Lecture 2 : Convex functions and sets

1 Introduction

So far, we already saw that it is necessary for a (local/global) solution $x^\star$ to be a critical point, i.e., $\nabla f(x^\star) = 0$. In general, the set of critical points of an optimization problem is strictly larger than the set of local/global minima.

In this lecture, we will investigate a family of optimiation problems where these two sets coincide. It is known as convex optimization. We will proceed two main notions in this lecture: convex sets and convex functions.

2 Convex sets

Definition 2.1 (Convex sets)**.** A set A is convex if for all elements $x, y \in A$, we have:

$$tx + (1-t)y \in A, \qquad \forall t \in [0,1]$$

Example 2.2. Below are several examples of convex (and nonconvex) sets:

1. The empty set $\emptyset$, singleton $\{x\}, x \in \mathbb{R}^d$ and the whole space $\mathbb{R}^d$ are convex.
2. Any hyperplane of the form $\{x \mid a^\top x = b\} \subseteq \mathbb{R}^d$ is convex.
3. Any halfspace of the form $\{x \mid a^\top x \leq b\} \subseteq \mathbb{R}^d$ is convex.
4. The sphere $\{x \mid \|x\|^2 = 1\} \subseteq \mathbb{R}^d$ is not convex.

Definition 2.3 (Convex hull of a set)**.** The convex hull of a set A is the set of all convex combination of finite points of A, i.e.,

$$\mathbf{conv}(A) = \{\theta_1 x_1 + \ldots + \theta_k x_k \mid x_i \in C, \theta_i \geq 0, \sum_{i=1}^{k} \theta_k = 1\}.$$

Proposition 2.4 (Convex hull of a set is convex)**.** *The convex hull of a set A is the smallest convex set containing A.*

Proof. For any pair $(x_1, x_2) \in \mathbf{conv}(A) \times \mathbf{conv}(A)$, we want to prove that:

$$tx_1 + (1-t)x_2 \in \mathbf{conv}(A), \quad \forall t \in [0,1].$$

WLOG, one can assume that:

$$x_i = \sum_{j=1}^{k} \theta_{i,j} a_j, \qquad \theta_{i,j} \geq 0, \forall 1 \leq j \leq k, \quad \sum_{j=1}^{k} \theta_{i,j} = 1.$$

Then,

$$\begin{aligned} tx_1 + (1-t)x_2 &= t\left(\sum_{j=1}^{k} \theta_{1,j} a_j\right) + (1-t)\left(\sum_{j=1}^{k} \theta_{2,j} a_j\right) \\ &= \sum_{j=1}^{k} \underbrace{(t\theta_{1,j} + (1-t)\theta_{2,j})}_{\theta_j} a_j \end{aligned}$$

where $\theta_j \geq 0$ and $\sum_{j=1}^{k} \theta_j = t(\sum_{j=1}^{k} \theta_{1,j}) + (1-t)(\sum_{j=1}^{k} \theta_{2,j}) = t + (1-t) = 1$. By definition, $tx_1 + (1-t)x_2$ is an element of $\mathbf{conv}(A)$. Therefore, $\mathbf{conv}(A)$ is convex.

The next claim is proved by using the following result (which will be discussed in the tutorial session).

Lemma 2.5 (Convex combination)**.** *If A is a convex set, then any convex combination of finite points of A is an element of A.*

To this end, assume that B to be a convex set that contains A. By the previous lemma, we have: $\mathbf{conv}(A) \subseteq B$. Hence, the claim is proved. □

Example 2.6. Using Proposition 2.4, we can construct even more convex sets.

1. The hypercube $\{x \mid 0 \leq x_i \leq 1\}$ is the convex hull of $A = \{0,1\}^d$.
2. The ball $\{x \mid \|x\| \leq 1\}$ is the convex hull of the sphere of the same dimension.
3. The probability simplex $\{x \mid x_i \geq 0, \sum_i x_i = 1\}$ is the convex hull of $A = \{e_i \mid i = 1, \ldots, d\} \cup \{0\}$.

Proposition 2.7 (Operation preserving convexity)**.** *The following operations preserve convexity:*

1. *(Infinite) intersection of convex sets is convex.*
2. *Affine transformation of a convex set A, i.e., a set of the form $\{\mathbf{A}x + b \mid x \in A\}$, is convex.*

Proof. We prove two claims one by one:

1. Consider $S = \cup_{i \in \mathcal{I}} S_i$ where $S_i, i \in \mathcal{I}$ are convex sets. For any pair $(x,y) \in S \times S$, we also have $(x,y) \in S_i \times S_i, \forall i \in \mathcal{I}$. Since S_i is convex (by assumption), we have:

$$tx + (1-t)y \in S_i, \forall i \in \mathcal{I}, \forall t \in [0,1].$$

Therefore, $tx + (1-t)y \in S$ and S is convex.

2. By the fact that $t(\mathbf{A}x + b) + (1-t)(\mathbf{A}y + b) = \mathbf{A}(tx + (1-t)y) + b$.

□

Example 2.8. We revisit previous examples and prove their convexity by Proposition 2.7:

1. The hypercube $\{x \mid 0 \leq x_i \leq 1\}$ is the intersection of the half spaces $x_i \geq 0$ and $x_i \leq 1, i = 1, \ldots, d$.

2. The ball $\{x \mid \|x\| \leq 1\}$ is the intersection of the the half spaces $\{x \mid x^\top d \geq 1\}, d \in \mathbb{R}^d, \|d\| = 1$.
3. The probability simplex $\{x \mid x_i \geq 0, \sum_i x_i = 1\}$ is the intersection of the half spaces $x_i \geq 0$ and $\sum_i x_i \leq 1$.
4. The projection of a convex set to some linear subspaces is convex.
5. Scaling or translation of a convex set is convex.

Remark 2.9. In fact, the examples demonstrates the following result (not proved): any closed convex set is equal to the intersection of (infinite) half spaces.

3 Convex functions

Definition 3.1 (Convex functions)**.** A function $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if its domain A is convex and for all $x, y \in A$, we have:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y), \quad \forall t \in [0,1].$$

Example 3.2. Below are examples of convex functions:

1. Affine function: $f(x) = a^\top x + b$.
2. Any norm function: $f(x) = \|x\|$.
3. ReLU: $f : \mathbb{R} \to \mathbb{R} : x \mapsto \max(x, 0)$.

In the following, we present three equivalent definitions of convex functions $f : \mathbb{R}^d \to \mathbb{R}$.

Definition 3.3 (Equivalent definitions of convex functions)**.** Three following conditions are equivalent:

1. $f : \mathbb{R}^d \to \mathbb{R}$ is convex.
2. If f is C^1: $f(y) \geq f(x) + \nabla f(x)^\top (y - x), \forall x, y \in \mathbb{R}^d$.
3. If f is C^2: $\nabla^2 f(x) \succeq 0$.

Proof. We prove that $1 \iff 2$ and $2 \iff 3$.

1. If f is C^1: Since $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$, we have:

$$\frac{f(x + t(y-x)) - f(x)}{t} \leq f(y) - f(x), \forall x, y \in \mathbb{R}, \forall t \in (0,1].$$

Taking the limit of the LHS, we obtain:

$$\nabla f(x)^\top (y - x) \leq f(y) - f(x).$$

The converse can be proved by taking the weighted sum of the two following equations:

$$f(x) \geq f(tx + (1-t)y) + (1-t)\nabla f(tx + (1-t)y)^\top (x - y).$$
$$f(y) \geq f(tx + (1-t)y) - t\nabla f(tx + (1-t)y)^\top (x - y).$$

2. If f is C^2: By contradiction, we suppose that $\nabla^2 f(x)$ is not positive semidefinite. By consequence, there exists a vector d of norm 1 such that:

$$c := d^\top \nabla^2 f(x) d < 0.$$

We consider the Taylor formula in a neighborhood of x, we get:

$$f(\underbrace{x+td}_{y}) = f(x) + \nabla f(x)^\top (y-x) + t^2 \underbrace{d^\top \nabla^2 f x) d}_{c} + o(t^2)$$

This equation shows that for sufficiently small t, we have: $f(y) < f(x) + \nabla f(x)^\top (y-x)$, which contradicts the second characterization of the convexity of f.

The converse can be proved by using an alternative form of the Taylor: Given a function $g : \mathbb{R} \to \mathbb{R}$ of class C^2 and two points $x, y \in \mathbb{R}$, there exists $z \in [x, y]$ such that:

$$g(y) = g(x) + g'(x)(y-x) + \frac{1}{2} g''(z)(y-x)^2.$$

In the general case $\mathbb{R}^d$, for two points $x, y \in \mathbb{R}^d$, we consider the following C^2 function:

$$g(t) = f(ty + (1-t)x) = f(x + t(y-x)).$$

By applying the previous form, we obtain:

$$\begin{aligned} f(y) = g(1) &= g(0) + g'(0) + \frac{1}{2} g''(z) \qquad \text{for certain } z \in [0,1] \\ &= f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla f(z)(y-x) \\ &\geq f(x) + \nabla f(x)^\top (y-x). \end{aligned}$$

□

Example 3.4. All these properties can be used to verfiy the convexity of a function. Consider examples such as $f(x) = \|x\|_2^2$ or $f(x) = \|Ax - b\|_2^2$.

Theorem 3.5 (Optimial conditions of a convex function)**.** *Consider the unconstrained optimization problem:*

$$\underset{x \in \mathbb{R}^d}{\textit{Minimize}} \quad f(x)$$

where f is C^1 and convex. A point x is a global solution of $f(x)$ if and only if $\nabla f(x) = 0$, i.e., x is a critical point of f.

Proof. If x is a global solution, then x has to be a critical point of f.

Conversely, if $\nabla f(x) = 0$, by applying the second characterization of the convexity of f, we have:

$$f(y) \geq f(x) + \nabla f(x)^\top (y-x) = f(x), \forall y \in \mathbb{R}^d,$$

which shows that x is indeed a global solution of f. □

Remark 3.6. A related notion of convexity is concavity. A function f is concave if $-f$ is convex. People usually consider minimizing a convex function and maximizing a concave one.

Lemma 3.7 (Operation preserving the convexity)**.** *The following operations preserve the convexity:*

1. *If f is convex, then αf is also convex ($\alpha > 0$).*

2. *If f and g are convex, then $f + g$ is also convex.*
3. *If $f_i, i = 1, \ldots, n$ are convex and $\alpha_i \geq 0, \forall i = 1, \ldots, n$, then $\sum_i \alpha_i f_i$ is also convex.*
4. *If f, g are convex, then $\max(f, g)$ is convex.*
5. *If f is convex, then $g(x) = f(Ax + b)$ is also convex.*
6. *if $f : \mathbb{R}^d \to \mathbb{R}$ is convex and $g : \mathbb{R} \to \mathbb{R}$ is non-decreasing and convex, then $g \circ f$ is convex.*