210709 Machine Learning **Spring 2020**

Lecture 2. Convex Sets and Convex Functions

Lecturer: Jie Wang Date: March 5, 2020

The major reference of this lecture is [1, 2, 3].

1 Mathematical Background

1.1 Norms

Definition 1. A function $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbf{dom}\, f = \mathbb{R}^n$ is called a norm if

- f is nonnegative: $f(x) \geq 0$ for all $x \in \mathbb{R}^n$;
- f is definite: $f(x) = 0$ only if $x = 0$;
- f is homogeneous: $f(tx) = |t|f(x)$, for all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$;
- f satisfies the triangle inequality: $f(x + y) \leq f(x) + f(y)$, for all $x, y \in \mathbb{R}^n$.

We often use the notation $f(x) = \|x\|$ to denote the norm function.

Example 1. For $x \in \mathbb{R}^n$, the commonly seen ℓ_p norm is defined by

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

The ℓ_1-norm and ℓ_2-norm (the Euclidean norm) are commonly-used regularization terms. Moreover, the Chebyshev or ℓ_∞-norm is given by

$$\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

Example 2. For $X \in \mathbb{R}^{m \times n}$, the Frobenius norm is

$$\|X\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j}^2}.$$

1.2 Analysis

1.2.1 Sequences

Definition 2. A sequence $\{x_k : k = 1, 2, \ldots\}$ of vectors in $\mathbb{R}^n$ is said to converge to $x \in \mathbb{R}^n$ if for any $\epsilon > 0$, there exists a positive integer N such that

$$\|x_k - x\|_2 < \epsilon, \forall\, k \geq N.$$

Symbolically, $x_k \to x$ or $\lim_{k \to \infty} x_k = x$.

Definition 3. A vector $x \in \mathbb{R}^n$ is a *limit (cluster/accumulation)* point of a sequence $\{x_k\}$ if there exists a subsequence of $\{x_k\}$ that converges to x.

1.2.2 Open and closed sets

Definition 4. An element $x \in C \subseteq \mathbb{R}^n$ is called a *closure* point of C if there exists a sequence $\{x_k\} \subseteq C$ that converges to x. The closure of C, denoted by **cl** C, is the set of all closure points of C.

Definition 5. A subset C of $\mathbb{R}^n$ is called *closed* if it is equal to its closure.

Definition 6. An element $x \in C \subseteq \mathbb{R}^n$ is called an *interior* point of C if there exists an $\epsilon > 0$ such that

$$\{y : \|y - x\|_2 \leq \epsilon\} \subseteq C.$$

The set of interior points of C is called the *interior* of C and is denoted by **int** C.

Definition 7. A set $C \subseteq \mathbb{R}^n$ is *open* if **int** $C = C$, i.e., every point in C is an interior point. It is *closed* if its complement

$$\mathbb{R}^n \setminus C = \{x \subseteq \mathbb{R}^n : x \notin C\}$$

is open. It is called *bounded* if there exists a scalar c such that

$$\|x\|_2 \leq c, \forall x \in C.$$

It is called *compact* if it is closed and bounded.

Definition 8. The *boundary* of the set C is defined as

$$\mathbf{bd}\ C = \mathbf{cl}\ C \setminus \mathbf{int}\ C.$$

For any *boundary* point, i.e., $x \in \mathbf{bd}\ C$, we have the following holds.

$$\forall \epsilon > 0,\ \exists y \in C \text{ and } z \notin C, \text{ such that } \|y - x\|_2 \leq \epsilon, \|z - x\|_2 \leq \epsilon.$$

1.2.3 Supremum and infimum

Definition 9. A set $C \subseteq \mathbb{R}$ is *bounded above* if there exists a number $u \in \mathbb{R}$ such that $c \leq u$ for all $c \in C$. The number u is called an *upper bound* for C.

Similarly, the set C is *bounded below* if there exists a *lower bound* $l \in \mathbb{R}$ such that $l \leq c$ for all $c \in C$.

Definition 10. The real number u is the least upper bound for a set $C \subseteq \mathbb{R}$ if

1. u is an upper bound for A;
2. if u' is any upper bound for C, then $u \leq u'$.

The least upper bound is called the *supremum* of the set C, which is denoted by

$$u = \sup C.$$

If $u \in C$, then u is called the *maximum* point of C, i.e.,

$$u = \max C.$$

Definition 11. The real number l is the greatest lower bound for a set $C \subseteq \mathbb{R}$ if

1. l is a lower bound for C;
2. if l' is any lower bound for C, then $l \geq l'$.

The greatest lower bound is called the *infimum* of the set C, which is denoted by

$$l = \inf C.$$

If $l \in C$, then l is called the *minimum* point of C, i.e.,

$$l = \min C.$$

1.2.4 Continuous functions on compact sets

Proposition 1 (**Bolzano-Weierstrass Theorem**). *A bounded sequence in $\mathbf{R}^n$ has at least one limit point.*

Theorem 1 (**Extreme Value Theorem**). *Let C be a compact subset of $\mathbb{R}^n$ and $f : C \to \mathbb{R}$ be continuous. Then, there exist $a, b \in C$ such that*

$$f(a) \leq f(x) \leq f(b), \forall x \in C.$$

In other words, f attains a maximum and minimum value in C.

2 Convex Sets

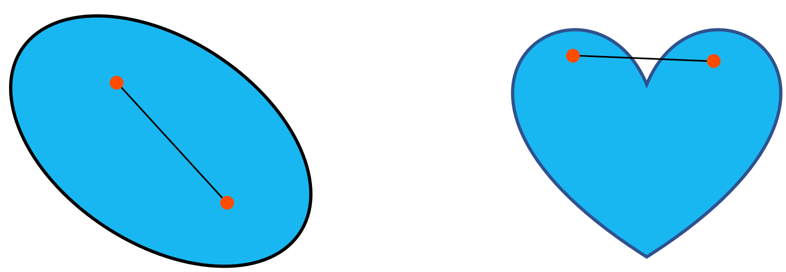

Figure 1: Convex and nonconvex sets.

Definition 12. A point $x \in \mathbb{R}^n$ is a *convex combination* of the points $\{x_1, \ldots, x_k\} \subseteq \mathbb{R}^n$ if

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k,$$

where $\theta_i \geq 0$ for $i = 1, \ldots, k$ and

$$\theta_1 + \theta_2 + \ldots + \theta_k = 1.$$

Definition 13. The *convex hull* of a set $C \subseteq \mathbb{R}^n$, denoted by **conv** C, is the set of all convex combinations of points in C:

$$\mathbf{conv}\ C = \left\{ \sum_{i=1}^{k} \theta_i x_k : x_i \in C, \theta_i \geq 0, \sum_{i=1}^{k} \theta_k = 1 \right\}.$$

Definition 14. A set C is convex if the line segment between any two points in C lies in C; that is, if $\forall\, x_1, x_2 \in C$ and $\forall\, \theta \in [0,1]$, we have

$$\theta x_1 + (1-\theta)x_2 \in C.$$

Example 3. Suppose $p : \mathbb{R}^n \to \mathbb{R}$ satisfies $p(x) \geq 0$ for all $x \in C$ and $\int_C p(x)dx = 1$, where $C \subseteq \mathbb{R}^n$ is convex. Then

$$\int_C p(x)x dx \in C,$$

if the integral exists.

Proposition 2.

1. *The intersection $\cap_{i\in I} C_i$ of any collection $\{C_i : i \in I\}$ of convex sets is convex.*
2. *The closure and the interior of a convex set are convex.*
3. *The image and the inverse image of a convex set under an affine function ($f(x) = Ax + b$) are convex.*

Example 4.

1. hyperplane: $\{x : a^\top x = b\}$, where $a \neq 0$ and $b \in \mathbb{R}$.
2. halfspace: $\{x : a^\top x \leq b\}$, where $a \neq 0$ and $b \in \mathbb{R}$.
3. norm ball: $\{x : \|x - x_0\| \leq r\}$, where $r > 0$.
4. polyhedron: $\{x : a_i^\top x \leq b_i, i = 1, \ldots, m\}$, where $a_i \neq 0$ and $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$.
5. positive semi-definite matrices

3 Convex Functions

3.1 Definition

Definition 15. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if **dom** f is a convex set and if for all $x, y \in$ **dom** f, and $\theta \in [0,1]$, we have

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

Definition 16. We have several variants of convexity.

- A function f is *strictly convex* if strict inequality 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 5. 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 $\mathbb{R}_{++}$.

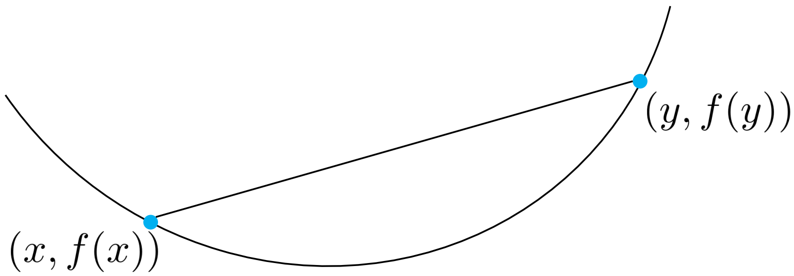

Figure 2: Convex function.

3.2 Extended-value extensions

Definition 17. 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 6. 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 1. The inequality in (1) holds for $\tilde{I}_C$ for all $x, y \in \mathbb{R}^n$.

3.3 First-order conditions

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

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

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 3. *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,

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

□

3.4 Second-order conditions

Theorem 4. *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,

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

The result follows by letting t tend to zero.

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

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

□

3.5 Epigraph

Definition 18 (Sublevel sets). The α-*sublevel set* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

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

Proposition 3. *Sublevel sets of a convex function are convex, for any value of α.*

Definition 19. The *graph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

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

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

Definition 20. The *epigraph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

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

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

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

Proposition 4. *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 **epi** f (of course, $x, y \in$ **dom** f). To show that **epi** f is convex, it suffices to show that the line segment joining (x,t) and (y,s) belongs to **epi** f, which is equivalent to

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

This can be seen easily from the convexity of f:

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

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

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

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

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

4 Operations that Preserve Convexity

Proposition 5. *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 6. *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 7. *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.

References

[1] S. Abbott. *Understanding Analysis, 2ed*. Springer, 2015.

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

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