Serie 1

Optimal transport, Fall semester

EPFL, Mathematics section, Dr. Xavier Fernández-Real

Remark: Any reference made to equations or statements (theorems, propositions, lemmas, etc) in the series of exercises refer to the book followed by this course:

A. Figalli, F. Glaudo, *An Invitation to Optimal Transport, Wasserstein Distances and Gradient Flows.*

The notion of convex function is fundamental in the course, since one of the main results of the theory represents optimal transport maps as gradients of convex functions, under suitable assumptions. For this reason, we devote the first exercise sheet to deduce useful properties of convex functions.

Definition 1. A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is said to be convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \qquad \forall \lambda \in [0,1].$$

We recall hereafter some basic properties of convex functions. They should be known to the students from bachelor courses in analysis and may be taken from granted.

- Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Then, for every finite collection of points $x_1, \ldots, x_n \in \mathbb{R}^d$, and any choice of $\lambda_1, \ldots, \lambda_n \in [0,1]$ with $\lambda_1 + \cdots + \lambda_n = 1$, we have:

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n).$$

- $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if its epigraph $\{(x,y) \in \mathbb{R}^{d+1} : y \geq f(x)\}$ is convex.
- $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if it can be written as the supremum of affine functions[1].
- If $f : \mathbb{R}^d \to \mathbb{R}$ is convex, then it satisfies the monotonicity of difference quotients, namely: for any triple of pairwise distinct points $x, y, z \in \mathbb{R}^d$ such that $y \in \{(1-\lambda)x + \lambda z : \lambda \in [0,1]\}$, we have

$$\frac{f(y) - f(x)}{|y-x|} \leq \frac{f(z) - f(x)}{|z-x|} \leq \frac{f(z) - f(y)}{|z-y|}.$$

  In particular, if f is C^1, then, for every direction $e \in \mathcal{S}^{d-1}$, $\partial_e f$ is non-decreasing in direction e.
- A C^2 function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if, for every $x \in \mathbb{R}^d$, $D^2 f(x)$ is a nonnegative definite matrix.

[1] If you did not see this fact before, you can prove it with the following hint: By the assumption and the first bullet, we know that the epigraph of φ is convex. We need to show that, for any point (x,y) with $y < f(x)$, there is a line passing through (x,y) and which lies below x. To this end, take the point at minimal distance to (x,y) in the epigraph (why does it exist?) and consider the plane passing from (x,y) and perpendicular to the segment which realizes the minimal distance.

Exercise 1.1 (Convex functions are locally Lipschitz)**.** Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Let $D := \{f < +\infty\}$ be its finiteness domain. Show that f is locally Lipschitz in the interior of D. More precisely, for each ball $B_R(x_0)$ compactly contained in $\mathrm{int}(D)$ show the following:

(i) f is bounded in $B_R(x_0)$.

(ii) For every $0 < r < R$ and for every $x, y \in B_r(x_0)$,

$$|f(x) - f(y)| \leq \frac{\sup_{B_R(x_0)} f - \inf_{B_R(x_0)} f}{R - r} |x - y|.$$

Hints: For (i), you may assume without loss of generality that the hypercube $Q_{2R}(x_0)$ with side length $2R$ centered in x_0 is compactly supported in $\mathrm{int}(D)$. Deduce from the finiteness of f in the vertices of $Q_{2R}(x_0)$ that f is bounded from above in $B_R(x_0)$. The bound from below follows from a characterization described before the exercise. To prove (ii) use appropriately the monotonicity of difference quotients.

By the previous exercise and the Rademacher's Theorem stated below, we deduce that convex functions are almost everywhere differentiable on their finiteness domain. This fact will be used in the course, for instance to give a meaning to Brenier theorem.

Theorem 1 (Rademacher)**.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz function. Then the set of points where f is not differentiable is negligible for the Lebesgue measure.

Second differentiability results are also known. In fact, for a general convex function $f : \mathbb{R}^d \to \mathbb{R}$ it can be proven that f is almost everywhere twice differentiable (Alexandrov's Theorem), moreover, in the sense of distributions, $D^2 f$ turns out to be a matrix-valued nonnegative measure. This complementary material won't be proved during the course, but the interested student is invited to ask for a proof of these results in the form of a guided exercise.

Definition 2. Given $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convex, we define the subdifferential of f at $x \in \mathbb{R}^d$ as

$$\partial f(x) = \{y \in \mathbb{R}^d : f(z) \geq f(x) + \langle y, z - x\rangle \ \forall z \in \mathbb{R}^d\}.$$

Exercise 1.2. Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function, $D = \{f < +\infty\}$ be its finiteness domain and $x_0 \in \mathrm{int}(D)$.

(i) Show that $\partial f(x_0)$ is not empty.

(ii) Show that $\partial f(x_0)$ is a closed convex set.

(iii) Compute the subdifferential of the following functions defined on $\mathbb{R}^2$:

$$f_1(x, y) = \sqrt{x^2 + y^2}, \qquad\qquad f_2(x, y) = |x - 1| + \frac{|x + 1|}{2}.$$

(iv) Give an example of a nonconvex function $f : \mathbb{R} \to \mathbb{R}$ whose subdifferential is empty at every point $x \in \mathbb{R}$. Can you make it C^1?

Hint: To prove point (i) it may be useful to recall the Hahn-Banach Theorem (first geometric form): Let $A, B \subset \mathbb{R}^n$ be two disjoint convex sets, with A open. Then there exists an hyperplane which separates A and B. More precisely, there exists a vector $v \in \mathbb{R}^n$ and a number $\alpha \in \mathbb{R}$ for which

$$\langle v, x\rangle \le \alpha \le \langle v, y\rangle \quad \text{for every } x \in A \text{ and every } y \in B.$$

To show the existence of an element in the subdifferential, choose appropriately A and B!

Exercise 1.3 (Monotonicity of the subdifferential)**.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Show the following facts:

(i) For every $x_1, x_2 \in \mathbb{R}^d$ and every $\xi_1 \in \partial f(x_1), \xi_2 \in \partial f(x_2)$ we have[2]:

$$\langle \xi_2 - \xi_1, x_2 - x_1\rangle \ge 0.$$

(ii) For every collection of points $x_1, \dots, x_n \in \mathbb{R}^d$, given any $\xi_1 \in \partial f(x_1), \dots, \xi_n \in \partial f(x_n)$, and any permutation σ of $\{1, \dots, n\}$, we have

$$\sum_{i=1}^{n} \langle \xi_i, x_i\rangle \ge \sum_{i=1}^{n} \langle \xi_{\sigma(i)}, x_i\rangle.$$

Exercise 1.4 (A characterization of differentiability)**.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Prove the following facts:

(i) If f is differentiable at a point x, then $\partial f(x) = \{\nabla f(x)\}$.

(ii) (💣) If $\partial f(x)$ is a singleton, then f is differentiable at x.

Hint: To prove (ii) you can assume that $x = 0$, $f(0) = 0$ and $\partial f(0) = \{0\}$. Then argue by contradiction: suppose that f is not differentiable at 0 and find some direction along which f grows linearly. Then use the Hahn-Banach Theorem to obtain a non-zero element in the subdifferential.

Exercise 1.5 ((💣) Boundedness and continuity of the subdifferential)**.** Given a convex function $f : \mathbb{R}^d \to \mathbb{R}$ and a set $E \subset \mathbb{R}^d$, we call

$$\partial f(E) := \bigcup_{x \in E} \partial f(x).$$

(i) Prove that if $f, g : \mathbb{R}^d \to \mathbb{R}$ are convex functions and $\Omega \subset \mathbb{R}^d$ is a bounded open set such that $f = g$ on $\partial\Omega$ and $f \le g$ in Ω, then

$$\partial g(\Omega) \subseteq \partial f(\Omega).$$

(ii) Prove that ∂f is locally bounded, that is to say, for every bounded set $E \subset \mathbb{R}^d$, $\partial f(E)$ is also bounded.

[2]This property can also be restated by saying that the multi-valued function ∂f is "monotone".

(iii) Given a sequence $x_j \in \mathbb{R}^d$ and $\xi_j \in \partial f(x_j)$, assume that $x_j \to x$. Prove that up to subsequences, $\xi_j \to \xi$, for some $\xi \in \partial f(x)$.

(iv) Prove that f is C^1 if and only if $\partial f(x)$ is a singlet for every $x \in \mathbb{R}^d$.

Hints: For i) take any hyperplane touching g from below at some point $x \in \Omega$ and translate it down vertically until it touches f from below for the first time. For ii) compare f locally with a suitably chosen convex paraboloid and use point i).

Exercise 1.6 ((💣) Extended gradient and descending slope)**.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function.

(i) Prove that for every $x \in \mathbb{R}^d$ there exists a unique vector $\xi \in \partial f(x)$ with minimal norm. Such vector is often called the "extended gradient" of f at x. In this exercise we will denote it by $\overline{\nabla} f(x)$.

(ii) Prove that x is a minimum of f if and only if $\overline{\nabla} f(x) = 0$.

(iii) Prove that the modulus of the extended gradient equals the so called "descending slope" of f:

$$|\overline{\nabla} f(x)| = \sup_{y \neq x} \frac{[f(x) - f(y)]^+}{|y - x|},$$

where the superscript "+" stands for the positive part, i.e. $a^+ := \max\{a, 0\}$. Deduce from it that the map $x \mapsto |\overline{\nabla} f(x)|$ is lower semi-continuous.

Hints: In point iii), to prove the $\leq$ inequality it may be a good idea to use the Hahn-Banach Theorem as follows. Call m the descending slope of f at x and prove that the following convex subsets of $\mathbb{R}^d \times \mathbb{R}$ are disjoint:

$$A := \{(y, r) \in \mathbb{R}^d \times \mathbb{R} : r < -m|y - x|\}, \quad B := \{(y, r) \in \mathbb{R}^d \times \mathbb{R} : f(y) - f(x) \leq r\}.$$

Deduce that there exists an hyperplane in $\mathbb{R}^d \times \mathbb{R}$ which separates A and B. Finally, find a vector $\xi \in \partial f(x)$ such that $|\xi| \leq m$.

Exercise 1.7. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function in one space dimension. Then there exists a countable set $Z \subset \mathbb{R}$ such that f is differentiable at x for every $x \in \mathbb{R} \setminus Z$.
Hint: Consider the map that associates to each point $x \in \mathbb{R}$ the length of $\partial f(x)$.