Exercice 1. *(Variation totale)*
Soit $b = (b_1, \dots, b_n)^T$ un vecteur de $\mathbb{R}^n$. On se pose le problème suivant :

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

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 ?*

   ▶ Par analogie avec la régularisation par la norme 1, ce deuxième terme favorisera les solutions vérifiant $x_{i+1} - x_i = 0$ pour une majorité de i. La solution de ce problème aura donc tendance à être constante par morceaux même si b ne l'était pas. Elle aura aussi moins de pics de grande amplitude.

2. *Montrer que le problème (1) 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.

   ▶ On pose $f(x) = \frac{1}{2}\|x - b\|^2$, $g(y) = \eta \sum_{i=1}^{n-1} |y_i|$ et $M \in \mathbb{R}^{(n-1)\times n}$ telle que $(Mx)_i = x_{i+1} - x_i$ pour tout $i \in \{1, \dots n-1\}$.

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

   ▶ L'opérateur proximal de g est le seuillage doux.
   Pour l'opérateur proximal de f, on cherche x qui minimise $\frac{1}{2}\|x - b\|^2 + \frac{1}{2}\|y - x\|^2$, c'est à dire $\mathrm{prox}_f(y) = \frac{y+b}{2}$.

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

   ▶

$$M = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \qquad M^T M = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

5. *Écrire les itérations de l'ADMM de pas $\sigma > 0$ pour la résolution de (1).*

   ▶ Initialisation : $\lambda_0 \in \mathbb{R}^{n-1}$ et $z_0 \in \mathbb{R}^{n-1}$.

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

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

   ▶ L'étape de calcul de x_{k+1} revient à chercher un x solution de

   $$(x-b)+M^T\lambda_k+\sigma M^T(Mx-z_k)=0$$

   c'est-à-dire $(I+\sigma M^TM)x_{k+1}=M^T(z_k-\lambda_k)+b$. Cette étape revient à la résolution d'un système linéaire.

   L'étape de calcul de z_{k+1} revient à chercher z qui minimise

   $$g(z)-\langle\lambda_k,z\rangle+\frac{\sigma}{2}\|Mx_{k+1}-z\|^2=\sigma\Big(\frac{1}{\sigma}g(z)+\frac{1}{2}\|z-Mx_{k+1}-\frac{1}{\sigma}\lambda_k\|^2\Big)-\frac{1}{2\sigma}\|\lambda_k\|^2+\langle Mx_{k+1},\lambda_k\rangle$$

   Ainsi, z_{k+1} minimise $\frac{1}{\sigma}g(z)+\frac{1}{2}\|z-Mx_{k+1}-\frac{1}{\sigma}\lambda_k\|^2$. C'est

   $$z_{k+1}=\text{prox}_{\frac{1}{\sigma}g}(Mx_{k+1}+\frac{1}{\sigma}\lambda_k).$$

   Cette étape revient donc à un seuillage doux.

Exercice 2. (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

$$\min F(x_1,\dots,x_N)\quad s.t.\quad x_1=\dots=x_N\,. \tag{2}$$

1. *State that problem (2) 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.*

   ▶ Let $C_N=\{z\in\mathcal{X}^N:z_1=\ldots=z_n\}$. $x\in C_N$ if and only if it satisfies the constraint so both problems are equivalent.

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

   ▶ The ADMM is defined as follows.

   Initialisation : $\lambda_0\in\mathcal{X}^N$ and $z_0\in\mathcal{X}^N$.

   $$\begin{aligned}x^{k+1}&=\arg\min_{x\in\mathcal{X}^N}F(x)+\langle\lambda^k,x\rangle+\frac{\sigma}{2}\|x-z^k\|^2\\ z^{k+1}&=\arg\min_{z\in\mathcal{X}^N}\iota_{C_N}(z)-\langle\lambda^k,z\rangle+\frac{\sigma}{2}\|x^{k+1}-z\|^2\\ \lambda^{k+1}&=\lambda^k+\sigma(x^{k+1}-z^{k+1})\end{aligned}$$

The function F is separable so, given λ_k and z_k, we can compute x_{k+1} element-wise and in parallel without communication :

$$x_i^{k+1} = \arg\min_{x\in\mathcal{X}} f_i(x) + \langle \lambda_i^k, x\rangle + \frac{\sigma}{2}\|x - z_i^k\|^2, \qquad \forall i \in \{1,\dots,N\}$$

The update of z_{k+1} amounts to the projection of $x^{k+1} - \frac{1}{\sigma}\lambda^k$ onto C_N. This is given for all j by

$$z_j^{k+1} = \frac{1}{N}\sum_{i=1}^N \big(x_i^{k+1} - \frac{1}{\sigma}\lambda_i^k\big).$$

This second step requires communication between the computing agents.
The update for λ^{k+1} is a sum of vectors and requires no communication.

3. *Explicit the algorithm in the case where*

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

►

$$\begin{aligned} &x_i^{k+1} \text{ solution to } (\sigma I + A_i^T A_i)x = A_i^T b + z_i^k - \lambda_i^k, && \forall i \in \{1,\dots,N\} \\ &z_j^{k+1} = \frac{1}{N}\sum_{i=1}^N \big(x_i^{k+1} - \frac{1}{\sigma}\lambda_i^k\big), && \forall j \in \{1,\dots,N\} \\ &\lambda_i^{k+1} = \lambda_i^k + \sigma(x_i^{k+1} - z_i^{k+1}), && \forall i \in \{1,\dots,N\} \end{aligned}$$

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*

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

► We have this equality if the graph is connected.

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.*

► We denote $g : \mathcal{X}^{2|E|} \to \mathbb{R}\cup\{+\infty\}$ such that $g(z) = \sum_{e\in E} \iota_{C_2}(z_{e,-}, z_{e,+})$ where $z_{e,-}$ is the first coordinate of z_e and $z_{e,+}$ is its second coordinate (z has two coordinates per edge by definition). Then

$$g(Mx) = \sum_{e\in E} \iota_{C_2}((Mx)_{e,-}, (Mx)_{e,+}) = \sum_{i,j\in E} \iota_{C_2}(x_i, x_j) = \iota_{C_N}(x)$$

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

▶ The ADMM is defined as follows.
Initialisation : $\lambda_0 \in \mathcal{X}^{2|E|}$ and $z_0 \in \mathcal{X}^{2|E|}$.

$$\begin{aligned} x^{k+1} &= \arg\min_{x \in \mathcal{X}^N} F(x) + \langle \lambda^k, Mx \rangle + \frac{\sigma}{2}\|Mx - z^k\|^2 \\ z^{k+1} &= \arg\min_{z \in \mathcal{X}^{2|E|}} g(z) - \langle \lambda^k, z \rangle + \frac{\sigma}{2}\|Mx^{k+1} - z\|^2 \\ \lambda^{k+1} &= \lambda^k + \sigma(Mx^{k+1} - z^{k+1}) \end{aligned}$$

The variables x are located on nodes and the variables λ and z are located on edges (2 per edge). Hence, the node i may take care of the variable x_i located on itself and half of the variables λ_e and z_e such that $i \in e$ (that is one side of the edge). If one node needs to access a variable located at an other side of an adjacent edge, it must communicate with its neighbour.

The update for λ^{k+1} writes as

$$\lambda_e^{k+1} = \lambda_e^k + \sigma(Mx^{k+1})_e - z_e^{k+1}$$

Hence, it can be performed without any communication.

The update for z^{k+1} relies on the function $z \mapsto g(z) - \langle \lambda^k, z \rangle + \frac{\sigma}{2}\|Mx^{k+1} - z\|^2$, which is block-separable with blocks of size 2 corresponding to each edge. Hence,

$$z_e^{k+1} = \arg\min_{z \in \mathcal{X}^2} \iota_{C_2}(z) - \langle \lambda_e^k, z_e \rangle + \frac{\sigma}{2}\|(Mx^{k+1})_e - z\|^2$$

$$z_{e,-}^{k+1} = z_{e,+}^{k+1} = \frac{1}{2}\Big((Mx^{k+1})_{e,-} + \frac{1}{\sigma}\lambda_{e,-}^k + (Mx^{k+1})_{e,+} + \frac{1}{\sigma}\lambda_{e,+}^k\Big)$$

This update features communication between the agents on each side of the edge.

The update for x^{k+1} is the minimizer to $(x \mapsto F(x) + \langle \lambda^k, Mx \rangle + \frac{\sigma}{2}\|Mx - z^k\|^2)$. This function does not look separable but in fact it is.
F and $\langle \lambda^k, Mx \rangle = \langle M^T\lambda^k, x \rangle$ are clearly separable. Let us now denote $\sigma(i)$ the set of neighbours of node i.

$$\begin{aligned} \|Mx - z^k\|^2 = \sum_{e \in E}\|(Mx)_e - z_e^k\|^2 &= \sum_{e \in E}\|(Mx)_{e,-} - z_{e,-}^k\|^2 + \|(Mx)_{e,+} - z_{e,+}^k\|^2 \\ &= \frac{1}{2}\sum_{i=1}^N \sum_{j \in \sigma(i)} \|(Mx)_{\{i,j\},-} - z_{\{i,j\},-}^k\|^2 + \|(Mx)_{\{i,j\},+} - z_{\{i,j\},+}^k\|^2 \\ &= \frac{1}{2}\sum_{i=1}^N \sum_{j \in \sigma(i)} \|x_i - z_{\{i,j\},+}^k\|^2 + \|x_j - z_{\{i,j\},+}^k\|^2 \\ &= \sum_{i=1}^N \sum_{j \in \sigma(i)} \|x_i - z_{\{i,j\},+}^k\|^2 \end{aligned}$$

In the third equality, we counted each edge twice, so we needed to divide by 2. In the fourth equality, we used the fact that $z^k_{\{i,j\},-} = z^k_{\{i,j\},+}$.
This shows that for all i, and for all $e \in E$,

$$
\begin{aligned}
x_i^{k+1} &= \arg\min_{x\in\mathcal{X}} f_i(x) + \left\langle \sum_{j\in\sigma(i), i<j} \lambda^k_{\{i,j\},-} + \sum_{j\in\sigma(i), i>j} \lambda^k_{\{i,j\},+}, x \right\rangle + \sum_{i=1}^{N} \sum_{j\in\sigma(i)} \|x - z^k_{\{i,j\},+}\|^2 \\
z^{k+1}_{e,-} &= z^{k+1}_{e,+} = \frac{1}{2}\Big((Mx^{k+1})_{e,-} + \frac{1}{\sigma}\lambda^k_{e,-} + (Mx^{k+1})_{e,+} + \frac{1}{\sigma}\lambda^k_{e,+}\Big) \\
\lambda^{k+1}_e &= \lambda^k_e + \sigma(Mx^{k+1})_e - z^{k+1}_e
\end{aligned}
$$

Exercice 3. *The goal of this exercise is to define a variant of ADMM able to solve the following problem*

$$
\begin{aligned}
&\min_{x,z} f(x) + g(z) \\
&st : Ax + Bz = c
\end{aligned}
\tag{3}
$$

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*

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

▶

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

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

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

▶ Let us begin with (3). The Lagrangian is

$$
L(x, z, \lambda) = f(x) + g(z) + \lambda^T(Ax + Bz - c)
$$

Minimizing with respect to x and z, we get

$$
\inf_{x,z} L(x, z, \lambda) = -f^*(-A^T\lambda) - g^*(-B^T\lambda) - \langle \lambda, c\rangle
$$

Up to a change in variables $\lambda' = -\lambda$, this is $D(\lambda)$. (Note that we could also have defined $L(x, z, \lambda) = f(x) + g(z) + \lambda^T(-Ax - Bz + c)$ to get directly the expected formula.)

For the other formulation, the Lagrangian is

$$L'(x, y, \phi) = f(x) + h(y) + \phi^T(y - Ax)$$

$$\begin{aligned}\inf_{x,y} L'(x, y, \phi) &= \inf_{x,y,z} f(x) + g(z) + \iota_{\{0\}}(Bz + y - c) + \phi^T(y - Ax) \\ &= \inf_{x,z} f(x) + g(z) + \phi^T(c - Bz - Ax) \\ &= -f^*(A^T\phi) - g^*(B^T\phi) + \phi^T c = D(\phi)\end{aligned}$$

3. *Write the ADMM for the problem*

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

▶

$$\begin{aligned}x^{k+1} &= \arg\min_x f(x) + \langle Ax, \lambda^k\rangle + \frac{\sigma}{2}\|Ax - y^k\|^2 \\ y^{k+1} &= \arg\min_y h(y) - \langle y, \lambda^k\rangle + \frac{\sigma}{2}\|Ax^{k+1} - y\|^2 \\ \lambda^{k+1} &= \lambda^k + \sigma(Ax_{k+1} - y^{k+1})\end{aligned}$$

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

▶

$$\inf_y h(y) - \langle y, \lambda^k\rangle + \frac{\sigma}{2}\|Ax^{k+1} - y\|^2 = \inf_{y,z} g(z) + \iota_{\{0\}}(Bz + y - c) - \langle y, \lambda^k\rangle + \frac{\sigma}{2}\|Ax^{k+1} - y\|^2$$
$$= \inf_z g(z) + \langle Bz - c, \lambda^k\rangle + \frac{\sigma}{2}\|Ax^{k+1} + Bz - c\|^2$$

Moreover, if z is a minimizer to the problem in g, then $y = c - Bz$ is a minimizer to the problem in h. Hence, we can write the ADMM by replacing every occurence of y by $c - Bz$ and we get

$$\begin{aligned}x^{k+1} &= \arg\min_x f(x) + \langle Ax, \lambda^k\rangle + \frac{\sigma}{2}\|Ax + Bz^k - c\|^2 \\ z^{k+1} &= \arg\min_z g(z) + \langle Bz, \lambda^k\rangle + \frac{\sigma}{2}\|Ax^{k+1} + Bz - c\|^2 \\ \lambda^{k+1} &= \lambda^k + \sigma(Ax_{k+1} + Bz^{k+1} - c)\end{aligned}$$