INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE

Optimization-I

Correction of tutorials

4th year

Dr. Frédéric de Gournay

Correction of tutorials

1 Reminders

Exercise 1.1 will test your computational skills, if you know Leibniz formula and the relationship between the gradient and the Jacobian.

(TD1 : 15 mins) Exercise 1.1

Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$.

1. Let $\Psi : \mathbb{R}^m \mapsto \mathbb{R}$ be defined as

$$\Psi(y) = \sum_{i=1}^{n} y_i^4 + 1.$$

   Compute $\nabla\Psi(y)$ and $H[\Psi]$. Is Ψ convex ?
2. Let $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $\varphi(x) = Ax - b$. Compute the jacobian of φ
3. Let $f(x) = \Psi(Ax - b)$ show that $\nabla f(x) = A^T \nabla\Psi(Ax - b)$ by two different methods. The first method is a direct Taylor expansion (compute $f(x+h) - f(x)$ for small h), for the second method use the chain rule (Leibniz formula).

Solution of Exercise 1.1

1. The function Ψ is a polynomial function on $\mathbb{R}^m$, it is differentiable on $\mathbb{R}^m$ (and even C^∞ on $\mathbb{R}^m$). Denoting $(e_i)_i$ the canonical basis, we have

$$\nabla\Psi(y) = \left(\frac{\partial\Psi}{\partial y_i}\right)_{1\le i\le m} = \begin{pmatrix} 4y_1^3 \\ \vdots \\ 4y_m^3 \end{pmatrix} = \sum_{i=1}^{m} 4y_i^3 e_i$$

$$H[\Psi](y) = \left(\frac{\partial^2\Psi}{\partial y_i \partial y_j}\right)_{1\le i,j\le m} = \begin{pmatrix} 12y_1^2 & & 0 \\ & \ddots & \\ 0 & & 12y_m^2 \end{pmatrix} = \sum_{i=1}^{m} 12y_i^2 e_i \otimes e_i.$$

   For any $y \in \mathbb{R}^m$, the eigenvalues of the hessian are $12y_1^2, \ldots, 12y_m^2 \ge 0$, so that Ψ is convex on $\mathbb{R}^m$.
2. The function φ is an affine function, hence it is differentiable and even C^∞ on $\mathbb{R}^n$. Its Jacobian is given by

$$\forall x \in \mathbb{R}^n,\ J_\varphi(x) = A.$$

   If we really need to be convinced of the result, we can perform a Taylor expansion:

$$\begin{aligned} \forall h \in \mathbb{R}^n,\ \varphi(x+h) &= A(x+h) - b = (Ax - b) + Ah \\ &= \varphi(x) + \underbrace{Ah}_{\text{linear part in } h = J_\varphi(x)h} \end{aligned}$$

   And we obtain $J_\varphi(x) = A$.
3. We compute the gradient of f by two different methods
   (a) **By Taylor expansions**

$$\begin{aligned} \forall h \in \mathbb{R}^n,\ f(x+h) &= \Psi(A(x+h) - b) = \Psi((Ax - b) + Ah) \\ &= \Psi(Ax - b) + \langle \nabla\Psi(Ax - b), Ah\rangle + o(\|h\|) \\ &= f(x) + \langle A^\top \nabla\Psi(Ax - b), h\rangle + o(\|h\|) \end{aligned}$$

Hence we obtain

$$\nabla f(x) = A^\top \nabla\Psi(Ax - b).$$

(b) **By Leibniz formula** We have $f = \Psi \circ \varphi$, indeed

$$f : x \overset{\varphi}{\longmapsto} Ax - b \overset{\Psi}{\longmapsto} \Psi(Ax - b).$$

By the chain rule, we have

$$J_f(x) \quad = \quad J_\Psi(\varphi(x))J_\varphi(x) = J_\Psi(Ax - b)J_\varphi(x).$$

Or $f : \mathbb{R}^n \to \mathbb{R}$ and $\Psi : \mathbb{R}^m \to \mathbb{R}$, hence:

$$J_f(x) = \nabla f(x)^\top, \quad \text{et} \quad J_\Psi(x) = \nabla\Psi(x)^\top.$$

We then have

$$\nabla f(x) = J_\varphi(x)^\top \nabla\Psi(Ax - b),$$

which gives by previous computations:

$$\nabla f(x) = A^\top \Psi(Ax - b).$$

In Exercise 1.2, we prove a very important formula for least-square approximation. The result of this exercise and the method of proof must be learned by heart, as if your life depends on it[1].

(TD1 : 15 mins) Exercise 1.2

Let $A_1 \in \mathcal{M}_{n,n}(\mathbb{R})$ be a symetric matrix, $b_1 \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Let

$$f_1 : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle A_1 x, x\rangle - \langle b_1, x\rangle + c.$$

1. Show that $\nabla f_1(x) = A_1 x - b_1$ and $H[f_1](x) = A_1$.
2. Under which condition is f_1 convex ? strictly convex ?
3. Let $A_2 \in \mathcal{M}_{n,m}(\mathbb{R})$, $b_2 \in \mathbb{R}^m$ and

$$f_2 : x \in \mathbb{R}^n \mapsto \frac{1}{2}\|A_2 x - b_2\|^2$$

4. Show that $\nabla f_2(x) = A_2^T(A_2 x - b_2)$ and $H[f_2](x) = A_2^T A_2$.
5. Under which condition is f_2 convex ? strictly convex ?

[1]Your life may not **really** depend on knowing Exercise 1.2 by heart, but your success in class certainly will.

Solution of Exercise 1.2

1. We perform a first order Taylor expansion of f_1, for any $h \in \mathbb{R}^n$, we obtain

$$\begin{aligned} f_1(x+h) &= \frac{1}{2}\langle A_1(x+h),(x+h)\rangle - \langle b_1,(x+h)\rangle + c \\ &= \frac{1}{2}\langle A_1 x, x\rangle - \langle b_1, x\rangle + c \\ &\quad + \frac{1}{2}\langle A_1 x, h\rangle + \frac{1}{2}\langle A_1 h, x\rangle - \langle b_1, h\rangle + o(\|h\|) \\ &= f_1(x) + \langle \frac{1}{2}A_1 x + \frac{1}{2}A_1^\top x - b_1, h\rangle + o(\|h\|)\,. \end{aligned}$$

   We then have $\nabla f_1(x) = \frac{1}{2}A_1 x + \frac{1}{2}A_1^\top x - b_1 = A_1 x - b_1$.
2. We remark that ∇f_1 is an affine mapping hence the Jacobian of the gradient of f_1 (which is the hessian of f) is given by the linear part of the affine mapping. We then have

$$H[f_1](x) = A_1.$$

   then f_1 is convex iff $H[f_1] \succeq 0$ at each point hence iff A_1 is semi-definite positive. The function f_1 is strictly convex iff $A_1 \succ 0$.
3. It is sufficient to remark that

$$f_2(x) = \frac{1}{2}\langle A_2 x, A_2 x\rangle - \langle A_2 x, b_2\rangle + \langle b_2, b_2\rangle.$$

   We just need to apply the previous result to $A_1 = A_2^\top A_2$ and $b_1 = A_2^\top b_2$.

In Exercise 1.3, we revise the natural scalar product on the space of matrices (Frobenius), we reprove some formula for the gradient of the determinant and the trace.

(15 mins) Exercise 1.3

We recall that $\mathcal{M}_{n,n}(\mathbb{R})$, the vector space of real square matrices of size n is Euclidean when endowed with the scalar product $\langle A, B\rangle = \mathrm{tr}(A^\top B)$. Show quickly that each of the following map is differentiable on its domain, compute their differential and their gradient.

1. $\mathrm{tr} : A \mapsto \mathrm{tr}(A)$
2. $\det : A \mapsto \det(A)$ when A is invertible (hint, first prove that $\nabla \det(I_n) = I_n$).
3. $f : A \mapsto \ln|\det(A)|$ when A is invertible

Solution of Exercise 1.3

1. For any $H \in \mathcal{M}_{n,n}(\mathbb{R})$, we have

$$\begin{aligned} \mathrm{tr}(A+H) &= \mathrm{tr}(A) + \mathrm{tr}(H) \qquad \text{because tr is linear} \\ &= \mathrm{tr}(A) + \langle I_n, H\rangle \end{aligned}$$

   Hence we obtain $\nabla \mathrm{tr}(A) = I_n$.
2. For any H, compute $\det(I_n + H)$, it is a polynom in the different components of H, the zero order term is 1 and the term of order 1 is $\mathrm{tr}(H)$. In short, we have

$$\det(Id + H) = 1 + \mathrm{tr}(H) + o(\|H\|)\,.$$

Hence $\nabla \det(I_n) = I_n$. Now for any invertible matrix A, we have

$$\begin{aligned} \det(A+H) &= \det(A(Id + A^{-1}H)) = \det(A)\det(Id + A^{-1}H) \\ &= \det(A) + \det(A)\operatorname{tr}(A^{-1}H) + \mathcal{O}(\|H\|) \end{aligned}$$

Hence $\nabla \det(A) = \det(A) A^{-\top}$.

3. We have $\ln(1+u) = u + \mathcal{O}(u)$, hence

$$\begin{aligned} \ln|\det(A+H)| &= \ln|\det(A)\det(Id + A^{-1}H))| \\ &= \ln|\det(A)| + \ln\left(1 + \operatorname{tr}(A^{-1}H) + \mathcal{O}(\|H\|)\right) \\ &= \ln|\det(A)| + \operatorname{tr}(A^{-1}H) + \mathcal{O}(\|H\|) \end{aligned}$$

Hence $\nabla f(A) = A^{-\top}$.

Exercise 1.4 has exactly the same pedagogical goal as Exercise 1.3, it can be used to see if one masters the techniques of Exercise 1.3.

(15 mins) Exercise 1.4

We work in the space $\mathcal{M}_{m,n}(\mathbb{R})$ of real matrices of size $m \times n$ with the scalar product $\langle U, V\rangle = \operatorname{tr}(U^\top V)$. Let $B \in \mathcal{S}_n$ a symetric $n \times n$ real matrix. Let

$$f : A \in \mathcal{M}_{m,n}(\mathbb{R}) \mapsto \operatorname{tr}(ABA^\top)$$

Give the differential and the gradient of f, you must find $\nabla f(A) = 2AB$.

Solution of Exercise 1.4

$\forall H \in \mathcal{M}_{m,n}(\mathbb{R})$, it holds that

$$\begin{aligned} f(A+H) &= tr((A+H)B(A+H)^\top) \\ &= tr(ABA + ABH^\top + HBA^\top + HBH^\top) \\ &= \underbrace{tr(ABA)}_{=f(A)} + \underbrace{tr(ABH^\top) + tr(HBA^\top)}_{\text{linear part in } H} + \underbrace{tr(HBH^\top)}_{\text{term of order 2 in } H} \\ &= f(A) + tr(ABH^\top) + tr(HBA^\top) + o(\|H\|) \end{aligned}$$

We have the linear part in H (the differential), we now want to write it as $\langle \nabla f(A), H\rangle$. To that extent, we recall the following properties of the trace function, valid for any function $M \in \mathcal{M}_{m,n}(\mathbb{R})$ and $N \in \mathcal{M}_{n,m}(\mathbb{R})$:

$$tr(MN) = tr(NM)$$

For any square matrix P, we have:

$$tr(P^\top) = tr(P).$$

We obtain :

- $tr(ABH^\top) = tr(H^\top AB) = \langle H, AB\rangle = \langle AB, H\rangle$.
- $tr(HBA^\top) = tr(BA^\top H) = \langle AB^\top, H\rangle = \langle AB, H\rangle$ (because B is symetric).

Hence:

$$f(A+H) = f(A) + 2\langle AB, H\rangle + o(\|H\|).$$

and we obtain $\nabla f(A) = 2AB$.

Correction of tutorials

In Exercise 1.5, the goal is to get a quick intuition on convexity, the idea is to rely on sketches of the solutions and to design an argument from an intuition.

(TD1 : 10 mins) Exercise 1.5

Draw the following functions and prove that they are convex or prove that they are not

1. $f : x \mapsto x^2$
2. $f : x \mapsto e^{-x^2}$
3. $f : x \mapsto e^x$
4. $f : x \mapsto \sqrt{x}$
5. $f : x \mapsto e^{-x}$
6. $f : x \mapsto \frac{1}{x}$

Solution of Exercise 1.5

1. We have $f''(x) = 2 > 0$, the function is strictly convex.
2. We have $f'(x) = -2xe^{-x^2}$ and $f''(x) = (-2 + 4x^2)e^{-x^2}$. We have $f''(x) < 0$ for small x, hence f is not convex.
3. We have $f''(x) = e^x > 0$. Hence f is convex.
4. We have $f''(x) = -\frac{1}{4}x^{-3/2}$, hence f'' is not convex, it is concave.
5. We have $f''(x) = e^{-x} > 0$. Hence f is convex.
6. We have $f''(x) = 2x^{-3} > 0$. Hence f is convex on $]0, +\infty[$.

The right approach to Exercise 1.6 is to compute Hessians and to discuss according to the sign of the eigenvalues.

(TD1 : 10 mins) Exercise 1.6

Discuss about the convexity or concavity of the following functions

1. $f(x, y) = (y - x^2)^2 - x^2$
2. $f(x, y) = x^4 + y^4 - (x - y)^2$
3. $f(x, y, z) = x^4 + 2y^2 + 3z^2 - yz - 23y + 4x - 5$

Solution of Exercise 1.6

1. We have:
$$H[f](x, y) = \begin{pmatrix} -2 - 4y + 12x^2 & -4x \\ -4x & 2 \end{pmatrix}$$
For instance, at the point $(0, 0)$, the matrix
$$H_{f_1}(0, 0) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$$
has two eigenvalues : 2 and -2 which have opposite sign. Hence f is neither convex nor concave.
2. We have :
$$H[f](x, y) = \begin{pmatrix} 12x^2 - 2 & 2 \\ 2 & 12y^2 - 2 \end{pmatrix}$$
For instance at the point $(0, 0)$, we have:
$$H[f](0, 0) = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}.$$
The determinant of the hessain is equal to 0. Hence 0 is an eigenvalue and the other one is equal to $\mathrm{tr}(H[f](0, 0)) = -4 < 0$. Hence f is not convex. Similarly, taking $(x, y) = (1, 1)$, the determinant and the trace are positive, hence f has two positive eigenvalues and is not concave.

3. We have :

$$\nabla f(x,y,z) = \begin{pmatrix} 4x^3+4 \\ 4y-z-23 \\ 6z-y \end{pmatrix}, \quad H[f](x,y) = \begin{pmatrix} 12x^2 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 6 \end{pmatrix}$$

The eigenvalues of the Hessian at the point (x,y,z) are $12x^2 \geq 0$ and the eigenvalues of the matrix $\begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$ which are striclty positive ($det > 0$, $Tr > 0$) for any choice of (x,y,z). Hence the function f is convex.

Exercise 1.7 is very simple, just go back to the definitions.

(5 mins) Exercise 1.7

Let E be a vector space with a norm $\| \bullet \|$, show that the norm is a convex function.

Solution of Exercise 1.7

Let any x, y in E and $\theta \in]0,1[$, then

$$\|\theta x + (1-\theta)y\| \quad \leq \quad \|\theta x\| + \|(1-\theta)y\| = \theta\|x\| + (1-\theta)\|y\|$$

For Exercise 1.8, you need to come back to the definitions.

(5 mins) Exercise 1.8

Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a convex non-decreasing function and $h : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function. Show that $f = g \circ h$ is convex. Give an example where g and h are convex but $g \circ h$ is non-convex.

Solution of Exercise 1.8

Let x and y in $\mathbb{R}^n$ and $\theta \in]0,1[$. Then, by convexity of h

$$h(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta)h(y).$$

By non-decreasingness of g, we have

$$f(\theta x + (1-\theta)y) = g(h(\theta x + (1-\theta)y)) \leq g\left(\theta h(x) + (1-\theta)h(y)\right).$$

Finally by convexity of g, we have

$$g\left(\theta h(x) + (1-\theta)h(y)\right) \leq \theta g(h(x)) + (1-\theta)g(h(y)) = \theta f(x) + (1-\theta)f(y).$$

For the counter example take $g : x \mapsto e^{-x}$ and $h : x \mapsto x^2$, then $f : x \mapsto e^{-x^2}$ is non-convex.

Exercise 1.9 is a little bit involved in the computational side. Hopefully we give you the result for the gradient and the Hessian, so that you can skip the first question if you can't solve it.

(15 mins) Exercise 1.9

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be defined as $f(x) = \log(e^{x_1} + \ldots e^{x_n})$. For any $x \in \mathbb{R}^n$, denote $s \in \mathbb{R}^n$ such that $s_i = \frac{e^{x_i}}{e^{x_1} + \cdots + e^{x_n}}$.

1. Show that

$$\nabla f(x) = s \quad \text{and} \quad \begin{cases} (H[f](x))_{ii} = s_i^2 - s_i s_i \\ (H[f](x))_{ij} = -s_i s_j & \text{if } i \neq j \end{cases}$$

2. Show that $(H[f](x)h, h) = \sum_i s_i h_i^2 - (\sum_i s_i h_i)^2$.
3. From $\sum s_i = 1$ and $s_i > 0$, conclude that f is (strictly) convex.

Solution of Exercise 1.9

1. First, f is regular. We compute the first order derivative of f, we obtain $\partial_i f(x) = s_i$. We now compute the second order derivative with respect to j, there are two cases $j = i$ or $j \neq i$. We obtain

$$\begin{cases} \partial_{ij} f = -\frac{e^{x_i} e^{x_j}}{(\sum_j e^{x_j})^2} = -s_i s_j & \text{if } j \neq i \\ \partial_{ii} f = \frac{e^{x_i}}{\sum_j e^{x_j}} - \frac{e^{x_i} e^{x_i}}{(\sum_j e^{x_j})^2} = s_i - s_i^2 & \text{if } j = i \end{cases}.$$

2. For any h we have

$$\begin{aligned} (H[f](x)h, h) &= \sum_i H[f](x)_{ii} h_i^2 + \sum_{i \neq j} H[f](x)_{ij} h_i h_j \\ &= \sum_i s_i h_i^2 - \sum_i s_i s_i h_i h_i + \sum_{i \neq j} s_i s_j h_i h_j \\ &= \sum_i s_i h_i^2 - \sum_{i,j} s_i s_j h_i h_j = \sum_i s_i h_i^2 - (\sum_i s_i h_i)^2 \end{aligned}$$

3. It is immediate that $\sum_i s_i = 1$ and $s_i > 0$. Denote $g : x \mapsto x^2$, then

$$(H[f](x)h, h) = \sum_i s_i g(h_i) - g(\sum_i s_i h_i) > 0.$$

The last inequality is Jensen's inequality applied to the strictly convex function g. Hence $H[f](x) \succ 0$ for all x and f is strictly convex .

For Exercise 1.10, remember the following result : If
- for every $i \in \mathcal{I}$, the function g_i is convex,
- for every $j \in \mathcal{E}$, the function g_i is affine.

then the set

$$\{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in \mathcal{I} \text{ and } g_i(x) = 0, i \in \mathcal{E}\}$$

is convex. The goal of this exercise is to remember this result and to focus on the fact that there is no equivalence. Most of the ideas must originate from drawings.

(TD1 : 10 mins) Exercise 1.10

Draw the following sets, are they convex ?
1. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y \geq x^2\}$
2. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1,\ y \geq x^2\}$
3. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y \leq x^2,\ y \leq 0\}$
4. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y = 2x\}$

Solution of Exercise 1.10

1. Let

$$g_1 : (x,y) \in \mathbb{R}^2 \mapsto x^2 + y^2 - 1 \quad \text{et} \quad g_2 : (x,y) \in \mathbb{R}^2 \mapsto x^2 - y.$$

   We have $X = \{(x,y) \in \mathbb{R}^2 \mid g_1(x,y) \leq 0,\ g_2(x,y) \leq 0\}$. The function g_1 is C^2 (even C^∞) on $\mathbb{R}^2$ as a polynomial function and its hessian is $H[g_1](x,y) = 2I_2 \succ 0$ for every $(x,y) \in \mathbb{R}^2$. Hence g_1 is (strictly) convex. Similarly g_2 is C^2 and its hessian is equal to :

$$H[g_2](x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \succeq 0$$

   Hence g_2 is convex and X is convex
2. We have

$$\begin{aligned} X &= \{(x,y) \in \mathbb{R}^2 \mid g_1(x,y) \geq 0,\ g_2(x,y) \leq 0\} \\ &= \{(x,y) \in \mathbb{R}^2 \mid -g_1(x,y) \leq 0,\ g_2(x,y) \leq 0\}. \end{aligned}$$

   The function g_1 is convex, hence $-g_1$ is concave. With a drawing, we observe that X is not convex . For instance take $x_1 = ((1-\varepsilon)^2, (1-\varepsilon))$ and $x_2 = ((1-\varepsilon)^2, -(1-\varepsilon))$ then $\frac{1}{2}x_1 + \frac{1}{2}x_2$ does not belong to X for small ε.
3. Here one of the inequalities is not convex but X is convex.
4. Here X is convex by a direct application of the result above.

Exercise 1.11 is solved by going back to the basis.

(5 mins) Exercise 1.11

Let g be a concave positive function. Show that $f = \frac{1}{g}$ is convex

Solution of Exercise 1.11

Let $h(z) = \frac{1}{z}$, we have $f = h \circ g$ and h is convex and decreasing on $\mathbb{R}^{+*}$. For any x, y and $\theta \in]0,1[$, we have

$$\begin{aligned} g(\theta x + (1-\theta)y) &\geq \theta g(x) + (1-\theta)g(y) \quad \text{(concavity of } g) \\ h \circ g(\theta x + (1-\theta)y) &\leq h(\theta g(x) + (1-\theta)g(y)) \quad (h \text{ is decreasing}) \\ f(\theta x + (1-\theta)y) &\leq \theta h(g(x)) + (1-\theta)h(g(y))) \quad (h \text{ is convex}) \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

2 Existence

The difficulty in showing existence of solution is the coercivity. Exercise 2.1 is in finite dimension and we have to prove coercivity by different ways.

(TD2 : 15 mins) Exercise 2.1

Show that the following problem $\inf_{M \in X} f(M)$ admits minima

1. $f : (x, y) \mapsto x^2 + y^2$, $X = \mathbb{R}^2$
2. $f : (x, y) \mapsto 3x$, $X = \{(x, y) \text{ such that } x^2 + y^2 \leq 1\}$
3. $f : (x, y) \mapsto x^4 + y^6 + x^2y^2$, $X = \mathbb{R}^2$
4. $f : (x, y) \mapsto 4x^2 + 6y^2 - 5xy$, $X = \mathbb{R}^2$

Solution of Exercise 2.1

1. The function f is continuous, and X is closed but unbounded. we have $f(M) = \|M\|^2$ and f is coercive.
2. The function f is continuous, and X is closed and bounded.
3. The function f is continuous, and X is closed but unbounded. Suppose that $M_n = (x_n, y_n)$ is a sequence of points such that $\|M_n\| \to +\infty$. We can suppose that $\|M_n\|_\infty > 1$, for each k, we have either $|x_k| = \|M_k\|_\infty$ or $|y_k| = \|M_k\|_\infty$.
   (a) In the case $|x_k| = \|M_k\|_\infty$, we have $f(M_k) \geq x_k^4 = \|M_k\|_\infty^4$.
   (b) In the case $|y_k| = \|M_k\|_\infty$, we have $f(M_k) \geq y_k^6 = \|M_k\|_\infty^6 \geq \|M_k\|_\infty^4$.
   In any case we have $f(M_k) \geq \|M_k\|_\infty^4$, hence f is coercive.
4. This case is a little bit more annoying because the term $-5xy$ may be very negative. We use
$$-\frac{1}{2}(x^2 + y^2) \leq xy \leq \frac{1}{2}(x^2 + y^2),$$
to obtain for any $M = (x, y)$
$$f(M) \geq 4x^2 + 6y^2 - \frac{5}{2}(x^2 + y^2) = \frac{3}{2}x^2 + \frac{7}{2}y^2 \geq \frac{3}{2}\|M\|_2^2.$$
Hence f is coercive

For Exercise 2.2, the coercivity is not trivial, you can use the following formula valid for each $x, y \in \mathbb{R}$ and each $\lambda > 0$
$$|xy| \leq \frac{1}{2}\left(\frac{x^2}{\lambda} + \lambda y^2\right)$$

(TD2 : 5 mins) Exercise 2.2

Let $f : (x, y) \mapsto \frac{1}{100}x^2 + 10^{10}y^2 - 100xy$, show that the problem $\inf_{M \in \mathbb{R}^2} f(M)$ admits minima.

Solution of Exercise 2.2

The formula
$$|xy| \leq \frac{x^2}{\lambda} + \lambda y^2$$
stems from $ab \leq \frac{a^2+b^2}{2}$ with $a = \frac{|x|}{\sqrt{\lambda}}$ and $b = \sqrt{\lambda}|y|$. Take $\lambda = 10^4$, we then have
$$-100xy \geq -\frac{100}{2}\left(\frac{1}{10^4}x^2 + 10^4y^2\right) = -\left(\frac{1}{200}x^2 + 5.10^5y^2\right)$$
and then
$$f(x, y) \geq \left(\frac{1}{100} - \frac{1}{200}\right)x^2 + \left(10^{10} - 5.10^5\right)y^2 \geq \frac{1}{200}(x^2 + y^2) = \frac{1}{200}\|(x, y)\|^2$$
and the function is coercive. It is a continuous function and $\mathbb{R}^2$ is obviously closed hence there exists a global minimum.

Correction of tutorials

In Exercise 2.3, the function is not coercive on $\mathbb{R}^2$ but is coercive on X. The notion of coercivity depends on the domain, even when the domain is unbounded.

(TD2 : 10 mins) Exercise 2.3

Let $X = \{M = (x, y) \text{ such that } x^2 \geq y^2\}$ and $f(x, y) = x^2 - y$

1. Show that f is not coercive on $\mathbb{R}^2$ and that $\inf_{\mathbb{R}^2} f = -\infty$.
2. Show that for every $(x, y) \in X$, then $f(x, y) \geq \frac{x^2+y^2}{2} - y$. Conclude that f is coercive on X and that there exists a global minimizer to $\inf_X f$.

Solution of Exercise 2.3

1. We look at $f(M)$ for $M = (0, y)$ and we let $y \to +\infty$, we obtain a sequence of point for which $\|M\|$ tends to $+\infty$ whereas $f(M)$ tends to $-\infty$.
2. From $x^2 \geq y^2$, we deduce $2x^2 \geq x^2 + y^2$ and then $f(x, y) \geq \frac{x^2+y^2}{2} - y$. We then have, for all $(x, y) \in X$

$$f(x, y) \geq \frac{1}{2}\|(x, y)\|^2 - \|(0, 1)\|\|(x, y)\| = r(\|(x, y)\|) \text{ with } r(t) = \frac{1}{2}t^2 - t.$$

It is clear that $r(t)$ goes to $+\infty$ as t goes to $+\infty$, and then f is coercive on X. For the existence of the minimizers, it is sufficient to note that f is obviously continuous (hence l.s.c) and that X is defined via an inequality constraint $g(x, y = y^2 - x^2$ which is also l.s.c.

Exercise 2.4 is a little bit on the theoretical side, but it shows that some standard problems of statistics have a solution.

(TD2 : 20 mins) Exercise 2.4

Let $A \in \mathcal{M}_{pn}(\mathbb{R})$ and $C \in \mathcal{M}_{qn}(\mathbb{R})$ and $b \in \mathbb{R}^n$. We consider the problem

$$\inf_X f \text{ if } f(x) = \frac{1}{2}\|Ax\|^2 - \langle b, x\rangle \text{ and } X = \{x \text{ such that } Cx = 0\}.$$

We suppose that both $\ker A$ and $\ker C$ are not trivial (not equal to $\{0\}$).

1. Suppose that there exists $u \in \ker A \cap \ker C$ such that $\langle u, b\rangle \neq 0$. Show that the problem does not admit minimizers.
2. Suppose that $\ker A \cap \ker C = \{0\}$
   (a) Show that the problem is equivalent to minimizing $g(x) = f(x) + \frac{1}{2}\|Cx\|^2$ on X.
   (b) Show that g can be written $g(x) = \frac{1}{2}\langle Bx, x\rangle - \langle b, x\rangle$ with $B = A^T A + C^T C$. Prove that B is a symetric definite positive matrix.
   (c) Show that there exists a minimizer to g over X, hence a minimizer of f over X.
3. If $I \in \mathbb{N}$ (number of classes) and for each $1 \leq i \leq I$, we are given $n_i \in \mathbb{N}$ (number of elements in each class). If we are given $Y_{ij} \in \mathbb{R}$ (random value of the (i, j) individual) for each $1 \leq i \leq I$ and $1 \leq j \leq n_i$. For any $\alpha \in \mathbb{R}^I$ and $\mu \in \mathbb{R}$, let

$$f(\alpha, \mu) = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \alpha_i - \mu)^2.$$

Apply the above result to show that the problem $\inf_X f$ admits a solution, where :

(a) $X = \left\{(\alpha, \mu) \in \mathbb{R}^I \times \mathbb{R} \text{ such that } \sum_{i=1}^I \alpha_i = 0\right\}$

(b) $X = \left\{(\alpha, \mu) \in \mathbb{R}^I \times \mathbb{R} \text{ such that } \sum_{i=1}^I n_i\alpha_i = 0\right\}$

Solution of Exercise 2.4

1. For any $t \in \mathbb{R}$, we have $tu \in X$ and we have $f(tu) = t\langle u, b\rangle$, depending on the sign of $\langle u, b\rangle$ we let t go to either $+\infty$ or $-\infty$, we obtain $\inf_X f = -\infty$.
2. We suppose that $\ker A \cap \ker C = \{0\}$
   (a) It is clear that $g = f$ on X so that minimizing g amounts to minimizing f.
   (b) We have

$$\begin{aligned} g(x) &= \frac{1}{2}\langle Ax, Ax\rangle + \frac{1}{2}\langle Cx, Cx\rangle - \langle b, x\rangle \\ &= \frac{1}{2}\langle A^T Ax, x\rangle + \frac{1}{2}\langle C^T Cx, x\rangle - \langle b, x\rangle \\ &= \frac{1}{2}\langle Bx, x\rangle - \langle b, x\rangle \end{aligned}$$

   The matrix B is clearly symmetric, it is positive because $\langle Bx, x\rangle = \|Ax\|^2 + \|Cx\|^2 \geq 0$. The matrix B is definite because

$$Bx = 0 \Rightarrow \langle Bx, x\rangle = 0 \Rightarrow (Ax = 0 \text{ and } Cx = 0) \Rightarrow x = 0$$

   (c) Because B is s.d.p, there exists $\lambda > 0$ such that $\langle Bx, x\rangle \geq \lambda\|x\|^2$ for all x and we have

$$g(x) \geq \frac{\lambda}{2}\|x\|^2 - \|b\|\|x\|.$$

   The function g is coercive, the set X is closed and g is continuous. Hence g admits a minimizer over X. So does f.
3. If $x = (\alpha, \mu)$, then f is of the form above with $\langle Ax, x\rangle = \sum_{i=1}^I \sum_{j=1}^{n_i} (\alpha_i + \mu)^2$. We just have to check that $\ker A \cap \ker C = \{0\}$. Set $\bar{\alpha} = (1, \dots, 1)$ and $\bar{\mu} = -1$ and $\bar{x} = (\bar{\alpha}, \bar{\mu})$ we see that $\ker A = Vect(\bar{x})$.
   (a) In this case $C\bar{x} = \sum_{i=1}^I 1 \neq 0$
   (b) In this case $C\bar{x} = \sum_{i=1}^I n_i \neq 0$

In Exercise 2.5, we showcase the problems that one may encounter in infinite dimension.

(TD3 : 10 mins) Exercise 2.5

For the following problems, show that the function f is coercive and continuous and that X is closed but that the problems do not admit a minimizer.

- $X = \ell^2(\mathbb{R})$ and $f(x) = (\|x\|^2 - 1)^2 + \displaystyle\sum_{i=0}^{+\infty} \frac{x_i^2}{i+1}$
- $X = \{x \in \ell^2(\mathbb{R}) \text{ such that } \|x\| = 1\}$ and $f(x) = \displaystyle\sum_{i=0}^{+\infty} \frac{x_i^2}{i+1}$

Solution of Exercise 2.5

In both cases, for each $x \in X$, we have $f(x) > 0$. It is sufficient to exhibit a sequence $(x^n)_n \in \ell^2(\mathbb{R})$ such that $f(x^n)$ converges to 0. In order to achieve this goal, use the following sequence defined for every n by

$$\begin{cases} x_i^n = 0 & \text{if } i \neq n \\ x_i^n = 1 \text{ if } i = n \end{cases}.$$

Note that in both cases, we have $f(x^n) = \frac{1}{n+1}$.

3 First order conditions

The following exercise is about minimizing a quadratic function with linear constraints, it is very similar to the exercise done in class.

(TD3 : 20 mins) Exercise 3.1

Show that the following problem admits a solution and find it using KKT theorem.

$$\min_{2x+y\geq 5} x^2 + y^2$$

Solution of Exercise 3.1

- **Step -1 : Standard form** Let $M = (x, y) \in \mathbb{R}^2$, define the functions $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ as

$$f(M) = x^2 + y^2 \quad g(M) = -2x - y + 5.$$

  If $X = \{M \in \mathbb{R}^2 \text{ s.t. } g(M) \leq 0\}$, then the problem reads $\min_X f$.
- **Step 0 : Explanatory Figure**
- **Step 1 : Existence of a minimum** The function f is continuous and trivially coercive (because $f(M) = \|M\|^2$) and the set X is closed (because g is continuous).
- **Step 2 : Qualification of constraints** The function g is affine hence the constraints are qualified everywhere by Slater.
- **Step 3 : Solving KKT** The Lagrangian reads

$$\mathcal{L}(M, \lambda) = x^2 + y^2 + \lambda(-2x - y + 5)$$

  KKT equations are given by

$$\begin{cases} (1) : & \begin{pmatrix} 2x \\ 2y \end{pmatrix} + \lambda \begin{pmatrix} -2 \\ -1 \end{pmatrix} = 0 \\ (2) : & \lambda \geq 0 \text{ and } g(M) \leq 0 \\ (3) : & \text{either } \lambda = 0 \text{ or } g(M) = 0 \end{cases}$$

  We discuss according to the cases in (3).
  - **If $g(M) \neq 0$:** Then we must have $\lambda = 0$ by (3). Equation (1) yields $x = y = 0$ which is a contradiction with $g(M) \leq 0$ because $g((0, 0)) = 5 > 0$.

- **If $g(M) = 0$:** Then (3) is verified and (2) boils down to $\lambda \geq 0$. From (1) we can see that $\lambda \neq 0$ and then $x = \lambda$ and $y = \frac{\lambda}{2}$. From $g(M) = 0$, we have

$$2\lambda + \frac{\lambda}{2} = 5$$

And we have $\lambda = 2$.We check that $\lambda \geq 0$ and we have a KKT point

We have only one KKT point.

- **Step 4 : Conclusion** Our global reasonning is as follows
  1. There exists a least one global minimizer, it is a local minimizer.
  2. The constraints are qualified everywhere, each global minimizer is a KKT point.
  3. There exists only one KKT point at point $\lambda = 2$ and $x = 2$ and $y = 1$.
  4. There exists only one local minimizer at point $\lambda = 2$ and $x = 2$ and $y = 1$. It is a global minimizer.

Exercise 3.2 is a very important exercise to bear in mind. It is an example of a convex problem (convex function and convex inequality constraints) for which the constraints are not qualified. In this example, there exists a global minimum but no KKT points. This exercise is a very good example of poor design in modelisation. Indeed, could we think of a worst way to describe the set X ? Do we need optimization theory to solve the problem ? Everything goes wrong because of poor design, and it shows how important it is to describe correctly the problem you want to solve. This exercise will be seen again in Exercise 5.2.

(TD3 : 15 mins) Exercise 3.2

Let $X = \{(x, y)$ such that $(x-1)^2 + y^2 \leq 1$ and $(x+1)^2 + y^2 \leq 1\}$.

Show that $X = \{0, 0\}$ the constraints are not qualified and show that the conclusions of KKT theorem do not apply in this case (check that $\inf_X y$ is attained at 0 and that this point fails to be a KKT point).

Solution of Exercise 3.2

We show easily that $X = \{(0, 0)\}$. Then $T_{(0,0)} = \{0\}$ but at $\bar{x} = (0, 0)$ both constraints are active and $\nabla g_1(\bar{x}) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$ and $\nabla g_2(\bar{x}) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$. The set of directions d that verifies

$$\langle d, \nabla g_i(\bar{x}) \rangle \leq 0 \forall i \in \mathcal{A}_{\bar{x}} \text{ and } \langle d, \nabla g_i(\bar{x}) \rangle = 0 \forall i \in \mathcal{E},$$

is exactly $Vect(\begin{pmatrix} 0 \\ 1 \end{pmatrix}) \neq T_{(0,0)}$. Hence the constraints are not qualified. For the function $f(x, y) = 0$, the KKT equation at 0 reads

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} -2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 0,$$

which is impossible.

In the following exercise, we do not apply KKT because the sets are open. It is a nice little reminder of minimization on open sets.

(20 mins) Exercise 3.3

Let: $f : (x, y) \in \mathbb{R}^2 \mapsto x^3 + 2xy + y^2$.

1. Is the function f convex ? concave ?
2. Give every local extrema of f and determine if they are local or global.
3. Define the following sets:

$$X_1 = \{(x, y) \in \mathbb{R}^2 \mid y - x < 0\} \quad \text{et} \quad \widetilde{X}_1 = \{(x, y) \in \mathbb{R}^2 \mid y - x > 0\}.$$

Give every local extrema of f on X_1, and then on $\widetilde{X}_1$.

Solution of Exercise 3.3

1. The function f is C^2 (even C^∞) on $\mathbb{R}^2$ since it is a polynom and $\forall (x, y) \in \mathbb{R}^2$, we have

$$\nabla f(x, y) = \begin{pmatrix} 3x^2 + 2y \\ 2x + 2y \end{pmatrix}, \quad H[f](x, y) = \begin{pmatrix} 6x & 2 \\ 2 & 2 \end{pmatrix}$$

Take for instance the point $(x, y) = (0, 0)$, then:

$$det\ H_f(0, 0) = -4 < 0.$$

The matrix $H[f](0, 0)$ admits a positive and a negative eigeinvalue. The function f is neither convex nor concave.

2.
   - We look for critical points, i.e. $(x, y) \in \mathbb{R}^2$ such that $\nabla f(x, y) = 0$.

$$\nabla f(x, y) = 0 \quad \Leftrightarrow \quad \begin{cases} 3x^2 + 2y & = 0 \\ 2x + 2y & = 0 \end{cases} \Leftrightarrow \begin{cases} 3x^2 - 2x = 0 \\ y = -x \end{cases}$$
$$\Leftrightarrow \quad \begin{cases} x = 0 \text{ or } x = \frac{2}{3} \\ y = -x \end{cases}$$

   The critical points of f are $(0, 0)$ and $(\frac{2}{3}, -\frac{2}{3})$.
   - Since f is neither convex, nor concave, we study second order optimality condition
     - $H[f](0, 0) = \begin{pmatrix} 0 & 2 \\ 2 & 2. \end{pmatrix}$. Or $det\ H[f](0, 0) = -4 < 0$. Hence the matrix $H[f](0, 0)$ admits a positive and a negative eigeinvalue. The point $(0, 0)$ is neither a local minimum or a local maximum, it is a saddle point of f.
     - $H[f](\frac{2}{3}, -\frac{2}{3}) = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$. We have $det\ H[f](\frac{2}{3}, -\frac{2}{3}) = 4 > 0$ and $tr(H[f](\frac{2}{3}, -\frac{2}{3}) = 6 > 0$. The matrix $H[f](\frac{2}{3}, -\frac{2}{3})$ has two positives eigenvalues. The point $(\frac{2}{3}, -\frac{2}{3})$ is therefore a local minimum of f on $\mathbb{R}^2$. It is not a globla minimum f on $\mathbb{R}^2$ because

$$\lim_{x \to -\infty} f(x, 0) = -\infty.$$

3. The sets X_1 and $\widetilde{X}_1$ are open in $\mathbb{R}^2$, hence, looking for extrema of f on X_1 (resp. $\widetilde{X}_1$) amounts to look for extrema of f on $\mathbb{R}^2$, and to verify *a posteriori* that the optimality condition is verified. The function f admits a unique local minimum in $(x, y) = (\frac{2}{3}, -\frac{2}{3})$ et:

$$y - x = -\frac{2}{3} - \frac{2}{3} = -\frac{4}{3} < 0,$$

hence $(\frac{2}{3}, -\frac{2}{3}) \in X_1$ and $(\frac{2}{3}, -\frac{2}{3}) \notin \widetilde{X}_1$. Therefore the function f admits a unique local minimum point in $(\frac{2}{3}, -\frac{2}{3})$ on X_1 and no extremum on $\widetilde{X}_1$.

Exercise 3.4 is another application of KKT, it is in $3d$, so it is a little bit more involved on a computational point of view. Moreover, we study minimizers and maximizers, hence this exercise allows recapping the KKT formulas for maximisation. It also emphasizes the fact that, when looking for solution to minimisation problems by the KKT method, the solution to maximisation problems is almost given for free.

(TD4 : 25 mins) Exercise 3.4

$$\text{Let } X = \{x \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 1,\ x_1^2 + x_2^2 + x_3^2 \leq 1\}.$$

Give the extrema of the function $f : x \in \mathbb{R}^3 \mapsto x_1$ on X.

Solution of Exercise 3.4

First f is continuous and X is closed and bounded. There exists global minimizers and maximizers.

Start by noticing that f is linear, hence convex and concave $\mathbb{R}^3$. Moreover, setting:

$$g_1(x) = x_1^2 + x_2^2 + x_3^2 - 1, \qquad g_2(x) = x_1 + x_2 + x_3 - 1,$$

we verify (easily) that g_1 is convex whereas g_2 is affine, hence X, the set of constraints is convex.

The constraint are qualified everywhere by Slater conditions, indeed the point $\bar{x} = (\frac{1}{2}, \frac{1}{2}, 0)$ verifies $g_2(\bar{x}) = 0$ and $g_1(\bar{x}) < 0$. The Lagrangian is defined for $\lambda \in \Lambda = \{\lambda \in \mathbb{R}^2, \lambda_1 \geq 0\}$ by:

$$\mathcal{L}(x, \lambda) = x_1 + \lambda_1(x_1^2 + x_2^2 + x_3^2 - 1) + \lambda_2(x_1 + x_2 + x_3 - 1), \lambda \in \Lambda.$$

KKT equations are

$$\begin{cases} \nabla_x \mathcal{L}(x, \lambda) &= 0 \\ x_1 + x_2 + x_3 &= 1 \\ \lambda_1(x_1^2 + x_2^2 + x_3^2 - 1) &= 0 \\ x_1^2 + x_2^2 + x_3^2 - 1 &\leq 1 \end{cases}$$

that is :

$$\begin{cases} 1 + 2\lambda_1 x_1 + \lambda_2 &= 0 \\ 2\lambda_1 x_2 + \lambda_2 &= 0 \\ 2\lambda_1 x_3 + \lambda_2 &= 0 \\ x_1 + x_2 + x_3 &= 1 \\ \lambda_1(x_1^2 + x_2^2 + x_3^2 - 1) &= 0 \\ x_1^2 + x_2^2 + x_3^2 - 1 &\leq 1 \end{cases}$$

The case $\lambda_1 \geq 0$ correspond to minimisers, whereas $\lambda_1 \leq 0$ correspond to maximisers.
We discuss according to the case $\lambda_1 = 0$ or $x_1^2 + x_2^2 + x_3^2 - 1 = 0$

- First case: The constraint g_1 is not active, $x_1^2 + x_2^2 + x_3^2 < 1$ and $\lambda_1 = 0$. In this case, KKT rewrites into

$$\begin{cases} 1 + \lambda_2 &= 0 \\ \lambda_2 &= 0 \\ x_1 + x_2 + x_3 &= 1 \end{cases}$$

Which is impossible, there is no solution in this case.

- 2nd case: the constraint g_1 is active, hence $x^2+y^2+z^2=1$.The optimality conditions are

$$\left\{\begin{array}{rcl} 1+2\lambda_1 x_1+\lambda_2 &=& 0\\ 2\lambda_1 x_2+\lambda_2 &=& 0\\ 2\lambda_1 x_3+\lambda_2 &=& 0\\ x_1+x_2+x_3 &=& 1\\ x_1^2+x_2^2+x_3^2 &=& 1 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} 1+2\lambda_1 x_1+\lambda_2 &=& 0\\ 2\lambda_1 x_2+\lambda_2 &=& 0\\ x_3 &=& x_2\\ x_1 &=& 1-2x_2\\ (1-2x_2)^2+2x_2^2 &=& 1 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{rcl} 1+2\lambda_1 x_1+\lambda_2 &=& 0\\ 2\lambda_1 x_2+\lambda_2 &=& 0\\ x_3 &=& x_2\\ x_1 &=& 1-2x_2\\ 3x_2^2-2x_2 &=& 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} x_2 &=& 0 \text{ ou } \frac{2}{3}\\ x_1 &=& 1-2x_2\\ x_3 &=& x_2\\ 2\lambda_1(x_1-x_2) &=& -1\\ \lambda_2 &=& -2\lambda_1 x_2 \end{array}\right.$$

The KKT points of f on X are given by:

$$M=(x,\lambda)=((1,0,0),(-\frac{1}{2},0)) \quad \text{and} \quad \tilde{M}=(\tilde{x},\tilde{\lambda})=((-\frac{1}{3},\frac{2}{3},\frac{2}{3}),(\frac{1}{2},-\frac{2}{3})).$$

Because $\lambda_1=-\frac{1}{2}<0$, the KKT point M is a KKT point for maximisation of f over X. Similarly $\tilde{M}$ is a KKT point for minimization. There exists a global minimizer and a global maximizer, they are KKT points, there exists only one KKT point of each kind, hence $\tilde{x}$ is a global minimizer of f over X whereas x is a global maximizer of f over X.

In the following exercise, the computations are a little bit on the abstract side, the final formula looks awful but is not *that* difficult to obtain.

(TD4 : 25 mins) Exercise 3.5

Let A be a real $m\times n$ matrix and $b\in\mathbb{R}^n$. Let:

$$\inf_{x\in\mathbb{R}^n}\frac{1}{2}\|x\|^2-\langle b,x\rangle \qquad \text{such that: } Ax=0.$$

1. Show that this problem is equivalent to a projection problem on a certain convex set. Make explicit this convex set. Show existence and uniqueness of a global minimizer.
2. Show that the constraints are qualified everywhere and show that if A is of full rank (i.e., $A^\star$ is injective), then the optimal solution is given by:

$$x^\star=(I_n-A^\star(AA^\star)^{-1}A)b.$$

Solution of Exercise 3.5

1. Add the constant $1/2\|b\|^2$ to the objective function, this doesn't change the minimizers (if they exist) although it changes the value of the infimum. The problem can be rewritten :

$$\inf \tfrac{1}{2}||x-b||_2^2 \qquad \text{with} \qquad Ax=0$$

The set $\mathcal{C} = \{x \in \mathbb{R}^n \ : \ Ax = 0\}$ is convex because the mapping $x \mapsto Ax$ is linear. The problem is then to project b on the set $\mathcal{C}$. The set $\mathcal{C}$ is closed, hence existence and uniqueness by the projection theorem. If you do not want to use the projection theorem, it is easy to show that the function $x \mapsto \frac{1}{2}||x-b||_2^2$ is continuous, coercive (this gives existence) and strongly convex on $\mathcal{C}$ which is also convex (hence uniqueness).

2. For each $1 \leq i \leq m$, denote $g_i(x) = (Ax)_i$ (the i^{th} coordinate of the vector Ax). Then each the set of constraints is

$$X = \{x \text{ s.t. } g_i(x) = 0 \quad \forall 1 \leq i \leq m\}$$

The constraints are affine, hence by Slater, they are qualified. Let $\mathcal{L}$ be the Lagrangian:

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\|x\|^2 - \langle b, x\rangle + \sum_{i=1}^m \lambda_i g_i(x) = \frac{1}{2}\|x\|^2 - \langle b, x\rangle + \langle Ax, \lambda\rangle, \ \lambda \in \mathbb{R}^m.$$

The KKT equations read

$$\begin{cases} \nabla_x \mathcal{L}(x,\lambda) = 0 \\ Ax = 0 \end{cases},$$

Rewriting $\mathcal{L}(x,\lambda) = \frac{1}{2}\|x\|^2 - \langle b, x\rangle + \langle A^\star \lambda, x\rangle$, we obtain :

$$\begin{cases} x - b + A^\star\lambda = 0 \\ Ax = 0 \end{cases}.$$

Multiply the first equation by A (on the left): we obtain:

$$-Ab + AA^\star\lambda = 0.$$

If $A^\star$ is injective, the matrix $AA^\star$ is invertible. Indeed take any x such that $AA^\star x = 0$ then $\langle AA^\star x, x\rangle = 0$. But

$$0 = \langle AA^\star x, x\rangle = \langle A^\star x, A^\star x\rangle = \|A^\star x\|^2,$$

and then $A^\star x = 0$ and then $x = 0$ by injectiviy of $A^\star$.
Because $AA^\star$ is invertible, we then have

$$\lambda = (AA^\star)^{-1}Ab.$$

Reinjecting the equation of λ, we have:

$$x^\star = (I - A^\star(AA^\star)^{-1}A)b,$$

This point indeed belongs to $\mathcal{C}$ because

$$Ax^\star = A(I - A^\star(AA^\star)^{-1}A)b = (A - AA^\star(AA^\star)^{-1}A)b = 0$$

This exercise solves the problem of finding the symmetric matrix closest to a given matrix A. It is an orthogonal decomposition, when one proves that symmetric matrices are orthogonal to anti-symmetric matrices, one can see that this exercise fits nicely with the decomposition formula

$$X = \underbrace{\frac{X + X^\top}{2}}_{symmetric} + \underbrace{\frac{X - X^\top}{2}}_{anti-symmetric} .$$

(20 mins) Exercise 3.6

Let A be a square matrix of size n. Solve the following problem:

$$\inf_{X \in \mathcal{M}_{nn}(\mathbb{R})} f(X) = \frac{1}{2}\|X - A\|_F^2 \qquad \text{with: } X^\top = X,$$

in the case where the matrix A is (a) symmetric and then (b) non-symmetric.

Solution of Exercise 3.6

This is a projection problem, we are looking for the symmetric matrix X which is the closest to A. The space of symmetric matrices is a vector space, hence it is closed and we have existence and uniqueness of the solution.

(a) **If A is symmetric.** The problem as a unique solution $X^\star = A$, indeed $f(A) = 0$, $A^\top = A$ and for any $X \neq A$, then $f(X) > 0$.

(b) **If A is not symmetric.** The constraints are qualified at any point because they are affine. The Lagrangian is defined for any X, Λ, square matrices of size n by

$$\mathcal{L}(X, \Lambda) = \frac{1}{2}\|X - A\|_F^2 + \langle \Lambda, X^\top - X\rangle_F,$$

And the KKT equations are

$$\nabla_X \mathcal{L}(X, \Lambda) = 0 \quad \text{and} \quad X^\top = X.$$

We rewrite

$$\mathcal{L}(X, \Lambda) = \frac{1}{2}\|X - A\|_F^2 + \langle \Lambda^\top - \Lambda, X\rangle_F,$$

Hence the optimality conditions can be written as

$$\begin{cases} X - A + \Lambda^\top - \Lambda = 0 \\ X^\top = X \end{cases} \quad \text{hence:} \quad \begin{cases} X = A + \Lambda - \Lambda^\top \\ X^\top = X. \end{cases}$$

Hence injecting the first equation into the second, we have

$$(A + \Lambda - \Lambda^\top)^\top = A + \Lambda - \Lambda^\top,$$

hence:

$$\Lambda - \Lambda^\top = \frac{A^\top - A}{2}.$$

And then

$$X = A + \frac{A^\top - A}{2} = \frac{A + A^\top}{2}.$$

Note here that we have an infinite number of Λ, indeed any $\Lambda = \frac{A^\top - A}{2} + B$ with $B^\top = -B$ will verify equation :

$$\Lambda - \Lambda^\top = \frac{A^\top - A}{2}.$$

We note that for any matrix A, we can decompose

$$A = \left(\frac{A+A^T}{2}\right) + \left(\frac{A-A^T}{2}\right)$$

The matrix $\frac{A+A^T}{2}$ is symmetric and the matrix $\frac{A-A^T}{2}$ is . Moreover the space of symmetric and antisymmetric matrices are orthogonal. Indeed, if B is symmetric and C is antisymmetric, then

$$\langle B, C\rangle = \langle B^\top, C^\top\rangle = \langle B, -C\rangle = -\langle B, C\rangle.$$

Hence $\langle B, C\rangle = 0$

This problem is all about solving KKT equations, a striking feature of to show that the computations for a certain problem can be helpful with another. If the student masters the following exercise, he can be assured he understands first order conditions.

(♠ 20 mins) Exercise 3.7

Let $f(x, y) = y - 2x$ and:

$$X = \{(x, y) \in \mathbb{R}^2 \mid y \geq x^2 \text{ et } x^2 + y^2 \leq 1\}.$$

1. Give the extremal points of f on X. Determine if they are local or global.
2. What happens if we replace $y \geq x^2$ by $y \leq x^2$?

Solution of Exercise 3.7

1. The set X is bounded and closed and the function f is continuous, there exists global minimum and maximum points if $g(x, y) = \begin{pmatrix} x^2 - y \\ x^2 + y^2 - 1 \end{pmatrix}$, then both g_1 and g_2 are convex. Moreover $M = (0, \frac{1}{2})$ is a point such that $g_i(M) < 0$ for every i so that the constraints are qualified by Slater's condition. We study KKT points, the Lagrangien is given by

$$\mathcal{L}(M, \lambda) = y - 2x + \lambda_1(x^2 - y) + \lambda_2(x^2 + y^2 - 1).$$

The KKT equations are

$$\begin{cases} -2 + 2(\lambda_1 + \lambda_2)x & = 0 \\ 1 - \lambda_1 + 2\lambda_2 y & = 0 \\ \lambda_1 = 0 \text{ or } y = x^2 & \\ \lambda_2 = 0 \text{ or } x^2 + y^2 = 1 & \end{cases}$$

We denote $M = (x, y)$

   (a) The case $\lambda_1 = \lambda_2 = 0$ is impossible.
   (b) The case $y = x^2$ and $\lambda_2 = 0$ gives $\lambda_1 = 1, x = 1, y = 1$ which yields $x^2 + y^2 = 2$ which is impossible.
   (c) The case $\lambda_1 = 0$ and $x^2 + y^2 = 1$ gives $x = \frac{1}{\lambda_2}$ and $y = \frac{-1}{2\lambda_2}$. The condition $x^2 + y^2 = 1$ turns into $\lambda_2^2 = \frac{5}{4}$ and the condition $y \geq x^2$ gives $\lambda_2 \leq 0$. Hence the only candidate is $M = (-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$ with $\lambda = (0, -\frac{\sqrt{5}}{2})$. Sadly we do not have $y \geq x^2$ in this case.

(d) The case $y = x^2$ and $x^2 + y^2 = 1$ yields $y^2 + y = 1$ and $y \geq 0$, this gives $y = \frac{-1+\sqrt{5}}{2}$ and we have $x = \pm\sqrt{y}$. We solve the linear system in λ given by

$$\begin{cases} \lambda_1 x + \lambda_2 x = 2 \\ -\lambda_1 + 2\lambda_2 y = -1 \end{cases} \Rightarrow \begin{cases} \lambda_2(x + 2xy) = 2 - x & (L_1 + xL_2) \\ \lambda_1(2yx + x) = 4y + x & (2yL_1 - xL_2) \end{cases}$$

Set $\phi = \frac{-1+\sqrt{5}}{2}$, the KKT points are $M = (\sqrt{\phi}, \phi)$ with $\lambda = (\frac{4\sqrt{\phi}+1}{2\phi+1}, \frac{2-\sqrt{\phi}}{\sqrt{\phi}(1+2\phi)})$ or $\tilde{M} = (-\sqrt{\phi}, \phi)$ with $\tilde{\lambda} = (\frac{-4\sqrt{\phi}+1}{2\phi+1}, \frac{-2-\sqrt{\phi}}{\sqrt{\phi}(1+2\phi)})$. From $\sqrt{\phi} \simeq 0.78$, we deduce that $\lambda \geq 0$ and $\tilde{\lambda} \leq 0$.

We only have two KKT points, one for minimization : (M, λ) and one for maximization : $(\tilde{M}, \tilde{\lambda})$. Hence M is the global minimum and $\tilde{M}$ is the global maximum.

2. If $y \geq x^2$ is replaced by $y \leq x^2$, then essentially the sign of λ_1 is changed. We then have
   (a) This does not change, the case is impossible
   (b) This does not change, the case is impossible
   (c) We still have $x = \frac{1}{\lambda_2}$ and $y = \frac{-1}{2\lambda_2}$ and $\lambda_2 = \pm\frac{\sqrt{5}}{2}$. Note that in both cases we have $y \leq x^2$, we then have two KKT points $\tilde{M} = (-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$ with $\tilde{\lambda} = (0, -\frac{\sqrt{5}}{2})$ and $M = (\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$ with $\lambda = (0, \frac{\sqrt{5}}{2})$.
   (d) The computation are the same except that λ_1 is replaced by $-\lambda_1$. The candidate KKT points obtained at the end of the computation do not obey $\lambda \geq 0$ or $\lambda \leq 0$, hence are not valid KKT points.

   We only have two KKT points, one for minimization : (M, λ) and one for maximization : $(\tilde{M}, \tilde{\lambda})$. Hence M is the global minimum and $\tilde{M}$ is the global maximum.

4 Second order conditions

The exercise below is tedious but rather easy. It is a non-convex problem and the idea is to study thoroughly second order necessary and sufficient conditions. Note that in the following problem, we could avoid second order analysis by comparing the value of the function at the different KKT points, but it wouldn't be fun.

(TD5 : 60 mins) Exercise 4.1

Denote $X = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 2\}$ and consider the following problems

$$\begin{array}{lll} (P_1) & \min_{x \in X} f_1(x) & \text{with } f_1(x) = x_1 + x_2. \\ (P_2) & \min_{x \in X} f_2(x) & \text{with } f_2(x) = x_1 x_2. \end{array}$$

1. Show that the problems (P_1) et (P_2) admit solutions.
2. Show that the constraint $x_1^2 + x_2^2 = 2$ is qualified in any point of X.
3. Determine every KKT point of f_1, by second order conditions determine if they are local minima or maxima of f_1 sur X. Are they global ?
4. Same question for f_2.

Solution of Exercise 4.1

1. The functions f_1 and f_2 are continuous on X which is a bounded closed set (it is the circle of center $(0,0)$ and radius $\sqrt{2}$). Hence (P_1) and (P_2) have solutions. We remark that X is not convex, hence the problems are not convex.
2. Let $g(x) = x_1^2 + x_2^2 - 2$. The polynom g is C^∞ everywhere, hence differentiable on any point of X and
$$\forall x \in \mathbb{R}^2,\ \nabla g(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}.$$
Hence the only critical point of g is $(0,0) \notin X$ and therefore $\forall x \in X,\ \nabla g(x) \neq 0$. The above property implies LICQ and hence constraint qualification.
3. We construct the Lagrangian associated to the problem (P_1):
$$\mathcal{L}(x,\lambda) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2),\ \lambda \in \mathbb{R}$$
The first order necessary conditions are
$$\left\{\begin{array}{rcl} \nabla_x \mathcal{L}(x,\lambda) & = & 0 \\ x_1^2 + x_2^2 - 2 & = & 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} 1 + 2\lambda x_1 & = & 0 \\ 1 + 2\lambda x_2 & = & 0 \\ x_1^2 + x_2^2 & = & 2 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} x_1 & = & x_2 \\ 1 + 2\lambda x_1 & = & 0 \\ x_1^2 & = & 1 \end{array}\right.$$
$$\Leftrightarrow \left\{\begin{array}{rcl} x_1 & = & 1 \text{ or } x_1 = -1 \\ x_2 & = & x_1 \\ \lambda & = & -\frac{1}{2x_1} \end{array}\right.$$
The stationnary points of problem (P_1) are $(x = (1,1), \lambda = -\frac{1}{2})$ and $(x = (-1,-1), \lambda = \frac{1}{2}$. Since problem (P_1) is not convex, we have to look at second order optimality conditions. We have
$$H_x[\mathcal{L}](x,\lambda) = \begin{pmatrix} 2\lambda & 0 \\ 0 & 2\lambda \end{pmatrix}.$$
For the stationnary points found earlier, we have
   - $H_x[\mathcal{L}]((1,1), -\frac{1}{2}) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. The matrix $H_x[\mathcal{L}]((1,1), -\frac{1}{2})$ is definite negative. This implies that $(1,1)$ is a local maximizer of f_1 on X. It is even a global maximizer (because there exists a global maximum).
   - $H_x[\mathcal{L}]((-1,-1), \frac{1}{2}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The matrix $H_x[\mathcal{L}]((-1,-1), \frac{1}{2})$ is definite positive. This implies that $(1,1)$ is a local minimizer of f_1 on X. It is even the global minimizer (because there exists a global minimum).
4. We construct the Lagrangian associated to the problem (P_2):
$$\mathcal{L}(x,\lambda) = x_1 x_2 + \lambda(x_1^2 + x_2^2 - 2),\ \lambda \in \mathbb{R}$$
The first order necessary conditions are
$$\left\{\begin{array}{rcl} \nabla_x \mathcal{L}(x,\lambda) & = & 0 \\ x_1^2 + x_2^2 & = & 2 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} x_2 + 2\lambda x_1 & = & 0 \\ x_1 + 2\lambda x_2 & = & 0 \\ x_1^2 + x_2^2 & = & 2 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} x_2 & = & -2\lambda x_1 \\ (1 - 4\lambda^2) x_1 & = & 0 \\ (1 + 4\lambda^2) x_1^2 & = & 2 \end{array}\right.$$
$$\Leftrightarrow \left\{\begin{array}{rcl} \lambda & = & \pm\frac{1}{2} \\ x_1^2 & = & 1 \\ x_2 & = & -2\lambda x_1 \end{array}\right.$$

The stationnary points of the problem (P_2) are the four points (x,λ) given by

$$((1,-1),\frac{1}{2}),\quad ((-1,1),\frac{1}{2}),\quad ((1,1),-\frac{1}{2}),\quad ((-1,-1),-\frac{1}{2}).$$

The problem (P_2) is not convex, we study second order optimality conditions, they are given by :

$$H_{xx}[\mathcal{L}](x,\lambda)=\begin{pmatrix}2\lambda & 1\\ 1 & 2\lambda\end{pmatrix}.$$

For the four KKT points, we have

- $H_x[\mathcal{L}]((1,1),-\frac{1}{2}) = H_x[\mathcal{L}]((-1,-1),-\frac{1}{2}) = \begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}$, which is a negative matrix which is not definite (its eigenvalues are 0 et -2).We study the linearized cone of constraints and the sign of $\langle H_x[\mathcal{L}](\pm 1,\pm 1)v,v\rangle$. Let us begin with the point $x=(1,1)$, we have

  $$\begin{aligned} V^*_{(1,1)}(X) &= \{v\in\mathbb{R}^2 \mid \langle \nabla g(1,1),v\rangle = 0\} = \{v\in\mathbb{R}^2 \mid \langle \begin{pmatrix}2\\2\end{pmatrix},v\rangle = 0\}.\\ &= \{v\in\mathbb{R}^2 \mid v_1+v_2=0\} = Vect(\begin{pmatrix}1\\-1\end{pmatrix})\end{aligned}$$

  Hence, for every $v\in V^*_{(1,1)}(X)$, there exists $\alpha\in\mathbb{R}$ such that $v=\alpha\begin{pmatrix}1\\-1\end{pmatrix}$, it is sufficient to study

  $$\langle H_x[\mathcal{L}]((1,1),-\frac{1}{2})\begin{pmatrix}1\\-1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\rangle = \langle\begin{pmatrix}-2\\2\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\rangle = -4<0.$$

  Hence $(1,1)$ is a local maximizer of f_2 on X. Similarly, we show that $(-1,-1)$ is also a local maximizer of f_2 on X.
- $H_x[\mathcal{L}]((1,-1),\frac{1}{2}) = H_x[\mathcal{L}]((-1,1),\frac{1}{2}) = \begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$, which is a positive matrix which is not definite (its eigenvalues are 0 et 2). We have to study the linearized cone of constraints and the sign of $\langle H_x[\mathcal{L}](\pm 1,\mp 1)v,v\rangle$. Let us begin with the point $x=(1,-1)$, we have

  $$\begin{aligned} V^*_{(1,-1)}(X) &= \{v\in\mathbb{R}^2 \mid \langle \nabla g(1,-1),v\rangle = 0\} = \{v\in\mathbb{R}^2 \mid \langle \begin{pmatrix}2\\-2\end{pmatrix},v\rangle = 0\}.\\ &= \{v\in\mathbb{R}^2 \mid v_1-v_2=0\} = Vect(\begin{pmatrix}1\\1\end{pmatrix})\end{aligned}$$

  Hence, for every $v\in V^*_{(1,1)}(X)$, there exists $\alpha\in\mathbb{R}$ such that $v=\alpha\begin{pmatrix}1\\1\end{pmatrix}$, it is sufficient to study

  $$\langle H_x[\mathcal{L}]((1,-1),\frac{1}{2})\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\rangle = \langle\begin{pmatrix}2\\2\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\rangle = 4>0.$$

  Hence $(1,-1)$ is a local minimizer of f_2 on X. Similarly, we show that $(-1,1)$ is also a local minimizer of f_2 on X.

To determine whether the local minimizer/maximizers are global, we compare the value of f and from $f(1,1)=f(-1,-1)$ and $f(1,-1)=f(-1,1)$ we conclude that the 4 KKT points are global extrema.

Correction of tutorials

Exercise 4.2 is kind of similar to Exercise 4.1, with a little twist here, (spoiler: the minimizers are not global).

(TD7 : 45 mins) Exercise 4.2

Compute the KKT points of

$$(P) \qquad \inf_{(x,y)\in\mathbb{R}^2} f(x,y) = -\frac{1}{2}(x-4)^2 + 2y^2 \ \text{ with } \ 1 - x^2 - y^2 \leq 0.$$

Compute the minimizers of (P) and determine if they are local or global.

Solution of Exercise 4.2

We will denote $M = (x, y)$. First remark that the problem is non-convex, indeed neither the objective function nor the domain are convex. The constraint "$1-x^2-y^2 \leq 0$" is qualified everywhere, indeed its gradient is given by

$$\nabla g(M) = \begin{pmatrix} -2x \\ -2y \end{pmatrix}$$

and the gradient of g never vanishes on X (because $(0,0)$ is the only point such that $\nabla g(M) = 0$ and this point does not belong to X since $g(0,0) = 1 > 0$. The Lagrangian associated to (P) is given by:

$$\mathcal{L}(M,\lambda) = -\frac{1}{2}(x-4)^2 + 2y^2 + \lambda(1-x^2-y^2), \quad \lambda \geq 0.$$

The Karush-Kuhn-Tucker (KKT) conditions (for minimization) are given by :

$$\left\{\begin{array}{rcl} -(x-4) - 2\lambda x & = & 0 \\ 4y - 2\lambda y & = & 0 \\ \lambda(1-x^2-y^2) & = & 0 \\ 1-x^2-y^2 & \leq & 0 \\ \lambda & \geq & 0. \end{array}\right.$$

- **First case : no active constraint** i.e.: $1 - x^2 - y^2 < 0$ and $\lambda = 0$. In this case, KKT conditions translate into

  $$x = 4, \quad y = 0,$$

  and $g(4,0) = -15 < 0$.
- **Second case: the constraint is active** i.e.: $1 - x^2 - y^2 = 0$. In this case, KKT conditions give:

  $$\left\{\begin{array}{rcl} x - 4 + 2\lambda x & = & 0 \\ y & = & 0 \text{ or } \lambda = 2 \\ 1-x^2-y^2 & = & 0 \\ \lambda & \geq & 0. \end{array}\right.$$

  - If $y = 0$, then $x^2 = 1$ and either:

    $$x = 1 \qquad \text{et} \qquad \lambda = \frac{3}{2} > 0$$

or

$$x = -1 \qquad \text{et} \qquad \lambda = -\frac{5}{2} < 0.$$

We are only interested in minimization problems, that is points such that $\lambda \geq 0$, we only keep $M = (1, 0)$ associated to $\lambda = \frac{3}{2}$.

- If $\lambda = 2$, then:

$$x = \frac{4}{5}, \quad y = \pm\frac{3}{5}, \quad \lambda = 2.$$

The KKT points for minimization associated to the problem (P) are given by:

$$(M, \lambda) = ((4, 0), 0), \quad (M, \lambda) = ((1, 0), \tfrac{3}{2}),$$

$$(M, \lambda) = ((\tfrac{4}{5}, \tfrac{3}{5}), 2), \quad (M, \lambda) = ((\tfrac{4}{5}, -\tfrac{3}{5}), 2).$$

The Hessian of the Lagrangian is given by:

$$H_{MM}[\mathcal{L}](M, \lambda) = \begin{pmatrix} -1 - 2\lambda & 0 \\ 0 & 4 - 2\lambda \end{pmatrix}$$

and the critical cone in any point $M = (x, y)$ such that $x^2 + y^2 = 1$ (i.e. on the boundary of the admissible domain) is given by:

$$\begin{aligned} V(M, \lambda) &= \left\{ v \in \mathbb{R}^2 \mid \langle \nabla g(M), v \rangle = 0 \right\}. \\ &= \left\{ v \in \mathbb{R}^2 \mid x v_1 + y v_2 = 0 \right\} = Vect \begin{pmatrix} -y \\ x \end{pmatrix}. \end{aligned}$$

For each KKT point, we discuss the second order conditions

- **Point $(4, 0)$, $\lambda = 0$:** In this case, the constraint is not active and $V(M, \lambda) = \mathbb{R}^2$.Moreover, the Hessian of f is given by:

$$H_{MM}[\mathcal{L}](M, \lambda) = H[f](M) = \begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix}.$$

  This Hessian is not positive, hence $(4, 0)$ is not a local minimizer of (P).

- **Point $(1, 0)$, $\lambda = \frac{3}{2}$:** In this case, the constraint is active. The Hessian is given by

$$H_{MM}[\mathcal{L}]((1, 0), \frac{3}{2}) = \begin{pmatrix} -4 & 0 \\ 0 & 1 \end{pmatrix}.$$

  This Hessian is not > 0, we cannot conlude. The critical cone is given by :

$$V(M, \lambda) = Vect \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

  We test the sufficient condition of optimality

$$\left\langle H_{MM}[\mathcal{L}]((1, 0), \frac{3}{2}) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} -4 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 1 > 0$$

  hence the point $(1, 0)$ is a local minimizer of problem (P).

- **Point** $(\frac{4}{5}, \pm\frac{3}{5})$**,** $\lambda = 2$**:** In this case, the constraint is active. The Hessian is given by

$$H_{MM}[\mathcal{L}]((\frac{4}{5}, \pm\frac{3}{5}), 2) = \begin{pmatrix} -5 & 0 \\ 0 & 0 \end{pmatrix}.$$

We have to compute the critical cone in order to conclude, it is given by

$$V((\frac{4}{5}, \pm\frac{3}{5}), 2) = Vect \begin{pmatrix} \mp\frac{3}{5} \\ \frac{4}{5} \end{pmatrix}$$

$$\begin{aligned} &\left\langle H_{MM}[\mathcal{L}](\frac{4}{5}, \pm\frac{3}{5}; 2) \begin{pmatrix} \mp\frac{3}{5} \\ \frac{4}{5} \end{pmatrix}, \begin{pmatrix} \mp\frac{3}{5} \\ \frac{4}{5} \end{pmatrix} \right\rangle \\ = &\left\langle \begin{pmatrix} -5 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mp\frac{3}{5} \\ \frac{4}{5} \end{pmatrix}, \begin{pmatrix} \mp\frac{3}{5} \\ \frac{4}{5} \end{pmatrix} \right\rangle \\ = & -\frac{9}{5} < 0. \end{aligned}$$

Hence the points $(\frac{4}{5}, \pm\frac{3}{5})$ are not solutions of (P).

As a conclusion f admits a unique local minimizer on $\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1\}$. It is not a global minimizer because if x is taken large enough, then $(x, 0)$ is an admissible point and $f(x, 0) = -\frac{1}{2}(x-4)^2 \to -\infty$, $x \to +\infty$.

The following exercise is a fun one, it is easy to be lost in one's own calculations... good luck !

(60 mins) Exercise 4.3

For any $k \in \mathbb{R}^\star$, we consider the following problem:

$$(P_k) \qquad \min_{(x,y)\in\mathbb{R}^2} f(x,y) = (x-1)^2 + y^2 \qquad \text{with} \qquad 2kx - y^2 \leq 0.$$

We denote $X_k = \{(x,y) \in \mathbb{R}^2 \mid 2kx - y^2 \leq 0\}$.

1. Show that the problem is not convex.
2. Show that the problem admits at least one global minimizer.
3. Solve the unconstrained problem. Show that the unique global minimizer of P_k is $(1, 0)$ for $k \leq 0$
4. Show that the constraints are qualified at each point.
5. Find the KKT points of (P_k). You must obtain

	(x,y)	λ
$k<0$	$(1,0)$	0
$0<k<1$	$(0,0)$	$\frac{1}{k}$
	$(1-k,\sqrt{2k(1-k)})$	1
	$(1-k,-\sqrt{2k(1-k)})$	1
$k\geq 1$	$(0,0)$	$\frac{1}{k}$

6. Without computing the Hessian of the Lagrangian, prove that except for the point $(0,0)$ in the case $0<k<1$, every KKT point is a global minimum.
7. By computing second order conditions, prove that the point $(0,0)$ in the case $0<k<1$ is not a local minimum.

Solution of Exercise 4.3

1. We let $k>0$ (the case $k<0$ is dealt in a similar way). Consider the points $A=(1,\sqrt{2k})$ and $B=(1,-\sqrt{2k})$ which belong to X_k. Their middle point $M=(A+B)/2=(1,0)$ does not belongs to X_k, hence X_k is not convex.
2. The function f is regular and coercive on $\mathbb{R}^2$ and $X=\{g(x,y)\leq 0\}$ with $g(x,y)=2kx-y^2$ is continuous. Hence there always exists a global minimizer.
3. Without any constraints, it is clear that f admits a unique minimizer at point $M^\star=(1,0)$. The point $M^\star$ belongs to X_k iff $k\leq 0$. So that for every $k\leq 0$ the global minimizer of f on X^k is actually the minimizer ot the unconstrained problem, that is $(1,0)$.
4. We have $\nabla g_k(x,y)=\begin{pmatrix}2k\\-2y\end{pmatrix}\neq(0,0)$. Hence the constraints are qualified everywhere.
5. The Lagrangian assocated to the problem (P_k) is, if $M=(x,y)$:

$$\mathcal{L}(M,\lambda)=(x-1)^2+y^2+\lambda(2kx-y^2),\ \lambda\geq 0.$$

and the KKT points solve

$$\begin{cases}\frac{\partial\mathcal{L}}{\partial x}(M,\lambda) & =0\\ \frac{\partial\mathcal{L}}{\partial y}(M,\lambda) & =0\\ \lambda(2kx-y^2) & =0\\ 2kx-y^2 & \leq 0\\ \lambda & \geq 0\end{cases}\Leftrightarrow\begin{cases}x-1+k\lambda & =0\\ y-\lambda y & =0\\ \lambda(2kx-y^2) & =0\\ 2kx-y^2 & \leq 0\\ \lambda & \geq 0\end{cases}$$

- **First case : no active constraints**, we have $2kx-y^2<0$ and $\lambda=0$. Without any surprise, we have $M=(1,0)$ which verifies $2kx-y^2<0$ iff $k<0$.

- **Second case : one active constraints**, then $2kx - y^2 = 0$, we have

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x}(M,\lambda) & = 0 \\ \frac{\partial \mathcal{L}}{\partial y}(M,\lambda) & = 0 \\ 2kx - y^2 & = 0 \\ \lambda & \geq 0 \end{cases} \Leftrightarrow \begin{cases} y = 0 & \text{or } \lambda = 1 \\ x & = 1 - k\lambda \\ 2kx - y^2 & = 0 \\ \lambda & \geq 0 \end{cases}$$

The above system is equivalent to

$$\begin{cases} (x,y) & = (0,0) \\ \lambda & = \frac{1}{k} \\ \lambda & \geq 0 \end{cases} \quad \text{or} \quad \begin{cases} x & = 1-k \\ y^2 & = 2k(1-k) \\ \lambda & = 1 \end{cases}$$

Note that the equation $y^2 = 2k(1-k)$ implies that $2k(1-k) \geq 0$, i.e. $0 \leq k \leq 1$.

Hence we obtain the following result

	(x,y)	λ
$k < 0$	$(1,0)$	0
$0 < k < 1$	$(0,0)$	$\frac{1}{k}$
	$(1-k, \sqrt{2k(1-k)})$	1
	$(1-k, -\sqrt{2k(1-k)})$	1
$k \geq 1$	$(0,0)$	$\frac{1}{k}$

$$H_{x,y}[L](x,y;\lambda) = \begin{pmatrix} 2 & 0 \\ 0 & 2(1-\lambda) \end{pmatrix}.$$

et le cône critique: soit $(\bar{x},\bar{y})$ tel que $2k\bar{x} - \bar{y}^2 = 0$. Alors:

$$\begin{aligned} V^*_{(\bar{x},\bar{y})}(X_k) &= \left\{ v \in \mathbb{R}^2 \mid \langle \begin{pmatrix} 2k \\ -2\bar{y} \end{pmatrix}, v \rangle = 0 \right\} \\ &= \left\{ v \in \mathbb{R}^2 \mid kv_1 - \bar{y}v_2 = 0 \right\} = Vect \begin{pmatrix} \bar{y} \\ k \end{pmatrix} \end{aligned}$$

$$H_{x,y}[L](0,0;\frac{1}{k}) = \begin{pmatrix} 2 & 0 \\ 0 & 2\frac{k-1}{k} \end{pmatrix},$$

dont les valeurs propres sont: 2 et $2\frac{k-1}{k}(< 0)$. La matrice n'est donc ni positive, ni négative, on ne peut pas encore conclure. Calculons le cône critique en $(0,0)$:

$$V^*_{(0,0)}(X_k) = Vect \begin{pmatrix} 0 \\ k \end{pmatrix} = Vect \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad (\text{puisque } k \neq 0)$$

On calcule:

$$\begin{aligned} \langle H_{x,y}[L](0,0;\frac{1}{k}) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle &= \langle \begin{pmatrix} 2 & 0 \\ 0 & 2\frac{k-1}{k} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle \\ &= 2\frac{k-1}{k} < 0. \end{aligned}$$

5 Duality

The following exercise can be solved directly or by looking at its dual. This exercise is a good introduction to dual problems.

(TD8 : 15 mins) Exercise 5.1

Consider the following problem

$$(P_1) \qquad \left| \begin{array}{ll} \inf\limits_{(x)\in\mathbb{R}^2} & f(x) = 2x_1 + x_2 \\ \text{s.c.} & x_1^2 + x_2^2 \leq 1 \end{array} \right.$$

1. Show that the problem admits a solution and that the dual approach is possible.
2. Compute $(P_1^\star)$, the dual problem of (P_1).
3. Solve $(P_1^\star)$.
4. From the solutions of $(P_1^\star)$, deduce the solutions of (P_1)

Solution of Exercise 5.1

1. We first remark that both f and $g : x \mapsto x_1^2 + x_2^2 - 1$ are convex, hence since $X = \{x \in \mathbb{R}^2 \mid g(x) \leq 0\}$ is convex, the problem is convex. Moreover Slater's condition holds because
$$g(0) = -1 < 0.$$
Moreover X is bounded and f is continuous, there exists a minimum, it is a KKT point (qualification of constraints) and it is a saddle-point. The dual approach is possible.
2. We start with the Lagrangian of (P_1):
$$\mathcal{L}(x, \lambda) = 2x_1 + x_2 + \lambda(x_1^2 + y_2^2 - 1), \ \lambda \geq 0.$$
The primal problem reads:
$$(P_1) \qquad \inf_{x\in\mathbb{R}^2} \sup_{\lambda\geq 0} \mathcal{L}(x, \lambda).$$
The dual problem reads:
$$(P_1^\star) \qquad \sup_{\lambda\geq 0} \inf_{x\in\mathbb{R}^2} \mathcal{L}(x, \lambda).$$
Let $f^\star$ be the dual function defined for all $\lambda \geq 0$.
$$f^\star(\lambda) = \inf_{x\in\mathbb{R}^2} \mathcal{L}(x, \lambda).$$
The above minimisation problem is **without** constraints. We search the critical points of $x \mapsto \mathcal{L}(x, \lambda)$, ie, $\nabla_x \mathcal{L}(x, \lambda) = 0$:
$$\left\{ \begin{array}{l} 2 + 2\lambda x_1 = 0 \\ 1 + 2\lambda x_2 = 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} x_1 = -\dfrac{1}{\lambda} \\[2ex] x_2 = -\dfrac{1}{2\lambda} \end{array} \right. \quad \text{if: } \lambda \neq 0.$$
   - **First case, $\lambda = 0$:** Solving $\nabla_x \mathcal{L}(x, 0) = 0$ is impossible, because the conditions read $2 = 0$ and $1 = 0$! Considering the Lagrangian, we have in this case:
$$\inf_{x\in\mathbb{R}^2} \mathcal{L}(x, 0) = \inf_{x\in\mathbb{R}^2} 2x_1 + x_2 = -\infty.$$

- **Second case, $\lambda > 0$:** The point $x = (-\frac{1}{\lambda}, -\frac{1}{2\lambda})$ is a critical point of the **convex** function $x \mapsto \mathcal{L}(x, \lambda)$, hence is the unique global minimizer. The dual function is

$$\begin{aligned}\forall \lambda > 0,\ f^\star(\lambda) &= \mathcal{L}((-\frac{1}{\lambda}, -\frac{1}{2\lambda}), \lambda) = -\frac{2}{\lambda} - \frac{1}{2\lambda} + \lambda(\frac{1}{\lambda^2} + \frac{1}{4\lambda^2} - 1)\\ &= -\lambda - \frac{5}{4\lambda}\end{aligned}$$

Hence we have

$$\forall \lambda \geq 0,\ f^\star(\lambda) = \begin{cases} -\lambda - \dfrac{5}{4\lambda} & \text{if: } \lambda > 0, \\ -\infty & \text{if: } \lambda = 0. \end{cases}$$

The dual problem is then

$$(P_1^*) \qquad \sup_{\lambda \geq 0} f^*(\lambda) = \sup_{\lambda > 0} \left(-\lambda - \frac{5}{4\lambda} \right).$$

3. By definition, the dual function f^* is concave. It is then sufficient to compute the stationary points of the dual function to obtain the global maximizers of f. We check that

$$f^{\star'}(\lambda) = 0 \ \Leftrightarrow\ -1 + \frac{5}{4\lambda^2} = 0 \ \Leftrightarrow\ \lambda = \pm\frac{\sqrt{5}}{2}.$$

We only keep $\lambda > 0$, we have :

$$\lambda^\star = \frac{\sqrt{5}}{2} > 0$$

is the unique global maximum of f^* and the primal solution associated to it is the unique solution to $\nabla_x \mathcal{L}(x^\star, \lambda^\star) = 0$, that is :

$$x^\star = (-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}).$$

4. We know there exists a global minimizer. We know that each minimizer is a KKT point and a saddle point (convexity), there exists only one possible $\lambda^\star$ and for this $\lambda^\star$, only one possible $x^\star$, hence we found the saddle point and the unique global minimizer.

The following exercise is a follow-up on Exercise 3.2, where the constraints are not qualified and the global minimum is not a KKT point. We study in this exercise the behavior of the dual function.

(TD8 : 20 mins) Exercise 5.2

We consider the problem $\min_{x \in X} f(x)$, where $X \subset \mathbb{R}^2$ and

$$f(x_1, x_2) = x_2 \quad X = \{x = (x_1, x_2) \text{ such that } (x_1-1)^2 + x_2^2 \leq 1 \text{ and } (x_1+1)^2 + x_2^2 \leq 1\}$$

1. Draw X, show that the problem is convex.
2. Show that $X = \{0\}$, give the global minimum of f over X.

3. Show that there is no KKT point. Show that the constraints are not qualified anywhere.
4. Compute the Lagrangian and the dual function, show that the latter is equal to

$$f^\star(\lambda) = -\frac{1}{2(\lambda_1+\lambda_2)} - \frac{(\lambda_2-\lambda_1)^2}{\lambda_1+\lambda_2}$$

5. Solve the dual problem, show that there is no duality gap but that there is no saddle point.

Solution of Exercise 5.2

1. X is the intersection of two circles, it should be clear by a graphic that $X = \{0\}$. The function f is convex and $X = \{g(x) \preceq 0\}$ with $g(x) = \begin{pmatrix} x_1^2 - 2x_1 + x_2^2 \\ x_1^2 + 2x_1 + x_2^2 \end{pmatrix}$. Both functions g_1 and g_2 are convex, hence the problem is convex.
2. Let $x = (x_1, x_2) \in X$. From $x_1^2 - 2x_1 + x_2^2 \leq 0$, we conclude $x_1 \geq 0$, similarly from $g_2(x) = 0$ we can conclude $x_1 \leq 0$. Then $x_1 = 0$ and hence $x_2 = 0$ (from $x_2^2 \leq 0$). Since X is a singleton, the global minimum of f over X is attained at 0.
3. We write $\nabla\mathcal{L} = 0$ which gives

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} -2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 0$$

This is not solvable, there is no KKT point. Hence the constraints are not qualified at 0 (or else the global minimum would be a KKT point). We can check that at 0, both constraints are active and that their gradient are colinear. We can also check that the Slater condition cannot hold.
4. The Lagrangian is $\mathcal{L}(x,\lambda) = x_2 + (\lambda_1+\lambda_2)(x_1^2+x_2^2) + 2(\lambda_2-\lambda_1)(x_1)$ If $\lambda_1 = \lambda_2 = 0$, then $\inf_x \mathcal{L}(x,\lambda) = -\infty$. If this is not the case, then

$$\inf_x \mathcal{L}(x,\lambda) = \mathcal{L}(x^\star,\lambda),$$

With $x^\star = -(\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}, \frac{1}{2}\frac{1}{\lambda_1+\lambda_2})$ and hence

$$f^\star(\lambda) = -\frac{1}{2(\lambda_1+\lambda_2)} - \frac{(\lambda_2-\lambda_1)^2}{\lambda_1+\lambda_2}.$$

5. The dual problem is $\sup f^\star(\lambda)$ which is equal to 0 as $\lambda_1 = \lambda_2$ and as they tend to $+\infty$. The minimum to the primal problem is 0, hence there is no duality gap. There is no saddle point because there is no maximum to the dual problem, only a supremum. Notice that if we had a saddle point, then we would have a KKT point.

The following exercise is an abstract exercise that is solved twice, first using a primal approach and then using a dual approach. The objective is to understand the common computations between the two approaches.

(TD8 : 15 mins) Exercise 5.3

Consider the following problem with $a \neq 0$

$$\inf_{x \in \mathbb{R}^n} \langle a, x \rangle \quad \text{with } \|x\|_2^2 \leq 1.$$

1. Show that the problem is convex and admits a global minimizer. Show that the constraints are qualified.
2. Find every KKT point.
3. Compute the dual function and solve the dual problem.
4. Discuss the difference between the KKT search and the solution of the dual problem.

Solution of Exercise 5.3

1. The function f is linear hence convex, the constraint $g(x) \leq 0$ with $g(x) = \|x\|^2 - 1$ is also convex because the Hessian of g is twice the identity (an other argumentation is that the set of constraints is the unit ball for the 2-norm). We have $g(0) < 0$, hence the constraints are qualified by Slater's conditions. There exists a global minimizer because both f and g are continuous and the domain is bounded. This global minimizer is a KKT point which is also a saddle point. The Lagrangian is given by

$$\mathcal{L}(x, \lambda) = \langle a, x \rangle + \lambda(\|x\|^2 - 1), \quad \lambda \geq 0.$$

2. We have $\nabla_x \mathcal{L}(x, \lambda) = a + 2\lambda x$, the case $\lambda = 0$ is impossible, hence $x = -\frac{a}{2\lambda}$. Because $\lambda \neq 0$, we must have $\|x\|^2 = 1$ and hence $\lambda = \pm\frac{\|a\|}{2}$, we have to take $\lambda \geq 0$, hence there is a unique KKT point and hence a unique minimizer of the primal problem given by $x = -\frac{a}{\|a\|}$ associated to $\lambda = \frac{\|a\|}{2}$.
3. The primal problem reads

$$\inf_{x \in \mathbb{R}^n} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda).$$

The dual problem is

$$\sup_{\lambda \geq 0} \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda).$$

We first compute the dual function

$$f^*(\lambda) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda), \quad \lambda \geq 0.$$

It is a convex optimization problem without constraint, the critical point are given by

$$a + 2\lambda x = 0, \quad \text{hence:} \quad x = -\frac{a}{2\lambda}$$

The critical points are global solution, because for fixed $\lambda \geq 0$, the function $x \mapsto \mathcal{L}(x, \lambda)$ is convexe, hence:

$$f^*(\lambda) = \begin{cases} \mathcal{L}(-\frac{a}{2\lambda}, \lambda) = -\lambda - \dfrac{\|a\|^2}{4\lambda} & \text{if } \lambda > 0 \\ -\infty & \text{if } \lambda = 0 \end{cases}$$

The dual problem is

$$\sup_{\lambda \geq 0} f^*(\lambda) = \sup_{\lambda > 0} \left(-\lambda - \frac{\|a\|^2}{4\lambda} \right).$$

Because the constraint $\lambda > 0$ defines an open admissible set, the dual problem is a concave optimization problem whose solution are caracterized by the necessary and sufficient first order optimality condition:

$$-1 - \frac{\|a\|^2}{4\lambda^2} = 0 \quad \text{that is:} \qquad \lambda = \pm\frac{\|a\|}{2}.$$

Adding the constraint $\lambda > 0$, we have:

$$\lambda^\star = \frac{\|a\|}{2} \text{ solution of the dual problem.}$$

The solution to the associate primal problem is:

$$x^\star = -\frac{a}{\|a\|}$$

The optimal value of the primal (and dual) is $-\|a\|$. We know that $(x^\star, \lambda^\star)$ is a saddle-point of the Lagrangian, hence $x^\star$ is the solution to the primal problem and $\lambda^\star$ is the corresponding Lagrange multiplier.

This exercise shows that, when there is no convexity, KKT points are not necessary saddle-points.

(TD9 : 20 mins) Exercise 5.4

Let

$$(P) \quad \left|\begin{array}{ll} \min\limits_{(x,y)\in\mathbb{R}^2} & -x^2 + y^2 \\ \text{s.c.} & x + y \geq 0 \\ & y \geq x \end{array}\right.$$

1. Draw X, the set of constraints of (P). Show **very simply** that problem (P) admits an infinity of solutions. You will show that this set of solutions is $\{(x, |x|), x \in \mathbb{R}\}$.
2. Put the problem in standard form, compute the dual function and show that the associated Lagrangian does not admit saddle points.

Solution of Exercise 5.4

1.

The admissible domain is given by

$$\begin{aligned} X &= \{(x,y) \in \mathbb{R}^2 \mid x + y \geq 0,\ y \geq x\} \\ &= \{(x,y) \in \mathbb{R}^2 \mid y \geq -x,\ y \geq x\} \\ &= \{(x,y) \in \mathbb{R}^2 \mid y \geq |x|\}. \end{aligned}$$

We then have:

$$\forall (x,y) \in X,\ f(x,y) = -x^2 + y^2 \geq 0.$$

Moreover, for every $(x,y) \in \mathbb{R}^2$, we have:

$$f(x,y) = 0 \Leftrightarrow -x^2 + y^2 = 0 \Leftrightarrow y = |x|.$$

Hence the points $(x, |x|)$, $x \in \mathbb{R}$, (which form the boundary of X) are the global minimum points of f on $\mathbb{R}^2$.

2. The problem (P) is written in standard form:

$$\left|\begin{array}{ll} \min\limits_{(x,y)\in\mathbb{R}^2} & -x^2+y^2 \\ \text{s.c.} & -x-y\leq 0 \\ & \quad x-y\leq 0 \end{array}\right.$$

The associated Lagrangian is if $M=(x,y)$:

$$\mathcal{L}(M,\lambda) = -x^2+y^2-\lambda_1(x+y)+\lambda_2(x-y),\ \lambda_1\geq 0,\ \lambda_2\geq 0.$$

The dual problem of (P) is :

$$\sup_{\lambda\geq 0} f^\star(\lambda),\ \text{with } f^\star(\lambda) = \inf_{M\in\mathbb{R}^2}\mathcal{L}(M,\lambda).$$

We solve the critical point equation

$$\nabla_M\mathcal{L}(M,\lambda)=0 \Leftrightarrow \left\{\begin{array}{rcl} -2x-\lambda_1+\lambda_2 &=& 0 \\ 2y-\lambda_1-\lambda_2 &=& 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} x &=& \dfrac{\lambda_2-\lambda_1}{2} \\ y &=& \dfrac{\lambda_1+\lambda_2}{2} \end{array}\right.$$

We check the second order necessary condition for this critical point to be a local minimum

$$\forall M\in\mathbb{R}^2,\ H_M[\mathcal{L}](M,\lambda) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}.$$

The second order necessary condition is not satisfied, this means that the problem $\inf_{M\in\mathbb{R}^2}\mathcal{L}(M,\lambda)$ has no minimizers. Indeed, a closer inspection shows that if $M=(x,0)$, and $x\to+\infty$ then $\mathcal{L}(M,\lambda)$ tends to $-\infty$. This means that $f^\star(\lambda)=-\infty$ for every λ. The Lagrangian cannot have any saddle-point, even if the primal problem admits (an infinite number of) global solutions.

The next exercise is a $3d$ convex optimization problem.

(20 mins) Exercise 5.5

Let

$$(P)\quad \left|\begin{array}{ll} \min\limits_{(x,y,z)\in\mathbb{R}^3} & x-y \\ \text{with:} & x+y+z\leq 1,\quad x^2+y^2+z^2\leq 1 \end{array}\right.$$

1. Show that problem (P) has a solution.
2. Solve problem (P) by a dual approach.

Solution of Exercise 5.5

1. Let $M=(x,y,z)$, $f(M)=x-y$ and $g_1(M)=x+y+z-1$ and $g_2(M)=x^2+y^2+z^2-1$. The function f continuous on $X=\{M\in\mathbb{R}^3 \mid g_i(M)\leq 0\}$ which is a closed (because g_i is continuous) and bounded set on $\mathbb{R}^3$. Hence f admits at least a global minimizer and a global maximizer. By taking $M=0$, we see that $g_i(M)<0$ for all i, hence Slater condition is true and the constraints

are qualified everywhere. Moreover the g_i are convex and so is f, so that the problem is convex. We now that the global minimizer is a saddle point.

2. The Lagrangian is defined by

$$\mathcal{L}(M,\lambda) = x - y + \lambda_1(x+y+z-1) + \lambda_2(x^2+y^2+z^2-1), \quad \lambda \geq 0.$$

We recall that the primal problem is:

$$\min_{M\in\mathbb{R}^3} \sup_{\lambda\geq 0} \mathcal{L}(M,\lambda).$$

The dual problem is :

$$\sup_{\lambda\geq 0} \inf_{M\in\mathbb{R}^3} \mathcal{L}(M,\lambda).$$

We first compute the dual function $f^\star$ which is given by

$$f^\star(\lambda) = \inf_{M\in\mathbb{R}^3} \mathcal{L}(M,\lambda).$$

Finding the infimum of the Lagrangian is a convex (for $\lambda \geq 0$) unconstrained optimization problem, whose solution (if they exist) are given by $\nabla_M \mathcal{L}(M,\lambda) = 0$, this yields

$$\left\{\begin{array}{rcl} 1+\lambda_1+2\lambda_2 x & = & 0 \\ -1+\lambda_1+2\lambda_2 y & = & 0 \\ \lambda_1+2\lambda_2 z & = & 0. \end{array}\right. \Rightarrow \left\{\begin{array}{rcl} x & = & -\dfrac{1+\lambda_1}{2\lambda_2} \\ y & = & \dfrac{1-\lambda_1}{2\lambda_2} \\ z & = & -\dfrac{\lambda_1}{2\lambda_2}. \end{array}\right.,$$

With the condition $\lambda_2 > 0$. We note that in the case $\lambda_2 = 0$, we have $f^\star(\lambda) = -\infty$. The dual function is given by:

$$\begin{aligned} f^\star(\lambda) &= \left\{\begin{array}{ll} \mathcal{L}\left(\left(-\dfrac{1+\lambda_1}{2\lambda_2}, \dfrac{1-\lambda_1}{2\lambda_2}, -\dfrac{\lambda_1}{2\lambda_2}\right), \lambda\right) & \text{if } \lambda_2 > 0, \\ -\infty & \text{otherwise.} \end{array}\right. \\ &= \left\{\begin{array}{ll} -\dfrac{1}{4\lambda_2}\left(3\lambda_1^2 + 4\lambda_1\lambda_2 + 2 + 4\lambda_2^2\right) & \text{if } \lambda_2 > 0, \\ -\infty & \text{otherwise.} \end{array}\right. \end{aligned}$$

The dual problem is now given by

$$\sup_{\lambda_1\geq 0, \lambda_2 > 0} f^\star(\lambda_1,\lambda_2).$$

Because the set of constraints is separable between λ_1 et λ_2, we start by solving, for every possible λ_2

$$\sup_{\lambda_1\geq 0} f^\star(\lambda_1,\lambda_2)$$

We notice that the constraints are qualified and we write KKT equations for this concave maximization problem, we have

$$-\frac{1}{4\lambda_2}(6\lambda_1+4\lambda_2) + s = 0 \text{ with } s \leq 0, \text{ and } s\lambda_1 = 0$$

The case $s = 0$ implies $\lambda_1 < 0$ which is impossible, hence :

$$\lambda_1^\star = 0.$$

We now solve

$$\sup_{\lambda_2 > 0} f^\star(0, \lambda_2) = -\frac{1}{2\lambda_2} - \lambda_2$$

We solve $\partial_2 f^\star(0, \lambda_2) = 0$, this gives

$$\lambda_2^\star = \frac{1}{\sqrt{2}}.$$

Then the Lagrangian admits a unique saddle point $(M^\star, \lambda^\star)$ with

$$\lambda^\star = (0, \frac{1}{\sqrt{2}}) \text{ and } M^\star = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0).$$

For some extra fun, we can check that

- The primal solution is admissible (i.e. $M^\star \in X$) (check it !!).
- Complementary conditions are verified:

$$\lambda_1^\star(x\star + y^\star + z^\star - 1) = 0 \text{ because } \lambda_1^* = 0$$

$$\lambda_2^\star(x^{\star 2} + y^{\star 2} + z^{\star 2} - 1) = \frac{1}{\sqrt{2}} \times 0 = 0.$$

The next exercise is all about linear programming in standard duality.

(TD9 : 25 mins) Exercise 5.6: Linear programming

Consider the linear optimization problem

$$(\mathcal{LP}) \quad \left| \begin{array}{ll} \displaystyle\min_{M=(x,y)\in\mathbb{R}^2} & 2x + y \\ \text{s.t.} & x + y = 1, x \geq 0, y \geq 0 \end{array} \right.$$

1. Show that the problem admits a solution and that the constraints are qualified.
2. Compute $\mathcal{L}$, the Lagrangian of the problem and show that $\nabla_M \mathcal{L}$ does not depend on M.
3. Solve the KKT equations.
4. Compute the dual function and solve the dual problem.
5. How do we recover the solution of the primal problem from the solution of the dual problem ?

Solution of Exercise 5.6

1. The set of constraints $X = \{g_1 \leq 0, g_2 \leq 0, g_3 = 0\}$ is bounded. Here we set $g_3(x, y) = 1 - x - y$ and $g_1(x, y) = -x$ and $g_2(x, y) = -y$. The functions g are continuous, hence X is closed. The function f is continuous. The constraints are qualified because they are affine.
2. The Lagrangian is equal to

$$\mathcal{L}(M, \lambda) = 2x + y - \lambda_1 x - \lambda_2 y + \lambda_3(1 - x - y)$$

3. The KKT equations are

$$\begin{pmatrix}2\\1\end{pmatrix} + \lambda_1 \begin{pmatrix}-1\\0\end{pmatrix} + \lambda_2 \begin{pmatrix}0\\-1\end{pmatrix} + \lambda_3 \begin{pmatrix}-1\\-1\end{pmatrix} = 0$$

   - The case $\lambda_1 = \lambda_2 = 0$ is impossible.
   - The case $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ is also impossible because in this case we have $g_1(x, y) = 0$ and $g_2(x, y) = 0$, hence $x = y = 0$ and then $g_3(x, y) \neq 0$. Hence this case is impossible.
   - We study the case $\lambda_1 \neq 0$, this yields $x = 0$ and $g_3(x, y) = 0$ implies $y = 1$. Moreover we have $\lambda_2 = 0$, hence $\lambda_3 = 1$ and then $\lambda_1 = 1$.
   - We study the case $\lambda_2 \neq 0$, this yields $y = 0$ and $g_3(x, y) = 0$ implies $x = 1$. Moreover we have $\lambda_1 = 0$, hence $\lambda_3 = 2$ and then $\lambda_2 = -1$. Which is impossible.

   There is only one KKT point is given by $(x, y) = (0, 1)$ associated to the Lagrange multipliers $(1, 0, 1)$.
4. The function $M \mapsto \mathcal{L}(M, \lambda)$ is affine. Hence either its linear part is 0 or its infimum is equal to $-\infty$. Hence $f^\star(\lambda) = -\infty$ except on the points λ that verifies

$$\begin{cases} 2 - \lambda_1 - \lambda_3 &= 0 \\ 1 - \lambda_2 - \lambda_3 &= 0 \end{cases}$$

   For all these points, we have $f^\star(\lambda) = \lambda_3$.
5. We want to maximize $f^\star(\lambda)$ when $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ and $\begin{cases} 2 - \lambda_1 - \lambda_3 &= 0 \\ 1 - \lambda_2 - \lambda_3 &= 0 \end{cases}$

   The second equation states that $\lambda_3 \leq 1$, and the choice $\lambda_2 = 0$ and $\lambda_1 = 1$ is the only choice for which $f^\star(\lambda) = 1$.
6. The solution of the dual problem is $\lambda = (1, 0, 1)$. Since $\lambda_1 \neq 0$ we know that $x = 0$ and then from $g_3(x, y) = 0$ we recover $y = 1$.

The following exercise is left as a training exercise

(25 mins) Exercise 5.7

Consider the optimization problem

$$(P) \quad \left| \begin{array}{ll} \displaystyle\min_{(x,y)\in\mathbb{R}^2} & x^2 + y^2 - x \\ \text{s.t.} & x + y \geq 1 \end{array} \right.$$

1. Put the problem in standard form and give $(P^\star)$, its dual.
2. Solve $(P^\star)$.
3. Give the solutions of (P).

This exercise is a little bit on the theoritical side, but it covers any quadratic function with affine equality constraints.

(TD9 : 30 mins) Exercise 5.8: Quadratic optimization

Let A be a square real symetric and invertible matrix of size n. Let $b \in \mathbb{R}^n$ and C be a real matrix of size $p \times n$ and of full rank $(p < n)$. Let $d \in \mathbb{R}^p$. We consider the following problem:

$$(P) \quad \left| \begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle \\ \text{s.c.} & Cx - d = 0 \end{array} \right.$$

1. Under which conditions is $f^\star$, the dual function associated to the problem (P) is well defined (not always equal to $-\infty$)? Compute it.
2. Solve the dual problem $(P^\star)$. Do we have existence and uniqueness of a solution of (P) ? Give the analytical expression of this solution and of the associated multiplicator
3. Replace the equality constraint by an inequality $Cx - d \leq 0$. Write the dual problem associated to the new constraint.

Solution of Exercise 5.8

1. First remark that (P) is convex if and only if the matrix A is positive. The Lagrangian is given by

$$\mathcal{L}(x, \lambda) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \langle \lambda, Cx - d\rangle, \lambda \in \mathbb{R}^p.$$

The dual problem of (P) is:

$$(P^\star) \qquad \sup_{\lambda\in\mathbb{R}^p} \left(\inf_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda) \right).$$

We first compute the dual function: $f^\star(\lambda) = \inf_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda)$. It is an optimization problem without constraint, the critical points are denoted x_λ and are the solutions to:

$$\nabla_x \mathcal{L}(x_\lambda, \lambda) = 0 \Longleftrightarrow Ax_\lambda + b + C^\top \lambda = 0,$$

Hence

$$x_\lambda = -A^{-1}(C^\top \lambda + b),$$

because the matrix A is invertible. Then $H_x[\mathcal{L}](x_\lambda, \lambda) = A$, hence the critical point is a local minimum iff the matrix A is definite positive (if the problem is convex). If A is not positive definite, then $f^\star(\lambda) = -\infty$ for every λ. We now suppose that A is definite positive. We have in this case :

$$\begin{aligned}
f^\star(\lambda) &= \mathcal{L}(x_\lambda, \lambda) = \frac{1}{2}\langle Ax_\lambda, x_\lambda\rangle + \langle b, x_\lambda\rangle + \langle \lambda, Cx_\lambda - d\rangle \\
&= \frac{1}{2}\langle Ax_\lambda, x_\lambda\rangle + \langle b, x_\lambda\rangle + \langle C^\top\lambda, x_\lambda\rangle - \langle \lambda, d\rangle \\
&= \frac{1}{2}\langle \underbrace{Ax_\lambda + b + C^\top\lambda}_{=0}, x_\lambda\rangle + \frac{1}{2}\langle C^\top\lambda, x_\lambda\rangle + \frac{1}{2}\langle b, x_\lambda\rangle - \langle \lambda, d\rangle \\
&= -\frac{1}{2}\langle C^\top\lambda + b, A^{-1}(C^\top\lambda + b)\rangle - \langle \lambda, d\rangle \\
&= -\frac{1}{2}\langle C^\top\lambda, A^{-1}C^\top\lambda\rangle - \frac{1}{2}\langle C^\top\lambda, A^{-1}b\rangle \\
&\quad -\frac{1}{2}\langle b, A^{-1}C^\top\lambda\rangle - \frac{1}{2}\langle b, A^{-1}b\rangle - \langle \lambda, d\rangle \\
&= -\frac{1}{2}\langle (CA^{-1}C^\top)\lambda, \lambda\rangle - \langle CA^{-1}b + d, \lambda\rangle - \frac{1}{2}\langle b, A^{-1}b\rangle
\end{aligned}$$

2. Suppose A definite positive. The dual problem $(P^\star)$ is:

$$\sup_{\lambda\in\mathbb{R}^p} -\frac{1}{2}\langle (CA^{-1}C^\top)\lambda,\lambda\rangle - \langle CA^{-1}b+d,\lambda\rangle - \frac{1}{2}\langle b, A^{-1}b\rangle,$$

or equivalently:

$$\inf_{\lambda\in\mathbb{R}^p} \frac{1}{2}\langle (CA^{-1}C^\top)\lambda,\lambda\rangle + \langle CA^{-1}b+d,\lambda\rangle + \frac{1}{2}\langle b, A^{-1}b\rangle.$$

This is a convex optimization problem without constraint. The solutions $\lambda^\star$ of the dual problem are solution to the system:

$$(CA^{-1}C^\top)\lambda^\star + CA^{-1}b + d = 0, \quad \text{soit:} \quad \lambda^\star = -(CA^{-1}C^\top)^{-1}(CA^{-1}b+d),$$

Indeed, since C is supposed to be of full rank, the matrix $CA^{-1}C^\top$ is invertible. We then have:

$$x^\star = x_{\lambda^\star} = -A^{-1}(C^\top\lambda^\star + b).$$

With a simple computation we check that

$$Cx^\star - d = 0,$$

3. The Lagrangien is the same, but we impose a condition on the sign of the Lagrange mutlipliers. The dual function is the same and the dual program is given by

$$\min_{\lambda\geq 0} \frac{1}{2}\langle (CA^{-1}C^\top)\lambda,\lambda\rangle + \langle CA^{-1}b+d,\lambda\rangle + \frac{1}{2}\langle b, A^{-1}b\rangle.$$

6 First order descent methods

The next exercise proves that the Wolfe step bisection algorithm converges in a finite number of iterations. It also proves the existence of a Wolfe step.

(TD10 : 45 mins) Exercise 6.1

The objective is to prove the convergence of the Wolfe step bisection algorithm. We recall that we are given a function ϕ which is C^1, bounded from below, with $\phi(0) = 0$ and $\phi'(0) < 0$. Morevoer, we are given $0 < \varepsilon_1 < \varepsilon_2 < 1$. If we denote

$$W_1 = \{s \geq 0 \text{ such that } \phi(s) \leq \varepsilon_1 s\phi'(0)\} \text{ and } W_2 = \{s \geq 0 \text{ such that } \phi'(s) \geq \varepsilon_2\phi'(0)\}$$

A Wolfe step is any element in $W_1 \cap W_2$. The bisection algorithm is the following

```
def Wolfe(s) # s is the initial step, it is >0.
  sm,sp=0,np.inf
  while True :
    if not s in W_1 :
      sp=s
      s=0.5*(sp+sm)
    elif not s in W_2 :
      sm=s
```

```
    s=2s if sp==np.inf else 0.5*(sp+sm)
  else :
    return s
```

We denote s^- the value of `sm` and s^+ the value of `sp` The objective is to prove that this algorithm converges in a finite number of iterations in the `while` loop, suppose it spends an infinite number of iterations. We denote s_k^+, s_k^-, s_k the values of s^+, s^- and s at iteration k

1. Show that for k, $s_k^+ \in W_1^c$ and $s_k^- \in W_2^c \cap W_1$
2. Show that for any k, $s_k^- \leq s_k \leq s_k^+$ and $(s_k^+)_k$ is decreasing and $(s_k^-)_k$ is increasing.
3. Show that it is not possible to have $s_k^+ = +\infty$ for each k.
4. Show that there exists $s^\star$ such that $s_k^- \leq s^\star \leq s_k^+$ and both $(s_k^+)_k$ and $(s_k^-)_k$ converge to $s^\star$.
5. By using $s^+ \in W_1^c$ and $s^- \in W^1$, show that $\phi(s^\star) = \varepsilon_1 s^\star \phi'(0)$.
6. By computing $\frac{\phi(s^\star)-\phi(s)}{s^\star - s}$ for $s = s_k^+$ or for $s = s_k^-$ (or both), show that
$$\phi'(s^\star) \geq \varepsilon_1 \phi'(0)$$
7. Show that the algorithm cannot spend an infinite number of iterations.

Solution of Exercise 6.1

1. Carefully check that at $k = 0$, we indeed have $s_0^+ \in W_1^c$ and $s_0^- \in W_2^c \cap W_1$. Each time s^+ (resp. s^-) is updated, it is by an s that belongs to W_1^c (resp. $W_2^c \cap W_1$).
2. Check by recurence that we always have $s^- < s < s^+$, so that when s^+ (resp. s^-) is updated by $s^+ = s$ (resp. $s^- = s$), it decreases (resp. increases).
3. If $s_k^+ = +\infty$ for each k, then s_k belongs to W^1 for each k and $s_k = 2^s$. Because s_k belongs to W^1, we have $\phi(s_k) \leq \varepsilon_1 s_k \phi'(0)$. As k goes to $+\infty$, this means that ϕ is unbounded from below.
4. Wait until $s_k^+ \neq +\infty$, from this iteration, then $(s_k^+)_k$ is decreasing and $(s_k^-)_k$ is increasing and the length $(s_k^+ - s_k^-)_k$ is positive and divided by two at each iteration, hence both sequence converge to the same $s^\star$, one from below and one from above.
5. Because $s_k^+ \in W_1^c$ for all k and s_k^+ converges to $s^\star$, we have $\phi(s^\star) \geq \varepsilon_1 s^\star \phi'(0)$. Similarly, with s_k^-, we have $\phi(s^\star) \leq \varepsilon_1 s^\star \phi'(0)$
6. We have
$$\begin{array}{rcl} \phi(s_k^+) & > & \varepsilon_1 s_k^+ \phi'(0) \\ \phi(s_k^+) - \phi(s^\star) & > & \varepsilon_1 (s_k^+ - s^\star)\phi'(0) \\ \dfrac{\phi(s_k^+) - \phi(s^\star)}{s_k^+ - s^\star} & > & \varepsilon_1 \phi'(0) \\ \phi'(s^\star) & \geq & \varepsilon_1 \phi'(0) \end{array}$$

or, using s^-.

$$\begin{array}{rcl} \phi(s_k^-) & \leq & \varepsilon_1 s_k^- \phi'(0) \\ \phi(s_k^-) - \phi(s^\star) & \leq & \varepsilon_1 (s_k^- - s^\star)\phi'(0) \\ \dfrac{\phi(s_k^-) - \phi(s^\star)}{s_k^- - s^\star} & \geq & \varepsilon_1 \phi'(0) \\ \phi'(s^\star) & \geq & \varepsilon_1 \phi'(0) \end{array}$$

7. Because $s_k^- \in W_2^c$, we must have $\phi'(s^\star) \leq \varepsilon_2 \phi'(0)$ and this is not possible with $\varepsilon_1 < \varepsilon_2$. Hence the algorithm finishes in a finite number of iterations

The exercise below is an introduction to the theory of convergence of the stochastic gradient.

(60 mins) Exercise 6.2: Stochastic gradient

The objective is to minimize the function $x \mapsto F(x) = \frac{1}{N}\sum_{i=1}^N F_i(x)$ with N very large. We suppose that at each iteration n, we draw batch of indices $\mathcal{B}_n \subset [1, N]$ uniformly and we are given the function $f_n = \frac{1}{|\mathcal{B}_n|}\sum_{i\in\mathcal{B}_n} F_i$. We suppose that the draws are independent. We then have, forall x deterministic

$$\mathbb{E}(f_n(x)) = F(x) \quad \mathbb{E}(\nabla f_n(x)) = \nabla F(x).$$

- We say that a random variable is $\mathcal{F}_n$-measurable if it only depends on $(\omega_1, \ldots, \omega_n)$, where ω_i is the randomness that defines $\omega_i \mapsto \mathcal{B}_i(\omega_i)$. In simple words, X is $\mathcal{F}_n$-measurable iff it is determined by the first n draws. In probabilistic words, $(\mathcal{F}_p)_p$ is the filtration generated by the batches $(\mathcal{B}_n)_n$.
- For any n, m, if X is $\mathcal{F}_m$-mesurable, we define $\mathbb{E}_n(X)$,a $\mathcal{F}_n$-measurable random variable as

$$\mathbb{E}_n(X) = \begin{cases} X & \text{if } m \leq n \\ \int X(\omega_1, \ldots \omega_m) d\mathbb{P}(w_{n+1}) \ldots d\mathbb{P}(w_m) & \text{if } m > n \end{cases}.$$

  In other words $\mathbb{E}_n(X) = \mathbb{E}(X|\mathcal{F}_n)$ is the conditional expectation with respect to the σ-algebra $\mathcal{F}_n$.
- We have $\mathbb{E}_0 = \mathbb{E}$ and for every X and Y random variables, if Y is $\mathcal{F}_n$-measurable, then :

$$\mathbb{E}_n(XY) = \mathbb{E}_n(X)Y \text{ and } \mathbb{E}_n(\mathbb{E}_m(X)) = \mathbb{E}_{\min(n,m)}(X)$$

We suppose F has an L-Lipschitz gradient and is bounded from below, we suppose that $(x_n)_n$ is a sequence of random variables given by the following gradient method:

$$x_{n+1} = x_n - \alpha_{n+1}\nabla f_{n+1}(x_n), \text{ with } \quad \alpha_n < c.$$

Here α_n is the step and is not a random variable. Suppose that there exists constants A and B such that, for every θ

$$\mathbb{V}(\nabla f_n(x)) \leq A + B\|\nabla F(x)\|^2,$$

Where

$$\mathbb{V}(\nabla f_n(x)) = \mathbb{E}(\|\nabla f_n(x)\|^2) - \|\nabla F(x)\|^2$$

1. If x_0 is deterministic, show that x_n is $\mathcal{F}_n$-measurable for each n.
2. Obtain the following inequality

$$F(x_{n+1}) - F(x_n) \leq -\alpha_n \langle \nabla f_{n+1}(x_n), \nabla F(x_n) \rangle + \frac{L\alpha_n^2}{2} \|\nabla f_{n+1}(x_n)\|^2$$

3. Apply $\mathbb{E}_n$ on both sides of the inequality to obtain

$$\mathbb{E}_n(F(x_{n+1})) - F(x_n) \leq \left(-\alpha_n + \frac{L\alpha_n^2(B+1)}{2}\right) \|\nabla F(x_n)\|^2 + \frac{L\alpha_n^2}{2} A$$

4. Show that there exists a $c > 0$ such that if $0 < \alpha_n < c$ for all n, then

$$\mathbb{E}_n(F(x_{n+1})) - F(x_n) \leq -\frac{\alpha_n}{2} \|\nabla F(x_n)\|^2 + \frac{C_2 \alpha_n^2}{2}$$

5. Apply the expectation $\mathbb{E}$ and prove that there exists $C, C_2 > 0$ such that

$$\sum_{n=1}^{K} \alpha_n \mathbb{E}\left(\|\nabla F(x_n)\|^2\right) \leq C + C_2 \sum_{n=1}^{K} \alpha_n^2$$

6. Suppose that $\sum_{n=1}^{+\infty} \alpha_n^2 < +\infty$ and $\sum_{n=1}^{+\infty} \alpha_n = +\infty$. Show that

$$\lim_{K \to +\infty} \left(\frac{1}{\sum_{n=1}^{K} \alpha_n} \sum_{n=1}^{K} \alpha_n \mathbb{E}\left(\|\nabla F(x_n)\|^2\right) \right) = 0$$

7. At each iteration n, draw a number between 1 and n, call τ_n this drawing process, it must be independent of the drawing of the batches $\mathcal{B}$ and the law is given by

$$\mathbb{P}(\tau_n = k) = \frac{\alpha_k}{\sum_{k=1}^{n} \alpha_n} \quad \forall k = 1, \ldots n$$

Show that

$$\lim_{n \to +\infty} \mathbb{E}\left(\|\nabla F(x_{\tau_n})\|^2\right) = 0$$

Solution of Exercise 6.2

1. The proof is done by recurrence.
2. Remember that F has L-Lipschitz gradient

$$\begin{aligned} F(x_{n+1}) - F(x_n) &\leq \langle x_{n+1} - x_n, \nabla F(x_n) \rangle + \frac{L}{2} \|x_{n+1} - x_n\|^2 \\ &\leq -\alpha_n \langle \nabla f_{n+1}(x_n), \nabla F(x_n) \rangle + \frac{L\alpha_n^2}{2} \|\nabla f_{n+1}(x_n)\|^2 \end{aligned}$$

3. Apply the conditional expectation

$$\begin{aligned}\mathbb{E}_n(F(x_{n+1})) - F(x_n) &\leq -\alpha_n \mathbb{E}_n\left(\langle \nabla f_{n+1}(x_n), \nabla F(x_n)\rangle\right) \\ &\quad + \frac{L\alpha_n^2}{2}\mathbb{E}_n(\|\nabla f_{n+1}(x_n)\|^2) \\ &\leq \left(-\alpha_n + \frac{L\alpha_n^2}{2}\right)\|\nabla F(x_n)\|^2 \\ &\quad + \frac{L\alpha_n^2}{2}\left(\mathbb{E}_n(\|\nabla f_{n+1}(x_n)\|^2) - \|\nabla F(x_n)\|^2\right) \\ &\leq \left(-\alpha_n + \frac{L\alpha_n^2(B+1)}{2}\right)\|\nabla F(x_n)\|^2 + \frac{L\alpha_n^2}{2}A\end{aligned}$$

4. If $\alpha_n \leq c$ and c is small enough, then $\frac{L\alpha_n^2(B+1)}{2} \leq \frac{\alpha_n}{2}$. Hence

$$\mathbb{E}_n(F(x_{n+1})) - F(x_n) \leq -\frac{\alpha_n}{2}\|\nabla F(x_n)\|^2 + \frac{L\alpha_n^2}{2}A$$

5. Apply the full expectation, we obtain

$$\mathbb{E}(F(x_{n+1})) - \mathbb{E}\left(F(x_n)\right) \quad \leq \quad -\frac{\alpha_n}{2}\mathbb{E}\left(\|\nabla F(x_n)\|^2\right) + \frac{C_2}{2}\alpha_n^2$$

sum from $n = 1$ to K all these inequalities to obtain

$$\sum_{n=1}^{K}\alpha_n \mathbb{E}\left(\|\nabla F(x_n)\|^2\right) \leq 2\mathbb{E}(F(x_1)) - 2\mathbb{E}\left(F(x_{K+1})\right) + C_2\sum_{n=1}^{K}\alpha_n^2$$

Use the fact that F is bounded from below, set

$$C = 2\mathbb{E}(F(x_1)) - 2\inf(F(x)),$$

to obtain the required inequality

6. We have

$$\frac{1}{\sum_{n=1}^K \alpha_n}\sum_{n=1}^{K}\alpha_n \mathbb{E}\left(\|\nabla F(x_n)\|^2\right) \leq \frac{1}{\sum_{n=1}^K \alpha_n}(C + C_2\sum_{n=1}^{+\infty} a_n) \to 0$$

7. Just note that

$$\begin{aligned}\frac{1}{\sum_{n=1}^K \alpha_n}\sum_{n=1}^{K}\alpha_n \mathbb{E}\left(\|\nabla F(x_n)\|^2\right) &= \sum_{n=1}^{K}\mathbb{P}(\tau_K = n)\mathbb{E}\left(\|\nabla F(x_n)\|^2\right) \\ &= \mathbb{E}\left(\|\nabla F(x_{\tau_K})\|^2\right)\end{aligned}$$

7 Second order descent methods

This exercise proves the formula for the BFGS algorithm. You will implement this formula.

(TD10 : 30 mins) Exercise 7.1

We want to solve $-B_k \nabla f(x_k)$ if B_k is defined by the BFGS recurence relationship

$$B_k = \left(I - \rho_k \sigma_k y_k^T\right) B_{k-1} \left(I - \rho_k y_k \sigma_k^T\right) + \rho_k \sigma_k \sigma_k^T. \text{ with } \rho_k = \frac{1}{\langle y_k, \sigma_k \rangle}$$

1. Let $q_k = -\nabla f(x_k)$ and define the sequences $(\alpha_i)_{1 \leq i \leq k}$ and $(q_i)_{0 \leq i \leq k}$ by the inverse recurrence for all $i = k, \ldots 1$:

$$\begin{cases} \alpha_i &= \rho_i \langle \sigma_i, q_i \rangle \\ q_{i-1} &= q_i - \alpha_i y_i \end{cases}$$

   Set $z_0 = B_0 q_0$ and compute the sequence $(z_i)_{i=1,k}$ by the following forward recurrence for all $i = 1, \ldots, k$:

$$\begin{cases} \beta_i &= \rho_i \langle y_i, z_{i-1} \rangle \\ z_i &= z_{i-1} + (\alpha_i - \beta_i)\, \sigma_i \end{cases}$$

   Show that for all i, we have $z_i = B_i q_i$
2. If $L = (\sigma_k, y_k, \rho_k)_k$ is saved in a sequence. Justify the following algorithm for computing $-B_k \nabla f(x_k)$:
   (a) $q = -\nabla f(x_k)$ and create an empty list called L_α
   (b) For (σ, y, ρ) in reversed order of L :
      i. Compute $\alpha = \rho \langle \sigma, q \rangle$ and append α to L_α
      ii. Set $q = q - \alpha y$
   (c) Reverse the list of L_α.
   (d) Set $q = B_0 q$.
   (e) For $(\sigma, y, \rho), \alpha$ in (L, L_α) :
      i. Compute $\beta = \rho \langle y, q \rangle$.
      ii. Set $q = q + (\alpha - \beta)\sigma$

Solution of Exercise 7.1

1. We suppose that $z_{i-1} = B_{i-1} q_{i-1}$ and we show that $z_i = B_i q_i$

$$\begin{aligned} B_i q_i &= \left(I - \rho_i \sigma_i y_i^T\right) B_{i-1} \underbrace{\left(I - \rho_i y_i \sigma_i^T\right) q_i}_{=q_{i-1}} + \rho_i \sigma_i \sigma_i^T q_i \\ &= \left(I - \rho_i \sigma_i y_i^T\right) \underbrace{B_{i-1} q_{i-1}}_{=z_{i-1}} + \alpha_i \sigma_i \\ &= z_{i-1} - \beta_i \sigma_i + \alpha_i \sigma_i = z_i \end{aligned}$$

2. The proposed algorihtm computes the sequence q_k in step (b) and the sequence α_k which is only needed in step $(e).ii$. Then the algorithm computes in step (e) the sequence z_k .Remark that the sequence z_k is computed in the same dummy variable q.

Correction of tutorials

8 Convex non-smooth optimization : I

(TD 11 : 20 mins) Exercise 8.1

Let $\mathcal{C}$ be a non-empty convex subset of $\mathbb{R}^n$.

1. Recall the definition of $i_{\mathcal{C}}$, the characteristic function of $\mathcal{C}$
2. Under which condition is $i_{\mathcal{C}}$ lower-semi-continuous ?

Solution of Exercise 8.1

1. The definition of $i_{\mathcal{C}}$ is

$$\forall x \in \mathbb{R}^n, \quad i_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{if not} \end{cases}$$

2. We know that $i_{\mathcal{C}}$ is l.s.c. iff its levelsets are closed. We recall that the levelsets of i_C are the sets $L_\alpha(i_{\mathcal{C}})$ defined by :

$$L_\alpha(i_{\mathcal{C}}) = \{x \in \mathbb{R}^n \text{ such that } i_{\mathcal{C}}(x) \leq \alpha\}$$

We have
- If $\alpha < 0$, then $L_\alpha(i_{\mathcal{C}}) = \emptyset$ wich is closed
- If $\alpha \geq 0$, then $L_\alpha(i_{\mathcal{C}}) = \mathcal{C}$ which is closed iff $\mathcal{C}$ is closed.

Hence $i_{\mathcal{C}}$ is lower-semi-continuous iff $\mathcal{C}$ is closed.

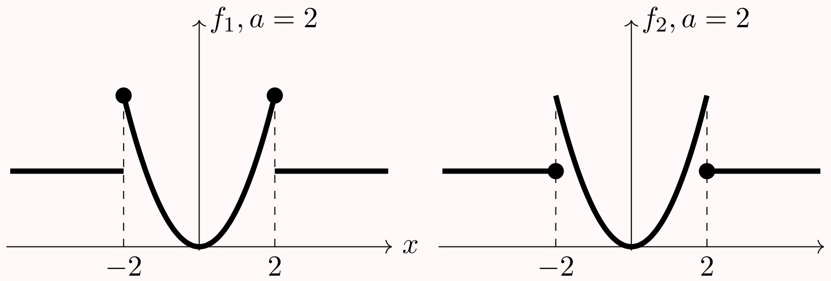
(TD 11 : 40 mins) Exercise 8.2

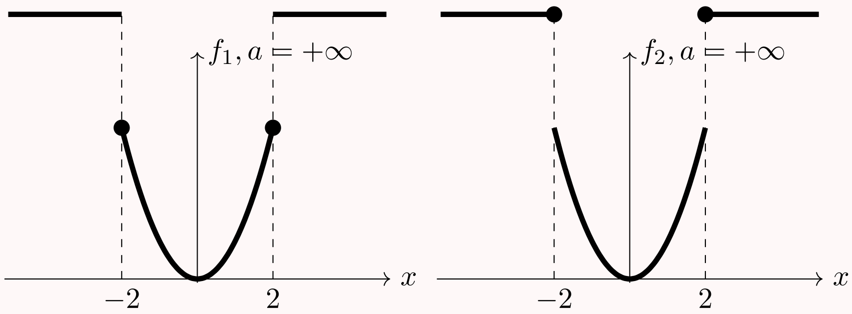
Let $a \in \{+\infty, 2\}$. Are the following functions l.s.c on $\mathbb{R}$?

$$f_1(x) = \begin{cases} x^2 & \text{if } |x| \leq 2 \\ a & \text{if } |x| > 2 \end{cases} \quad \text{and } f_2(x) = \begin{cases} x^2 & \text{if } |x| < 2 \\ a & \text{if } |x| \geq 2 \end{cases}$$

Solution of Exercise 8.2

We sketch the functions





We study the levelsets $L_\alpha(f_i)$ of the functions. $\alpha \in \mathbb{R}$.

1. If $a = +\infty$ In this case, we have:

$$\{x \in \mathbb{R}^n \mid f_1(x) \leq \alpha\} = \begin{cases} [-2, 2] & \text{if } \alpha \geq 4. \\ [-\sqrt{\alpha}, \sqrt{\alpha}] & \text{if } 0 \leq \alpha < 4. \\ \emptyset & \text{if } \alpha < 0. \end{cases}$$

$$\{x \in \mathbb{R}^n \mid f_2(x) \leq \alpha\} = \begin{cases}]-2, 2[& \text{if } \alpha \geq 4. \\ [-\sqrt{\alpha}, \sqrt{\alpha}] & \text{if } 0 \leq \alpha < 4. \\ \emptyset & \text{if } \alpha < 0. \end{cases}$$

Hence the function f_1 is l.s.c on $\mathbb{R}$ whereas f_2 is not.

2. If $a = 2$

In this case, we have :

$$\{x \in \mathbb{R}^n \mid f_1(x) \leq \alpha\} = \begin{cases} \mathbb{R} & \text{if } \alpha \geq 4. \\ ]-\infty, -2[\cup[-\sqrt{\alpha}, \sqrt{\alpha}]\cup]2, +\infty[& \text{if } 2 \leq \alpha < 4. \\ [-\sqrt{\alpha}, \sqrt{\alpha}] & \text{if } 0 \leq \alpha < 2. \\ \emptyset & \text{if } \alpha < 0. \end{cases}$$

$$\{x \in \mathbb{R}^n \mid f_2(x) \leq \alpha\} = \begin{cases} \mathbb{R} & \text{if } \alpha \geq 4. \\ ]-\infty, -2] \cup [-\sqrt{\alpha}, \sqrt{\alpha}] \cup [2, +\infty[& \text{if } 2 \leq \alpha < 4. \\ [-\sqrt{\alpha}, \sqrt{\alpha}] & \text{if } 0 \leq \alpha < 2. \\ \emptyset & \text{if } \alpha < 0. \end{cases}$$

This time, the function f_1 is not l.s.c on $\mathbb{R}$ whereas f_2 is.

(15 mins) Exercise 8.3

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the function defined by:

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & \text{si} \quad x \neq 0 \\ -1 & \text{si} \quad x = 0. \end{cases}$$

1. Show that the function f is l.s.c on $\mathbb{R}$.
2. If $f(0) = 1$, is f still l.s.c ?

Solution of Exercise 8.3

(TD 12 : 60 mins) Exercise 8.4

Consider the following functions:

- $f_1(x) = \frac{1}{2}(x-1)^2 + |3x-2|$, $x \in \mathbb{R}$,
- $f_2(x) = \max(|x|, x^2 - 2)$, $x \in \mathbb{R}$,
- $f_3(x,y,z) = x^2 + xy + y^2 + \frac{1}{2}z + |y-z|$.

1. Show that f_i, $i = 1, \ldots, 3$, is convex and l.s.c on $\mathbb{R}$.
2. For any $x \in \mathbb{R}$, compute the subdifferential $\partial f_1(x)$. Deduce the global minimizer(s) of f_1.
3. For any $x \in \mathbb{R}$, compute the subdifferential $\partial f_2(x)$. Deduce the global minimizer(s) of f_2.
4. For any $M = (x,y,z) \in \mathbb{R}^3$, compute the subdifferential $\partial f_3(M)$. Deduce the global minimizer(s) of f_3.

Solution of Exercise 8.4

1. - The function f_1 is convex as the composition of convex functions with affine functions. Moreover, f_1 is continuous on its domain $dom(f_1) = \mathbb{R}$, hence f_1 is l.s.c.
   - The function f_2 is the maximum of $g_1 : x \mapsto |x|$ and $g_2 : x \mapsto x^2 - 2$ which are two convex l.s.c functions with domain equal to $\mathbb{R}$. Hence f_2 is a l.s.c. convex function with $dom(f_2) = dom(g_1) \cap dom(g_2) = \mathbb{R}$.
   - La fonction $M = (x,y,z) \mapsto |y-z|$ is convex and l.s.c. as the composition of an affine function and the function $|\cdot|$ which is l.s.c and convex. The function $\varphi : (x,y,z) \mapsto x^2 + xy + y^2 + \frac{1}{2}z$ is l.s.c because it is C^∞on $\mathbb{R}^3$. We check the convexity of φ by computing its Hessian :

$$\forall M \in \mathbb{R}^3,\ H_M[\varphi](M) = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

     This matrix has an eigenvalue which is equal to 0 with the two others being strictly positive. Hence φ is convex on $\mathbb{R}^3$. Now the function f_3 is convex l.s.c. as a positive sum of two convex l.s.c functions.
2. The function f_1 is the sum of two convex l.s.c functions, hence:

$$\begin{aligned}
\forall x \in \mathbb{R},\ \partial f_1(x) &= \{x-1\} + 3\partial|.|(3x-2) \\
&= \begin{cases} \{x-1\} + \{3\} & \text{if: } x > \frac{2}{3} \\ \{x-1\} + \{-3\} & \text{if: } x < \frac{2}{3} \\ \{x-1\} + 3[-1,1] & \text{if: } x = \frac{2}{3} \end{cases} \\
&= \begin{cases} \{x+2\} & \text{if: } x > \frac{2}{3} \\ \{x-4\} & \text{if: } x < \frac{2}{3} \\ x-1+3[-1,1] & \text{if: } x = \frac{2}{3} \end{cases}
\end{aligned}$$

   Because f_1 is continuous and obviously coercive on $\mathbb{R}$, we know that f_1 admits at least a global minimizer on $\mathbb{R}$. The necessary and sufficient optimality condition for an $x^\star$ to be a minimizer is

$$0 \in \partial f_1(x^\star).$$

First case: suppose that $x^\star > \frac{2}{3}$. Then the optimality condition is

$$x^\star + 2 = 0, \qquad \text{that is:} \qquad x^\star = -2,$$

which does not fulfills our assumption $x^\star > \frac{2}{3}$. Hence, $x^\star \leq \frac{2}{3}$.

Second case: suppose that $x^\star < \frac{2}{3}$. Then the optimality condition translates into

$$x^\star - 4 = 0, \qquad \text{that is:} \qquad x^\star = 4,$$

which is in contradiction with $x^\star < \frac{2}{3}$.

Last case: suppose that $x^\star = \frac{2}{3}$. We check the optimality condition,

$$\partial f_1(\frac{2}{3}) = \frac{2}{3} - 1 + 3[-1,1] = -\frac{1}{3} + [-3,3] = [-\frac{10}{3}, \frac{8}{3}].$$

And we have $0 \in \partial f_1(\frac{2}{3})$.

The function f_1 admits a unique global minimizer, it is $x^* = \frac{2}{3}$.

3. The function f_2 is the supremum of $|\cdot|$ and $g : x \mapsto x^2 - 2$ which are convex and l.s.c. and whose subdifferential is known, we have:

$$\partial f_2(x) = \begin{cases} \partial|\cdot|(x) & \text{if: } |x| > x^2 - 2, \\ \partial g(x) & \text{if: } |x| < x^2 - 2, \\ conv(\partial|\cdot|(x), \partial g(x)) & \text{if: } |x| = x^2 - 2. \end{cases}$$

We have:

$$|x| = x^2 - 2 \iff x = -2 \quad \text{ or } x = 2,$$

and

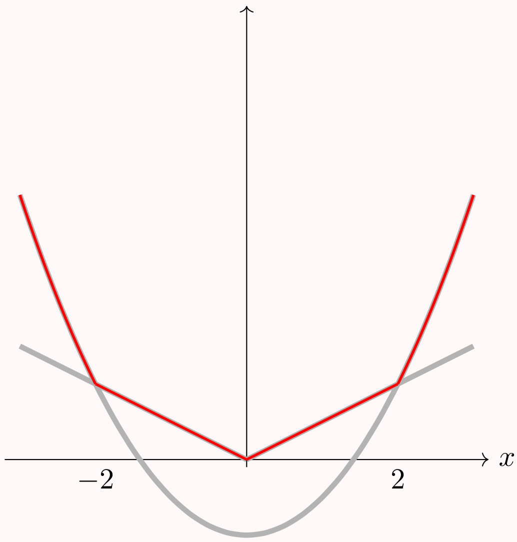
$$x > x^2 - 2 \iff -2 < x < 2.$$

Hence:

$$\partial f_2(x) = \begin{cases} \partial|\cdot|(x) & \text{if: } -2 < x < 2, \\ \partial g(x) & \text{if: } x < -2 \text{ ou } x > 2, \\ conv(\partial|\cdot|(x), \partial g(x)) & \text{if: } x \in \{-2, 2\}. \end{cases}$$

Moreover, the function g is differentiable for any $x \in \mathbb{R}$ with $\partial g(x) = \{2x\}$, so that we have:

$$\partial f_2(x) = \begin{cases} \{2x\} & \text{if: } x < -2, \\ [-4,-1] & \text{if: } x = -2, \\ \{-1\} & \text{if: } -2 < x < 0, \\ [-1,1] & \text{if: } x = 0, \\ \{1\} & \text{if: } 0 < x < 2, \\ [1,4] & \text{if: } x = 2, \\ \{2x\} & \text{if: } x > 2. \end{cases}$$



We now search the global minimizers of f_2 on $\mathbb{R}$, that is we try to find $x^\star \in \mathbb{R}$ such that :

$$0 \in \partial f_2(x^\star).$$

- Do we have $x^\star < -2$ or $x^\star > 2$? In this case the optimality condition is:

$$0 \in \{2x^\star\} \qquad \text{i.e.:} \qquad x^\star = 0,$$

which is in contradiction with $x^\star < -2$ or $x^\star > 2$.
- Do we have $-2 < x^\star < 0$ (resp. $0 < x^\star < 2$) ? The optimality condition is:

$$0 \in \{-1\}, \quad (\text{resp. } 0 \in \{1\})$$

which is impossible.
- Finally we check if $x^\star = -2$, $x^\star = 0$ or $x^\star = 2$. We have

$$\partial f_2(-2) = [-4, -1], \quad \partial f_2(0) = [-1, 1], \quad \partial f_2(2) = [1, 4].$$

Hence, $0 \notin \partial f_2(-2)$, $0 \in \partial f_2(0)$ and $0 \notin \partial f_2(-1)$.

There is only one global minizer of f_2, it is given by $x^\star = 0$ (which is easy to check on a simple sketch).

4. We have, for $M = (x, y, z)$:

$$\partial f_3(M) = \left\{ \begin{pmatrix} 2x + y \\ x + 2y + s \\ \frac{1}{2} - s \end{pmatrix} \; ; \; s \in \partial |\cdot|(y - z) \right\}$$

D'où:

$$\partial f_3(M) = \left\{ \begin{pmatrix} 2x + y \\ x + 2y + 1 \\ -\frac{1}{2} \end{pmatrix} \right\} \text{ if: } y > z$$

$$\partial f_3(M) = \left\{ \begin{pmatrix} 2x + y \\ x + 2y - 1 \\ \frac{3}{2} \end{pmatrix} \right\} \text{ if: } y < z$$

$$\partial f_3(M) = \left\{ \begin{pmatrix} 2x+y \\ x+2y+s \\ \frac{1}{2}-s \end{pmatrix} \; ; \; s \in [-1,1] \right\}$$

The optimality condition is
$$0 \in \partial f_3(M).$$

After carefull inspection of the above computation of $\partial f_3(M)$, any point (x, y, z) that verifies $y \neq z$, cannot be a global minimum of f_3 (because the third component of gradient is nonzero !). In the case $y = z$, the condition reads

$$\left\{ \begin{array}{rcl} 2x+y &=& 0 \\ x+2y &=& -s \\ s &=& \frac{1}{2} \end{array} \right.$$

so that: $x = \frac{1}{6}$, $y = z = -\frac{1}{3}$. The function f_3 admits global minimizer at $(\frac{1}{6}, -\frac{1}{3}, -\frac{1}{3})$.

(TD 13: 60 mins) Exercise 8.5

Define the following minimization problem

$$(P) \qquad \inf_{(x,y)\in\mathbb{R}^2} g(x,y) = \frac{1}{2}(x-2)^2 + |y| \quad \text{such that:} \quad x^2+y^2 \leq 1.$$

And denote $X = \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 \leq 1\}$.

1. Transform the problem (P) in a unconstrained minimization problem :
$$(P') \qquad \inf_{(x,y)\in\mathbb{R}^2} f(x,y) = g(x,y) + h(x,y)$$
where $h : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ is a convex function on $\mathbb{R}^2$ that has to be given.
2. Compute the subdifferential $\partial f(M)$ in each point $M = (x,y) \in dom(f)$.
3. Show that g does not admit minimizers on the interior of X (denoted $\mathring{X}$).
4. Give the global minimizers of g on X.
5. Can we apply the same method to solve the following maximization problem ?
$$\sup_{(x,y)\in\mathbb{R}^2} g(x,y) = \frac{1}{2}(x-2)^2 + |y| \quad \text{such that:} \quad x^2+y^2 \leq 1.$$

Solution of Exercise 8.5

1. We introduce the characteristic function of the set X:
$$i_X(M) = \left\{ \begin{array}{ll} 0 & \text{if } M \in X, \\ +\infty & \text{otherwise.} \end{array} \right.$$
The set X is closed and convex (because the function $(x,y) \mapsto x^2+y^2-1$ is convex and l.s.c), hence i_X is convex and l.s.c and the problem (P) is equivalent to the convex l.s.c problem:
$$(P') \qquad \min_{(x,y)\in\mathbb{R}^2} f(x,y) = g(x,y) + i_X(x,y)$$

2. Because f is the sum of two l.s.c functions and $dom(f) = X$ has non empty interior, we have:

$$\forall M \in X,\ \partial f(M) = \partial g(M) + \partial i_X(M).$$

The subdifferential of g is easy to compute:

$$\partial g(M) = \begin{pmatrix} x-2 \\ \partial|\cdot|(y) \end{pmatrix} = \begin{cases} \left\{\begin{pmatrix} x-2 \\ 1 \end{pmatrix}\right\} & \text{if: } y>0, \\ \left\{\begin{pmatrix} x-2 \\ -1 \end{pmatrix}\right\} & \text{if: } y<0, \\ \begin{pmatrix} x-2 \\ [-1,1] \end{pmatrix} & \text{if: } y=0, \end{cases}$$

the hardest part is to compute $\partial i_X(M)$ which is the normal cone to X at the point $M \in \partial X$. We can use a sketch to find the normal cone

$$\forall M \in \mathring{X},\ \partial i_X(M) = \{0\},$$

and

$$\forall M \in \partial X,\ \partial i_X(M) = \{t\begin{pmatrix} x \\ y \end{pmatrix}\ ;\ t \geq 0\}.$$

We recap our result so far

$$\forall M \in \mathring{X},\ \partial f(M) = \partial g(M) = \begin{cases} \left\{\begin{pmatrix} x-2 \\ 1 \end{pmatrix}\right\} & \text{if: } y>0, \\ \left\{\begin{pmatrix} x-2 \\ -1 \end{pmatrix}\right\} & \text{if: } y<0, \\ \begin{pmatrix} x-2 \\ [-1,1] \end{pmatrix} & \text{if: } y=0, \end{cases}$$

et

$$\forall M \in \partial X,\ \partial f(M) = \begin{cases} \left\{\begin{pmatrix} x-2+tx \\ 1+ty \end{pmatrix}\ ;\ t \geq 0\right\} & \text{if: } y>0, \\ \left\{\begin{pmatrix} x-2+tx \\ -1+ty \end{pmatrix}\ ;\ t \geq 0\right\} & \text{if: } y<0, \\ \left\{\begin{pmatrix} x-2+tx \\ s \end{pmatrix}\ ;\ t \geq 0, s \in [-1,1]\right\} & \text{if: } y=0. \end{cases}$$

3. First remark that g is l.s.c on a X which is a bounded closed set. Hence g admits at least one global minimizer $M^\star$ (because g is u.s.c it also admits a global maximizer). Due to the introduction of $f = g + i_X$, we now that minimizing g over X is equivalent to minimizing the unconstrained problem $\min_{M\in\mathbb{R}^2} f(M)$, we study the optimality condition

$$0 \in \partial f(M^\star),$$

   - Suppose that the constraint is not active at $M^\star = (x^\star, y^\star)$, that is $x^{\star 2} + y^{\star 2} < 1$ or equivalently $M \in \mathring{X}$. We have

$$\partial f(M^\star) = \left\{\begin{pmatrix} x^\star - 2 \\ 1 \end{pmatrix}\right\} \qquad \text{if } y^\star > 0,$$

$$\partial f(x^\star, y^\star) = \left\{ \begin{pmatrix} x^\star - 2 \\ -1 \end{pmatrix} \right\} \qquad \text{if } y^\star < 0,$$

$$\partial f(x^\star, 0) = \left\{ \begin{pmatrix} x^\star - 2 \\ s \end{pmatrix} \; ; \; s \in [-1, 1] \right\}.$$

In the case $y^\star \neq 0$, the second components of the subgradients is non-zero, hence $0 \notin \partial f(x^\star, y^\star)$. In the case $y^\star = 0$, in order to have $0 \in \partial f(x^\star, 0)$, one needs $x^\star = 2$ but $(2, 0) \notin X$.

- Suppose that $M^\star \in \partial X$, that is $x^{\star 2} + y^{\star 2} = 1$ and write $0 \in \partial f(M^\star)$. Suppose $y^\star > 0$. In this case :

$$0 \in \partial f(M^\star) \quad \text{i.e.: exists } t \geq 0, \left\{ \begin{array}{rcl} x^\star - 2 + tx^\star & = & 0 \\ 1 + ty^\star & = & 0 \end{array} \right. ,$$

which is impossible because $1 + ty^\star \geq 1$. Suppose now that $y^\star < 0$. The optimality condition is :

$$0 \in \partial f(M^\star) \quad \text{i.e.: exists } t \geq 0, \left\{ \begin{array}{rcl} x^\star - 2 + tx^\star & = & 0 \\ -1 + ty \star * & = & 0 \end{array} \right. ,$$

which is impossible because $-1 + ty^\star \leq -1$.

- At this point we know that $y^\star = 0$ and that the constraint is active, this implies

$$x^\star = \pm 1.$$

We have

$$\partial f(1, 0) = \left\{ \begin{pmatrix} -1 + t \\ g \end{pmatrix} \; ; \; t \geq 0, g \in [-1, 1] \right\}.$$

And we have $0 \in \partial f(1, 0)$ (take $t = 1 \geq 0$ and $g = 0 \in [-1, 1]$). The point $(1, 0)$ is a global minimizer of g on X Similarly:

$$\partial f(-1, 0) = \left\{ \begin{pmatrix} -3 - t \\ g \end{pmatrix} \; ; \; t \geq 0, g \in [-1, 1] \right\}.$$

For every $t \geq 0$, we have: $-3 - t \leq -3$ and then $0 \notin \partial f(-1, 0)$. The point $(1, 0)$ is not a global minimizer of g on X

The only global minimizer of g on X is $(1, 0)$.

4. The maximisation problem can be transformed into a minimization problem with:

$$\min_{(x,y) \in \mathbb{R}^2} -g(x, y) \quad \text{with:} \quad x^2 + y^2 \leq 1.$$

But we are stopped in our tracks here, because we cannot conclude anything about minimization of a concave function

(KKT Theorem-I : 20 mins) Exercise 8.6

The objective of this exercise is to prove KKT theorem for convex non differentiable functions. We will prove it in two different ways. We aim at minimizing a function f over a set of constraints X which is given by a finite number of inequalities only

$$X = \{x \text{ s.t. } g_i(x) \leq 0, \forall i \in \mathcal{I}\}.$$

We suppose that the functions f and $(g_i)_i$ are convex and that they are finite everywhere (hence they are continuous everywhere). We also suppose that there exists a point $x_0 \in X$ such that $g_i(x_0) < 0$ for every i (this is a variant of Slater's condition). The KKT theorem is :

If $x^\star$ is a global minimizer of f over X, there exists $\lambda \geq 0$ such that $\lambda_i g_i(x^\star) = 0$ for all i and

$$0 \in \partial f(x^\star) + \sum_i \lambda_i \partial g_i(x).$$

1. Denote $\phi(x) = \max_i(f(x) - f(x^\star), g_i(x))$. Prove that $\phi \geq 0$ everywhere and that $\phi(z) = 0$ iff z is a global minimizer of f on X.
2. Show that for any x with $\phi(x) = 0$, we have

$$\partial\phi(x) = \left\{ \alpha\partial f(x) + \sum_{i\in\mathcal{A}_x} \gamma_i \partial g_i(x), \text{ where } \alpha + \sum_{i\in\mathcal{A}_x} \gamma_i = 1 \right\}$$

3. Conclude that if $x^\star$ is a global minimizer of f over X, there exists $\alpha \geq 0$ and $\gamma \geq 0$ such that $\gamma_i g_i(x^\star) = 0$ for all i and $\alpha + \sum_i \gamma_i = 1$

$$0 \in \alpha\partial f(x^\star) + \sum_i \lambda_i \partial g_i(x).$$

4. Let $x^\star$ be a global minimizer and α, γ be as above, we prove that $\alpha \neq 0$, to this end we suppose that $\alpha = 0$. Consider now the function $\psi : x \mapsto \sum_{i\in\mathcal{A}_x} \gamma_i g_i(x)$. Compute $\partial\psi(x)$ and prove that $x^\star$ is a global minimizer of ψ. Prove this is absurd and hence $\alpha \neq 0$.
5. Set $\lambda = \frac{\gamma}{\alpha}$ and prove KKT theorem.

Solution of Exercise 8.6

1. If $x \in X$ then $f(x) \geq f(x^\star)$ and then $\phi(x) \geq 0$. If $x \notin X$, then there exists and i such that $g_i(x) > 0$, hence $\phi(x) > 0$. Hence if $\phi(z) = 0$ then $z \in X$ and $f(z) = f(x^\star)$, that is z is a global minimizer of f on X.
2. The functions f and g_i are finite everywhere, so that their domain is $\mathbb{R}^d$, hence the domain of ϕ is $\mathbb{R}^d$. If $\phi(x) = 0$, then $x \in X$ and $f(x) - f(x^\star) = \phi(x)$. If $\phi(x) = 0$, then $g_i(x) = \phi(x)$ if and only if $i \in \mathcal{A}_x$. Moreover, because the interior of $dom(\phi)$ is $\mathbb{R}^d$, we have, if $\phi(x) = 0$

$$\partial\phi(x) = \mathrm{ConvHull}(\partial f(x), \partial g_i(x))_{i\in\mathcal{A}_x},$$

that is

$$\partial\phi(x) = \left\{ \alpha\partial f(x) + \sum_{i\in\mathcal{A}_x} \gamma_i \partial g_i(x), \text{ where } \alpha + \sum_{i\in\mathcal{A}_x} \gamma_i = 1 \right\}$$

3. If $x^\star$ is a global minimizer of f over X then $\phi(x^\star) = 0$ and $0 \in \partial\phi(x^\star)$, that is there exists $\alpha \geq 0$ and $\gamma \geq 0$ such that

$$0 \in \alpha\partial f(x^\star) + \sum_{i\in\mathcal{A}_{x^\star}} \lambda_i \partial g_i(x), \text{ and } \alpha + \sum_{i\in\mathcal{A}_{x^\star}} \gamma_i = 1.$$

By setting $\gamma_i = 0$ when $i \notin \mathcal{A}_{x^\star}$, or equivalently if we enforce $\gamma_i g_i(x^\star) = 0$ for all i, we can transform the above equation into

$$0 \in \alpha \partial f(x^\star) + \sum_i \lambda_i \partial g_i(x) \text{ and } \alpha + \sum_i \gamma_i = 1.$$

4. We suppose that $\alpha = 0$, we note that ψ is the sum of l.s.c convex functions and that the domain of ψ is $\mathbb{R}^d$, hence the interior of the domain of ψ is $\mathbb{R}^d$, it is non-empty. Because $\alpha = 0$, we have

$$0 \in \sum_i \lambda_i \partial g_i(x) = \partial \psi,$$

and then $x^\star$ is a global minimizer of ψ, but $\psi(x^\star) = 0$ and then we must have $\psi(x_0) = \sum_i \gamma_i g_i(x_0) \geq 0$. But each $g_i(x_0)$ is < 0 and every γ_i is ≥ 0. Hence $\psi(x_0) \geq 0$ implies $\gamma_i = 0$ for every i. But in this case $\alpha + \sum_i \gamma_i = 1 \neq 0$). This leads to a contradiction, and $\alpha \neq 0$.
5. Because $\alpha \neq 0$ we can divide by α.

(Subdifferential of a norm : 0 mins) Exercise 8.7

1. Show that the dual of the norm $\|\cdot\|_1$ is $\|\cdot\|_\infty$.
2. Give the subdifferential of the norm $\|\cdot\|_1$.
3. **Additional question:** Answer the same questions by switching the roles of $\|\cdot\|_\infty$ and $\|\cdot\|_1$.

Solution of Exercise 8.7

1. Par définition, on a:

$$\forall x \in \mathbb{R}^n, \|y\|_{1*} = \sup_{\|x\|_1 \leq 1} \langle x, y\rangle = \sup_{\sum_{i=1}^n |x_i| \leq 1} \sum_{i=1}^n x_i y_i.$$

On remarque alors que:

$$\forall i = 1, \dots, n, \ |y_i| \leq \|y\|_\infty,$$

d'où

$$\begin{aligned} \forall x \in \mathbb{R}^n \text{ tq: } \|x\|_1 \leq 1, \sum_{i=1}^n x_i y_i &\leq \left(\sum_{i=1}^n |x_i|\right) \|y\|_\infty, \\ &\leq \|y\|_\infty \end{aligned}$$

ce qui implique:

$$\forall y \in \mathbb{R}^n, \ \|y\|_{1*} = \sup_{\|x\|_1 \leq 1} \langle x, y\rangle \leq \|y\|_\infty.$$

Cherchons maintenant un x qui réalise l'égalité: soit $i_0 \in \{1, \dots, n\}$ tq: $\|y\|_\infty = |y_{i_0}|$. On choisit le vecteur x tel que:

$$\forall i \neq i_0, \ x_i = 0, \qquad x_{i_0} = sign(y_{i_0}).$$

On a alors:

$$\|x\|_1 = 1 \text{ et } \langle x, y\rangle = x_{i_0} y_{i_0} = \|y\|_\infty.$$

Autrement dit,

$$\forall y \in \mathbb{R}^n, \ \|y\|_{1*} = \sup_{\|x\|_1 \leq 1} \langle x, y\rangle = \|y\|_\infty.$$

2. Par définition, on a:

$$\partial \|\cdot\|_1(x) \quad = \quad \arg \max_{\|y\|_{1*} \leq 1} \langle x, y\rangle = \arg \max_{\|y\|_\infty \leq 1} \langle x, y\rangle.$$

On cherche donc l'ensemble des solutions du problème:

$$\max_{\|y\|_\infty \leq 1} \langle x, y\rangle$$

i.e.:

$$\max_{y \in \mathbb{R}^n} \sum_{i=1}^n x_i y_i \qquad \text{sous} \qquad |y_i| \leq 1, \ i = 1, \dots, n.$$

Le problème est séparable: il suffit donc de résoudre les problèmes:

$$\max_{y_i \in \mathbb{R}} x_i y_i \qquad \text{sous} \qquad |y_i| \leq 1$$

pour $i = 1, \dots, n$, dont l'unique solution est:

$$y_i^* = sign(x_i)$$

si $x_i \neq 0$. Dans le cas où $x_i = 0$, alors l'ensemble des solutions est $[-1, 1]$. Dans les deux cas, on a donc:

$$\langle x, y^*\rangle = \sum_{i=1}^n |x_i| = \|x\|_1.$$

Au final:

$$\begin{aligned} \partial \|\cdot\|_1(x) &= \{y \in \mathbb{R}^n \mid \langle x, y\rangle = \|x\|_1\}. \\ &= \left\{ y \in \mathbb{R}^n \mid \forall i = 1, \dots, n, \ \begin{cases} y_i = sign(x_i) & \text{si } x_i \neq 0, \\ y_i \in [-1, 1] & \text{si } x_i = 0 \end{cases} \right\} \end{aligned}$$

3. To be done

9 Convex non-smooth optimization : II

(Preparing labwork, TD 14 : 60 mins) Exercise 9.1

Let $\gamma > 0$, the objective is to compute the proximal operator to $x \mapsto \gamma \|x\|_1$.

1. Compute the proximal operator of the function $h : x \mapsto \gamma |x|$, $x \in \mathbb{R}$.
2. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be defined as:

$$f(x_1, \dots, x_n) = f_1(x_1) + f_2(x_2) + \dots + f_n(x_n),$$

where the functions $f_i : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ are all convex and l.s.c.

(a) Show that:

$$\partial f(x) = \partial f_1(x_1) \times \partial f_2(x_2) \times \cdots \times \partial f_n(x_n).$$

(b) Deduce that:

$$prox_f(x) = (prox_{f_1}(x_1), prox_{f_2}(x_2), \ldots, prox_{f_n}(x_n)).$$

3. Using the previous questions, compute the proximal operator of the function $f : x \mapsto \gamma \|x\|_1$.

Solution of Exercise 9.1

1. On cherche à calculer $\text{prox}_h(x)$, $x \in \mathbb{R}$: soit $x \in \mathbb{R}$ fixé, on veut résoudre le problème suivant:

$$(P) \qquad \min_{y \in \mathbb{R}} \gamma |y| + \frac{1}{2}(y-x)^2.$$

Remarquons que le problème (P) est convexe. La CNS d'optimalité est donc:

$$0 \in \gamma \partial |\cdot|(y) + y - x,$$

soit:

$$y \in x - \gamma \partial |\cdot|(y).$$

Cherchons s'il existe des solutions $y > 0$: dans ce cas, la condition d'optimalité nous donne la solution:

$$y = x - \gamma,$$

à condition que: $x - \gamma > 0$. Sinon pas de solution strictement négative.
Existe-t-il des solutions $y < 0$? dans ce cas, la condition d'optimalité nous donne la solution:

$$y = x + \gamma,$$

à condition que: $x + \gamma < 0$. Sinon pas de solution strictement positive.
Regardons enfin si $y = 0$ est solution: ce sera le cas si la CNS d'optimalité est satisfaite i.e. si:

$$0 \in x - \gamma[-1, 1] = [x - \gamma, x + \gamma].$$

Autrement dit, $y = 0$ est solution de (P) ssi:

$$-\gamma \leq x \leq \gamma.$$

En résumé, on a donc:

$$\text{prox}_h(x) = \begin{cases} x + \gamma & \text{si: } x < -\gamma \\ x - \gamma & \text{si: } x > \gamma \\ 0 & \text{si: } -\gamma \leq x \leq \gamma. \end{cases}$$

ce qui s'écrit de façon plus synthétique sous la forme:

$$\text{prox}_h(x) = \text{sign}(x) \max(|x| - \gamma, 0).$$

(a) On va raisonner par double inclusion.

i. Soit $g = (g_1, \ldots, g_n) \in \partial f_1(x_1) \times \partial f_2(x_2) \times \cdots \times \partial f_n(x_n)$. Cela signifie que:
$$\forall i = 1, \ldots, n,\ g_i \in \partial f_i(x_i),$$
i.e.:
$$\forall i = 1, \ldots, n,\ \forall y \in \mathbb{R},\ f_i(y) \geq f_i(x_i) + g_i(y - x_i).$$
Montrons que $g \in \partial f(x)$.
$$\begin{aligned} \forall y \in \mathbb{R}^n,\ f(y) &= \sum_{i=1}^{n} f_i(y_i) \\ &\geq \sum_{i=1}^{n} (f_i(x_i) + g_i(y_i - x_i)) = \underbrace{\sum_{i=1}^{n} f_i(x_i)}_{=f(x)} + \underbrace{\sum_{i=1}^{n} g_i(y_i - x_i)}_{=\langle g, y-x\rangle} \\ &\geq f(x) + \langle g, y - x\rangle \end{aligned}$$
D'où: $g \in \partial f(x)$, et par suite:
$$\partial f(x) \supset \partial f_1(x_1) \times \partial f_2(x_2) \times \cdots \times \partial f_n(x_n).$$

ii. Soit $g = (g_1, \ldots, g_n) \in \partial f(x)$. Soit $i \in \{1, \ldots, n\}$. Montrons que $g_i \in \partial f_i(x_i)$.
Par définition de $\partial f(x)$:
$$\forall y \in \mathbb{R}^n,\ f(y) \geq f(x) + \langle g, y - x\rangle. \tag{1}$$
En particulier pour $y = (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)$, $y_i \in \mathbb{R}$, on a:
$$\begin{aligned} f(y) &= \sum_{j=1, j\neq i}^{n} f_j(x_j) + f_i(y_i) \\ f(x) &= \sum_{j=1, j\neq i}^{n} f_j(x_j) + f_i(x_i), \\ \langle g, y - x\rangle &= g_i(y_i - x_i) \end{aligned}$$
d'où (1) s'écrit:
$$\forall y_i \in \mathbb{R},\ f_i(y_i) \geq f_i(x_i) + g_i(y_i - x_i).$$
Autrement dit, on a bien: $g_i \in \partial f_i(x_i)$ d'où:
$$\partial f(x) \subset \partial f_1(x_1) \times \partial f_2(x_2) \times \cdots \times \partial f_n(x_n).$$

(b) On rappelle que calculer $prox_f(x)$ revient à chercher les solutions optimales du problème:
$$\min_{y \in \mathbb{R}^n} f(y) + \frac{1}{2}\|y - x\|^2,$$
$$\begin{aligned} y = prox_f(x) &\Leftrightarrow 0 \in \partial f(y) + y - x, \\ &\Leftrightarrow 0 \in \partial f_1(y_1) \times \cdots \times \partial f_n(y_n) + y - x, \\ &\Leftrightarrow \forall i = 1, \ldots, n,\ 0 \in \partial f_i(y_i) + y_i - x_i, \\ &\Leftrightarrow \forall i = 1, \ldots, n,\ y_i = prox_{f_i}(x_i), \end{aligned}$$
d'où le résultat attendu.

2. On cherche à calculer l'opérateur proximal associé à la fonction $f(x) = \gamma\|x\|_1$. On observe que:

$$\forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n,\ f(x) = \sum_{i=1}^{n} \gamma|x_i| = \sum_{i=1}^{n} h(x_i),$$

où h est la fonction définie à la question 1: elle est bien convexe s.c.i. sur $\mathbb{R}$. D'après la question 2, on a donc:

$$prox_f(x) \quad = \quad (prox_h(x_1), prox_h(x_2), \ldots, prox_h(x_n)),$$

avec: $prox_h(x_i) = sign(x_i)\max(|x_i| - \gamma, 0)$.

(TD 15 : 60 mins) Exercise 9.2

Compute the proximal operators of the following functions

$$\begin{aligned}
(a)\ f : x &\mapsto x_1^2 - 3x_1x_2 + 3x_2^2,\ x = (x_1, x_2) \in \mathbb{R}^2,\\
(b)\ g : x &\mapsto \frac{1}{2}\|Ax - b\|^2,\ x \in \mathbb{R}^n,\\
(c)\ h : x &\mapsto \max(x^2 - 1, 0)\ x \in \mathbb{R}.
\end{aligned}$$

Solution of Exercise 9.2

(a) f est un polynôme de degré 2 en x, donc de classe C^∞ sur $\mathbb{R}^2$ qui est son domaine. Donc f est s.c.i. sur $\mathbb{R}^2$. On peut calculer la hessienne de f pour vérifier que f est convexe (et même strictement convexe) sur $\mathbb{R}^2$.

$$\begin{aligned}
prox_f(x) &= \arg\min_{y\in\mathbb{R}^2} f(y) + \frac{1}{2}\|y - x\|^2\\
&= \arg\min_{y\in\mathbb{R}^2} \left(y_1^2 - 3y_1y_2 + 3y_2^2 + \frac{1}{2}(y_1 - x_1)^2 + \frac{1}{2}(y_2 - x_2)^2\right).
\end{aligned}$$

Le problème étant un problème d'optimisation convexe différentiable, la CNS d'optimalité s'écrit:

$$\nabla f(y) + y - x = 0$$

soit:

$$\begin{cases} 2y_1 - 3y_2 + y_1 - x_1 &= 0\\ -3y_1 + 6y_2 + y_2 - x_2 &= 0 \end{cases}$$

d'où:

$$prox_f(x) = (\frac{7x_1 + 3x_2}{12}, \frac{x_1 + x_2}{4}).$$

(b) La fonction g est convexe s.c.i. sur $\mathbb{R}^n$, donc l'opérateur proximal est bien défini. Soit $x \in \mathbb{R}^n$. On cherche à calculer $prox_g(x)$ i.e. à résoudre le problème:

$$\min_{y\in\mathbb{R}^n} \frac{1}{2}\|Ay - b\|^2 + \frac{1}{2}\|y - x\|^2.$$

La condition d'optimalité s'écrit:

$$0 \in \{A^\top(Ay - b) + y - x\} \Leftrightarrow (I_n + A^\top A)y = x + A^\top b,$$

soit: $y = (I_n + A^\top A)^{-1}(x + A^\top b)$. On a donc:

$$prox_g(x) = (I_n + A^\top A)^{-1}(x + A^\top b).$$

(c) La fonction h est convexe s.c.i., donc l'opérateur proximal est bien défini. Soit $x \in \mathbb{R}$. On cherche à calculer $prox_h(x)$ i.e. à résoudre le problème:

$$\min_{y \in \mathbb{R}} h(y) + \frac{1}{2}\|y - x\|^2.$$

On a:

$$\partial h(y) = \begin{cases} 2y & \text{si} \quad |y| > 1, \\ 0 & \text{si} \quad |y| < 1, \\ 2y[0,1] & \text{si} \quad |y| = 1. \end{cases}$$

La condition d'optimalité s'écrit:

$$0 \in \partial h(y) + y - x.$$

- Existe-t-il des solutions y telles que $|y| > 1$? dans ce cas, la condition d'optimalité devient:

  $$2y + y - x = 0, \qquad \text{soit: } y = \frac{x}{3},$$

  Le point $y = \frac{x}{3}$ est donc solution à condition que $|x| > 3$.
- Existe-t-il des solutions y telles que $|y| < 1$? dans ce cas, la condition d'optimalité devient:

  $$y - x = 0, \qquad \text{soit: } y = x,$$

  Le point $y = x$ est donc solution à condition que $|x| < 1$.
- Le point $y = 1$ est-il solution ? on regarde la condition d'optimalité qui s'écrit alors:

  $$0 \in [0,2] + 1 - x, \qquad \text{soit:} \qquad x \in [1,3].$$

  Le point $y = 1$ est donc solution à condition que $1 \leq x \leq 3$.
- Le point $y = -1$ est-il solution ? on regarde la condition d'optimalité qui s'écrit alors:

  $$0 \in [-2,0] - 1 - x, \qquad \text{soit:} \qquad x \in [-3,-1].$$

  Le point $y = -1$ est donc solution à condition que $-3 \leq x \leq -1$.

En résumé, on a:

$$prox_h(x) = \begin{cases} \frac{x}{3} & \text{si} \quad x < -3 \text{ ou } x > 3, \\ x & \text{si} \quad -1 < x < 1, \\ 1 & \text{si} \quad 1 \leq x \leq 3, \\ -1 & \text{si} \quad -3 \leq x \leq -1, \end{cases}$$

(Minimization of the difference of convex functions, TD16 : 60 mins) Exercise 9.3

Let g and h be two convex l.s.c functions from $\mathbb{R}^n$ into $\mathbb{R}$. We consider the following problem:

$$(P) \qquad \text{Minimize } f(x) = g(x) - h(x), \ x \in \mathbb{R}^n.$$

1. Without additional hypothesis, is the problem (P) a convex problem.
2. We say that x is a T-critical point of f iff:

$$\partial g(x) \cap \partial h(x) \neq \emptyset.$$

Show that if $x^\star$ is a global minimizer of f, then:
   (i) $\partial h(x^\star) \subset \partial g(x^\star)$.
   (ii) $x^\star$ is a T-critical point of f.
3. From now, on we suppose that both g and h are strongly convex on $\mathbb{R}^n$. We are interested in the following algorithm: start with any $x_0 \in \mathbb{R}^n$. If x_k is given,
   - Pick up $s_k \in \partial h(x_k)$.
   - x_{k+1} is the solution to the following problem:

$$(P_k) : \min_{x \in \mathbb{R}^n} g(x) - \langle s_k, x\rangle.$$

   (a) Show that the problem (P_k) is strongly convex.
   (b) Write down the optimality condition for (P_k) which is satisfied for x_{k+1}.
   (c) Show that for any $k \in \mathbb{N}$,

$$g(x_k) + h(x_{k+1}) \geq g(x_{k+1}) + h(x_k) + \frac{c_g + c_h}{2}\|x_{k+1} - x_k\|^2,$$

   where c_g and c_h are the constants of strong convexity of g and h respectively.
   (d) En déduire que l'algorithme proposé est bien un algorithme de descente et que, à condition que f soit bornée inférieurement, on a:

$$\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\|^2 < +\infty.$$

   Pour conclure, si les suites $(x_k)_{k\in\mathbb{N}}$ et $(s_k)_{k\in\mathbb{N}}$ sont bornées, on peut montrer que tout point d'accumulation $\bar{x}$ de la suite $(x_k)_{k\in\mathbb{N}}$, alors $\bar{x}$ est un point T-critique de (P) (mais c'est un peu technique...).

Solution of Exercise 9.3

1. Soit $g = 0$ et $h(x) = x^2$ sur $\mathbb{R}$. Les fonctions g et h sont convexes mais la fonction $f = g - h$ définie par: $f(x) = -x^2$, $x \in \mathbb{R}$, ne l'est pas.
2. (i) Soit $s \in \partial h(\bar{x})$. Montrons que $s \in \partial g(\bar{x})$.

$$\begin{aligned} s \in \partial h(\bar{x}) \quad &\Rightarrow \quad \forall x \in \mathbb{R}^n, \ h(x) \geq h(\bar{x}) + \langle s, x - \bar{x}\rangle \\ &\Rightarrow \quad \forall x \in \mathbb{R}^n, \ f(x) + h(x) \geq f(x) + h(\bar{x}) + \langle s, x - \bar{x}\rangle \quad (2) \end{aligned}$$

   Or sachant que $\bar{x}$ est un point de minimum global de f, on a également:

$\forall x \in \mathbb{R}^n, f(x) - f(\bar{x}) \geq 0$, d'où:

$$\forall x \in \mathbb{R}^n, \underbrace{f(x) + h(x)}_{=g(x)} \geq \underbrace{f(\bar{x}) + h(\bar{x})}_{=g(\bar{x})} + \langle s, x - \bar{x}\rangle$$

On en déduit alors:

$$\forall x \in \mathbb{R}^n,\ g(x) \geq g(\bar{x}) + \langle s, x - \bar{x}\rangle.$$

D'où: $s \in \partial g(\bar{x})$.

(ii) D'après la question précédente, on a:

$$\partial g(\bar{x}) \cap \partial h(\bar{x}) = \partial h(\bar{x}).$$

Or h est supposée convexe s.c.i. ce qui implique que son sous-différentiel est non vide. Donc

$$\partial g(\bar{x}) \cap \partial h(\bar{x}) = \partial h(\bar{x}) \neq \emptyset.$$

Autrement dit, $\bar{x}$ est bien un point T-critique de f.

(a) La fonction g est convexe et la fonction $x \mapsto -\langle s_k, x\rangle$ est affine. Donc leur somme est convexe.

(b) Puisque g est convexe et que le problème (P_k) est convexe, on peut écrire:

$$0 \in \partial g(x_{k+1}) - s_k, \qquad \text{i.e.:} \qquad s_k \in \partial g(x_{k+1}).$$

(c) Rappelons la définition de la forte convexité: pour tout $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, on a:

$$\begin{aligned}
\forall s \in \partial g(x),\ g(y) &\geq g(x) + \langle s, y - x\rangle + \frac{c}{2}\|y - x\|^2,\\
\forall t \in \partial h(x),\ h(y) &\geq h(x) + \langle t, y - x\rangle + \frac{d}{2}\|y - x\|^2,
\end{aligned}$$

Appliquons la forte convexité de g aux points $x = x_{k+1}$ et $y = x_k$ pour le sous-gradient $s_k \in \partial g(x_{k+1})$ (question 3(b)):

$$g(x_k) \geq g(x_{k+1}) + \langle s_k, x_k - x_{k+1}\rangle + \frac{c}{2}\|x_{k+1} - x_k\|^2.$$

Appliquons la forte convexité de h aux points $x = x_k$ et $y = x_{k+1}$ pour le sous-gradient $s_k \in \partial h(x_k)$ (par construction):

$$h(x_{k+1}) \geq g(x_k) + \langle s_k, x_{k+1} - x_k\rangle + \frac{d}{2}\|x_{k+1} - x_k\|^2.$$

En sommant les deux dernières inégalités, on obtient l'inégalité cherchée:

$$g(x_k) + h(x_{k+1}) \geq g(x_{k+1}) + h(x_k) + \frac{c+d}{2}\|x_{k+1} - x_k\|^2.$$

(d) On a démontré à la question précédente que:

$$\forall k,\ g(x_k) + h(x_{k+1}) \geq g(x_{k+1}) + h(x_k) + \frac{c+d}{2}\|x_{k+1} - x_k\|^2,$$

soit:

$$\forall k,\ 0 \leq \frac{c+d}{2}\|x_{k+1} - x_k\|^2 \leq f(x_k) - f(x_{k+1}).$$

On en déduit donc:

$$\forall k,\ f(x_k) - f(x_{k+1}) \geq 0.$$

L'algorithme proposé est donc bien un algorithme de descente et en sommant les inégalités, on a également:

$$\forall N \in \mathbb{N},\ \sum_{k=0}^{N} \|x_{k+1} - x_k\|^2 \leq \frac{2}{c+d}(f(x_0) - f(x_{N+1})).$$

En supposant que les courbes de niveau de f sont bornées (ou que f est bornée inférieurement), on obtient:

$$\forall N \in \mathbb{N},\ \sum_{k=0}^{N} \|x_{k+1} - x_k\|^2 < +\infty.$$

10 Previous Exams

(CC1-2024-2025 : 45 mins) Exercise 10.1

Soit $a > 0$, on veut résoudre

$$(P) \qquad \inf_{x \in X} ax_1 - x_1x_2$$

avec: $X = \left\{x = (x_1, x_2) \in \mathbb{R}^2 \text{ et } x_1 \geq 0 \text{ et } x_2 \geq 0 \text{ et } x_1 + x_2^2 = 3\right\}$. *Pour information, il s'agit d'un problème en économie connu sous le nom "d'optimisation de la courbe d'offre". Il y a une demande de h unités d'un produit, x_1 est la quantité de produit que vous allez produire et x_2 est le prix auquel il est vendu. Le coût de production de x_1 unités est ax_1. De plus on sait que pour un prix x_2, les concurrents sont capable d'offrir αx_2^2 unités de produit. Dans cet exercice, on a choisi: $\alpha = 1$ et $h = 3$.*

1. Ecrire le problème sous forme standard et montrer qu'il n'est pas convexe.
2. **Existence d'un point de minimum global.**
   (a) Montrer que X est fermé borné.
   (b) En déduire l'existence d'un point de minimum global du problème (P).
3. Montrer que les trois contraintes ne peuvent pas être nulles/actives en même temps, et montrer que les contraintes sont qualifiées en tout point.
4. Ecrire le Lagrangien et les conditions vérifiées par les points de KKT.
5. On cherche tout d'abord les points de KKT de la forme $(0, x_2)$ (i.e. $x_1 = 0$).
   (a) Montrer que $(0, \bar{x}_2)$ soit un point de KKT de (P) si et seulement si $a^2 \geq 3$, et calculer $\bar{x}_2$.
   (b) Dans le cas où $a^2 > 3$, étudier les conditions d'ordre 2 en ce point de KKT montrer qu'il est un minimum local.
   Pour information, l'interprétation économique est la suivante : votre prix de production est a et si $\alpha a^2 \geq h$, alors les concurrents sont capable de

saturer le marché pour ce prix là, il n'y a pas de place pour vous et vous produisez une quantité $x_1 = 0$ de marchandise.

6. On recherche les points de KKT où $x_1 \neq 0$. Montrer qu'il en existe un et qu'il vérifie $x_1 = 3 - x_2^2$ et $x_2 = \frac{a}{3} + \sqrt{\frac{a^2}{9} + 1}$. *Ici on n'étudie pas les conditions de second ordre pour conclure sur l'optimalité du point trouvé.*
7. On revient maintenant au cas $a^2 < 3$. Sans calcul supplémentaire, donner le point de minimum global de (P) dans ce cas.

Solution of Exercise 10.1

1. On introduit $g_1 : x \mapsto -x_1$ et $g_2 : x \mapsto -x_2$ et $g_3 : x \mapsto x_1 + x_2^2 - 3$ et $f : x \mapsto ax_1 - x_1x_2$. Pour montrer que le problème n'est pas convexe, le mieux est de montrer que f n'est pas convexe en calculant sa Hessienne qui vaut $H[f](x) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$ et dont le déterminant est < 0. Attention, ici g_3 n'est pas convexe mais cela n'entraîne pas forcément que X n'est pas convexe.
2. Ici il faut faire attention au fait que l'ensemble des x tel que $g_3(x) = 0$ n'est pas borné. Par contre en utilisant $x_1 \geq 0$ et $x_2 \geq 0$ alors on a forcément $0 \leq x_1 \leq 3$ et $0 \leq x_2 \leq \sqrt{3}$. Ainsi X est borné et comme les fonctions g_i sont continues alors X est fermé et comme la fonction à minimiser est continue alors il existe un min global.
3. Si $x_1 = x_2 = 0$ alors $g_3(x) \neq 0$, donc les trois contraintes ne peuvent être nulles en même temps. Les trois gradients sont respectivement $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ et $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ et $\begin{pmatrix} 1 \\ 2x_2 \end{pmatrix}$. Quand on en prend au plus 2 parmi les trois, ils sont forcément linéairement indépendants (notons que si on prend ∇g_1 et ∇g_3 alors forcément $x_2 \neq 0$ car les trois contraintes ne peuvent être nulles en même temps). Les contraintes sont qualifiées pour LICQ.
4. $\mathcal{L}(x, \lambda) = ax_1 - x_1x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \lambda_3(x_1 + x_2^2 - 3)$ et KKT est

$$\begin{cases} a - x_2 - \lambda_1 + \lambda_3 & = 0 \\ -x_1 - \lambda_2 + 2\lambda_3 x_2 & = 0 \\ x_1 + x_2^2 - 3 & = 0 \\ \lambda_1 x_1 = 0 & \lambda_1 \geq 0 \\ \lambda_2 x_2 = 0 & \lambda_2 \geq 0 \end{cases}$$

   Ici il faut faire très attention à ne pas poser $\lambda_3 \geq 0$.
5. (a) Dans le cas $x_1 = 0$, alors $x_2 = \sqrt{3}$, $\lambda_2 = 0$ puis $\lambda_3 = 0$ et finalement on doit avoir $\lambda_1 = a - x_2$. Il ne reste plus qu'à vérifier $\lambda_1 \geq 0$. Donc $a \geq x_2$ est bien une condition nécessaire et suffisante pour avoir un point de KKT en $x_1 = 0$.
   (b) On a

$$H\mathcal{L}(x, \lambda) = \begin{pmatrix} 0 & -1 \\ -1 & 2\lambda_3 \end{pmatrix}.$$

   Au point qui nous intéresse, on ainsi $H\mathcal{L}(M) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$ et le cône linéarisé est l'ensemble des directions d telles que $(d|\nabla g_3) = 0$ et $(d|\nabla g_1) =$

0 (ou ≤ 0 dans le cas $a = x_2$), le cône linéarisé est réduit à $\{0\}$ donc les conditions suffisantes d'ordre 2 sont vérifiées.

6. Dans le cas $x_1 \neq 0$, alors $x_2 \neq 0$ (sinon $x_1 = -\lambda_2$ ce qui est impossible) et donc $\lambda_1 = \lambda_2 = 0$. On a

$$\begin{cases} a - x_2 + \lambda_3 & = 0 \\ -x_1 + 2\lambda_3 x_2 & = 0 \\ x_1 + x_2^2 - 3 & = 0 \end{cases} \Rightarrow \begin{cases} \lambda_3 & = x_2 - a \\ x_2^2 - 3 + 2(x_2 - a)x_2 & = 0 \\ x_1 & = 3 - x_2^2 \end{cases}$$

pour trouver x_2 il suffit de trouver les racines positives du polynôme $3x_2^2 - 2ax_2 - 3$ qui sont données par $\frac{a}{3} \pm \sqrt{\frac{a^2}{9} + 1}$, il n'y en a qu'une positive.

7. Dans le cas $a^2 < 3$, il n'y a qu'un point de KKT possible. c'est celui pour $x_1 \neq 0$. C'est donc le minimum global car on sait que ce minimum existe.

(CC1-2024-2025 : 45 mins) Exercise 10.2

Soit $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ une matrice symétrique définie positive, $a \in \mathbb{R}^n$ avec $a \neq 0$ et $c \in \mathbb{R}$, on cherche à résoudre le problème suivant

$$(P_2) \qquad \inf_{x \in X} \langle a, x \rangle \text{ avec } X = \left\{ x \in \mathbb{R}^n \text{ tel que } \frac{1}{2}\langle Qx, x \rangle \leq c \text{ et } \langle a, x \rangle \leq 0 \right\}$$

1. *(0.5 pt)* Que se passe-t-il si on prend $c < 0$? et $c = 0$?
   Dans la suite, on suppose: $c > 0$.
2. **Quelques propriétés de (P_2).**
   (a) *(0.5 pt)* Démontrer l'existence d'un point de minimum global.
   (b) *(0.5 pt)* Montrer que le problème (P_2) est convexe.
   (c) *(0.5 pt)* Montrez que les contraintes sont qualifiées en tout point.
3. **Résolution par KKT.**
   (a) *(1 pt)* Ecrire le Lagrangien du problème et les équations de KKT.
   (b) *(3 pt)* Résoudre les équations de KKT, montrer qu'il n'y a qu'un seul point de minimum global au problème donné. *Indication, vous devrez montrer que le minimum global est (x, λ) avec*

$$x = -\frac{1}{\lambda_1} Q^{-1} a, \quad \lambda = (\lambda_1, 0) \quad \lambda_1 = \sqrt{\frac{\langle Q^{-1}a, a \rangle}{2c}}$$

.

4. **Résolution par l'approche duale.**
   (a) *(2 pt)* Ecrire la fonction duale.
   (b) *(2 pt)* Résoudre le problème dual.
   (c) *(1 pt)* Pourquoi est-ce que la résolution du problème dual permet de résoudre le problème primal sans résoudre les points de KKT ?

Solution of Exercise 10.2

1. Comme Q est définie positive, si on prend $c < 0$, alors X est vide. Si on prend $c = 0$ alors X se réduit à $\{0\}$.
2. (a) La matrice Q est définie positive et avec $\langle Qx, x \rangle \geq \lambda_{min} \|x\|^2$ où λ_{min} est la plus petite valeur propre de Q, on montre que X est borné. Comme

les fonctions $x \mapsto \langle Qx, x\rangle$ et $x \mapsto \langle a, x\rangle$ sont continues alors X est fermé. De plus on minimise une fonction continue. Ainsi il existe un mininmum global.

(b) Comme les fonctions $x \mapsto \langle Qx, x\rangle$ et $x \mapsto \langle a, x\rangle$ sont convexes alors X est convexe. De plus on minimise une fonction convexe. Le problème est convexe.

(c) Les contraintes sont convexes, par Slater il suffit de trouver un point tel que $\frac{1}{2}\langle Qx, x\rangle < c$ et $\langle a, x\rangle \leq 0$ on prend $x = ta$ avec t négatif et très petit ($x = 0$ convient aussi).

3. (a) Le Lagrangien est

$$\mathcal{L}(x, \lambda) = \langle a, x\rangle(1 + \lambda_2) + \lambda_1(\frac{1}{2}\langle Qx, x\rangle - c)$$

Les équations de KKT sont

$$\begin{cases} a(1+\lambda_2) + \lambda_1 Qx & = 0 \\ \lambda_2\langle a, x\rangle = 0 & \lambda_2 \geq 0 \\ \lambda_1(\frac{1}{2}\langle Qx, x\rangle - c) = 0 & \lambda_1 \geq 0 \end{cases}$$

(b) Le cas $\lambda_1 = 0$ est impossible sinon $a = 0$. On a donc $x = -tQ^{-1}a$ avec $t = \frac{1+\lambda_2}{\lambda_1}$. Ensuite $0 = \langle a, x\rangle = t\langle a, Q^{-1}a\rangle$ entraîne $t = 0$ ce qui donne $\lambda_2 = -1$ ce qui est impossible. On a donc $\langle a, x\rangle \neq 0$ et donc $\lambda_2 = 0$. Finalement on a $x = \frac{-1}{\lambda_1}Q^{-1}a$. En remplaçant dans $\frac{1}{2}\langle Qx, x\rangle = c$ ont trouve $\lambda_1 = \sqrt{\frac{\langle Q^{-1}a, a\rangle}{2c}}$

4. (a) La résolution de $\inf_x \mathcal{L}(x, \lambda)$ donne $x = -tQ^{-1}a$ avec $t = \frac{1+\lambda_2}{\lambda_1}$. On obtient ensuite

$$f^\star(\lambda) = -\langle Q^{-1}a, a\rangle\frac{(1+\lambda_2)^2}{2\lambda_1} - c\lambda_1$$

(b) On résout d'abord $\sup_\lambda f^\star(\lambda)$ en λ_2, ce qui donne $\lambda_2 = 0$, puis on résout

$$\sup_{\lambda_1}\langle Q^{-1}a, a\rangle\frac{1}{2\lambda_1} - c\lambda_1$$

Ce qui se résout en $\lambda_1 = \sqrt{\frac{\langle Q^{-1}a, a\rangle}{2c}}$.

(c) Comme le problème est convexe et les contraintes qualifiées, le min global correspond à un point-selle, il n'y a qu'une seule solution au problème dual, on a trouvé les multiplicateurs de Lagrange $\lambda^\star$ et on a un seul point tel que $f^\star(\lambda^\star) = \mathcal{L}(x^\star, \lambda^\star)$, on a donc trouvé le point-selle.

(CC2-2024-2025 : 90 mins) Exercise 10.3

Exercice I:

1. Soit $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ une matrice symétrique définie positive, $a \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

(a) On note f la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ donnée par

$$f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle a, x\rangle.$$

Montrer que f est coercive sur $\mathbb{R}^n$

(b) Montrer que $X = \{x \text{ tel que } \frac{1}{2}\langle Qx, x\rangle \leq c\}$ est borné.

2. Montrer que les contraintes sont qualifiées en tout point pour les ensembles X suivants
   (a) $X = \{(x, y) \text{ tel que } x^2 - y^4 - 1 \leq 0 \text{ et } x \geq 0\}$
   (b) $X = \{x \in \mathbb{R}^n \text{ tel que, pour tout } i, \sum_{j=1}^n x_j^2 \leq x_i + 1\}$

Exercice II: Soit le problème :

$$\max_{(x,y,z)\in\mathbb{R}^3} f(x, y, z) := xyz \quad \text{sous} \quad x + y + z = 3.$$

1. Trouver les points de KKT.
2. En étudiant les conditions d'ordre 2, montrer que le point $(1, 1, 1)^T$ est maximiseur local. *Indication : En notant* $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$,

$$\forall d \in \mathbb{R}^3, \ si \ \left\langle d, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle = 0 \ alors \ Ad = -d.$$

Exercice III: Soit

$$(P) \quad \left| \begin{array}{ll} \min\limits_{(x,y)\in\mathbb{R}^2} & \frac{1}{2}x^2 - y \\ \text{sous:} & x + e^y \leq 1, \end{array} \right.$$

1. Déterminer la fonction duale du problème (P).
2. Résoudre le problème dual.
3. En déduire une solution de (P).
4. Que se passe-t-il si on remplace la contrainte $x + e^y \leq 1$ par $x + e^y = 1$?

Solution of Exercise 10.3

Exercice I:

1. (a) On sait que $\langle Qx, x\rangle \geq \lambda_{min}\|x\|^2$ où λ_{min} est la plus petite valeur propre de Q et est > 0 (car Q est définie positive). De plus par Cauchy-Schwarz, on a $\langle a, x\rangle \geq -\|a\|\|x\|$. Ainsi $f(x) \geq r(\|x\|)$ avec

$$r(t) = \frac{\lambda_{min}}{2}t^2 - \|a\|t.$$

avec $\lim_{t\to+\infty} r(t) = +\infty$.

(b) On sait que $\langle Qx, x\rangle \geq \lambda_{min}\|x\|^2$ où λ_{min} est la plus petite valeur propre de Q et est > 0 (car Q est définie positive). Ainsi, si $x \in X$ alors $\|x\|^2 \leq \frac{c}{\lambda_{min}}$ et donc X est borné.

2. (a) Si $M = (x, y)$, on note $g_1(M) = x^2 - y^4 - 1 \leq 0$ et $g_2(M) = -x$. On a $\nabla g_1(M) = \begin{pmatrix} 2x & -4y^3 \end{pmatrix}$ et $\nabla g_2(M) = \begin{pmatrix} -1 & -0 \end{pmatrix}$. On regarde dans

les différents cas si $(\nabla g_i(M))_{i \text{ active}}$ est libre.

- Si il y n'y a aucune contrainte active, il n'y a rien à montrer.
- Si seulement g_2 est active alors $\nabla g_2(M) \neq 0$ donc la famille est libre
- Si seulement g_1 est active alors $\nabla g_1(M) \neq 0$ (car $x \neq 0$ car $g_2(M) \neq 0$) donc la famille est libre
- Si g_1 et g_2 sont actives alors $x = 0$ et ensuite $g_1(M) = 0$ est impossible. Donc ce cas n'arrive jamais

Les contraintes sont qualifiées en tout point

(b) On note $g_i(x) = \|x\|^2 - x_i - 1$. Les g_i sont toutes convexes, et on a $g_i(0) < 0$. Par Slater les contraintes sont qualifiées en tout point.

Exercice II:

1. Si $M = (x, y, z)$ le Lagrangien est donné par

$$\mathcal{L}(M, \lambda) = xyz + \lambda(x + y + z - 3)$$

Les points de KKT sont donnés par

$$\begin{cases} yz + \lambda & = 0 \\ xz + \lambda & = 0 \\ yx + \lambda & = 0 \\ x + y + z & = 3 \end{cases}$$

- Supposons que l'une des coordonnée soit nulle (disons $x = 0$), alors forcément $\lambda = 0$ puis ensuite $yz = 0$, donc une deuxième coordonnée doit être nulle et on obtient que la troisième doit être égale à 3. Ainsi $(0, 0, 3)$,$(0, 3, 0)$ et $(3, 0, 0)$ sont points de KKT pour $\lambda = 0$.
- Supposons maintenant que aucune coordonnée n'est nulle, comme $x \neq 0$ les équations $xz + \lambda = xy + \lambda$ impliquent $z = y$, de même en utilisant $y \neq 0$, on obtient $x = y = z$. Puis de $x + y + z = 3$ on trouve que $(1, 1, 1)$ est point de KKT.

2. On a $H[\mathcal{L}](1, 1, 1) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ et $\nabla g(M) = (1, 1, 1)^T$.Comme g est une contrainte d'égalité on a $V(M) = E_2^{\perp}$ avec $E_2 = Vect(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix})$. Grâce à l'indication, pour tout $d \in V(M)$ alors $Ad = -d$ et, si $d \neq 0$

$$\langle H[\mathcal{L}](M)d, d\rangle = -\|d\|^2 < 0$$

Ainsi la condition suffisante est vérifiée et tous les points de KKT sont des maximums locaux.

Exercice III:

1. Si $M = (x, y)$, Le Lagrangien est donné par

$$\mathcal{L}(M, \lambda) = \frac{1}{2}x^2 - y + \lambda(x + e^y - 1)$$

Et la fonction duale est donnée par

$$f^\star(\lambda) = \inf_{M\in\mathbb{R}^2} \mathcal{L}(M,\lambda) = \inf_{M\in\mathbb{R}^2} \frac{1}{2}x^2 - y + \lambda(x + e^y - 1)$$

On s'aperçoit que le problème est décomposable en somme de deus sous-problèmes

$$f^\star(\lambda) = \inf_x \left(\frac{1}{2}x^2 + \lambda x\right) + \inf_y \left(-y + \lambda e^y\right) - \lambda$$

Le premier problème se résout en prenant $x^\star = -\lambda$ et le deuxième en prenant $y^\star = -\ln(\lambda)$ si $\lambda > 0$ (et $y^\star = +\infty$ si $\lambda \le 0$). Ainsi

$$f^\star(\lambda) = \begin{cases} -\infty \text{ si } \lambda \le 0 \\ -\frac{\lambda^2}{2} + \ln(\lambda) + 1 - \lambda \end{cases}$$

Le problème dual est

$$\sup_{\lambda>0} -\frac{\lambda^2}{2} + \ln(\lambda) + 1 - \lambda$$

2. On calcule $(f^\star)'(\lambda) = 0$ ce qui donne

$$-\lambda + \frac{1}{\lambda} - 1 = 0$$

ou encore $\lambda^2 + \lambda - 1 = 0$, dont les solutions sont $-\frac{1}{2} \pm \frac{\sqrt{5}}{2}$. On garde la solution positive et on obtient $\lambda^\star = -\frac{1}{2} + \frac{\sqrt{5}}{2}$.
3. Le point $x^\star = -\lambda$, $y^\star = -\ln(\lambda)$ et $\lambda = -\frac{1}{2} + \frac{\sqrt{5}}{2}$ est un point de KKT. Le problème est convexe, donc les points de KKT sont minimums globaux.
4. Cela ne change rien car on a forcément $\lambda > 0$ dans le calcul de $f^\star(\lambda)$.

INSA TOULOUSE
135 avenue de Rangueil
31400 Toulouse

Tél : + 33 (0)5 61 55 95 13
www.insa-toulouse.fr

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE