Convex functions, subdifferential, normal cones

Exercise 1. Derive the subdifferential of $x \mapsto |x|$ on $\mathbb{R} \to \mathbb{R}$.

Exercise 2. Let ι_C denote the indicator function of a non empty closed convex set C in $\mathcal{X}$. We set $N_C(x) = \{\varphi \in \mathcal{X} : \forall y \in C, \langle \varphi, y - x \rangle \leq 0\}$ for all $x \in C$, and $N_C(x) = \emptyset$ otherwise.

1. Show that for all $x \in C$, $N_C(x)$ is a cone, *i.e.* $\forall \varphi \in N_C(x)$, $\forall \lambda \geq 0$, $\lambda\varphi \in N_C(x)$.

2. Derive $N_C(x)$ when $x \in \mathrm{int}(C)$.

3. Derive $N_C(x)$ in the following cases: $C = (-\infty, 0]$, $C = (-\infty, 1]$, $C = \mathbb{R}_- \times \mathbb{R}_-$, $C = \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 \leq 1\}$.

Exercise 3. For every $x \in \mathbb{R}^n$, denote $x = (x_1, \ldots, x_n)$. Consider a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ s.t. $f(x) = \sum_{i=1}^n f_i(x_i)$ where $f_i : \mathbb{R}^n \to (-\infty, +\infty]$ for all i.

1. Express ∂f as a function of $\partial f_1, \ldots, \partial f_n$.

2. Application: $f(x) = \|x\|_1$.

Exercise 4. Let $f : \mathcal{X} \to [-\infty, +\infty]$ satisfy: $\exists x_0 \in \mathrm{relint}(\mathrm{dom}\, f)$, $f(x_0) \in \mathbb{R}$. Show that $-\infty \notin f(\mathcal{X})$.

Operations that preserve convexity

Exercise 5. Let $(f_\alpha)_{\alpha \in I}$ represent an arbitrary collection of functions on $\mathcal{X} \to (-\infty, +\infty]$. Set $f = \sup_\alpha f_\alpha$. Show that:

1. If f_α is convex for all α, then f is convex.

2. If f_α is closed for all α, then f is closed.

3. Let S denote the set of all symetric matrices in $\mathbb{R}^n$. Prove the function defined on S by $f(X) = \lambda_{max}(X)$ is convex (*Hint: write $f(X)$ as a supremum*).

Exercise 6. Let $F : \mathcal{X} \times \mathcal{Y} \to (-\infty, +\infty]$ be a convex function. Show that the mapping $y \mapsto \inf_{x \in \mathcal{X}} F(x, y)$ is convex.

Exercise 7.

1. Show that a nonnegative weighted sum of convex functions is convex.

2. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be convex. Let $A \in \mathbb{R}^{n \times m}$ be a matrix and b a vector. Show that $x \mapsto f(Ax + b)$ is convex.

Operations on multivalued functions

Let $A : \mathcal{X} \to 2^{\mathcal{X}}$ be a set-valued mapping, where $\mathcal{X} = \mathbb{R}^n$. We define the graph of A by $\mathrm{gr}A = \{(x, y) : y \in A(x)\}$. We define $A^{-1} : \mathcal{X} \to 2^{\mathcal{X}}$ as $A^{-1}(x) = \{y : x \in A(y)\}$ that is

$$\mathrm{gr}(A^{-1}) = \{(x, y) : (y, x) \in \mathrm{gr}A\}.$$

Its domain $\mathrm{dom}(A)$ is the set of $x \in \mathcal{X}$ such that $A(x) \neq \emptyset$.

Exercise 8. Let C be a non empty closed convex set of $\mathcal{X}$ and let $P_C(b)$ be the projection of a point $b \in \mathcal{X}$ onto C.

1. Show that $P_C(b)$ is the unique minimizer of the problem $\min_x \|x - b\|^2 + g(x)$ where $g(x)$ is a function to be determined.
2. Show that $x = P_C(b)$ iff (if and only if) $0 \in x - b + N_C(x)$..
3. Deduce that $P_C = (I + N_C)^{-1}$.
4. Application: Set $C = [-1, 1]$.
   (a) Calculate algebrically $P_C(b)$ as a function of b.
   (b) Plot the graph of the multifunction N_C.
   (c) Plot the graph of $I + N_C$.
   (d) Plot the graph of $(I + N_C)^{-1}$ and conclude.

Exercise 9. Let $\gamma > 0$ and $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$.

1. Draw the graph of ∂f and the graph of $\gamma\, \partial f$.
2. Draw the graph of $I + \gamma\partial f$ where $I : x \mapsto x$ is the identity function.
3. Draw the graph of $(I + \gamma\partial f)^{-1}$. Whose function is it the graph ?

Exercise 10. Let $f : \mathcal{X} \to (-\infty, +\infty]$ be a proper l.s.c. convex function. Let $x \in \mathcal{X}$ and $\gamma > 0$. Define

$$\mathrm{prox}_{\gamma f}(x) = \arg\min_{y \in \mathcal{X}} f(y) + \frac{\|y - x\|^2}{2\gamma}.$$

1. Show that the mapping $\mathrm{prox}_{\gamma f} : \mathcal{X} \to \mathcal{X}$ is well defined.
2. Show that $p = \mathrm{prox}_{\gamma f}(x)$ iff $p \in x + \gamma\partial f(x)$.
3. Deduce the identity $\mathrm{prox}_{\gamma f} = (I + \gamma\partial f)^{-1}$.
4. What are the fixed points of $\mathrm{prox}_{\gamma f}$?

Exercise 11. Evaluate $\mathrm{prox}_{\gamma f}$ when $f(x)$ coincides with $\iota_C(x)$ (C closed convex and non empty), $|x|$, $\|x\|_1$, $\|x\|^2$, $\|x\|$.

Fenchel-Legendre transform

Exercise 12. Let $f : \mathcal{X} \to (-\infty, +\infty]$ be a proper closed convex function.

1. Write Fenchel-Young inequality and the conditions for equality.
2. Show the equivalence
$$\phi \in \partial f(x) \ \Leftrightarrow \ x \in \partial f^*(\phi)\,.$$
3. Deduce from it the equality $(\partial f)^{-1} = \partial f^*$.

Exercise 13. The goal of this exercise is to show Moreau's identity: for every $x \in \mathbb{R}^n$,
$$\mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x) = x\,.$$

1. Let $x \in \mathbb{R}^n$ and $p = \mathrm{prox}_f(x)$. Show that $x - p \in \partial f(p)$.
2. Using the result of Exercise 12, show that $p \in \partial f^*(x - p)$.
3. Prove Moreau's identity.
4. Show that Moreau's formula generalizes the famous identity $\Pi_E + \Pi_{E^\perp} = I$, where Π_E and $\Pi_{E^\perp}$ are the orthogonal projectors onto some linear subspace $E \subset \mathbb{R}^n$ and its supplementary space $E^\perp$ respectively.
Hint: choose f as the indicator function of a properly chosen set.
5. *Homework.* For $\gamma > 0$, generalize the identity to:
$$\mathrm{prox}_{\gamma f}(x) + \gamma\, \mathrm{prox}_{\gamma^{-1} f^*}(\frac{x}{\gamma}) = x\,.$$

Fenchel-Legendre transforms and subdifferentials

Exercise 14. Calculate f^*, where $f : \mathbb{R}^n \to \mathbb{R}$ is provided below.

1. $f(x) = \|x\|_1$
2. $f(x) = \frac{1}{2}x^\top Qx$ where Q is a symmetric positive definite matrix.
3. $f(x) = \frac{1}{2}x^\top Qx$ where Q is a symmetric positive semi-definite matrix.
   Hint: Distinguish between the cases $\varphi \in Im(Q)$ and $\varphi \notin Im(Q)$.

Exercise 15. Let C be a convex subset of $\mathbb{R}^n$. The *normal cone* of C at x is the set

$$N_C(x) = \{\varphi \in \mathbb{R}^n : \forall y \in C,\ \langle \varphi, y - x\rangle \leq 0\},$$

if $x \in C$, and $N_C(x) = \emptyset$ otherwise.

1. Show that $N_C = \partial \iota_C$.
2. Let $x \in \operatorname{int}(C)$. Prove that $N_C(x) = \{0\}$.
3. Let $C = \{(u, v) \in (-\infty, 0]^2 : u + v \geq 0\}$. Draw C and $N_C((0, 0))$.
4. Let $C = \{x \in \mathbb{R}^n\ :\ \|x\|_2 \leq 1\}$. Calculate $N_C(x)$ for all x.

Exercise 16. Calculate the subdifferential of $f(x) = \sum_{i=1}^n g_i(x_i)$.

Exercise 17. Let $f : \mathbb{R}^n \to \mathbb{R}$.

1. Show that if f is differentiable at x, then $\partial f(x) \subseteq \{\nabla f(x)\}$.
2. Show that if f is convex and differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

Lagrangian duality

Exercise 18 (Lagrangian duality)**.** Let us consider the following optimisation problem with constraints :

$$\begin{aligned}
&\min_{x\in\mathbb{R}^n} f(x)\\
&\text{subject to: } g_j(x) \leq 0, \ \ \forall j \in \{1,\dots,p\}
\end{aligned}$$

We assume that $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is convex, $g_j : \mathbb{R}^n \to \mathbb{R}$ is convex and finite-valued for all j and that there exists $x_0 \in \operatorname{dom} f$ such that $g_j(x_0) < 0$ for all j.
We define the Lagrangian of the optimization problem as

$$\begin{aligned}
L :\ \mathbb{R}^n \times \mathbb{R}^p &\to \mathbb{R}\cup\{-\infty,+\infty\}\\
(x,\phi) &\mapsto f(x) + \langle \phi, g(x)\rangle - \iota_{\mathbb{R}^p_+}(\phi)
\end{aligned}$$

1. Show that

$$\inf_{x\in\mathbb{R}^n} \Big\{ f(x) \text{ subject to: } g_j(x) \leq 0, \ \ \forall j \in \{1,\dots,p\}\Big\} = \inf_{x\in\mathbb{R}^n} f(x) + \iota_{\mathbb{R}^p_-}(g(x))$$

   and that $h : x \mapsto f(x) + \iota_{\mathbb{R}^p_-}(g(x))$ is convex.

2. Show that

$$f(x) + \iota_{\mathbb{R}^p_-}(g(x)) = \sup_{\phi\in\mathbb{R}^p} L(x,\phi)$$

3. Show that

$$\inf_{x\in\mathbb{R}^n}\sup_{\phi\in\mathbb{R}^p} L(x,\phi) \geq \sup_{\phi\in\mathbb{R}^p}\inf_{x\in\mathbb{R}^n} L(x,\phi)$$

This property is called the *weak duality* property. The problem $\min_{x\in\mathbb{R}^n}\big(\sup_{\phi\in\mathbb{R}^p} L(x,\phi)\big)$ is the primal problem, the problem $\max_{\phi\in\mathbb{R}^p}\big(\inf_{x\in\mathbb{R}^n} L(x,\phi)\big)$ is the dual problem.

We say that (x^*,ϕ^*) is a *saddle point* of L if

$$\forall x \in \mathbb{R}^n, \forall \phi \in \mathbb{R}^p, \quad L(x^*,\phi) \leq L(x^*,\phi^*) \leq L(x,\phi^*).$$

4. Show that if (x^*,ϕ^*) is a saddle point

$$L(x^*,\phi^*) = \sup_{\phi\in\mathbb{R}^p}\inf_{x\in\mathbb{R}^n} L(x,\phi) = \inf_{x\in\mathbb{R}^n}\sup_{\phi\in\mathbb{R}^p} L(x,\phi)\,.$$

In that case, the *duality gap*

$$G = \inf_{x\in\mathbb{R}^n}\sup_{\phi\in\mathbb{R}^p} L(x,\phi) - \sup_{\phi\in\mathbb{R}^p}\inf_{x\in\mathbb{R}^n} L(x,\phi)$$

is equal to 0.

5. Show that (x^*, ϕ^*) is a saddle point of L if and only if

$$\begin{cases} 0 \in \partial_x L(x^*, \phi^*) \\ 0 \in \partial_\phi (-L)(x^*, \phi^*) \end{cases}$$

   where $\partial_\phi(-L)(x^*, \phi^*)$ means the sub-differential of the function $\big(\phi \mapsto -L(x^*, \phi)\big)$ at ϕ^*.

6. Calculate $\partial_\phi(-L)$ and deduce that (x^*, λ) is a saddle point of L if and only if

$$\begin{cases} \phi^* \geq 0 \\ g(x^*) \leq 0 \\ \phi_j^* g_j(x^*) = 0, \forall j \\ 0 \in \partial f(x^*) + \sum_{j=1}^{p} \phi_j^* \partial g_j(x^*) \end{cases}$$

   You may use the results of Exercises 2 and 3.

These conditions are called the Karush-Kuhn-Tucker (KKT) conditions.

Exercise 19. We consider the following basis pursuit problem:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \|x\|_1 \\ &\text{subject to: } \|Ax - b\|_2^2 \leq \delta^2 \end{aligned}$$

1. Show that if there is at least one feasible point then there exists a primal solution.
2. Write the Lagrangian of this problem.
3. Show that the duality gap is 0 and that there exists a dual solution.
4. Write the KKT conditions for this problem.
5. We take $n = 2$, $\delta = 0.5$, $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Solve the KKT system. It may be useful to distinguish between cases depending on the sign of the dual variable and eliminate the impossible alternatives.

Parametric duality

Exercise 20 (Parametric duality)**.** In this exercise, we consider a function $F : \mathcal{X} \times \mathcal{Y} \to (-\infty, +\infty]$ where $\mathcal{X}$ and $\mathcal{Y}$ are two Euclidean spaces. We assume that F is *jointly convex* on $\mathcal{X} \times \mathcal{Y}$. Without risk of ambiguity, we can denote with the same symbol $\langle\, .\,,\, .\,\rangle$ the scalar products in each of these spaces. We equip the product space $\mathcal{X} \times \mathcal{Y}$ with the scalar product $\langle (\nu, \phi), (x, y)\rangle := \langle \nu, x\rangle + \langle \phi, y\rangle$.
The goal of the exercise is to define a general notion of duality that will encompass the cases of inequality constraints and equality constraints.
We define the primal function by $(x \mapsto F(x, 0))$ and the primal value by

$$p := \inf F(\mathcal{X}, 0)\,.$$

We define the value function for all $y \in \mathcal{Y}$ by

$$\vartheta(y) := \inf F(\mathcal{X}, y)\,.$$

Clearly $p = \vartheta(0)$. The dual value is defined by $d := \vartheta^{**}(0) = \sup_\phi -\vartheta^*(\phi)$. We can write

$$\vartheta^*(\phi) = \sup_{x,y} \langle \phi, y\rangle - F(x, y) = F^*(0, \phi)\,.$$

We will call $(\phi \mapsto -\vartheta^*(-\phi))$ the dual function. To sum up,

$$\begin{aligned} p &= \vartheta(0) = \inf F(\mathcal{X}, 0) \\ d &= \vartheta^{**}(0) = \sup -F^*(0, \mathcal{Y})\,. \end{aligned}$$

We call an *optimal primal (resp. dual) point* or *primal (resp. dual) solution* any minimizer of $F(\,.\,, 0)$ (resp. $F^*(0,\, .\,)$). We denote $D = \arg\min \vartheta^*$ the set of dual solutions.

1. Let $F(x, y) = f(x) + \iota_{\{0\}}(Ax - y)$. What is the value function? the primal value? the dual value? the dual function?

2. Show that ϑ is convex.

3. Show that if $0 \in \text{relint}(\text{dom}\, \vartheta)$, then there exists $\lambda \in \mathcal{Y}$ such that

$$\vartheta(0) + \vartheta^*(\lambda) = 0$$

From now on, we assume that $0 \in \text{relint}(\text{dom}\, \vartheta)$.

4. Show that $p = d$ and that there exists at least one dual solution.

5. Show that if p is finite then ϑ is proper and that if $p = -\infty$ then $D = \mathcal{Y}$.

6. Suppose that $F(x, y) = f(x) + \iota_{\{0\}}(Ax - y)$.
   Show that $0 \in \text{relint}(\text{dom}\, \vartheta) \Leftrightarrow 0 \in \text{relint}(A(\text{dom}\, f))$.

Examples of dual problems

Exercise 21. Let us consider the Support Vector Machine problem

$$\min_{x\in\mathbb{R}^n, x_0\in\mathbb{R}} C\sum_{j=1}^{m} \max(0, 1 - b_j(a_j^\top x + x_0)) + \frac{1}{2}\|x\|_2^2 \,.$$

1. Show that strong duality holds.
2. Write a dual to the SVM problem.
3. Show that if we are given a dual solution, then we can recover a primal solution.

Exercise 22. Give a dual problem to the elastic net problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|^2 + \lambda_1\|x\|_1 + \frac{\lambda_2}{2}\|x\|_2^2 \,.$$

Exercise 23. Give a dual problem to the square-root Lasso problem

$$\min_{x\in\mathbb{R}^n} \|Ax-b\| + \lambda\|x\|_1$$

Continuity of convex functions

Exercise 24. The goal of this exercise is to show that a convex function $f : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ is continuous on the interior of its domain.

1. Let $x \in \operatorname{relint}(\operatorname{dom} f)$. Show that f is lower semi-continuous at x.
2. Let $\rho > 0$ and $x \in \mathbb{R}^m$. Show that there exists $(y_i)_{i\in I}$ with I finite and $\rho' > 0$ such that
$$B(x,\rho) \subseteq \operatorname{conv}((y_i)_{i\in I}) \subseteq B(x,\rho')$$
where $B(x,\rho)$ is the ball of radius ρ and center x and $\operatorname{conv}(A)$ is the convex hull of A.
3. Let $t \geq 1$ and $x, y \in \mathbb{R}^m$. Show that
$$f(x + t(y-x)) \geq tf(y) + (1-t)f(x)$$
4. Show that f is continuous on the interior of its domain.

Denote by $\Gamma_0(\mathcal{X})$ the set of proper convex l.s.c. functions on $\mathcal{X} \to (-\infty, +\infty]$.

Douglas-Rachford algorithm

Let A and B be maximal monotone operators on $\mathcal{X} \to 2^{\mathcal{X}}$. Denote by J_A, J_B their resolvents. Define

$$C_A = 2J_A - I$$

and $C_B = 2J_B - I$. The Douglas-Rachford (DR) operator is given by

$$T = \frac{I + C_A C_B}{2}\,.$$

Exercise 25 (Fixed points of the DR-operator)**.** The aim is to prove the following points.

i) If $\xi \in \operatorname{fix} T$, then $J_B(\xi) \in \operatorname{zer}(A+B)$.
ii) If $\operatorname{zer}(A+B) \neq \emptyset$, then $\operatorname{fix} T \neq \emptyset$.

1. Show that $\operatorname{fix} T = \operatorname{fix} C_A C_B$.
2. Choose $\xi \in \mathcal{X}$. Define $u = J_B(\xi)$, $b = \xi - u$. Prove that
$$C_B(\xi) = u - b\,.$$
3. Assume that $\xi \in \operatorname{fix} C_A C_B$. Prove that $u = J_A(u - b)\,.$
4. Deduce that $0 \in b + A(u)$.
5. Using that $J_B(\xi) = u$, show that $b \in B(u)$.
6. Deduce that $u \in \operatorname{zer}(A+B)$ and conclude about point *i)*.
7. Consider an arbitrary $\bar{u} \in \operatorname{zer}(A+B)$. Prove that $\exists \bar{b} \in B(\bar{u})$ s.t. $-\bar{b} \in A(\bar{u})$.
8. Using that $\bar{b} \in B(\bar{u})$, prove that $\bar{u} = J_B(\bar{u} + \bar{b})$. Deduce that $C_B(\bar{u} + \bar{b}) = \bar{u} - \bar{b}$.
9. Using that $-\bar{b} \in A(\bar{u})$, prove that $\bar{u} = J_A(\bar{u} - \bar{b})$. Deduce that $C_A(\bar{u} - \bar{b}) = \bar{u} + \bar{b}$.
10. From the previous two questions, deduce that $\bar{u} + \bar{b} \in \operatorname{fix} C_A C_B$ and conclude.

Exercise 26 (DR as an averaged operator)**.** The aim is to prove the following result.

$T = \frac{I + C_A C_B}{2}$ is an averaged operator.

We recall the inequality: $\|J_B(x) - J_B(y)\|^2 \leq \|x - y\|^2 - \|(I - J_B)(x) - (I - J_B)(y)\|^2$.

1. Using the above inequality, prove that C_B is non-expansive *i.e.*, $\|C_B x - C_B y\| \leq \|x - y\|$ for every x, y.

2. Show that $C_A C_B$ is non-expansive and conclude.

Exercise 27. We assume that $\text{zer}(A + B) \neq \emptyset$. We consider the iterates $\xi^{k+1} = T(\xi^k)$.

1. Show that the iterates ξ^k are generated by the following algorithm:

$$\boxed{\begin{aligned} u^k &= J_B(\xi^k) \\ v^k &= J_A(2u^k - \xi^k) \\ \xi^{k+1} &= \xi^k + x^k - u^k\,. \end{aligned}}$$

2. Prove that the above sequence (u^k) converges to a point in $\text{zer}(A + B)$.

3. Consider the problem
$$\text{minimize } f + g$$
where $f, g \in \Gamma_0(\mathcal{X})$. Let $\gamma > 0$. Using Question 2, propose an algorithm which requires one call to $\text{prox}_{\gamma f}$ and $\text{prox}_{\gamma g}$ at every iteration.

4. Under what sufficient condition on $f + g$ does this algorithm indeed converge to a minimizer?

Exercise 28 (Parallel programming)**.** Consider the problem

$$\text{minimize } \sum_{i=1}^{n} f_i(x) \text{ w.r.t.} x \in \mathcal{X} \tag{1}$$

where $f_1, \ldots, f_n \in \Gamma_0(\mathcal{X})$. Let $C \subset \mathcal{X}^n$ be the linear space $C := \{(x, \ldots, x) : x \in \mathcal{X}\}$.

1. In what sense is the following problem equivalent to (1) ?

$$\text{minimize } \sum_{i=1}^{n} f_i(x_i) + \iota_C(x_1, \ldots, x_n) \text{ w.r.t.}(x_1, \ldots, x_n) \in \mathcal{X}^n$$

2. Write the Douglas-Rachford algorithm associated to this problem.

3. In what sense does this algorithm deserves the name of *parallel algorithm*?

Smoothness, strong convexity and Baillon Haddad's lemma

- We say that $f : \mathcal{X} \to \mathbb{R}$ is L-smooth if it is differentiable and if ∇f is L-lipschitz continuous.
- We say that $f \in \Gamma_0(\mathcal{X})$ is μ-strongly convex ($\mu > 0$) if $f - \frac{\mu}{2}\|\,.\|^2$ is convex.
- We say that a operator $A : \mathcal{X} \to 2^{\mathcal{X}}$ is μ-strongly monotone if

$$\forall (x, y) \in \mathrm{dom}(A) \times \mathrm{dom}(A),\ \forall (u, v) \in A(x) \times A(y),\ \langle u - v, x - y\rangle \geq \mu\|x - y\|^2 .$$

Exercise 29. The aim is to prove the statement:

If $f \in \Gamma_0(\mathcal{X})$ is μ-strongly convex, then f^* is μ^{-1}-smooth.

1. Let $f \in \Gamma_0(\mathcal{X})$ be μ-strongly convex. Prove that ∂f is μ-strongly monotone.
   Hint: set $g := f - \frac{\mu}{2}\|\,.\|^2$ and use that ∂g is monotone.
2. Justify that $\mathrm{dom}(f^*) = \mathcal{X}$.
3. Show that for all $\varphi \in \mathcal{X}$, $\partial f^*(\varphi) = \arg\min_{x\in\mathcal{X}} f(x) - \langle \varphi, x\rangle$.
4. Deduce that f^* is differentiable.
5. Using 1), prove that for every $\varphi, \lambda \in \mathcal{X}$,
$$\langle \varphi - \lambda, \nabla f^*(\varphi) - \nabla f^*(\lambda)\rangle \geq \mu\|\nabla f^*(\varphi) - \nabla f^*(\lambda)\|^2 . \tag{2}$$
6. Conclude.

As a matter of fact, the above result has a converse, which we admit (the proof can be found in Hiriart-Urruty and LeMaréchal, Fundamentals of Convex Analysis.).

If f is convex and L-smooth, then f^* is L^{-1}-strongly convex

7. Using Question 5, recover the Baillon-Haddad Lemma.

ADMM and applications

Exercise 30 (Variation totale)**.** Soit $b = (b_1, \dots, b_n)^T$ un vecteur de $\mathbb{R}^n$. On se pose le problème suivant :

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x-b\|^2 + \eta \sum_{i=1}^{n-1} |x_{i+1} - x_i| \,. \tag{3}$$

1. Expliquer brièvement l'effet que peut avoir le deuxième terme (dit de régularisation). Autrement dit, intuitivement, en quoi la solution x^* de ce problème va-t-elle différer de b ?

2. Montrer que le problème (3) peut être réécrit

$$\min_{x\in\mathbb{R}^n} f(x) + g(Mx)$$

   où $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathbb{R}^m \cup \{+\infty\}$ sont des fonctions convexes et M est une matrice dont on donnera les dimensions.

3. Calculer les opérateurs proximaux de f et de g.

4. Pour $n = 5$, expliciter M et $M^T M$.

5. Écrire les itérations de l'ADMM de pas $\sigma > 0$ pour la résolution de (3).

6. Montrer que l'algorithme se réduit à une succession de seuillages doux et de résolution de systèmes linéaires.

Exercise 31 (Distributed optimization)**.** A database is distributed on a computer network composed of N parallel workers. Each worker i has a private cost function $f_i : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is a Euclidean space. The aim is to find a minimizer of the function

$$f(x) = \sum_{i=1}^{N} f_i(x) \,.$$

We define the function $F(x_1, \dots, x_N) = \sum_{i=1}^{N} f_i(x_i)$ on $\mathcal{X}^N \to \mathbb{R}$. One can therefore reformulate the problem as

$$\min F(x_1, \dots, x_N) \ \text{ s.t.} \quad x_1 = \dots = x_N \,. \tag{4}$$

1. State that problem (4) is equivalent to the minimization of $F(x) + \iota_{C_N}(x)$ on $x \in \mathcal{X}^N$ where C_N is the indicator function of a linear space C_N which you will specify.

2. Write the iterations of ADMM for that problem, making clear the communications between workers that are needed at each step of the algorithm.

3. Explicit the algorithm in the case where

$$f_i(x) = \frac{1}{2}\|A_i x - b\|^2 .$$

We now assume that the workers are connected through a graph structure. Let $G = (V, E)$ be a graph with $V = \{1, \dots, N\}$ and E is a set of edges such that $\{i, j\} \in E$ if and only if the workers i and j can communicate.

4. Under what condition on the graph have we

$$\iota_{C_N}(x) = \sum_{\{i,j\}\in E} \iota_{C_2}(x_i, x_j)\,?$$

5. For any $e = \{i, j\}$ in E ($i < j$), we define the matrix $M_e : \mathcal{X}^N \to \mathcal{X}^2$ such that $M_e x = (x_i, x_j)^T$. We define the matrix $M : \mathcal{X}^N \to \mathcal{X}^{2|E|}$ such that $Mx = (M_e x)_{e\in E}$. Show that $\iota_{C_N}(x) = g(Mx)$ where g is a function that will be specified.

6. Write and simplify the iterations of ADMM, making clear the communications between workers that are needed at each step of the algorithm.

Exercise 32. The goal of this exercise is to define a variant of ADMM able to solve the following problem

$$\begin{aligned} &\min_{x,z} f(x) + g(z) \\ &\text{st: } Ax + Bz = c \end{aligned} \tag{5}$$

where f and g are convex functions and A and B are matrices.

1. Let us define $h(y) = \inf_z g(z) + \iota_{\{0\}}(Bz + y - c)$. Show that

$$\begin{array}{ccc} \inf_{x,y} f(x) + h(y) & = & \inf_{x,z} f(x) + g(z) \\ \text{st: } y = Ax & & \text{st: } Ax + Bz = c \end{array}$$

2. Show that for both problems, the dual function is equal to

$$D(\lambda) = -f^*(A^T\lambda) - g^*(B^T\lambda) + \langle \lambda, c\rangle\ .$$

3. Write the ADMM for the problem

$$\begin{aligned} &\min_{x,z} f(x) + h(y) \\ &\text{st: } y = Ax \end{aligned}$$

4. How does the algorithm write in terms of the original function g and variable z, instead of h and y?

Vũ-Condat's algorithm

Exercise 33 (Régularisations multiples)**.** Soit $b = (b_1, \ldots, b_n)^T$ un vecteur de $\mathbb{R}^n$, soit $H \in \mathbb{R}^{p \times n}$, et soit ∇ une matrice $q \times n$. On se pose le problème suivant :

$$\min_{x \in \mathbb{R}^n,\, x \geq 0} \frac{1}{2} \|Hx - b\|^2 + \eta \, \|\nabla x\|_1 \,. \tag{6}$$

Proposer deux algorithmes pour résoudre ce problème. Le premier utilisera l'approche de Vũ-Condat, le second utilisera l'ADMM (dans le cas de l'ADMM, on devra d'abord transformer le problème en introduisant une variable auxiliaire).

Exercise 34 (Vũ-Condat généralise l'ADMM)**.** On considère l'algorithme de Vũ-Condat pour la résolution du problème convexe :

$$\min_{x \in \mathcal{X}} f_1(x) + f_2(x) + g(Mx)$$

où $\mathcal{X}$ est un espace euclidien quelconque, où $M : \mathcal{X} \to \mathcal{Y}$ est un opérateur linéaire, et où f_1 est la fonction lisse.

1. Rappeler l'expression de l'algorithme et la condition sur les pas σ, τ assurant la convergence vers un point selle du Lagrangien.

On considère désormais le cas où $f_1 = 0$. On admet que dans ce cas précis (voir l'article de Condat'13 pour la preuve), la condition sur le pas peut être généralisée à

$$\frac{1}{\tau} - \sigma \|M\|^2 \geq 0 \,.$$

2. On considère le cas où $M = I$. En utilisant l'identité de Moreau, montrer que l'algorithme de Vũ-Condat coïncide avec l'ADMM pour des valeurs de τ et σ que l'on précisera.

3. On suppose que M est injective (*i.e.* $M^T M$ est inversible) et on pose $M^{\#} = (M^T M)^{-1} M^T$ l'inverse de M sur $\mathrm{Im}(M) \to \mathbb{R}^n$. Montrer que le problème

$$\min_{x \in \mathbb{R}^n} f(x) + g(Mx)$$

revient à un changement de variable près au problème

$$\min_{y \in \mathrm{Im}(M)} f(M^{\#} y) + g(y) \,.$$

4. Appliquer l'algorithme de la question 2 au problème ci-dessus. Quel algorithme retrouve-t-on ?

Random coordinate descent

Exercise 35 (Random Coordinate descent)**.** Set $\mathcal{X} := \mathbb{R}^n$ and let $T : \mathcal{X} \to \mathcal{X}$ be an averaged operator admitting a fixed point. We use the notation x_i or $(x)_i$ to denote the ith component of a vector x. We define for every $i = 1 \dots n$:

$$T^{(i)}(x) = (x_1, \dots, x_{i-1}, (T(x))_i, x_{i+1}, \dots, x_n)\,.$$

We let $(i_k : k \in \mathbb{N}^*)$ be a iid sequence, uniformly distributed on $\{1, \dots, n\}$. We consider the sequence of r.v. defined by $x^{k+1} = T^{(i_{k+1})}(x^k)$. We let $\mathbb{E}$ be the expectation and $\mathbb{E}_k$ be the conditional expectation wrt the σ-field generated by $i_1, \dots, i_k$.

1. Show that for every fixed point x^* of T,

$$\mathbb{E}_k(\|x^{k+1} - x^*\|^2) = \frac{1}{n}\|Tx^k - x^*\|^2 + \frac{n-1}{n}\|x^k - x^*\|^2\,.$$

2. Deduce that there exists $\beta > 0$ s.t.

$$\mathbb{E}_k(\|x^{k+1} - x^*\|^2) \le \|x^k - x^*\|^2 - \beta\|(I - T)x^k\|^2\,.$$

3. Prove that

$$\mathbb{E}\left(\sum_{k=1}^{\infty} \|(I - T)x^k\|^2\right) < \infty.$$

4. Deduce that $(I - T)x^k$ converges a.s. to zero.

5. Conclude that, almost surely, every cluster point of the sequence x^k is a fixed point of T.

A random sequence $M_k \ge 0$ is called a nonnegative supermartingale if $\mathbb{E}_k(M_{k+1}) \le M_k$. Doobs' convergence theorem implies that any nonnegative supermartingale converges almost surely.

6. Justify that for every $x^* \in \mathrm{fix}(T)$, the r.v. $\lim_{k\to\infty} \|x^k - x^*\|^2$ exists almost surely.

7. Show that, almost surely, for every $x^* \in \mathrm{fix}(T)$, the r.v. $\lim_{k\to\infty} \|x^k - x^*\|^2$ exists. *Hint: Introduce a dense denumerable subset $S \subset$ fix(T).*

8. Conclude that, almost surely, there exists $x^* \in \mathrm{fix}(T)$ s.t. $\lim_{k\to\infty} x^k = x^*$.

Speed of Convergence

Exercise 36 (Gradient descent)**.**
The exercises 36 and 37 are aimed at proving that the gradient algorithm for minimizing f, where f is convex and differentiable has convergence rate $O(1/k)$ in general (where k is the number of iterations) and $O((\frac{Q-1}{Q})^k)$ when f is strongly convex (where Q is called the *condition number*).
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function whose gradient is L-Lipschitz continuous *i.e.* $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ for all x, y.

1. Prove that for all x, y, $\langle \nabla f(y) - \nabla f(x), y - x\rangle \leq L\|y - x\|^2$.
2. Set $\varphi(t) = f(x + t(y - x))$ for all $t \in [0, 1]$. Prove that
$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle = \varphi(1) - \varphi(0) - \varphi'(0)\,.$$
3. Deduce that
$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle = \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle dt$$
4. Using the first question, conclude that
$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2\,.$$

We consider the gradient algorithm *i.e.*, the sequence (x_k) defined by $x_{k+1} = x_k - \gamma\nabla f(x_k)$ where $\gamma > 0$ is a constant step size.

5. Show that
$$x_{k+1} = \arg\min_{y\in\mathbb{R}} \langle \nabla f(x_k), y - x_k\rangle + \frac{1}{2\gamma}\|x_k - y\|^2.$$
6. Prove that for all $z \in \mathbb{R}^n$,
$$\langle \nabla f(x_k), x_{k+1}-x_k\rangle+\frac{1}{2\gamma}\|x_k-x_{k+1}\|^2 = \langle \nabla f(x_k), z-x_k\rangle+\frac{1}{2\gamma}\|x_k-z\|^2-\frac{1}{2\gamma}\|x_{k+1}-z\|^2. \tag{7}$$
7. Deduce that $f(x_{k+1}) \leq f(x_k) - \frac{1}{\gamma}(1 - \frac{\gamma L}{2})\|x_{k+1} - x_k\|^2$.
8. Provide a condition on γ which ensures that when $x_{k+1} \neq x_k$, $f(x_{k+1}) < f(x_k)$.

From now on, we set $\gamma = \frac{1}{L}$.

9. Using (7), show that for all $z \in \mathbb{R}^n$,
$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), z - x_k\rangle + \frac{L}{2}\|x_k - z\|^2 - \frac{L}{2}\|x_{k+1} - z\|^2\,. \tag{8}$$

We assume from now on that f is convex and admits (at least) one minimizer $x^\star$.

10. Show that
$$f(x_{k+1}) \leq f(x^\star) + \frac{L}{2}\|x_k - x^\star\|^2 - \frac{L}{2}\|x_{k+1} - x^\star\|^2 \,.$$

11. Deduce that for all $k \geq 1$,
$$\sum_{i=1}^{k} f(x_i) \leq k f(x^\star) + \frac{L}{2}\|x_0 - x^\star\|^2 \,.$$

12. Show that
$$f(x_k) - f(x^\star) \leq \frac{L\|x_0 - x^\star\|^2}{2k} \,.$$

Exercise 37 (Gradient descent – strongly convex functions)**.**
We assume from now on that f is μ-strongly convex. Thus, for any x, y,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2 \,.$$

1. Using Eq. (8), prove that
$$f(x_{k+1}) \leq f(x^\star) + \frac{L-\mu}{2}\|x_k - x^\star\|^2 - \frac{L}{2}\|x_{k+1} - x^\star\|^2 \,.$$

2. Define $\Delta_{k+1} = f(x_{k+1}) - f(x^\star) + \frac{L}{2}\|x_{k+1} - x^\star\|^2$. Show that
$$\Delta_{k+1} \leq \left(1 - \frac{\mu}{L}\right)\Delta_k \,.$$

3. Conclude that
$$\begin{aligned} f(x_k) - f(x^\star) &\leq \left(1 - \frac{\mu}{L}\right)^k \Delta_0 \\ \|x_k - x^\star\|^2 &\leq \left(1 - \frac{\mu}{L}\right)^k \frac{2\Delta_0}{L} \,. \end{aligned}$$

4. The ratio $Q = L/\mu$ is called the *condition number* of f. Discuss the influence of Q on the convergence rate.

Exercise 38 (ADMM)**.**
We consider the ADMM

$$
\begin{aligned}
x_{k+1} &\in \arg\min_x f(x) + \langle \phi_k, Mx \rangle + \frac{\rho}{2}\|Mx - z_k\|^2 \\
z_{k+1} &= \arg\min_z g(z) - \langle \phi_k, z \rangle + \frac{\rho}{2}\|Mx_{k+1} - z\|^2 \\
\phi_{k+1} &= \phi_k + \rho(Mx_{k+1} - z_{k+1})
\end{aligned}
$$

for the resolution of the optimization problem

$$\min_x f(x) + g(Mx)$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are convex functions, $M : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map and there exists a solution x^* to the problem.

1. Denote $\varphi : x \mapsto F(x) + \frac{\rho}{2}\|Mx - z_0\|^2$, where F is convex and let $p \in \arg\min \varphi$. Show that $(x \mapsto \varphi(x) - \frac{\rho}{2}\|Mx - Mp\|^2)$ is convex and that

$$p \in \arg\min_x \varphi(x) - \frac{\rho}{2}\|Mx - Mp\|^2.$$

2. Deduce that for all x,

$$F(p) + \frac{\rho}{2}\|Mp - z_0\|^2 \leq F(x) + \frac{\rho}{2}\|Mx - z_0\|^2 - \frac{\rho}{2}\|Mx - Mp\|^2$$

3. Show that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, $\phi \in \mathbb{R}^m$,

$$
\begin{aligned}
&f(x_{k+1}) + \langle \phi_k, Mx_{k+1} \rangle + \frac{\rho}{2}\|Mx_{k+1} - z_k\|^2 \leq f(x) + \langle \phi_k, Mx \rangle + \frac{\rho}{2}\|Mx - z_k\|^2 - \frac{\rho}{2}\|Mx - Mx_{k+1}\| \\
&g(z_{k+1}) - \langle \phi_k, z_{k+1} \rangle + \frac{\rho}{2}\|Mx_{k+1} - z_{k+1}\|^2 \leq g(z) - \langle \phi_k, z \rangle + \frac{\rho}{2}\|Mx_{k+1} - z\|^2 - \frac{\rho}{2}\|z - z_{k+1}\|^2 \\
&- \langle \phi_{k+1}, Mx_{k+1} - z_{k+1} \rangle + \frac{1}{2\rho}\|\phi_{k+1} - \phi_k\|^2 = -\langle \phi, Mx_{k+1} - z_{k+1} \rangle + \frac{1}{2\rho}\|\phi - \phi_k\|^2 - \frac{1}{2\rho}\|\phi - \phi_{k+1}\|^2
\end{aligned}
$$

4. Denote the Lagrangian function of the problem as

$$L(x, z, \phi) = f(x) + g(z) + \langle \phi, Mx - z \rangle.$$

Show that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, $\phi \in \mathbb{R}^m$,

$$
\begin{aligned}
L(x_{k+1}, z_{k+1}, \lambda) - L(x, z, \phi_{k+1}) &\leq \frac{1}{2\rho}\Big[\|\phi_k - \phi\|^2 - \|\phi_{k+1} - \phi\|^2 - \|\phi_k - \phi_{k+1}\|^2\Big] \\
&\quad + \frac{\rho}{2}\Big[\|z_k - Mx\|^2 - \|z_{k+1} - Mx\|^2 - \|Mx_{k+1} - z_k\|^2 + \|Mx_{k+1} - z_{k+1}\|^2\Big]
\end{aligned}
$$

5. Denote $\bar{x}_k = \frac{1}{k}\sum_{i=1}^k x_i$, $\bar{z}_k = \frac{1}{k}\sum_{i=1}^k z_i$ and $\bar{\phi}_k = \frac{1}{k}\sum_{i=1}^k \phi_i$.
   Show that for all $\phi \in \mathbb{R}^m$ and for all solution x^*, we have
$$L(\bar{x}_k, \bar{z}_k, \phi) - L(x, z, \bar{\phi}_k) \leq \frac{\rho}{2k}\|z_0 - Mx^*\|^2 + \frac{1}{2\rho k}\|\phi - \phi_0\|^2$$

6. For $\beta > 0$, denote
$$S_\beta(\bar{x}, \bar{z}, \phi_0) = \sup_\phi f(\bar{x}) + g(\bar{z}) + \langle \phi, M\bar{x} - \bar{z}\rangle - f(x^*) - g(Mx^*) - \frac{\beta}{2}\|\phi_0 - \phi\|^2$$
   Show that
$$S_\beta(\bar{x}, \bar{z}, \phi_0) = f(\bar{x}) + g(\bar{z}) - f(x^*) - g(Mx^*) + \langle \phi_0, M\bar{x} - \bar{z}\rangle + \frac{1}{2\beta}\|M\bar{x} - \bar{z}\|^2$$
   and that
$$S_{\frac{1}{\rho k}}(\bar{x}_k, \bar{z}_k, \phi_0) \leq \frac{\rho}{2k}\|z_0 - Mx^*\|^2.$$

7. Show that for all $\bar{x}, \bar{z}, \phi_0, \beta$, and for a saddle point (x^*, ϕ^*),
   (a) $f(\bar{x}) + g(\bar{z}) - f(x^*) - g(Mx^*) \geq -\langle \phi^*, M\bar{x} - \bar{z}\rangle \geq -\|\phi^*\|.\|M\bar{x} - \bar{z}\|$
   (b) $f(\bar{x}) + g(\bar{z}) - f(x^*) - g(Mx^*) = S_\beta(\bar{x}, \bar{z}, \phi_0) - \langle \phi_0, M\bar{x} - \bar{z}\rangle - \frac{1}{2\beta}\|M\bar{x} - \bar{z}\|^2$
   $\leq S_\beta(\bar{x}, \bar{z}, \phi_0) + \|\phi_0\|.\|M\bar{x} - \bar{z}\|$
   (c) $\|M\bar{x} - \bar{z}\| \leq \beta\Big(\|\phi_0 - \phi^*\| + \big(\|\phi_0 - \phi^*\|^2 + \frac{2}{\beta}S_\beta(\bar{x}, \bar{z}, \phi_0)\big)^{1/2}\Big)$

8. Show that the ergodic sequence $(\bar{x}_k, \bar{z}_k, \bar{\phi}_k)$ has a convergence speed as $O(1/k)$ in function value $f(\bar{x}_k) + g(\bar{z}_k)$ and feasibility.

9. Assume that g is Lipschitz continuous. Show that $f(\bar{x}_k) + g(M\bar{x}_k)$ converges to $f(x^*) + g(Mx^*)$ as $O(1/k)$.

In this exercise, we showed the convergence speed for the ergodic mean of the iterates. We shall refer to "Damek Davis and Wotao Yin, (2016), Convergence rate analysis of several splitting schemes, in Splitting Methods in Communication, Imaging, Science, and Engineering (pp. 115-163), Springer International Publishing" for the non-ergodic speed of the ADMM.