Tutorial 6 : Projected Gradient algorithm

Operations research, 3rd semester.

2024

Exercice 1 — *A simple example.*

Let (P) be the following problem :

$$\min_{x\in\mathbb{R}^2} f(x) = x_1^2 + 4x_2^2 \ \text{ s.c. } \begin{cases} x_1 + 2x_2 \geq 1 \\ -x_1 + x_2 \leq 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

1. Draw the graphical representation of the problem.
2. Check that every feasible solution x of (P) (i.e. satisfying the constraints) satisfies the linear independance qualification. Deduce that only the optimale solutions satisfies the Karush Kuhn Tucker conditions.
3. Apply the algorithm from $P_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ (3 iterations).
4. Check that the point satisfies the Karush Khun-Tucker conditions.

▶ **Correction**

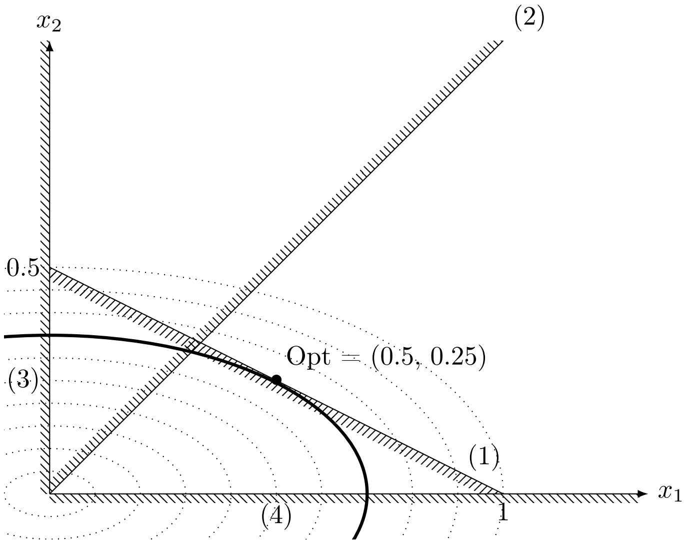
We rewrite the problem in the correct form :

$$\min_{x\in\mathbb{R}^2} f(x) = x_1^2 + 4x_2^2 \ \text{ s.c. } \begin{cases} -x_1 - 2x_2 \leq -1 & (1) \\ -x_1 + x_2 \leq 0 & (2) \\ -x_1 \leq 0 & (3) \\ -x_2 \leq 0 & (4) \end{cases}$$

1. The equation of an ellipse with semi-major axes a and b centered at the origin is $x^2/a^2 + y^2/b^2 = 1$.
   Thus, $x_1^2 + 4x_2^2 = R$ is the equation of an ellipse such that :

   $$\begin{cases} \frac{\sqrt{R}}{a^2} & = 1 \\ \frac{\sqrt{R}}{b^2} & = 4 \end{cases}$$

   Therefore, $a = 2b = \sqrt{R}$.
   We then plot ellipses around the origin that get progressively smaller with $a = 2b$ until we exit the set of feasible solutions ; the point at which we exit indicates the optimal solution.



2. The coefficients matrix is $A_s = \begin{pmatrix} -1 & -2 \\ -1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}$.

   We can see in the drawing that, among all the feasible solutions, only the following cases are possible for $I(x)$, the set of inequalities saturated by a point x :

   — $I(x) = \emptyset$
   — $I(x) = \{1\}$
   — $I(x) = \{2\}$
   — $I(x) = \{4\}$
   — $I(x) = \{1, 2\}$
   — $I(x) = \{1, 4\}$

   In the first 4 cases, $(\nabla g_i(x)|i \in I(x))$ contains 0 or 1 vector, so it is a linearly independent family.

   The other two cases are

   — $(\nabla g_i(x)|i \in I(x)) = (\begin{pmatrix} -1 \\ -2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix})$ which is linearly independent (the matrix is squared and has determinant -3, so is invertible).
   — $(\nabla g_i(x)|i \in I(x)) = (\begin{pmatrix} -1 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix})$ which is linearly independent (the matrix is squared and has determinant 1, so is invertible)..

     We can deduce that any local (or global) minimum satisfies (KKT). However, f is convex and the g_i are convex (since they are linear), so any point that satisfies (KKT) is a global minimum. Therefore, there is equivalence.

3. We start at $P_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. We have $I(x) = \{2\}$ et $\nabla f(x) = \begin{pmatrix} 2 \\ 8 \end{pmatrix}$.

   We compute the projection $L = \{y| - y_1 + y_2 = 0\}$:

   — via $d = (I - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(x)) = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} -2 \\ -8 \end{pmatrix} = \begin{pmatrix} -5 \\ -5 \end{pmatrix}$
   — ou via $y \in L \Leftrightarrow y_1 = y_2$; donc $d = \beta \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ et

   $$(-\nabla f(x)) \cdot d = d^2 \Rightarrow -2\beta - 8\beta = \beta^2 \cdot 2 \Rightarrow \beta = -5$$

   .

We can check that $d \in L$ and that $(-\nabla f(x)) - d = \begin{pmatrix} 3 \\ -3 \end{pmatrix} \perp d$,

$d \neq \vec{0}$, Thus, we move as far as possible in the indicated direction. We denote S as the feasible set.

$$\begin{aligned}
\alpha_1 &= \max_{0 \leq \alpha} \{\alpha | P_0 + \alpha \cdot d \in S\} \\
&= \max_{0 \leq \alpha} \left\{ \alpha | \begin{pmatrix} 1 - 5\alpha \\ 1 - 5\alpha \end{pmatrix} \in S \right\} \\
&= \max_{0 \leq \alpha} \left\{ \alpha | \begin{array}{c} -3 + 15\alpha \leq -1 \\ 0 \leq 0 \\ -1 + 5\alpha \leq 0 \\ -1 + 5\alpha \leq 0 \end{array} \right\} \\
&= \frac{2}{15}
\end{aligned}$$

$$\begin{aligned}
\alpha_2 &= \arg \min_{0 \leq \alpha \leq \alpha_1} \{g(\alpha) = f(P_0 + \alpha \cdot d)\} \\
&= \arg \min_{0 \leq \alpha \leq \alpha_1} \{(1 - 5\alpha)^2 + 4(1 - 5\alpha)^2\} \\
&= \arg \min_{0 \leq \alpha \leq \alpha_1} \{5(1 - 5\alpha)^2\} \\
g'(\alpha) &= 5 \cdot (-5) \cdot 2(1 - 5\alpha) \leq 0 \\
\Leftrightarrow \alpha &\leq \frac{1}{5}
\end{aligned}$$

Thus $g'(\alpha)$ is negative on $[0, \alpha_1]$, so

$$\alpha_2 = \alpha_1 = \frac{2}{15}$$

We move to $x = P_0 + \frac{2}{15} d = \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix}$.

We have $I(x) = \{1, 2\}$ et $\nabla f(x) = \begin{pmatrix} 2/3 \\ 8/3 \end{pmatrix}$.

We project on $L = \{y | -y_1 + y_2 = 0; -y_1 - 2y_2 = 0\} = \{0\}$, donc $d = \vec{0}$. (We can also recalculate the projection operator, which is the zero matrix).

We can then write $-\nabla f(x)$ as $\lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) = \lambda_1 \begin{pmatrix} -1 \\ -2 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

— either by solving the system

— or with $(A_s \cdot {}^t A_s)^{-1} \cdot A_s \cdot (-\nabla f(x))$ that gives λ.

$(A_s \cdot {}^t A_s)^{-1} \cdot A_s = \begin{pmatrix} -1/3 & -1/3 \\ -2/3 & 1/3 \end{pmatrix}$ then $\lambda = \begin{pmatrix} 10/9 \\ -4/9 \end{pmatrix}$

We remove the constraint (2) from $I(x)$. We compute the projection on $L = \{y | -y_1 - 2y_2 = 0\}$.

$d = (I - {}^t A_s \cdot (A_s \cdot {}^t A_s)^{-1} \cdot A_s) \cdot (-\nabla f(x)) = \begin{pmatrix} 0.8 & -0.4 \\ -0.4 & 0.2 \end{pmatrix} \cdot \begin{pmatrix} -2/3 \\ -8/3 \end{pmatrix} = \begin{pmatrix} 8/15 \\ -4/15 \end{pmatrix}$

We can check that $d \in L$ and that $(-\nabla f(x)) - d = \begin{pmatrix} -6/5 \\ -12/5 \end{pmatrix} \perp d$

$d \neq \vec{0}$, thus, we move as far as possible in the indicated direction.

$$\begin{aligned}
\alpha_1 &= \max_{0\leq\alpha}\{\alpha | P_1 + \alpha \cdot d \in S\}\\
&= \max_{0\leq\alpha}\left\{\alpha| \begin{pmatrix} 1/3 + 8/15\alpha \\ 1/3 - 4/15\alpha \end{pmatrix} \in S\right\}\\
&= \max_{0\leq\alpha}\left\{\alpha| \begin{matrix} -1 \leq -1 \\ -12/15\alpha \leq 0 \\ -1/3 - 8/15\alpha \leq 0 \\ -1/3 + 4/15\alpha \leq 0 \end{matrix}\right\}\\
&= \frac{5}{4}
\end{aligned}$$

$$\begin{aligned}
\alpha_2 &= \arg\min_{0\leq\alpha\leq\alpha_1}\{g(\alpha) = f(P_1 + \alpha \cdot d)\}\\
&= \arg\min_{0\leq\alpha\leq\alpha_1}\{(1/3 + 8/15\alpha)^2 + 4(1/3 - 4/15\alpha)^2\}\\
&= \frac{5}{16}
\end{aligned}$$

We move to $x = P_1 + \frac{5}{16}d = \begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix}$

We still have $I(x) = \{1\}$.

We compute the projection on $L = \{y | -y_1 - 2y_2 = 0\}$.

$d = (I - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(x)) = \begin{pmatrix} 0.8 & -0.4 \\ -0.4 & 0.2 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$d = \vec{0}$ (since we are in 2D and the projection operator is non-zero, this simply means that the gradient is orthogonal to the boundary we are on ; therefore, its projection on the boundary is zero).

We then write $-\nabla f(x) = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$ as $\lambda_1 \nabla g_1(x) = \lambda_1 \begin{pmatrix} -1 \\ -2 \end{pmatrix}$.

We have $\lambda_1 = 1$.

All the λ_i are positive, so we stop and return the solution. $\begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix}$

4. We have $\nabla f \begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix} + \nabla g_1 \begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix} = 0$ et $g_1 \begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix} = 0$

   we set $\lambda_1 = 1$ et $\lambda_2, \lambda_3, \lambda_4 = 0$ $\begin{cases} \nabla f(x) + \sum_{i=1}^{4} \lambda_i \nabla g_i(x) & = 0 \\ \lambda_1 \cdot g_1(x) & = 0 \\ \lambda_2 \cdot g_2(x) & = 0 \\ \lambda_3 \cdot g_3(x) & = 0 \\ \lambda_4 \cdot g_4(x) & = 0 \end{cases}$

   (KKT) conditions are satisfied.

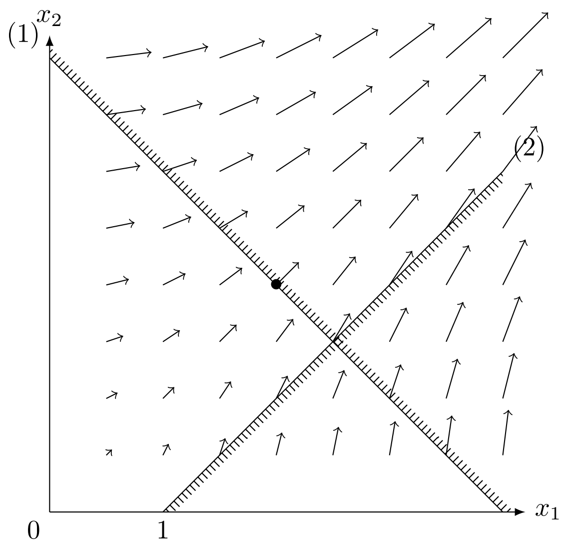
Exercice 2 — *With equalities*

Let (P) be the following problem :

$$\max_{x\in\mathbb{R}^2} f(x) = x_1 \cdot x_2 \quad \text{s.c.} \begin{cases} x_1 + x_2 + x_3 = 4 & (1) \\ x_1 - x_2 + x_4 = 1 & (2) \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

Same questions as exercice 1 except that the algorithm should be started from the point $\begin{pmatrix} 1.5 \\ 0.5 \\ 2 \\ 0 \end{pmatrix}$).

▶ **Correction**

1. 

   We plotted the gradient on the drawing. The optimum is at (2, 2). The function is not convex, but there is only one point satisfying the (KKT) conditions.

2. We can rewrite the problem as follows :

$$\min_{x\in\mathbb{R}^2} f(x) = -x_1 \cdot x_2 \quad \text{s.c.} \quad \begin{cases} x_1 + x_2 + x_3 = 4 & (1) \\ x_1 - x_2 + x_4 = 1 & (2) \\ -x_1, -x_2, -x_3, -x_4 \leq 0 & (3)...(6) \end{cases}$$

   The matrix is $A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$

   We can be in one of the following cases :
   - $I(x) = \emptyset$
   - $I(x) = \{3\}$
   - $I(x) = \{4\}$
   - $I(x) = \{5\}$
   - $I(x) = \{6\}$
   - $I(x) = \{3, 4\}$
   - $I(x) = \{3, 5\}$
   - $I(x) = \{4, 6\}$
   - $I(x) = \{5, 6\}$

   For example, we can check the linear independence in the last case directly. We then have the family of vectors $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}\right)$

   The associated square matrix has a determinant of -2, therefore it is invertible, hence the family is linearly independent. Thus, the subfamilies are also linearly independent. The qualification under linear independence is verified if $I(x) \subset \{5, 6\}$.

3. Here are the main outlines of the algorithm, the method being the same as in the previous exercise :

We start at $\begin{pmatrix} 1.5 \\ 0.5 \\ 2 \\ 0 \end{pmatrix}$

We have $\nabla f(x) = \begin{pmatrix} -0.5 \\ -1.5 \\ 0 \\ 0 \end{pmatrix}$

We have $I(x) = \{6\}$. We project on $L = \{y | y_4 = 0; y_1 + y_2 + y_3 = 0; y_1 - y_2 + y_4 = 0\} = \{y | y_4 = 0; y_3 = -2y_1; y_1 = y_2\}$.

We find $d = \begin{pmatrix} 1/3 \\ 1/3 \\ -2/3 \\ 0 \end{pmatrix}$.

$d \neq \vec{0}$. We move following the direction $x + \alpha d$. We find $\alpha_1 = \alpha_2 = 3$

We move to $x = \begin{pmatrix} 2.5 \\ 1.5 \\ 0 \\ 0 \end{pmatrix}$.

We have $\nabla f(x) = \begin{pmatrix} -1.5 \\ -2.5 \\ 0 \\ 0 \end{pmatrix}$

We have $I(x) = \{5, 6\}$. There are 2 saturated inequalities and 2 equalities, so the space onto which we project is necessarily $\{0\}$ (because 1. we are projecting onto a space orthogonal to a 4-dimensional space. Since we are in $\mathbb{R}^4$, this space has dimension 0, so it is $\{0\}$. 2. We have A_S which is a square and invertible matrix. If we calculate the projection operator, we will get the zero matrix.)

Thus we write $-\nabla f(x)$ as $\lambda_5 \nabla g_5(x) + \lambda_6 \nabla g_6(x) + \mu_1 \nabla h_1(x) + \mu_2 \nabla h_2(x)$. We find $\lambda_5 = 2$, $\lambda_6 = -0.5$, $\mu_1 = 2$, $\mu_2 = -0.5$. We remove 6 from $I(x)$.

We have $I(x) = \{5\}$. We project on $L = \{y | y_3 = 0; y_1 + y_2 + y_3 = 0; y_1 - y_2 + y_4 = 0\} = \{y | y_3 = 0; y_2 = -y_3; y_4 = y_2/2\}$. We find $d = \begin{pmatrix} -1/6 \\ 1/6 \\ 0 \\ 1/3 \end{pmatrix}$.

$d \neq \vec{0}$. We move with the direction d and we find $\alpha_1 = 15, \alpha_2 = 3$

We move to $x = \begin{pmatrix} 2 \\ 2 \\ 0 \\ 1 \end{pmatrix}$.

We have $\nabla f(x) = \begin{pmatrix} -2 \\ -2 \\ 0 \\ 0 \end{pmatrix}$

We project onto the same space as before, $I(x)$ has not changed and still equals $\{5\}$. We find $d = \vec{0}$.

Thus, we express $-\nabla f(x)$ in the form $\lambda_5 \nabla g_5(x) + \mu_1 \nabla h_1(x) + \mu_2 \nabla h_2(x)$. We find $\lambda_5 = 2$, $\mu_1 = 2$, $\mu_2 = 0$.

All the λ_i (here only λ_5) are positive. We stop here.

4. The point satisfies (KKT) with $\lambda_1 = \lambda_2 = \lambda_6 = 0$ and $\lambda_5 = 2$, $\mu_1 = 2$, $\mu_2 = 0$. We cannot deduce global optimality because the function is not convex. However, it can be noted that this is the only feasible point that meets these conditions. Therefore, there is no other local maximum in the feasible space. It is thus optimal.

Exercice 3 — *Projection operator.*

Given a vector $g \in \mathbb{R}^n$, we want to minimize the $\frac{1}{2}\|g-p\|^2$ where such that $p \in \mathbb{R}^n$ and $Ap = 0$. The size of the matrix A is $m \times n$ with rank $m < n$.

In other words, we search for the projection of g over the space $\{p \backslash Ap = 0\}$.

1. Write and solve the (KKT) conditions for this problem.
2. This way, find the formula of the projection operator on the space $L = \{p \backslash Ap = 0\}$.

▶ **Correction**

All the points satisfy the qualification of linear independence since the rank of the matrix is m ; the local minima thus all satisfy KKT, and $A\,{}^tA$ is invertible.

1.

$$(KKT) : \begin{cases} -(g-p) + {}^tA\mu & = 0 \quad (1) \\ Ap & = 0 \quad (2) \end{cases}$$

$$\begin{aligned} A \cdot (-(g-p) + {}^tA\mu) &= 0 \\ -Ag + Ap + A\,{}^tA\mu &= 0 \\ A\,{}^tA\mu &= Ag \\ \mu &= (A\,{}^tA)^{-1}Ag \end{aligned}$$

2. We rewrite (1)

$$\begin{aligned} -(g-p) + {}^tA(A\,{}^tA)^{-1}Ag &= 0 \\ p &= (I - {}^tA(A\,{}^tA)^{-1}A) \cdot g \end{aligned}$$

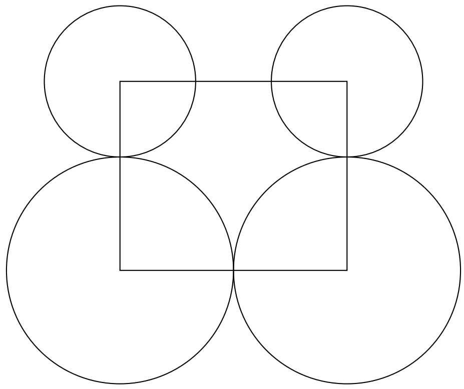
The function is convex, the equalities are linear, so (KKT) is sufficient to prove global optimality ; p is indeed the projection of g onto L.

Exercice 4 — *Non convex problem*

We consider the following problem :

Given a rectangle of size 10×12, we want to place 4 disks, centered in each of the 4 vertices of the rectangle such that the area is **maximum** and such that the interiors of the disks do not intersect ; in other words, the disks may only touch the boundaries of the others.

For example, the following drawing contains a maximal feasible solution :



1. Modelize the problem as a mathematical program. We set $\gamma = \sqrt{244} \simeq 15.62$.
2. Show that, if we apply the projected gradient algorithm from the point where every radius is nul, the algorithm stops at the first iteration.
3. Show that, if we apply the projected gradient algorithm from the point given on the drawing (we may assume that the radius of the two lower disks are equals), the algorithm stops at the first iteration.
4. Show that those solutions are not optimal.

▶ **Correction**

1. We can write the problem in the following form, with x_1, x_2, x_3, x_4 as the radii of the discs with center 1 at the bottom left, 2 at the bottom right, 3 at the top right, and 4 at the top left.

$$\max_{x \in \mathbb{R}^2} f(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 \text{ s.c. } \begin{cases} x_1 + x_2 \leq 12 \\ x_1 + x_3 \leq \gamma \\ x_1 + x_4 \leq 10 \\ x_2 + x_3 \leq 10 \\ x_2 + x_4 \leq \gamma \\ x_3 + x_4 \leq 12 \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

We can rewrite it that way :

$$\min_{x \in \mathbb{R}^2} f(x) = -x_1^2 - x_2^2 - x_3^2 - x_4^2 \text{ s.c. } \begin{cases} x_1 + x_2 \leq 12 \quad (1) \\ x_1 + x_3 \leq \gamma \quad (2) \\ x_1 + x_4 \leq 10 \quad (3) \\ x_2 + x_3 \leq 10 \quad (4) \\ x_2 + x_4 \leq \gamma \quad (5) \\ x_3 + x_4 \leq 12 \quad (6) \\ -x_1, -x_2, -x_3, -x_4 \leq 0 \quad (7), (8), (9), (10) \end{cases}$$

2. Note : This point is a local maximum, so the gradient is zero, and the algorithm will necessarily stop.

If $x = 0$, then we have $I(x) = \{7, 8, 9, 10\}$. We calculate the projection onto $L = \{y | -y_1 = -y_2 = -y_3 = -y_4 = 0\} = \vec{0}$.

Alternatively, the matrix A_s is $-I$, so ${}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s = I$, thus the projection operator is the zero matrix.
Another method, the gradient is zero, so its projection is also zero.
Thus, we express $-\nabla f(x) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ in the form $\sum_{i=7}^{10} \lambda_i \nabla g_i(x)$. We have trivially $\lambda_i = 0$ for all i, so we stop.

3. On the figure, we deduce the radiuses as follows
$$\begin{cases} x_1 + x_2 &= 12 \quad (1) \\ x_1 + x_4 &= 10 \quad (3) \\ x_2 + x_3 &= 10 \quad (4) \end{cases} \text{ with } x_1 = x_2 \text{ and } x_3 = x_4.$$
Then $\begin{cases} x_1 &= 6 \\ x_2 &= 6 \\ x_3 &= 4 \\ x_4 &= 4 \end{cases}$

Then $\nabla f(x) = \begin{pmatrix} -12 \\ -12 \\ -8 \\ -8 \end{pmatrix}$

We have $I(x) = \{1, 3, 4\}$ et $A_s = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$.
$$I - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s = \begin{pmatrix} 0.25 & -0.25 & 0.25 & -0.25 \\ -0.25 & 0.25 & -0.25 & 0.25 \\ 0.25 & -0.25 & -0.25 & 0.25 \\ -0.25 & 0.25 & 0.25 & -0.25 \end{pmatrix}$$
Then $d = (I - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(x)) = \vec{0}$.

So we write $-\nabla f(x)$ as $\sum_{i=1,3,4} \lambda_i \nabla g_i(x) = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \lambda_4 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$.

Then $\begin{cases} \lambda_1 + \lambda_3 &= 12 \\ \lambda_1 + \lambda_4 &= 12 \\ \lambda_4 &= 4 \\ \lambda_3 &= 4 \end{cases}$
So $\lambda_1 = 8$
All the λ_i are positive, so we stop and return the solution.

4. These two solutions are not optimal.
The first has a value of 0 and the second -104.
For example, the solution $x = \begin{pmatrix} 10 \\ 2 \\ \gamma - 10 \\ 0 \end{pmatrix}$ has a value of -135.59.

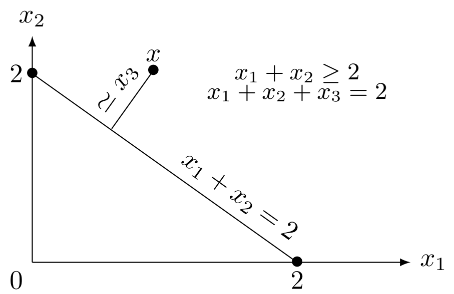
Exercice 5 — *Projection and standard form*

We consider the following programs :
$$\min_{x \in \mathbb{R}^2} f(x) = x_1^2 + x_2^2 \quad \text{s.c.} \quad \begin{cases} x_1 + x_2 \geq 2 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

$$\min_{x\in\mathbb{R}^2} f(x) = x_1^2 + x_2^2 \text{ s.c. } \begin{cases} x_1 + x_2 - x_3 = 2 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0 \end{cases}$$

1. Show that the two programs are equivalent. (They have the same optimal value and every optimal solution of hte first program can be deduced from an optimal solution of the other.)
2. Draw the graphical representation of the two programs on the same drawing.
3. Show that if we apply an iteration of the projected gradient algorithm on the first program from the point $x = (3, 2)$, the direction we follow is $(-6, -4)$.
4. Show that if we apply an iteration of the projected gradient algorithm on the second program from the point $x = (3, 2, 3)$, the direction we follow is $(-8/3, -2/3, -10/3)$.
5. How can we explain that two equivalent programs do not follow the same direction on the drawing ?

► **Correction**

1. They are equivalent since, if there exists (x_1, x_2) a feasible solution to the first problem, then $(x_1, x_2, x_1 + x_2 - 2)$ is a solution to the second one with the same objective value. Similarly, if (x_1, x_2, x_3) is a solution to the first problem, then $x_1 + x_2 = 2 + x_3 \geq 2$, so (x_1, x_2) is a solution to the first program with the same objective value.
2. 

3. In the first program, at the point $(3, 2)$, $I(x) = \emptyset$, we project onto $\mathbb{R}^2$, thus $d = -\nabla f(x) = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix} = \begin{pmatrix} -6 \\ -4 \end{pmatrix}$
4. In the second program, at the point $(3, 2, -3)$ we have $I(x) = \emptyset$ but we have an equality. Therefore, we project onto the space $L = \{y | y_1 + y_2 - y_3 = 0\}$.
   Using the projection operator, we have $A_S = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}$; so $PL = I - {}^tA_S \cdot (A_S \cdot {}^tA_S)^{-1} \cdot A_S = \begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \\ 1/3 & 1/3 & 2/3 \end{pmatrix}$ donc $d = PL \cdot \begin{pmatrix} -6 \\ -4 \\ 0 \end{pmatrix} = \begin{pmatrix} -8/3 \\ -2/3 \\ -10/3 \end{pmatrix}$
5. This can be explained by the fact that the projection operator is not equivalent in the two programs. In the second program, we project onto a slightly different space than in the first. Indeed, in the first program, we project onto $L_1 = \{y \in \mathbb{R}^2\} = \{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\}$ and in the second one, onto $L_2 = \{y | y_1 + y_2 - y_3 = 0\} = \left\{ \begin{pmatrix} y_1 \\ y_2 \\ y_1 + y_2 - 2 \end{pmatrix} \right\}$
   One can see that the first is, in a way, included in the second ; since one is free to choose any coordinates y_1 and y_2 in L_2. However, the non-free existence of this third coordinate means that the point closest to the gradient of f in L_2 does not have exactly the same first two coordinates, it is slightly deviated.
   There are, however, points where the projection does indeed contain $-\nabla f(x)$.

These are the points where $x_1 = x_2$. The gradient then has the form $\begin{pmatrix} \alpha \\ \alpha \\ 0 \end{pmatrix}$ et $PL \cdot \begin{pmatrix} -\alpha \\ -\alpha \\ 0 \end{pmatrix} =$

$\begin{pmatrix} -\alpha \\ -\alpha \\ -2\alpha \end{pmatrix}$

This difference can be explained differently. Instead of making the drawing in 2D, it should be done in 3D. We are in the plane $P = \{x_3 = x_1 + x_2 - 2\}$. This plane is tilted. If the gradient has the first two coordinates equal, then, as explained earlier, the projection also has its first two coordinates equal, so we are indeed following the same direction in both programs.

Conversely, if these coordinates are different, then the gradient is not in the direction of descent, and when projected onto the plane, the projection deviates a bit. The farther we move away from the line $x_1 = x_2$, the more the projection deviates.