Université Paris Cité
UFR de Mathématiques et Informatique
45, rue des Saints-Pères, 75006, Paris.

M1 MMA 2023-2024
OPTIMISATION SOUS-CONTRAINTES

Feuille de TD n°3

Approche géométrique de la dualité de Lagrange.

Exercice 2

On considère le problème

$$\inf_{(x_1,x_2)\in S} x_1^2 - 14x_1 + x_2^2 - 6x_2 - 7,$$

où $S = \left\{x \in \mathbb{R}^2 \,/\, x_1 + x_2 \leqslant 2 \text{ et } x_1 + 2x_2 \leqslant 3\right\}$.

1. Montrer que le problème est bien posé et admet une unique solution.
2. La déterminer à l'aide des conditions KKT.
3. Interpréter géométriquement le problème.

Correction.

1. Notons f la fonction objectif du problème et f_1, f_2 les fonctions telles que pour tout $x = (x_1, x_2) \in \mathbb{R}^2$

$$f_1(x) = x_1 + x_2 - 2 \quad \text{et} \quad f_2(x) = x_1 + 2x_2 - 3,$$

de sorte que $S = \{x \in \mathbb{R}^2 \,/\, f_1(x) \leqslant 0, f_2(x) \leqslant 0\}$.

Les fonctions f, f_1, f_2 sont $\mathcal{C}^\infty$ sur $\mathbb{R}^2$ donc en particulier continue sur $\mathbb{R}^2$. Ainsi S est fermé, non vide car contient $(0,0)$. De plus f se réécrit

$$\forall x \in \mathbb{R}^2, \quad f(x) = \frac{1}{2}\langle Ax,\, x\rangle - \langle b,\, x\rangle + c,$$

avec $A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $b = (14, 6)$, $c = -7$. Donc f est une fonctionnelle quadratique avec A définie positive donc f est coercive sur $\mathbb{R}^2$ donc sur S.

> Comme f est une fonctionnelle quadratique, on a pour tout $x \in \mathbb{R}^2$, $\nabla^2 f(x) = A$ et comme A est définie positive, on a donc $\nabla^2 f(x) \succ mI_2$ pour un certain $m > 0$ (on peut prendre ici trivialement $m = 2$). Donc f est m-fortement convexe sur $\mathbb{R}^2$ donc coercive sur $\mathbb{R}^2$.

Par conséquent on déduit que le problème est bien posée, en particulier admet un minimum global sur S.

Puis f est strictement convexe sur $\mathbb{R}^2$ donc sur S, qui est bien convexe car f_1 et f_2 le sont, d'où f admet un unique minimum global sur S.

2. Commençons par déterminer les points réguliers de S. On a pour tout $x \in \mathbb{R}^2$, $\nabla f_1(x) = (1,1)$ et $\nabla f_2(x) = (1,2)$. Or ces deux vecteurs forment une famille libre. Ainsi pour tout $x \in S$, la famille $(\nabla f_i(x))_{i \in I(x)}$ est une famille libre (potentiellement vide) en tant que sous famille de la famille libre $(\nabla f_1(x), \nabla f_2(x))$. Donc tout point de S est libre. Et donc en tout point $x \in S$, les contraintes sont qualifiées.

Notons $x^\star = (x_1^\star, x_2^\star)$ l'unique minimum global de f sur S. Alors puisque les contraintes sont qualifiées en $x^\star$, les conditions KKT sont vérifiées en $x^\star$. Ainsi il existe $\lambda_1, \lambda_2 \geqslant 0$ (conditions d'admissibilité duale) tels que

$$\lambda_1 f_1(x^\star) = 0, \tag{1}$$

$$\lambda_2 f_2(x^\star) = 0, \tag{2}$$

$$\nabla f(x^\star) + \lambda_1 \nabla f_1(x^\star) + \lambda_2 \nabla f_2(x^\star) = 0_{\mathbb{R}^2}. \tag{3}$$

> Les deux premières conditions sont les conditions de complémentarité, et la dernière la condition de stationnarité.
> La condition d'admissibilité primale ($x^\star \in S$), les conditions d'admissibilité duales et les conditions de complémentarité induisent une disjonction de cas naturelle en fonction de si $f_1(x^\star) < 0$, $f_1(x^\star) = 0$, $f_2(x^\star) < 0$, $f_2(x^\star) = 0$. Dans chacune de ces quatre situations, la condition de stationnarité va se simplifier différemment menant à une résolution différente et, espérons, à l'identification de $x^\star$. Étudions ces différents cas.

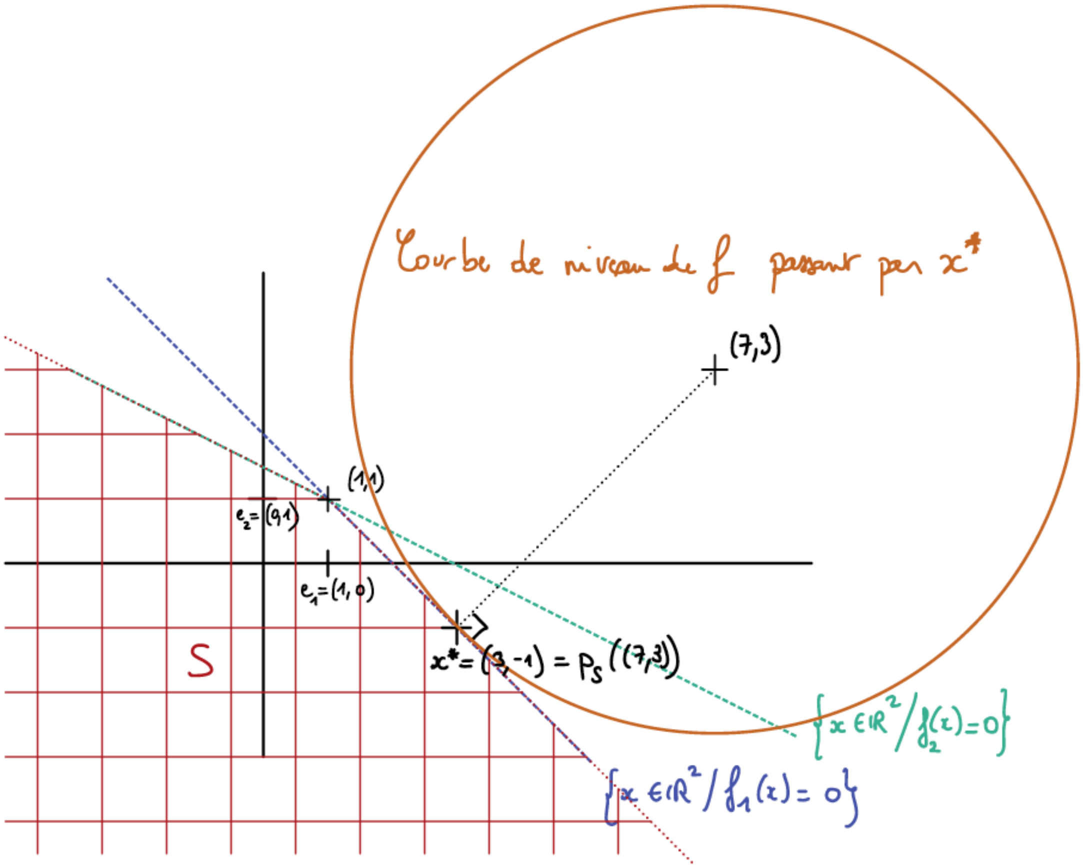

FIGURE 1. Représentation géométrique du problème d'optimisation de l'Exercice 2. On remarque bien que $x^\star$ est tel que $f_1(x^\star) = 0$ et $f_2(x^\star) < 0$.

Si $x^\star \in \mathring{S}$ alors on a simplement par les conditions d'optimalité d'ordre 1 que $\nabla f(x^\star) = 0$.

> Condition que l'on retrouve bien par les conditions KKT, puisque $\lambda_1 = \lambda_2 = 0$ grâce à (1) et (2) avec $f_1(x^\star) < 0$ et $f_2(x^\star) < 0$, et donc par (3) $\nabla f(x^\star) = 0$.

Or l'unique vecteur qui annule le gradient de f (qui est l'unique minimum globale de f sur $\mathbb{R}^2$) est $A^{-1}b = (7,3) \notin S$. Donc $x^\star \in \partial S = S \setminus \mathring{S}$, i.e. $f_1(x^\star) = 0$ ou $f_2(x^\star) = 0$.

Supposons que $f_1(x^\star) = 0$ et $f_2(x^\star) = 0$ alors en résolvant le système on obtient $x^\star = (1,1)$. On a alors $\nabla f(x^\star) = Ax - b = (-12,-4)$ et par la condition de stationnarité on obtient le système

$$\begin{cases} \lambda_1 + \lambda_2 = 12, \\ \lambda_1 + 2\lambda_2 = 4, \end{cases}$$

ce qui donne $(\lambda_1, \lambda_2) = (20, -8)$, car par conditions d'admissibilité duale $\lambda_2 \geqslant 0$. Contradiction, donc $x^\star$ ne satisfait pas $f_1(x^\star) = 0$ et $f_2(x^\star) = 0$. Il nous reste deux cas à étudier

— $f_1(x^\star) = 0$ et $f_2(x^\star) < 0$ (donc avec $\lambda_2 = 0$ par (2)),

— $f_1(x^\star) < 0$ (donc avec $\lambda_1 = 0$ par (1)) et $f_2(x^\star) = 0$.

Si $f_1(x^\star) < 0$ et $f_2(x^\star) = 0$, alors $\lambda_1 = 0$ et $x_1 = 3 - 2x_2$ donc la condition de stationnarité (3) mène à

$$(2(3-2x_2), 2x_2) - (14,6) + \lambda_2(1,2) = (0,0),$$

et en éliminant λ_2 du système on trouve $x^\star = (5,-1)$. Mais alors $f_1(x^\star) > 0$ i..e $x^\star \notin S$. Contradiction.

Si $f_1(x^\star) = 0$ et $f_2(x^\star) < 0$, alors $\lambda_2 = 0$ et $x_1 = 2 - x_2$ donc la condition de stationnarité (3) mène à

$$(2(2-x_2), 2x_2) - (14,6) + \lambda_1(1,1) = (0,0),$$

alors en retranchant la première équation dans la seconde, on trouve $x_2 = -1$ et donc $x_1 = 3$. On a bien $x^\star = (3,-1) \in S$. Puis on trouve $\lambda_1 = 8$ qui est bien positif. Ainsi on a trouvé qu'un seul candidat après l'exploitation des conditions KKT pour $x^\star$: $(3,-1)$.

Bilan : l'unique minimum global de f sur S est $x^\star = (3,-1)$.

3. On a pour tout $x \in \mathbb{R}^2$, $f(x) = (x_1-7)^2 + (x_2-3)^2 - 65 = \|x - (7,3)\|_2^2 - 65$. Ainsi le problème d'optimisation revient à minimiser la distance séparant les points de l'ensemble convexe fermé S au point

$(7,3)$. On sait donc que l'unique solution de ce problème d'optimisation est la projection orthogonale de $(7,3)$ sur S. On a donc montré que $P_S((7,3)) = (3,-1)$.

Voir la Figure 1 pour une représentation de ce qui a été discuté dans cet exercice.

Exercice 3

On s'intéresse au problème

$$\inf_{(x_1,x_2)\in S} x_1^2 - x_1x_2,$$

où $S = \{x \in \mathbb{R}^2 \,/\, x_2 \leqslant 2x_1 \text{ et } x_1^2 + x_2^2 = 1\}$.

1. Représenter graphiquement l'ensemble S sur lequel l'optimisation se fait. Quels sont les points réguliers pour les contraintes qui définissent S ?

2. Déterminer l'unique minimiseur sur S à l'aide des conditions KKT. *Indication : on pourra s'aider de l'outil informatique pour effectuer une discrimination entre les différentes possibilités.*

3. Illustrer les résultats sur Python en représentant S ainsi que les lignes de niveau de la fonction objectif.

<u>Correction</u>.

L'énoncé de cette exercice n'est pas incroyable. Il mérite une reformulation de la première question qui devrait demander au préalable de montrer que le problème est bien posé. Je vous propose de vous présenter cela dans un premier temps.

1. Soit $f_1 : x \in \mathbb{R}^2 \mapsto -2x_1 + x_2$ et $h_1 : x \in \mathbb{R}^2 \mapsto x_1^2 + x_2^2 - 1$. Alors $S = \{x \in \mathbb{R}^2 \,/\, f_1(x) = 0 \text{ et } h_1(x) = 0\}$. L'ensemble S est non vide car $e_1 = (1,0) \in S$: $f_1((1,0)) = -2 \leqslant 0$ et $h_1((1,0)) = 0$. De plus S est fermé comme f_1 et f_2 sont $\mathcal{C}^\infty$ sur $\mathbb{R}^2$ donc continues. L'ensemble S est borné car inclut dans le cercle unité. Donc S est compact. La fonction objectif $f : x \in \mathbb{R}^2 \mapsto x_1^2 - x_1x_2$ étant $\mathcal{C}^\infty$ sur $\mathbb{R}^2$ donc continue, on déduit ainsi que f admet au moins un minimum global sur S. En particulier le problème est bien posé.

> Attention : dans l'étude de ce problème d'optimisation, nous ne parlerons pas de convexité car l'ensemble S n'est pas convexe (et de toute façon f n'est pas convexe) !

Déterminons maintenant l'ensemble des points réguliers de S. Soit $x \in S$.

— Si la contrainte associée à f_1 est active, i.e. $f_1(x) = 0$, alors x est régulier par rapport aux contraintes si la famille $(\nabla f_1(x), \nabla h_1(x))$ est libre.

— Si la contrainte associée à f_1 n'est pas active, i.e. $f_1(x) < 0$, alors x est régulier si la famille $(\nabla h_1(x))$ est libre, i.e. si et seulement si $\nabla h_1(x) \neq 0_{\mathbb{R}^2}$.

Or $\nabla h_1(x) = 2x \neq 0_{\mathbb{R}^2}$ pour tout $x \in S$ (car $(0,0) \notin S$). Ainsi on sait déjà que tous les points de S dont la contrainte associée à f_1 n'est pas active sont des points réguliers de S.

Il n'y a que deux points de S satisfaisant $f_1 = 0$. On peut faire les calculs, trouver les coordonnées de ces points et observer alors que $(\nabla f_1(x), \nabla h_1(x))$ est bien libre. Alternativement, on peut observer géométriquement que des ces deux situations, les vecteurs $\nabla f_1(x) = (-2,1)$ et $\nabla h_1(x)$ sont orthogonaux. Ainsi tous les points de S sont réguliers.

2. Soit $x \in S$ un minimum global de f sur S. Alors comme x est régulier d'après la précédente question, les contraintes sont qualifiées en x et donc les conditions KKT sont nécessairement satisfaites.

> Pour rappel, il n'y a pas besoin de la convexité pour obtenir la nécessité des conditions KKT quand les contraintes sont qualifiées. La convexité permet d'obtenir l'autre sens : si les conditions KKT sont satisfaites en x (peu importe si les contraintes sont qualifiées ou pas), alors x est un minimum global, cf cours.

Ainsi il existe $\lambda \geqslant 0$ (condition d'admissibiltié duale) et $\mu \in \mathbb{R}$ tels que

$$\begin{aligned} &\lambda f_1(x) = 0, \\ &\nabla f(x) + \lambda \nabla f_1(x) + \mu \nabla h_1(x) = 0_{\mathbb{R}^2}. \end{aligned}$$

On distingue alors deux cas :

— si $f_1(x) < 0$ alors $\lambda = 0$ par condition de complémentarité. Et donc la condition de stationnarité devient

$$(2x_1 - x_2, -x_1) + 2\mu x = 0_{\mathbb{R}^2}.$$

On peut alors résoudre ce système d'équations en utilisant le fait que $h_1(x) = 0$. On trouve alors 4 points et seuls 2 points satisfont l'autre contrainte $f_1(x) \leqslant 0$. Les calculs sont un peu pénible, il faut utiliser pour se simplifier un peu la vie que $\cos(\frac{\pi}{8}) = \frac{\sqrt{2+\sqrt{2}}}{2}$.

— Si $f_1(x) = 0$, alors on a déjà les coordonnées de ces deux points de S (si on a fait les calculs...). Parmi ces deux points en résolvant en λ seul un permet d'assurer la condition d'admissibilité duale $\lambda \geqslant 0$.

Après ces calculs pas très intéressants, on se retrouve donc avec une liste de 3 points qui satisfont les conditions KKT. L'ensemble des minimums globaux est donc inclu dans cet ensemble de candidats. On peut alors soit faire les calculs à la main, soit utiliser l'ordinateur, pour évaluer f en ces 3 points et observer qu'il y en a un dont la valeur associée par f est strictement inférieure aux autres valeurs. Ce point de S est donc l'unique candidat pour être le minimum global de f sur S. Et comme on sait que f admet au moins un minimum global sur S, on a donc trouvé un unique minimum global de f sur S. Ouf !

3.

Exercice 4

Soit $n \in \mathbb{N}^*$. On considère le problème

$$\inf_{x \in S} \frac{1}{2} \|x\|_2^2 ,$$

où $S = \{x \in \mathbb{R}^n \,/\, Ax = b\}$ avec $A \in \mathcal{M}_{p,n}(\mathbb{R})$, $b \in \mathbb{R}^p$. On suppose que $p < n$ [1] et $\mathrm{rg}(A) = p$.

1. Montrer que le problème est bien posé et admet une unique solution.

2. La déterminer à l'aide des conditions KKT.

Correction.

1. L'ensemble S est non vide car l'application $x \in \mathbb{R}^n \mapsto Ax$ est surjective puisque $\mathrm{rg}(A) = p$ avec $p < n$. De plus S est fermé en tant que sous espace affine de $\mathbb{R}^n$. Notons $f : x \in \mathbb{R}^n \mapsto \frac{1}{2} \|x\|_2^2$ la fonction objectif du problème d'optimisation, qui est donc $\mathcal{C}^\infty$ sur $\mathbb{R}^n$ et donc en particulier continue. La fonction f est aussi trivialement coercive sur $\mathbb{R}^n$ donc sur S. Par conséquent le problème est bien posée et en particulier f admet un minimum global sur S.

   La fonction f est strictement convexe sur $\mathbb{R}^n$ (par exemple car pour tout $x \in \mathbb{R}^n$, $\nabla^2 f(x) = I_n$ qui est définie positive). L'ensemble S est convexe en tant que sous espace affine de $\mathbb{R}^n$ (si la justification ne vous convainc pas, on peut le vérifier facilement à la main grâce à la linéarité de A). Donc f est également strictement convexe sur S. Ainsi f admet un unique minimum global sur S. Notons le $x^\star$.

2. Commençons par réécrire l'ensemble S sous une forme appropriée au traitement par les outils que nous avons étudiés. On a $S = \{x \in \mathbb{R}^n \,/\, \forall j \in \{1, \ldots, p\},\, h_j(x) = 0\}$ où pour tout $j \in \{1, \ldots, p\}$, $h_j : x \in \mathbb{R}^n \mapsto (Ax-b)_j$ associe à tout $x \in \mathbb{R}^n$ le j-ème coefficient du vecteur $Ax-b \in \mathbb{R}^p$. Pour que cela soit plus agréable du point de vue du calcul différentiel, remarquons que h_j se réécrit : $h_j : x \in \mathbb{R}^n \mapsto \langle Ax - b,\, e_j \rangle$ où $(e_1, \ldots, e_p)$ est la base canonique de $\mathbb{R}^p$. Pour tout j, h_j est $\mathcal{C}^1$ sur $\mathbb{R}^n$ et pour tout $x \in \mathbb{R}^n$, on a $\nabla h_j(x) = A^T e_j \in \mathbb{R}^n$ (c'est la j-ème ligne de la matrice A).

   Recherchons les points réguliers de S (par rapport aux contraintes le définissant choisies au-dessus). Soit $x \in S$. On a donc x point régulier si la famille de vecteurs de $\mathbb{R}^n$: $(\nabla h_1(x), \ldots, \nabla h_p(x))$ est libre, i.e. $(A^T e_1, \ldots, A^T e_p)$ est libre. Or cette famille est la famille des lignes de A et A étant de rang p, l'espace vectoriel engendré par cette famille de vecteurs est donc de dimension p, ainsi elle est donc nécessairement libre (puisque de cardinal p).

> Voici une justification alternative si celle au-dessus ne vous convainc pas. Dire que la famille $(A^T e_1, \ldots, A^T e_p)$ est liée, c'est dire qu'il existe $v \in \mathbb{R}^p$, $v \neq 0$, tel que $A^T v = 0_{\mathbb{R}^n}$. Pour montrer qu'elle est libre, il faut et il suffit donc de prouver que $\ker(A^T) = \{0_{\mathbb{R}^p}\}$. Or $\ker(A^T) = \mathrm{Im}(A)^\perp$. En effet, soit $v \in \ker(A^T)$, alors $A^T v = 0$, mais pour tout $x \in \mathbb{R}^n$, on a
>
> $$\left\langle A^T v,\, x \right\rangle = \langle v,\, Ax \rangle ,$$
>
> d'où $\langle v,\, Ax \rangle = 0$, i.e. $v \in \mathrm{Im}(A)^\perp$. Réciproquement si $v \in \mathrm{Im}(A)^\perp$, alors par la relation au-dessus, on a pour tout $x \in \mathbb{R}^n$, $\left\langle A^T v,\, x \right\rangle = 0$ i.e. $A^T v \in (\mathbb{R}^n)^\perp = \{0_{\mathbb{R}^n}\}$, d'où $v \in \ker(A^T)$. On peut montrer par la même technique que $\mathrm{Im}(A^T) = \ker(A)^\perp$.
> Revenons au raisonnement qui nous intéresse. Puisque $\mathrm{rg}(A) = p$ et $\mathrm{Im}(A) \subset \mathbb{R}^p$, on a donc $\mathrm{Im}(A)^\perp = \{0_{\mathbb{R}^p}\}$. D'où la conclusion attendue.

Ainsi tous les éléments de S sont des points réguliers par rapport aux contraintes. C'est le cas en particulier de $x^\star \in S$.

---

1. puisque $\mathrm{rg}(A) = p$, par le théorème du rang on a nécessairement que $p \leqslant n$. Renforcer cette condition en $p < n$ permet d'éviter la situation triviale où $S = \{A^{-1} b\}$. Ainsi S est au moins une droite affine.

Par conséquent on sait que le minimum global $x^\star$ de f sur S satisfait les conditions KKT. Il existe donc $\mu_1, \dots, \mu_p \in \mathbb{R}$ tels que

$$\nabla f(x^\star) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^\star) = 0_{\mathbb{R}^n},$$

et notant $\mu = (\mu_1, \dots, \mu_p) \in \mathbb{R}^p$, cela revient ainsi à écrire que

$$x^\star = -A^T \mu.$$

> Notez qu'ici, puisqu'il n'y a pas de contraintes d'inégalités définissant S (seulement des contraintes d'égalité), il n'y a pas de conditions d'admissibilité duale et de conditions de complémentarité qui apparaissent.

Mais $x^\star \in S$, donc $Ax^\star = b$, d'où $-AA^T\mu = b$. Or $AA^T \in \mathcal{M}_p(\mathbb{R})$ est inversible comme $\mathrm{rg}(A) = p$.

> En effet, si $v \in \mathbb{R}^p$ tel que $v \in \ker(AA^T)$ alors $AA^Tv = 0_{\mathbb{R}^p}$ et donc $v^TAA^Tv = 0$. Mais $v^TAA^Tv = \langle A^Tv, A^Tv \rangle = \|A^Tv\|_2^2$, d'où $A^Tv = 0$. Or on a vu que $\ker(A^T) = \{0_{\mathbb{R}^p}\}$ (car $\mathrm{rg}(A) = p$), d'où $v = 0_{\mathbb{R}^p}$. Ainsi $\ker(AA^T) = \{0_{\mathbb{R}^p}\}$ et AA^T est bien inversible puisque c'est une matrice carré.

D'où $\mu = -(AA^T)^{-1}b$. Finalement, on obtient que

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

Complément. Géométriquement ce problème d'optimisation revient à trouver le point de S le plus proche de $0_{\mathbb{R}^n}$. S étant convexe fermé non vide, $x^\star$ est donc la projection sur S de $0_{\mathbb{R}^n}$. Ainsi $x^\star = P_S(0_{\mathbb{R}^n}) = A^T(AA^T)^{-1}b = A^T(AA^T)^{-1}Ax_0$, en notant $x_0 \in \mathbb{R}^n$ une solution particulière de l'équation linéaire $Ax = b$.

Or ! Nous avons déjà calculé cette projection P_S dans le TP1 par un raisonnement géométrique. Nous avons montré que pour tout $x \in \mathbb{R}^n$, $P_S(x) = x_0 + p(x - x_0)$ où $p : \mathbb{R}^n \to \mathbb{R}^n$ est la projection orthogonale sur $\ker(A)$ avec $p(x) = x - A^T(AA^T)Ax$. On retrouve bien par cette formule que $P_S(0_{\mathbb{R}^n}) = x_0 - p(x_0) = A^T(AA^T)^{-1}Ax_0$.