remark*** 1. The inequality in (1) holds for $\tilde{I}_C$ for all $x, y \in \mathbb{R}^n$.

## 4 Operations that Preserve Convexity

**Proposition 2.** *Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be a given function, let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, and let*

$$f(\mathbf{x}) = f(A\mathbf{x} + b),\ x \in \mathbb{R}^n.$$

*If $f$ is convex, then $F$ is also convex.*

**Proposition 3.** *Let $f_i : \mathbb{R}^n \to (-\infty, \infty]$, $i = 1, \dots, m$, be given functions, let $w_1, \dots, w_m$ be positive salars, and*

$$f(\mathbf{x}) = w_1 f_1(\mathbf{x}) + \cdots + w_m f_m(\mathbf{x}),\ \mathbf{x} \in \mathbb{R}^n.$$

*If $f_1, \dots, f_m$ are convex, then $f$ is also convex.*

**Proposition 4.** *Let $f_i : \mathbb{R}^n \to (-\infty, \infty]$ be given functions for $i \in I$, where $I$ is an arbitrary index set, and*

$$f(\mathbf{x}) = \sup_{i \in I} f_i(\mathbf{x}).$$

*If $f_i$, $i \in I$, are convex, then $f$ is also convex.*

**Example 5.** The weighted least squares

$$h(\mathbf{w}) = \frac{1}{n}\|\mathbf{y} - X\mathbf{w}\|^2 + \lambda\Omega(\mathbf{w}),$$

where $\Omega(\mathbf{w}) = \|\mathbf{w}\|_2^2$ or $\Omega(\mathbf{w}) = \|\mathbf{w}\|_1$, is convex for all $\lambda > 0$.

**Example 6.** For $\mathbf{x} \in \mathbb{R}^n$, let $x_{[i]}$ be the $i^{th}$ largest component of $\mathbf{x}$. Then, the function

$$f(\mathbf{x}) = \sum_{i=1}^{r} x_{[i]},$$

is convex.

**Example 7.** For $A \in \mathbb{S}^n$, its largest eigenvalue

$$f(A) = \lambda_{\max}(A)$$

is a convex function with respect to $A$, as

$$f(A) = \max_{\|\mathbf{v}\|=1} \langle \mathbf{v}, A\mathbf{v} \rangle,$$

and $\langle \mathbf{v}, A\mathbf{v} \rangle$ is linear with respect to $A$ for all $\mathbf{v}$.

**Example 8.** For a nonempty set $C \subset \mathbb{R}^n$, the support function of $C$ is defined as

$$f_C(\mathbf{v}) = \sup\{\langle \mathbf{v}, \mathbf{x} \rangle : \mathbf{x} \in C\}$$

with its domain $\mathbf{dom}\ f_C = \{\mathbf{v} : f_C(\mathbf{v}) < \infty\}$. The support function is convex.

# References


[1] D. Bertsekas. *Convex Optimization Theory*. Athena Scientific, 2009.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] A. Ruszczyński. *Nonlinear Optimization*. Princeton University Press, 2006.