Optimization for Machine Learning
Convex sets and convex functions

The exercises indicated with a star ($\star$) are regarded as part of the syllabus. The stated results should be known, and it is highly recommended to practice by trying to solve them.

Last revised on December, 9th, 2019.

The blue text denotes the addition of a question or indication.

The red text denotes the correction of a typo.

Convex functions

($\star$) **Exercise 1** (Convexity of the level sets).
Give an example of function $f : \mathbb{R}^N \to \mathbb{R}$ which is **not convex** but such that for all $t \in \mathbb{R}$, the level set $\{f \leq t\} \stackrel{\text{def}}{=} \{ x \in \mathbb{R}^N \mid f(x) \leq t \}$ is convex.

Exercise 2 (Convex functions which take the value $-\infty$ are not very interesting).
Let $f : \mathbb{R}^N \to \mathbb{R} \cup \{\pm\infty\}$ be convex. Assume that there is some x_0 such that $f(x_0) = -\infty$.
Prove that for all direction $v \in \mathbb{R}^N \setminus \{0\}$, there is at most one value of $t \in \mathbb{R}$ such that the function $t \mapsto f(x_0+tv)$ takes a value in $\mathbb{R}$ (it can only take the value $+\infty$ or $-\infty$ for the other values of t). If, moreover, f is lower semi-continuous it can only take the values $\pm\infty$.

Exercise 3 (Smooth convex functions).
Let $U \subset \mathbb{R}^N$ be an open convex set, and $f : \mathbb{R}^N \to \mathbb{R}$ a differentiable function and denote by $\nabla f(x)$ its gradient at x.

1. Prove that the following properties are equivalent.
   a) for all $x, y \in U$, for all $\theta \in [0, 1]$, $f((1-\theta)x + \theta)y) \leq (1-\theta)f(x) + \theta f(y)$.
   b) for all $x, y \in U$, $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle$.
   c) for all $x, y \in U$, $\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq 0$.

   Hint: *Observe that it is sufficient to study the univariate function $g : t \mapsto f(x + t(x-y))$ on $[0, 1]$. What is $g'(t)$? $g'(0)$? $g'(1)$?*
2. Assume that f is twice differentiable and denote by $\nabla^2 f(x)$ its Hessian matrix at x. Show that the conditions a), b), c) are equivalent to
   $$\forall x \in \mathbb{R}^N, \quad \nabla^2 f(x) \succeq 0 \tag{1}$$
   Hint: *Observe that it is equivalent to g' being increasing ("croissante" en français)*
3. We assume now that there is a continuous function $\bar{f}$ defined on $\overline{U}$ such that $\bar{f}(x) = f(x)$ for all $x \in U$. We define $\bar{\bar{f}} : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$,
   $$\forall x \in \mathbb{R}^N, \quad \bar{\bar{f}}(x) \stackrel{\text{def}}{=} \begin{cases} \bar{f}(x) & \text{if } x \in \overline{U}, \\ +\infty & \text{otherwise.} \end{cases}$$
   Prove that $\bar{\bar{f}}$ is convex if and only if f satisfies any of the equivalent properties a), b), c) in U.

($\star$) **Exercise 4** (Stability of convex functions).
Prove the following assertions.

1. If $\alpha > 0$, and $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is convex, then αf is convex.
2. If f_1, $f_2 : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ are convex, then $f_1 + f_2$ is convex.
3. If $\{f_i\}_{i \in I}$ is a (finite or infinite) collection of convex functions, then $\sup_{i \in I} f_i$ is convex.
4. If $f : \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \to \mathbb{R} \cup \{+\infty\}$ is convex, then $g : x \mapsto \inf_{y \in \mathbb{R}^{N_2}} f(x, y)$ is convex.
5. If $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is convex and $A \in \mathbb{R}^{m \times n}$, then $g : x \mapsto f(Ax)$ is convex.

Hint: *For the points 3 to 5, provide two proofs : direct verification or argument on the epigraph.*

Exercise 5.
Are the following functions convex ?

1. $f(x) = \|Ax - b\|$, for $x \in \mathbb{R}^N$, where $A \in \mathbb{R}^{m\times N}$, $b \in \mathbb{R}^m$.
2. (ReLU) $f(x) = \max\{x, 0\}$, for all $x \in \mathbb{R}$.
3. (Quadratic over linear function) $f(x, y) = x^2/y$ for all $x, y \in \mathbb{R}$ such that $y > 0$.
4. (Log-sum-exp) $f(x) = \log(e^{x_1} + \cdots + e^{x_N})$, for all $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$.
5. (Maximal eigenvalue) $f(M) = \lambda_n(M)$ for all $M \in S_n^+(\mathbb{R})$. ***Hint:*** $\lambda_n(M) = \sup\left\{ y^\top My \mid y \in \mathbb{R}^n, \|y\| = 1 \right\}$.
6. (Sum of the k largest components) $f(x) = x_{[1]} + \cdots + x_{[k]}$ where $1 \le k \le N$ and $x_{[1]} \ge \cdots \ge x_{[N]}$ are the ordered components of $x \in \mathbb{R}^N$. ***Hint:*** *Write f as the supremum of affine functions.*

($\star$) **Exercise 6** (Continuity of convex functions).
Let $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be a convex function such that $\mathrm{Int}(\mathrm{dom}\, f) \ne \emptyset$, and let $x_0 \in \mathrm{Int}(\mathrm{dom}\, f)$. We want to prove that f is continuous at x_0.

1. Assume that there is some $\delta > 0$ such that $B(x_0, 2\delta) \subset \mathrm{Int}(\mathrm{dom}\, f)$ and that there exists $m, M \in \mathbb{R}$ such that $m \le f(x) \le M$ for all $x \in B(x_0, 2\delta)$. Prove that f is Lipschitz in $B(x_0, \delta)$, more precisely
$$\forall y, y' \in B(x_0, \delta), \quad |f(y) - f(y')| \le \frac{M - m}{\delta}\|y - y'\|.$$
2. Show that there exist $v_0, \ldots, v_{\mathbf{N}} \in \mathrm{Int}(\mathrm{dom}\, f)$, affinely independent (*i.e.* $v_1 - v_0, \ldots, v_N - v_0$ are linearly independent), such that x_0 is in the interior of $\mathrm{conv}\{v_0, \ldots, v_N\}$, the convex hull of $\{v_0, \ldots, v_N\}$.
3. Let $\delta > 0$ small enough so that $\overline{B(x_0, 2\delta)} \subset \mathrm{conv}\{v_0, \ldots, v_N\}$. Prove that there exists $M \in \mathbb{R}$ (which depends only on $f(v_0), \ldots, f(v_N)$) such that $f(x) \le M$ for all $x \in B(x_0, \mathbf{2}\delta)$.
4. Prove that there exists $m \in \mathbb{R}$ such that for all $x \in \overline{B(x_0, 2\delta)}$, $f(y) \ge m$ (for instance one may choose $m = 2f(x_0) - M$).
5. Conclude.

Note : it is possible to prove that f is locally Lipschitz on $\mathrm{Int}(\mathrm{dom}\, f)$*, i.e. it is Lipschitz on every compact subset of* $\mathrm{Int}(\mathrm{dom}\, f)$.

Subdifferential

($\star$) **Exercise 7** (The subgradient of smooth functions).
Let $f : \mathbb{R}^N \to \mathbb{R}$ be a convex differentiable function. Prove that the subdifferential is single-valued everywhere, namely $\partial f(x) = \{\nabla f(x)\}$.
Hint: *Evaluate the subgradient inequality at $y = x + h$, with h "small".*

($\star$) **Exercise 8** (Projection onto a convex set (revisited)).
Let $C \subset \mathbb{R}^N$ be a nonempty closed convex set, and $\chi_C(x) \stackrel{\mathrm{def}}{=} 0$ if $x \in C$, $+\infty$ otherwise. Let $x_0 \in \mathbb{R}^N$ and consider the problem
$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|x - x_0\|^2 + \chi_C(x).$$

1. Prove that there is a unique minimizer to that problem.
2. Let $p \in \mathbb{R}^N$, what does $p \in \partial\chi_C(x)$ mean ?
3. Using the subdifferential, recover the characterization of the projection onto C.

Exercise 9 (ℓ^1 regularization).

1. Let $f_1, \ldots, f_N : \mathbb{R} \to \{+\infty\}$ be convex proper lower semi-continuous functions. Consider the function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ defined by $f(x) = \sum_{i=1}^N f_i(x_i)$ for all $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$. Prove that
$$\forall x = (x_1, \ldots, x_N) \in \mathbb{R}^N, \quad \partial f(x) = (\partial f_1(x_1)) \times \ldots \times (\partial f_N(x_N))$$

2. Consider the function f defined by $f(x) = \|x\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{N} |x_i|$ and let $p \in \mathbb{R}^N$. Prove that $p \in \partial f(x)$ if and only if

$$\begin{cases} p_i = \operatorname{sign}(x_i) & \text{for all } i \in \{1,\dots,N\} \text{ such that } x_i \neq 0, \\ p_i \in [-1,1] & \text{for all } i \in \{1,\dots,N\} \text{ such that } x_i = 0 \end{cases}$$

3. Consider the minimization problem, for fixed $y \in \mathbb{R}^N$, and $\lambda > 0$,

$$\min_{x \in \mathbb{R}^N} \lambda \|x\|_1 + \frac{1}{2} \|x - y\|_2^2 .$$

   Prove that there is a unique solution, and that it is given by the *soft thresholding* of y,

$$\forall i \in \{1,\dots,N\}, \quad x_i = \begin{cases} y_i + \lambda & \text{if } y_i < -\lambda, \\ 0 & \text{if } -\lambda \leq y_i \leq \lambda, \\ y_i - \lambda & \text{if } y_i > \lambda. \end{cases} \tag{2}$$

Exercise 10 (The Moreau-Yosida regularization and the proximal point).
Let f be a proper convex lower semi-continuous function, and $\lambda > 0$. Define the *Moreau-Yosida regularization* of f,

$$\forall x \in \mathbb{R}^N, \quad f_\lambda(x) \stackrel{\text{def}}{=} \inf_{y \in \mathbb{R}^N} \left(f(y) + \frac{1}{2\lambda} \|x - y\|^2 \right) \tag{3}$$

1. Draw f_λ for $f(x) = \chi_{[-1,1]}(x)$, $x \in \mathbb{R}$.
2. Prove that there is a unique minimizer y in (3). It is called the proximal point of f at x. It is often denoted by $\operatorname{prox}_{\lambda f}(x)$.
3. Prove that f_λ is convex proper, and that $\operatorname{dom} f_\lambda = \mathbb{R}^N$ (hence f_λ is continuous on $\mathbb{R}^N$).
4. Prove the following properties
   a) $f_\lambda(x) \leq f(x)$ for all $x \in \mathbb{R}^N$, $\lambda > 0$.
   b) $\lim_{\lambda \to 0^+} \operatorname{prox}_{\lambda f}(x) = x$ for all $x \in \operatorname{dom} f$.
   c) $\lim_{\lambda \to 0^+} f_\lambda(x) = f(x)$ for all $x \in \mathbb{R}^N$.

Conjugate function

($\star$) **Exercise 11.**
Compute the conjugate function of the ℓ^p norm : $x \mapsto \|\mathbf{x}\|_{\mathbf{p}}$.
Hint: *Remember the Hölder inequality :* $|\langle x, y\rangle| \leq \|x\|_p \|y\|_q$ *for* $p, q \in [1, +\infty]$ *such that* $1/p + 1/q = 1$*. What is the equality case ?*

Exercise 12 (Support function of a convex set).
Let $C \subset \mathbb{R}^N$ be a closed convex set. We define the *support function* of C by

$$\forall x \in \mathbb{R}^N, \quad \sigma_C(x) \stackrel{\text{def}}{=} \sup_{q \in C} \langle q, x \rangle .$$

0. Compute σ_C when C is the unit ball of the ℓ^p norm, $1 \leq p \leq +\infty$.
1. Show that σ_C is positively homogeneous, *i.e.* $\sigma_C(tx) = t\sigma_C(x)$ for all $x \in \mathbb{R}^N$, $t > 0$.
2. a) Let C_1, C_2 be two nonempty closed convex sets. Prove that $C_1 \subset C_2$ if and only if $\sigma_{C_1}(x) \leq \sigma_{C_2}(\mathbf{x})$ for all $x \in \mathbb{R}^N$.
   b) Prove that $|\sigma_C(x)| \leq R\|x\|$ for all x if and only if $C \subset \overline{B(0,R)}$.
   c) Prove that $|\sigma_C(x)| \geq r\|x\|$ for all x if and only if $\overline{B(0,r)} \subset C$.
3. Show that $(\sigma_C)^*$ is the indicator of some closed convex set,

$$\forall p \in \mathbb{R}^N, \quad (\sigma_C)^*(p) = \chi_B(p) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } p \in B, \\ +\infty & \text{otherwise.} \end{cases}$$

   and characterize the convex set B.
   Hint: *Observe that* σ_C *is already the convex conjugate of some function.*

4. Prove that $\partial\sigma_C(0) = B$, and characterize $\partial\sigma_C(x)$ for all $x \in \mathbb{R}^N$.

Exercise 13.
Using the results of Exercise 12. Characterize the subdifferential of

$$a) f(x) = \|x\|_p \text{ for } 1 \leq p \leq \infty \qquad\qquad b) f(M) = \lambda_n(M) \text{ for all } M \in S_n^+(\mathbb{R})$$

Exercise 14 (Infimal convolution).
Let $f, g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, be two proper convex lower semi-continuous functions. We define the infimal convolution of f and g by

$$h(x) \stackrel{\text{def}}{=} \inf \{ f(y) + g(x-y) \mid y \in \mathbb{R}^N \}.$$

We assume that $(\operatorname{dom} f^*) \cap (\operatorname{dom} g^*) \neq \emptyset$.

1. Prove that there is an affine function ℓ such that for all $x \in \mathbb{R}^N$, $\ell(x) \leq f(x)$ and $\ell(x) \leq g(x)$. Deduce that h is convex proper.
2. Prove that $h^*(p) = f^*(p) + g^*(p)$ for all $p \in \mathbb{R}^N$.
3. We want to prove that h is lower semi-continuous **under the additional assumption that** $\operatorname{ri}(\operatorname{dom} f^*) \cap \operatorname{ri}(\operatorname{dom} g^*) \neq \emptyset$.
   a) Let $x \in \operatorname{dom} h$, and let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence such that $\lim_{k\to+\infty} x_k = x$. Prove that there is a sequence $\{y_k\}_{k\in\mathbb{N}}$ such that
   $$\forall k \in \mathbb{N}, \quad f(y_k) + g(x_k - y_k) \leq h(x_k) + \frac{1}{k+1}.$$
   b) Assume that $\{y_k\}_{k\in\mathbb{N}}$ is **bounded**. Prove that $\liminf_{k\to+\infty} h(x_k) \geq h(x)$.
   c) In fact, we cannot assume that $\{y_k\}_{k\in\mathbb{N}}$ is bounded. However, let us define $\{q_k\}_{k\in\mathbb{N}}$ in the following way. Let $V \stackrel{\text{def}}{=} \operatorname{Span}(\operatorname{dom} f^* - \operatorname{dom} g^*)$, *i.e.* the vector space spanned by the set $(\operatorname{dom} f^* - \operatorname{dom} g^*)$, and let q_k be the orthogonal projection of y_k onto V.
   Prove that for all $k \in \mathbb{N}$,
   $$f(q_k) + g(x_k - q_k) = f(y_k) + g(x_k - y_k).$$
   Hint: *Write* $f(q_k) + g(x_k - q_k) = f^{**}(q_k) + g^{**}(x_k - q_k) = \sup_{p,s}(\cdots)$ *and use the fact that* $\langle q_k, p - s\rangle = \langle y_k, p - s\rangle$ *for all* $p \in \operatorname{dom} f^*$, $s \in \operatorname{dom} g^*$.
   d) Prove that for all $\varepsilon > 0$ small enough, $(B(0,\varepsilon) \cap V) \subset (\operatorname{dom} f^* - \operatorname{dom} g^*)$, where $(\operatorname{dom} f^* - \operatorname{dom} g^*) = \{ p - s \mid p \in \operatorname{dom} f^*, s \in \operatorname{dom} g^* \}$. ***Hint:*** *Prove that* $\operatorname{ri}(\operatorname{dom} f^*) \cap \operatorname{ri}(\operatorname{dom} g^*) \neq \emptyset$ *is equivalent to* $0 \in \operatorname{ri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$.
   e) We fix such an $\varepsilon > 0$. Prove that for all $z \in B(0,\varepsilon) \cap \operatorname{Span}(\operatorname{dom} f^* - \operatorname{dom} g^*)$, there exists $p \in \operatorname{dom} f^*$, $s \in \operatorname{dom} g^*$ such that $z = p - s$, and moreover,
   $$\forall k \in \mathbb{N}, \quad \langle q_k, z\rangle \leq h(x_k) + \frac{1}{k+1} + f^*(p) + g^*(s) - \langle \mathbf{x_k}, \mathbf{s}\rangle.$$
   f) Deduce that the sequence $\{q_k\}_{k\in\mathbb{N}}$ is bounded and conclude.
   Incidentally, note that this also proves that the infimum is attained, i.e. there is some y *such that* $h(x) = f(y) + g(x-y)$.

($\star$) **Exercise 15** (Subdifferential of a sum).
Let $f, g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be two proper convex functions. We consider the function $h = f + g$ and we assume that $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$

1. Prove that h is convex proper.
2. Prove that for all $x \in \mathbb{R}^N$, $(\partial f(x) + \partial g(x)) \subset \partial(f+g)(x)$.
3. We want to prove the converse inclusion, **under the additional hypothesis** that $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{ri}(\operatorname{dom} g) \neq \emptyset$. **To simplify the proof, we also assume that f and g are lower semi-continuous (though the result still holds without that assumption).**

a) Prove that

$$\forall p \in \mathbb{R}^N, \quad h^*(p) = \inf_{q \in \mathbb{R}^N} \left(f^*(q) + g^*(p-q) \right).$$

Hint: *Use the results of Exercise 14.*

b) Using the equality case in the Fenchel inequality, prove that $p \in \partial h(x)$ if and only if there is some $q \in \mathbb{R}^N$ such that $q \in \partial f(x)$ and $p - q \in \partial g(x)$. Conclude.

Hint: *Remember that the infimum in the definition of $h^*(p)$ is attained.*