

In this lecture, we study

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

1 Convex sets

1.1 Definition

Definition 1.1. A set $X \subseteq \mathbb{R}^d$ is *convex* if for any $u, v \in X$ and any $\lambda \in [0, 1]$,

$$\lambda u + (1 - \lambda)v \in X.$$

In words, the line segment joining any two points is entirely contained the set. In Figure 1.1, we have a convex set and a non-convex set.

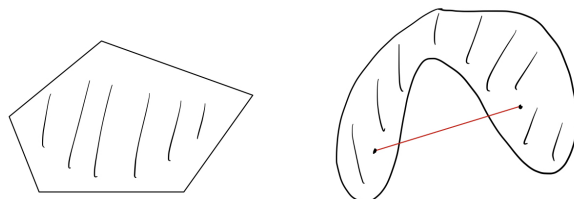


Figure 1.1: A convex set and a nonconvex set

Definition 1.2. Given $v^1, \dots, v^k \in \mathbb{R}^d$, any linear combination $\lambda_1 v^1 + \dots + \lambda_k v^k$ is a *convex combination* of $v^1, \dots, v^k$ if

$$\sum_{i=1}^k \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0 \text{ for } i = 1, \dots, k.$$

The convex combination of two distinct points u, v is the line segment $\{\lambda u + (1 - \lambda)v : 0 \leq \lambda \leq 1\}$ connecting them.

Definition 1.3. The *convex hull* of a set X , denoted $\text{conv}(X)$, is the set of all convex combinations of points in X . By definition,

$$\text{conv}(X) = \left\{ \sum_{i=1}^n \lambda_i v^i : \begin{array}{l} n \in \mathbb{N}, v^1, \dots, v^n \in X, \\ \sum_{i=1}^n \lambda_i = 1, \lambda_1, \dots, \lambda_n \geq 0 \end{array} \right\}.$$

Here, $\text{conv}(X)$ is always convex regardless of X . Figure 1.2 shows some examples of taking the convex hull of a set.

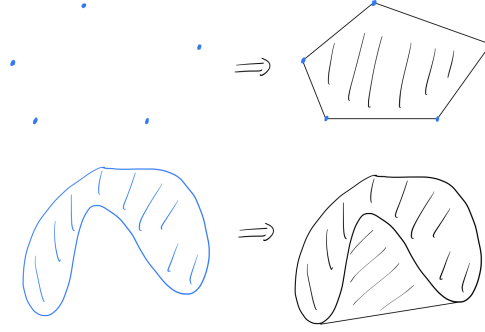


Figure 1.2: A convex set and a nonconvex set

1.2 Cones and affine subspaces

Definition 1.4. A set $C \subseteq \mathbb{R}^d$ is a *cone* if for any $v \in C$ and $\alpha > 0$, we have $\alpha v \in C$. Furthermore, if C is convex, then it is called a *convex cone*.

Note that not all cones are convex.

Definition 1.5. Given $v^1, \dots, v^k \in \mathbb{R}^d$, any point of the form $\alpha_1 v^1 + \dots + \alpha_k v^k$ is a *conic combination* of $v^1, \dots, v^k$ if $\alpha_1, \dots, \alpha_k \geq 0$.

In other words, any nonnegative linear combination is a conic combination.

Definition 1.6. The *conic hull* of a set X , denoted $\text{cone}(X)$, is the set of all conic combinations of points in X . By definition,

$$\text{conv}(X) = \left\{ \sum_{i=1}^n \lambda_i v^i : \begin{array}{l} n \in \mathbb{N}, v^1, \dots, v^n \in X, \\ \lambda_1, \dots, \lambda_n \geq 0 \end{array} \right\}.$$

As $\text{conv}(X)$, $\text{cone}(X)$ is always convex, Figure 1.3 shows an example taking the conic hull of a set in $\mathbb{R}^2$.

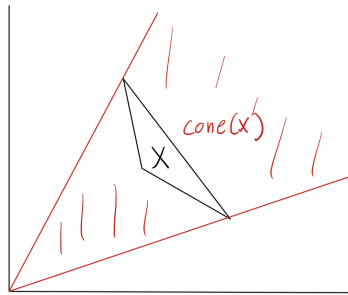


Figure 1.3: Taking the conic hull of a triangle in $\mathbb{R}^2$

Lastly, we define the notion of *affine subspaces*.

Definition 1.7. Given $v^1, \dots, v^k \in \mathbb{R}^d$, any point of the form $\theta_1 v^1 + \dots + \theta_k v^k$ is a *affine combination* of $v^1, \dots, v^k$ if $\theta_1 + \dots + \theta_k = 1$.

In contrast to convex combinations, affine combinations allow negative multipliers.

Definition 1.8. The *affine hull* of a set X is the set of all affine combinations of points in X .

The affine hull of X is also referred to as the *affine subspace* spanned by X . In the previous lecture, we defined the linear subspace spanned by a finite set of vectors, but we can extend the definition to an arbitrary set. The *linear hull* of a set X is equivalent to the linear subspace spanned by X .

In Figure 1.4, we have a set S of two points in $\mathbb{R}^2$. The red line segment is $\text{conv}(S)$, the green line through the two points is the affine subspace spanned by S , the blue cone depicts $\text{cone}(S)$, and lastly, the orange region (in fact, $\mathbb{R}^2$) is the linear subspace spanned by S .

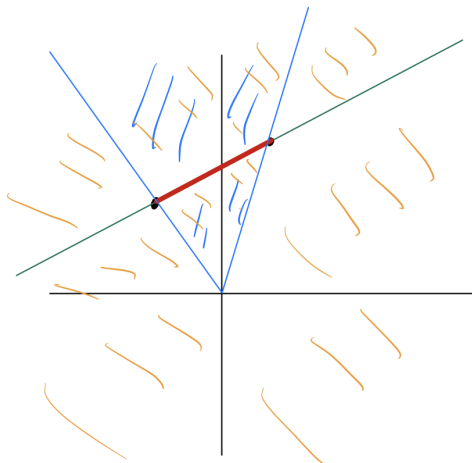


Figure 1.4: Comparing the linear subspace, the affine subspace, the convex hull, and the conic hull

Theorem 1.9. An affine subspace is a translation of a linear subspace. For an affine subspace $V \subseteq \mathbb{R}^d$, there exist matrices A and b such that $V = \{x \in \mathbb{R}^d : Ax = b\}$.

1.3 Examples

We saw that the convex hull and conic hull of a set are convex and that the linear subspace and affine subspace spanned by a set are convex. There are many more examples.

1. Empty set, singletons (sets of the form $\{v\}$),
2. Norm ball: $\{x \in \mathbb{R}^d : \|x - c\| \leq r\}$ where c is the center.
3. Ellipsoid: $\{x \in \mathbb{R}^d : (x - c)^\top P(x - c) \leq 1\}$ where P is positive definite and c is the center.
4. Hyperplane: $\{x \in \mathbb{R}^d : a^\top x = b\}$ where $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$.
5. Half-space: $\{x \in \mathbb{R}^d : a^\top x \leq b\}$ where $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$.
6. Polyhedron: A *polyhedron* is a finite intersection of half spaces, $\{x \in \mathbb{R}^d : Ax \leq b\}$ where $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. Here, $Ax \leq b$ is a short-hand notation for system $a_k^\top x \leq b_k$ for $k \in [m]$.
7. Polytope: A *polytope* is a polyhedron that is bounded. Equivalently, a polytope is the convex hull of some finite set of vectors.

8. Simplex: A set of the form $\{x \in \mathbb{R}^d : 1^\top x = 1, x \geq 0\}$, which is equal to the convex hull of $e^1, \dots, e^d$, the d -dimensional unit vectors.
9. Nonnegative orthant: $\mathbb{R}_+^d = \{x \in \mathbb{R}^d : x \geq 0\}$.
10. Positive orthant: $\mathbb{R}_{++}^d = \{x \in \mathbb{R}^d : x > 0\}$.

There are examples of convex cones, which are convex as well.

1. Norm cone: $\{(x, t) \in \mathbb{R}^d \times \mathbb{R} : \|x\| \leq t\}$. When $\|\cdot\|$ is the Euclidean norm, the cone is called the *second-order cone*.
2. Positive semidefinite cone: The set of all positive semidefinite matrices of a fixed dimension.

2 Convex functions

2.1 Definition

Definition 1.10. A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is *convex* if the domain, denoted $\text{dom}(f)$, is convex and for all $x, y \in \text{dom}(f)$, 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 $f(y)$.

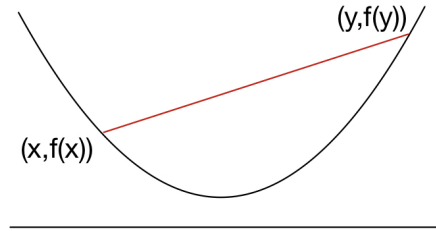


Figure 1.5: Illustration of a convex function in $\mathbb{R}^2$

Definition 1.11. We say that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is *concave* if $-f$ is convex.

Definition 1.12. A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is

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

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

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

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

2.3 Properties of convex functions

Definition 1.13. The *epigraph* of a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 1.14. Prove that f is a convex function if and only if the epigraph is a convex set.

Example 1.15. 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 1.16. A level set of a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 convex. However, the converse does not hold as Figure 1.6 demonstrates.

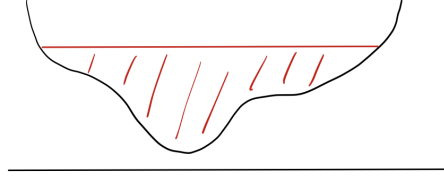


Figure 1.6: Convex level sets from a nonconvex function

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

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

Theorem 1.17. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function. Then f is convex if and only if $\text{dom}(f)$ is convex and*

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

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

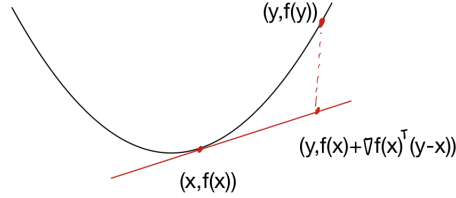


Figure 1.7: Illustration of the first-order characterization

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

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

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

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

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

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

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

Proof. ($\Rightarrow$) By Theorem 1.17, 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 1.17. □

Next, we consider the second-order characterization.

Theorem 1.19. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a twice differentiable function¹. Then f is convex if and only if $\text{dom}(f)$ is convex and

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

for all $x \in \text{dom}(f)$.

Proof. ($\Rightarrow$) We first consider the $d = 1$ case. By Theorem 1.17, 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 $y \rightarrow x$, we obtain $f''(x) \geq 0$.

Next, let us consider the general case. Let $x \in \text{dom}(f)$ and $v \in \mathbb{R}^d$. As $\text{dom}(f)$ is open, we have a sufficiently small $\epsilon > 0$ such that $x + \lambda v \in \text{dom}(f)$ 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 1.18. $\square$

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