210709 Machine Learning **Spring 2021**

Lecture 3. Elementary Convex Programming II

Lecturer: Jie Wang Date: April 2, 2021

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 first study convex functions, and then we give a general formulation of convex optimization problems. The major reference of this lecture is [1, 2].

1 Convex Functions

1.1 Definition

Definition 1. A function $f : D \to \mathbb{R}$ with $D \subseteq \mathbb{R}^n$ is ***convex*** if D—that is, the domain of f denoted by $\mathbf{dom}\ f$—is a convex set, and if for all $x, y \in \mathbf{dom}\ f$, and $\theta \in [0, 1]$, we have

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

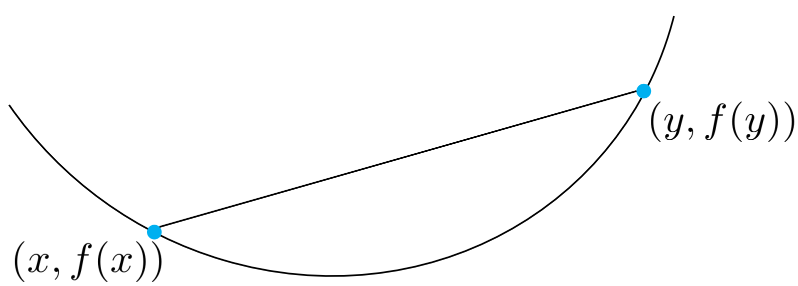

Figure 1: Convex function.

***Remark* 1.** Notice that, in Definition 1, we do not ask the *continuity* of f.

Question 1. What about the continuity of convex functions?

Definition 2. We have several variants of convexity.

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

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

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

1.2 First-order conditions

Theorem 1. *Suppose that f is continuously differentiable. Then, f is convex if and only if* **dom** f *is convex and*

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

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

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

This leads to

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

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

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

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

Theorem 2. *Suppose that f is continuously differentiable. Then, f is convex if and only if* **dom** f *is convex and*

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

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

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

Adding them together leads to desired result.

$\Leftarrow$ Let $x_t = x + t(y - x)$. Then,

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

□

1.3 Second-order conditions

Theorem 3. *Suppose that f is twice continuously differentiable. Then, f is convex if and only if* **dom** f *is convex and $\nabla^2 f(x) \succeq 0$.*

Proof. $\Rightarrow$ Let $x_t = x + ts$, $t > 0$. Then,

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

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

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

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

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

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

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

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

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

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

□

1.4 Extended-value extensions

Definition 3. If f is convex, we define its *extended-value extension* $\tilde{f} : \mathbf{R}^n \to \mathbb{R} \cup \{\infty\}$ by

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

Example 2. Let $C \subseteq \mathbb{R}^n$ be a convex set. Its *indicator function* $I_C : C \to \mathbb{R}$ is zero for all $x \in C$. The extended-value extension of I_C is

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

Remark 2. The inequality in (1) holds for $\tilde{I}_C$ for all $x, y \in \mathbb{R}^n$.

1.5 Epigraph

Definition 4 (Sublevel sets). The *α-sublevel set* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

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

Proposition 1. *Sublevel sets of a convex function are convex, for any value of α.*

Definition 5. The *graph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\{(x, f(x)) : x \in \mathbf{dom}\ f\},$$

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

Definition 6. The *epigraph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$f = \{(x, t) : x \in \textbf{dom}\ f, f(x) \le t\},$$

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

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

Proposition 2. *A function is convex if and only of its epigraph is a convex set.*

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

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

This can be seen easily from the convexity of f:

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y) \le \theta t + (1-\theta)s,$$

as $f(x) \le t$ and $f(y) \le s$ by the definition of epigraph.

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

$$(\theta x + (1-\theta)y, \theta f(x) + (1-\theta)f(y)) \in f.$$

The convexity of f follows immediately by the definition of f. □

2 Operations that Preserve Convexity

Proposition 3. *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*

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

If f is convex, then F is also convex.

Proposition 4. *Let $f_i : \mathbb{R}^n \to (-\infty, \infty]$, $i = 1, \ldots, m$, be given functions, let $w_1, \ldots, w_m$ be positive salars, and*

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

If $f_1, \ldots, f_m$ are convex, then F is also convex.

Proposition 5. *Let $f_i : \mathbb{R}^n \to (-\infty, \infty]$ be given functions for $i \in I$, where I is an arbitrary index set, and*

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

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

3 Basic Terminology

We consider the problem as follows.

$$\begin{aligned} \min_{\mathbf{x}} \; & f(\mathbf{x}) \\ \text{s.t.} \; & g_i(\mathbf{x}) \leq 0, \; i = 1, \ldots, m, \\ & h_i(\mathbf{x}) = 0, \; i = 1, \ldots, p, \\ & \mathbf{x} \in X, \end{aligned} \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, p$, and $X \subseteq \mathbb{R}^n$. To simplify notations, let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ be a vector function whose i^{th} component is g_i, and $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^p$ be a vector function whose i^{th} component is h_i. Then, the problem in (4) becomes

$$\begin{aligned} \min_{\mathbf{x}} \; & f(\mathbf{x}) \\ \text{s.t.} \; & \mathbf{g}(\mathbf{x}) \leq 0, \\ & \mathbf{h}(\mathbf{x}) = 0, \\ & \mathbf{x} \in X. \end{aligned} \tag{5}$$

We call the problem in (5) as the primal problem.

Definition 7.

- The *feasible set* is

$$D = \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \leq 0, \mathbf{h}(\mathbf{x}) = 0, \mathbf{x} \in X\}. \tag{6}$$

- Each element in D is called a *feasible solution*.
- The optimal function value is defined by

$$f^* = \inf_{\mathbf{x} \in D} f(\mathbf{x}). \tag{7}$$

Assumption 1. Feasibility and Boundedness *The feasible set is nonempty and the objective function is bounded from below, that is,*

$$-\infty < f^* = \inf_{\mathbf{x} \in D} f(\mathbf{x}) < \infty.$$

Definition 8. We say $\mathbf{x}^*$ is an *optimal point*, or solves the problem (5), if $\mathbf{x}^*$ is feasible and $f(\mathbf{x}^*) = f^*$. The set of all optimal points is the *optimal set*, denoted by

$$X^* = \{\mathbf{x}^* : \mathbf{x}^* \in D, f(\mathbf{x}^*) = f^*\}.$$

***Remark* 3.**

- If problem (5) has an optimal solution, we say the optimal value is *attained* or *achieved*, and the problem is *solvable*. Otherwise (X^* is empty), we say the optimal value is not attained or not achieved.

- A feasible point $\mathbf{x}$ with $f(\mathbf{x}) \leq f^* + \epsilon$ ($\epsilon > 0$) is called *ϵ-suboptimal*, and the set of all ϵ-suboptimal points is called *ϵ-suboptimal set* for the problem (5).

Definition 9. Consider the problem (5). Suppose that the functions f, g_i, $i = 1, \ldots, m$ are convex, h_i, $i = 1, \ldots, p$ are affine, and the set X is convex. Then, we say that the problem (5) is a *convex optimization problem.*

Remark 4. A general convex optimization problem takes the form of

$$\begin{aligned} \min_{\mathbf{x}} \; & f(\mathbf{x}) \\ \text{s.t.} \; & \mathbf{x} \in \mathbf{dom}\, f, \\ & \mathbf{x} \in X, \end{aligned}$$

where f is convex and the feasible set

$$D = \mathbf{dom}\, f \cap X$$

is convex. Throughout this lecture, we assume that D is nonempty.

4 The Optimal Set

Proposition 6. *Suppose that the problem (5) is a convex optimization problem and solvable. Then, the optimal set X^* is convex.*

Proof. If there is only one point in X^*, we can see that X^* is clearly convex. Thus, we consider the cases where there are multiple points in X^*.

Suppose that $\mathbf{x}, \mathbf{y} \in X^*$ and $\mathbf{x} \neq \mathbf{y}$. As $X^* \subseteq D$, the line segment connecting $\mathbf{x}$ and $\mathbf{y}$ belongs to the feasible set D as well. Let $\theta \in (0, 1)$. Then,

$$f^* \leq f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) = f^*,$$

which implies that

$$f^* = f(\theta\mathbf{x} + (1-\theta)\mathbf{y}).$$

Thus, the points on the segment joining $\mathbf{x}$ and $\mathbf{y}$ belong to X^*, and thus X^* is convex. This completes the proof. □

Definition 10. A feasible point $\mathbf{x}$ is *locally optimal* if there is a $\delta > 0$ such that

$$f(\mathbf{x}) = \inf\{f(\mathbf{z}) : \mathbf{x} \in D, \|\mathbf{z} - \mathbf{x}\| < \delta\}.$$

Proposition 7. *Suppose that the problem (5) is a convex optimization problem and solvable. Then, if $\mathbf{x}$ is a local optimum, it is also a global optimum.*

Proof. Let $\mathbf{y} \in D$ be an arbitrary feasible point other than $\mathbf{x}$. Thus, to show that the claim holds, it suffices to show that,

$$f(\mathbf{x}) \leq f(\mathbf{y}). \tag{8}$$

As $\mathbf{x}$ is a local optimum, we can find a $\delta > 0$ such that

$$f(\mathbf{x}) \leq f(\mathbf{z}), \forall \mathbf{z} \in D \cap B := \{\mathbf{z} : \|\mathbf{z} - \mathbf{x}\| < \delta\}.$$

Clearly, if $\mathbf{y} \in B$, the inequality (8) holds. Thus, we only need to consider the case where $\mathbf{y} \notin B$, i.e.,

$$\|\mathbf{y} - \mathbf{x}\| \geq \delta.$$

Due to the convexity of D, all the points on the line segment ℓ joining $\mathbf{x}$ and $\mathbf{y}$ belong to D. Let

$$\theta = 1 - \frac{\delta}{2\|\mathbf{y} - \mathbf{x}\|},$$

and

$$\mathbf{z}_0 = \theta \mathbf{x} + (1 - \theta)\mathbf{y}.$$

We can see that $\mathbf{z}_0$ is on the line segment ℓ as $\theta \in (0, 1)$, and

$$\|\mathbf{z}_0 - \mathbf{x}\| = \frac{\delta}{2}.$$

This implies that $\mathbf{z}_0 \in B$ and thus

$$f(\mathbf{x}) \leq f(\mathbf{z}_0). \tag{9}$$

Combining with the convexity of f, we have

$$f(\mathbf{x}) \leq f(\mathbf{z}_0) \leq \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y}).$$

By moving $\theta f(\mathbf{x})$ to the LHS, and dividing both sides by $(1 - \theta)$, we can see that the inequality (8) holds. This completes the proof. □

Proposition 8. *Suppose that the problem (5) is a convex optimization problem and solvable. Then, if f is strictly convex, the problem (5) has a unique global optimum.*

Proposition 9. *Suppose that the problem (5) is a convex optimization problem. If f is strongly convex and continuous over its domain, and the feasible set is closed, then the problem (5) is solvable and has a unique global optimum.*

Remark *5.* In view of Propositions 8 and 9, the problem in (5) has a unique global optimum if its objective function is strictly convex or strongly convex. However, the problem in (5) with a strongly convex objective function has a remarkable advantage over that with a strictly convex objective function, that is, the former is guaranteed to admit at least one global optimum, while the latter is not even its feasible set is closed (why?).

Definition 11. If $\mathbf{x}$ is feasible, and $g_i(\mathbf{x}) = 0$, we say the i^{th} inequality constraint $g_i(\mathbf{x}) \leq 0$ is *active* at $\mathbf{x}$; otherwise ($g_i(\mathbf{x}) < 0$), we say the constraint $g_i(\mathbf{x}) \leq 0$ is *inactive* at $\mathbf{x}$.

References

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

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