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 $x_1, \dots, x_n \in \mathbb{R}^d$, and any choice of $\lambda_1, \dots, \lambda_n \in [0,1]$ 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 $\operatorname{int}(D)$ 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 $\operatorname{int}(D)$. 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.

Solution:

(i) Assume without loss of generality that the hypercube $Q_{2R}(x_0)$ is compactly contained in $\operatorname{int}(D)$ and call $\{v_k\}_{k=1}^N$ its vertices ($N = 2^d$). Then clearly $M := \max_{k=1,\dots,N} |f(v_k)| < +\infty$ as the cube is compactly contained in $\operatorname{int}(D)$. Note that any point $x \in B_R(x_0)$ can be written as a convex combination of $\{v_k\}_{k=1}^N$, i.e. there are $\{\lambda_k\}_{k=1}^N$ with $\lambda_k \in [0, 1]$, such that $\sum_{k=1}^N \lambda_k = 1$ and $x = \sum_{k=1}^N \lambda_k v_k$. Now, since f is convex,

$$f(x) = f\left(\sum_{k=1}^N \lambda_k v_k\right) \leq \sum_{k=1}^N \lambda_k f(v_k) \leq M.$$

This proves that f is bounded from above in $B_R(x_0)$. To prove the bound from below, notice that since f can be written as the supremum of affine functions, in particular there must exist $c \in \mathbb{R}$ and $v \in \mathbb{R}^n$ such that

$$f(x) \geq c + \langle v, x \rangle.$$

Now, clearly the affine function on the right-hand side is bounded below in $B_R(x_0)$ by some constant $m \in \mathbb{R}$. We conclude that $m \leq f \leq M$ on $B_R(x_0)$.

(ii) Without loss of generality assume that $f(x) < f(y)$. Let z be the intersection between the half line starting at x passing through y and $\partial B_R(x_0)$. By the monotonicity of difference quotients, we have

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

Since $z \in \partial B_R(x_0)$ and $x \in B_r(x_0)$, $|z - x| > R - r$. In addition, $f(z) \leq \sup_{B_R(x_0)} f$, and $f(x) \geq \inf_{B_R(x_0)} f$. Therefore we conclude that

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

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 $x_0 \in \operatorname{int}(D)$.

(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 $\mathbb{R}^2$:

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

(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

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

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

Solution:

(i) Let us start proving that $\partial f(x_0)$ is not empty. Define $A = \{(x, y) \in \operatorname{int}(D) \times \mathbb{R} : y > f(x)\}$ which is clearly convex and open, since f is convex and hence continuous in $\operatorname{int}(D)$. Notice that $(x_0, f(x_0)) \notin A$, and therefore by the first geometric form of the Hahn-Banach theorem, there is a hyperplane that separates A and $(x_0, f(x_0))$. Now we need to exclude the case where the hyperplane is of the form $\{(x, y) \in \mathbb{R}^d \times \mathbb{R} : \langle w, x\rangle_{\mathbb{R}^d} = c\}$ for some $w \in \mathbb{R}^d \setminus \{0\}$

and $c \in \mathbb{R}$ (that is, it is vertical). Let us argue by contradiction. Assume that it does take this form. Since $\mathrm{int}(D)$ is open, there is $r > 0$ such that $B_r(x_0) \subset \mathrm{int}(D)$. Then there are $x_1, x_2 \in B_r(x_0)$ for which

$$\langle w, x_1 \rangle_{\mathbb{R}^d} < c \quad \text{and} \quad \langle w, x_2 \rangle_{\mathbb{R}^d} > c.$$

But then the hyperplane strictly separates $(x_1, f(x_1)+1)$ and $(x_2, f(x_2)+1)$, although both belong to A. This gives a contradiction. Hence the hyperplane separating A and $(x_0, f(x_0))$ is not vertical and takes the form $\{(x,y) \in \mathbb{R}^d \times \mathbb{R} : \langle w, x \rangle_{\mathbb{R}^d} + y = c\}$ for some $w \in \mathbb{R}^d$. Since $(x_0, f(x_0))$ is in the closure of A,

$$\langle w, x_0 \rangle_{\mathbb{R}^d} + f(x_0) = c \leq \langle w, x \rangle_{\mathbb{R}^d} + y \quad \text{for every } (x,y) \in A. \tag{1}$$

Taking $x \in \mathrm{int}(D)$ and $y = f(x) + \epsilon$, for an arbitrary $\epsilon > 0$ in the formula above we get

$$f(x) \geq f(x_0) + \langle w, x - x_0 \rangle_{\mathbb{R}^d} - \epsilon \quad \text{for every } \epsilon > 0,$$

from which we deduce that $w \in \partial f(x_0)$.

(ii) The convexity of $\partial f(x_0)$ follows from a simple computation. Let $y, z \in \partial f(x)$ and $t \in [0,1]$. Then for all $x' \in \mathbb{R}^d$ we have

$$\begin{aligned} f(x') = (1-t)f(x') + tf(x') &\geq (1-t)(f(x) + \langle y, x' - x \rangle) + t(f(x) + \langle z, x' - x \rangle) \\ &= f(x) + \langle (1-t)y + tz, x' - x \rangle. \end{aligned}$$

Hence $(1-t)y + tz \in \partial f(x_0)$. The closedness of $\partial f(x_0)$ is deduced immediately from the definition of subdifferential and the fact that the scalar product is continuous.

(iii) We have

$$\partial f_1(x,y) = \begin{cases} \left\{ \left(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}} \right) \right\} & \text{if } (x,y) \neq (0,0); \\ \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\} & \text{if } (x,y) = (0,0), \end{cases}$$

and

$$\partial f_2(x,y) = \begin{cases} \{(-\frac{3}{2}, 0)\} & \text{if } x < -1; \\ [-\frac{3}{2}, -\frac{1}{2}] \times \{0\} & \text{if } x = -1; \\ \{(-\frac{1}{2}, 0)\} & \text{if } -1 < x < 1; \\ [-\frac{1}{2}, \frac{3}{2}] \times \{0\} & \text{if } x = 1; \\ \{(\frac{3}{2}, 0)\} & \text{if } x > 1. \end{cases}$$

(iv) Take $f(x) = -x^2$. Let x_0 be arbitrary and assume for a contradiction that there is v such that $v \in \partial f(x_0)$, then

$$v(x - x_0) \leq x_0^2 - x^2 = (x_0 - x)(x_0 + x) \quad \forall x \in \mathbb{R},$$

so that for $x > x_0$, we get $v \leq -(x_0 + x)$. Letting $x \to \infty$, we deduce a contradiction. We conclude that the subdifferential is empty at every point.

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]:

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

(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

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

Solution:

(i) By the definition of subdifferential, we have

$$f(x_2) \geq f(x_1) + \langle \xi_1, x_2 - x_1 \rangle,$$
$$f(x_1) \geq f(x_2) + \langle \xi_2, x_1 - x_2 \rangle.$$

Summing the two inequalities above and rearranging terms we get precisely the monotonicity condition $\langle \xi_2 - \xi_1, x_2 - x_1 \rangle \geq 0$.

(ii) Let us first assume that σ is a cycle, that is to say

$$\{1, \dots, n\} = \{\sigma(1), \sigma^2(1), \dots, \sigma^n(1)\}.$$

By the definition of subdifferential we know that

$$f(x_{\sigma^i(1)}) \geq f(x_{\sigma^{i+1}(1)}) + \langle \xi_{\sigma^{i+1}(1)}, x_{\sigma^i(1)} - x_{\sigma^{i+1}(1)} \rangle \qquad \text{for every } i \in \{1, \dots, n\}.$$

Summing the above inequalities over $i \in \{1, \dots, n\}$ we get

$$\sum_{i=1}^{n} \langle \xi_{\sigma^{i+1}(1)}, x_{\sigma^{i+1}(1)} \rangle \geq \sum_{i=1}^{n} \langle \xi_{\sigma^{i+1}(1)}, x_{\sigma^i(1)} \rangle,$$

which, using the fact that σ is a cycle, can also be rewritten as

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

Once we have proven the result for cyclic permutations, the general case can be obtained by noticing that each permutation is the union of disjoint cycles.

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

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.

Solution:

(i) Remember the definition of differentiability: f is differentiable at x if there exists a vector $\nabla f(x) \in \mathbb{R}^d$ such that

$$\frac{|f(x+v) - f(x) - \langle \nabla f(x), v\rangle|}{|v|} \to 0 \quad \text{as } |v| \to 0.$$

The above condition can be rewritten in a compact form as

$$f(x+v) = f(x) + \langle \nabla f(x), v\rangle + o(|v|),$$

where we recall that the "little-o notation" $g = o(h)$ stands for $|g|/h \to 0$. Suppose that f is differentiable at x. Then, for every $y \neq x$ and $t \in (0,1)$, from the monotonicity of difference quotients:

$$f(y) - f(x) \geq \frac{f(x + t(y-x)) - f(x)}{t|y-x|}|y-x| = \langle \nabla f(x), y - x\rangle + o(1) \quad \text{as } t \to 0^+.$$

Hence, sending t to 0 we get $f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle$ for every $y \in \mathbb{R}^d$, which means that $\nabla f(x) \in \partial f(x)$. Now we want to prove that indeed $\partial f(x) = \{\nabla f(x)\}$. Given $\xi \in \partial f(x)$, by definition of subdifferential for every $v \in \mathbb{R}^d$,

$$f(x) + \langle \xi, v\rangle \leq f(x+v) = f(x) + \langle \nabla f(x), v\rangle + o(|v|).$$

Rearranging terms we get

$$\langle \xi - \nabla f(x), v\rangle \leq o(|v|).$$

Choosing $v = t(\xi - \nabla f(x))$ for $t > 0$ we obtain

$$t|\xi - \nabla f(x)|^2 \leq o(t) \quad \text{as } t \to 0^+$$

which implies that $\xi = \nabla f(x)$.

(ii) Up to a translation and the subtraction of the affine function $f(x) + \langle \xi, \cdot\rangle$, we may assume that $x = 0$, $f(0) = 0$ and $\partial f(0) = \{0\}$, so that we need to prove that

$$\frac{f(y)}{|y|} \to 0 \quad \text{as } y \to 0.$$

Suppose by contradiction that there exists a sequence $y_j \neq 0$, such that $y_j \to 0$ and

$$\frac{f(y_j)}{|y_j|} \geq \epsilon > 0 \quad \text{for every } j \in \mathbb{N}.$$

Up to subsequences we may assume that $y_j/|y_j| \to e \in \mathcal{S}^{d-1}$. By the monotonicity of difference quotients, for every $r > 0$ and every j large enough we have

$$\frac{f(ry_j/|y_j|)}{r} \geq \frac{f(y_j)}{|y_j|} \geq \epsilon,$$

and sending j to infinity, by the continuity of f we deduce

$$f(re) \geq r\epsilon \quad \text{for every } r > 0.$$

This means that the open convex set

$$A := \{(y,t) \in \mathbb{R}^d \times \mathbb{R} : t > f(y)\}$$

and the convex set

$$B := \{(re, r\epsilon) : r > 0\}$$

are separated in $\mathbb{R}^d \times \mathbb{R}$. Now the geometric form of Hahn-Banach Theorem provides us with an hyperplane touching A from below and containing B. This contradicts the fact that 0 is the unique element in the $\partial f(0)$, since $\epsilon > 0$, thus concluding the proof.

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 $f \leq g$ 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.

(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).

Solution:

(i) Let $x \in \Omega$. Take any $\xi \in \partial g(x)$. We wish to prove that $\xi \in \partial f(z)$ for some $z \in \Omega$. For every number $\alpha \in \mathbb{R}$, we consider the α-translation of the affine function $f(x) + \langle \xi, \cdot - x\rangle$,

$$\varphi_\alpha(y) := f(x) + \langle \xi, y - x\rangle + \alpha.$$

Notice that by the local boundedness of convex functions, for α small enough, $\varphi_\alpha < f$ in $\overline{\Omega}$. Hence, it is well-defined the minimum α for which φ_α touches f from below on $\overline{\Omega}$, i.e.

$$\alpha_0 := \sup\{\alpha \in \mathbb{R} : \varphi_\alpha < f \text{ in } \overline{\Omega}\}.$$

Notice that since φ_0 touches g from below and $g \geq f$ in Ω, we must have $\alpha_0 \leq 0$. Let $x_0 \in \overline{\Omega}$ be a point at which $\varphi_{\alpha_0}(x_0) = f(x_0)$. If $x_0 \in \Omega$, then $\xi \in \partial f(x_0)$ and we are done. Suppose instead that $x_0 \in \partial\Omega$. Then, since $f(x_0) = g(x_0)$, it must be the case that $\alpha_0 = 0$ and φ_0 touches also f from below at x, which means that $\xi \in \partial f(x)$ concluding again.

(ii) Consider the convex paraboloid $\varphi(x) = \alpha + \beta|x - x_0|^2$, for some $\alpha \in \mathbb{R}$ and $\beta > 0$. Since f is locally bounded above and below in $B_{2r}(x_0)$, we may choose α small enough and β large enough so that $f < \varphi$ in $\mathbb{R}^d \setminus B_{2r}(x_0)$ and $f > \varphi$ in $B_r(x_0)$. In particular the open set $\Omega := \{f > \varphi\}$ satisfies

$$\begin{cases} B_r(x_0) \subset \Omega \subset B_{2r}(x_0), \\ f = \varphi \quad \text{on } \partial\Omega, \\ f > \varphi \quad \text{in } \Omega. \end{cases}$$

Thus, from point i) we deduce that $\partial f(B_r(x_0)) \subset \partial\varphi(B_{2r}(x_0)) = B_{4\beta r}(0)$.

(iii) Since x_j is a bounded sequence, by point ii) also ξ_j is bounded, and so, up to subsequences $\xi_j \to \xi$. We only need to show that $\xi \in \partial f(x)$. Now, since $\xi_j \in \partial f(x_j)$, we have

$$f(y) \geq f(x_j) + \langle \xi_j, y - x_j\rangle \quad \text{for every } j \text{ and every } y \in \mathbb{R}^d.$$

Sending j to ∞, and using the convergences $x_j \to x$, $\xi_j \to \xi$ and $f(x_j) \to f(x)$ we get

$$f(y) \geq f(x) + \langle \xi, y - x\rangle \quad \text{for every } y \in \mathbb{R}^d$$

which means that $\xi \in \partial f(x)$, as desired.

(iv) If $f \in C^1$, in particular it is differentiable at each point, thus, from a previous exercise, $\partial f(x) = \{\nabla f(x)\}$ for every $x \in \mathbb{R}^d$. Conversely, if $\partial f(x)$ is a singlet for every $x \in \mathbb{R}^d$, then f is differentiable at each point and $\partial f(x) = \{\nabla f(x)\}$, so that we only need to prove that ∇f is a continuous function, but this easily follows from point iii).

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 $|\xi| \leq m$.

Solution:

(i) $\overline{\nabla} f(x)$ is simply obtained as the unique projection of 0 on the closed convex set $\partial f(x)$.

(ii) It is enough to notice that x is a minimum point for f if and only if $0 \in \partial f(x)$.

(iii) We prove the stated formula. The fact that $x \mapsto |\overline{\nabla} f(x)|$ is lower semi-continuous then follows from its representation as the supremum of continuous functions. First notice that, by $\overline{\nabla} f(x) \in \partial f(x)$, for every $y \neq x$,

$$f(x) - f(y) \leq \langle -\overline{\nabla} f(x), y - x \rangle \leq |\overline{\nabla} f(x)||y - x|.$$

Hence

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

Now we wish to prove the opposite inequality. Call

$$m := \sup_{y \neq x} \frac{[f(x) - f(y)]^+}{|y - x|}.$$

We want to prove that there exists an element $\xi \in \partial f(x)$ such that $|\xi| \leq m$. Observe that by definition of m we have

$$f(y) - f(x) \geq -m|y - x| \quad \text{for every } y \in \mathbb{R}^d.$$

Thus, the open convex set

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

and the closed convex set

$$B := \{(y, r) \in \mathbb{R}^d \times \mathbb{R} : f(y) - f(x) \leq r\}$$

are disjoint in $\mathbb{R}^d \times \mathbb{R}$. From the geometric form of Hahn-Banach theorem, we deduce that there exist an hyperplane in $\mathbb{R}^d \times \mathbb{R}$ separating A and B. In particular, there exist $\xi \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$ such that

$$f(y) - f(x) \geq \langle \xi, y\rangle + \alpha \geq -m|y - x| \quad \text{for every } y \in \mathbb{R}^d.$$

Now, taking $y = x$ in the formula above, we derive $\alpha = -\langle \xi, x\rangle$. Therefore, from the first and the second inequality we respectively get, for every $y \in \mathbb{R}^d$:

$$f(y) - f(x) \geq \langle \xi, y - x\rangle,$$
$$\langle \xi, y - x\rangle \geq -m|y - x|.$$

The first implies that $\xi \in \partial f(x)$, while choosing $y = x - \xi$ in the second we get $|\xi| \leq m$. This concludes the proof.

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 $\partial f(x)$.

Solution: For every $x \in \mathbb{R}$, $\partial f(x)$ is a non-empty bounded closed convex subset of $\mathbb{R}$, hence

$$\partial f(x) = [a(x), b(x)]$$

for some $b(x) \geq a(x)$. The result is proven if we show that $a(x) = b(x)$ out of a countable set, as in each such point $\partial f(x) = \{a(x)\}$ which implies that f is differentiable at x thanks to a previous exercise. From the monotonicity of ∂f we know that

$$a(x) \leq b(x) \leq a(y) \leq b(y) \quad \text{for every } x < y,$$

thus, in particular

$$\text{int}(\partial f(x)) \cap \text{int}(\partial f(y)) = \emptyset \quad \text{for every } x \neq y.$$

Define the function $\phi : \mathbb{R} \to \mathbb{R}$ as

$$\phi(x) = \mathscr{L}^1(\partial f(x)) = b(x) - a(x).$$

For every $n \in \mathbb{N}$ we consider the following sets

$$Z_n := \{x \in \mathbb{R} : \partial f(x) \subset (-n, n), \phi(x) > 1/n\}$$

and we call Z their union

$$Z := \bigcup_{n\in\mathbb{N}} Z_n.$$

Observe that $a(x) = b(x)$ for every $x \in \mathbb{R} \setminus Z$, therefore we only need to prove that each Z_n is finite. Indeed, we claim that $\#Z_n < 2n^2$. In fact, if $x_1, \dots, x_N$ are pairwise distinct points in Z_n,

$$2n = \mathscr{L}^1((-n,n)) \geq \mathscr{L}^1\left(\bigcup_{i=1}^{N} \mathrm{int}(\partial f(x_i))\right) = \sum_{i=1}^{N} \mathscr{L}^1\left(\mathrm{int}(\partial f(x_i))\right) > N/n.$$