Optimisation non linéaire : correction des TD

Grégory Bonnet
Rémi Douvenot
Nicolas Fezans
Emmanuel Rachelson
Stéphanie Roussel

17 janvier 2008

TD 1 : Généralités

Exercice 1 : étude des fonctions quadratiques

Les fonctions quadratiques sont très souvent rencontrées dans les problèmes d'optimisation car il s'agit d'une forme *standard* d'énergie. Soient Q une matrice de $\mathbb{R}^{n\times n}$, b un vecteur de $\mathbb{R}^n$ et $f : \mathbb{R}^n \mapsto \mathbb{R}$ la fonction définie par $f(x) = \frac{1}{2}x^TQx + b^Tx$, $x \in \mathbb{R}^n$.

Différentiabilité au sens de Fréchet

f est différentiable au sens de Fréchet en $\hat{x}$ si $\exists f'_F(\hat{x}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ tel que $\forall v \in \mathbb{R}^n$, $f(\hat{x}+v) = f(\hat{x}) + f'_F(\hat{x})(v) + o(v)$ avec $\lim\limits_{v\mapsto 0} \frac{||o(v)||}{||v||} = 0$.

Il nous faut donc exhiber $f'_F(\hat{x})$ et il suffit de poser l'équation précédente et de la développer :

$$\begin{aligned}
f(x+v) =& \frac{1}{2}(x+v)^TQ(x+v) + b^T(x+v)\\
=&(\frac{1}{2}x^T + \frac{1}{2}v^T)(Qx+Qv) + b^Tx + b^Tv\\
=&\frac{1}{2}x^TQx + \frac{1}{2}x^TQv + \frac{1}{2}v^TQx + \frac{1}{2}v^TQv + b^Tx + b^Tv\\
=&f(x) + \frac{1}{2}x^TQv + \frac{1}{2}v^TQx + \frac{1}{2}v^TQv + b^Tv
\end{aligned}$$

Nous pouvons remarquer que x^TQv est un scalaire (c'est-à-dire $x^TQv \in \mathbb{R}$). Or un scalaire est égal à sa transposée, donc $x^TQv = (x^TQv)^T = (v^TQ^Tx)$.

$$\begin{aligned}
f(x+v) =& f(x) + \frac{1}{2}x^TQv + \frac{1}{2}(v^TQx)^T + \frac{1}{2}v^TQv + b^Tv\\
=&f(x) + \frac{1}{2}x^TQv + \frac{1}{2}x^TQ^Tv + \frac{1}{2}v^TQv + b^Tv\\
=&f(x) + (\frac{1}{2}x^T(Q+Q^T) + b^T)v + \frac{1}{2}v^TQv
\end{aligned}$$

Les équations précédentes nous conduisent alors à poser :

$$f'_F(\hat{x})(v) = (\frac{1}{2}x^T(Q+Q^T) + b^T)v \quad \text{et} \quad o(v) = \frac{1}{2}v^TQv$$

Il nous faut alors vérifier que $f'_F(\hat{x})(v)$ est linéaire en v (ce qui est trivial) ; et il faut nous assurer que $\lim\limits_{v \mapsto 0} \frac{||o(v)||}{||v||} = 0$

En posant $q_m = \max\limits_{i,j \in \{1,\dots,n\}} q_{ij}$: $||o(v)|| = ||\frac{1}{2}v^T Q v|| = |\frac{1}{2}v^T Q v| \leq |\frac{1}{2}q_m ||v||^2|$
Nous avons donc : $0 \leq \frac{||o(v)||}{||v||} \leq |q_m| ||v||^2$
Or $\lim\limits_{v \to 0} |q_m| ||v||^2 = 0$. Donc $\lim\limits_{v \to 0} \frac{||o(v)||}{||v||} = 0$.

f est donc différentiable au sens de Fréchet (et par conséquent au sens de Gâteaux) et $f'_F(\hat{x})(v) = \frac{\partial f}{\partial x}(\hat{x})v$; et par identification :

$$\frac{\partial f}{\partial x}(\hat{x}) = \frac{1}{2}\hat{x}^T(Q + Q^T) + b^T$$

Calcul du Hessien

Calculer le Hessien de f est fondé sur le principe précédent, appliqué au gradient de f. En effet, montrer que f est deux fois différentiable revient à montrer que $f'_F(.)(v)$ est différentiable ; donc que $f'_F(\hat{x} + w)(v) = f'_F(\hat{x})(v) + f''(\hat{x})(v)(w) + o(w)$ avec $f''(\hat{x})(v)(.) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ et $\lim\limits_{w \to 0} \frac{||o(w)||}{w} = 0$.

$$\begin{aligned}
f'_F(\hat{x} + w)(v) =& \frac{\partial f}{\partial x}(\hat{x} + w)v \\
\frac{\partial f}{\partial x}(\hat{x} + w) =& \frac{1}{2}(\hat{x} + w)^T(Q + Q^T) + b^T \\
=& \frac{1}{2}\hat{x}^T(Q + Q^T) + b^T + \frac{1}{2}w^T(Q + Q^T) \\
=& \frac{\partial f}{\partial x}(\hat{x}) + \frac{1}{2}w^T(Q + Q^T)
\end{aligned}$$

On pose alors $f''(\hat{x})(v)(w) = \frac{1}{2}w^T(Q+Q^T)$ et $o(w) = 0$. Il faut alors s'assurer que $f''(\hat{x})(v)(w)$ est linéaire en w (ce qui est trivial) et que $\lim\limits_{w \to 0} \frac{||o(w)||}{||w||} = 0$ (ce qui est trivial aussi) ; ainsi le Hessien de f est

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{2}(Q + Q^T)$$

L'expression de $f''(x)(v)(v)$ est $< \frac{\partial^2 f}{\partial x^2}v, v > = \frac{1}{2}v^T(Q + Q^T)v$

Convexité

Si l'on suppose Q symétrique alors $Q = Q^T$; et $\frac{\partial f}{\partial x} = x^T Q + b^T$ et $\frac{\partial^2 f}{\partial x^2} = Q$. Rappelons qu'une matrice est définie positive (resp. semi-définie positive) si et seulement si ses valeurs propres sont toutes strictement positives (resp. positives ou nulles). Pour montrer que f est convexe, deux méthodes sont proposées.

Une fonction est strictement convexe (resp. convexe) si son Hessien est défini positif (resp. semi-défini positif). Dans notre application, le Hessien est Q et, donc, f est strictement convexe (resp. convexe) si Q est défini positive (resp. semi-défini positive).

En utilisant la proposition donnée dans l'exercice, nous pouvons vérifier ceci d'une autre manière. On pose :

$$\begin{aligned}
A(x,y) =& f(y) - f(x) - \frac{\partial f}{\partial x}(x)(y-x) \\
=& \frac{y^TQy}{2} + b^Ty - \frac{x^TQx}{2} - b^Tx - x^TQ(y-x) - b^T(y-x) \\
=& \frac{y^TQy}{2} - \frac{x^TQx}{2} + x^TQx - x^TQy \\
=& \frac{y^TQy}{2} + \frac{x^TQx}{2} - x^TQy
\end{aligned}$$

Or $x^TQy = \frac{x^TQy}{2} + \left(\frac{x^TQy}{2}\right)^T = \frac{x^TQy}{2} + \frac{y^TQx}{2}$.

Donc $A(x,y) = \frac{1}{2}(y-x)^TQ(y-x)$ et donc $A(x,y) \geq 0$ si Q est semi-définie positive ($A(x,y) > 0$ si Q est définie positive).

Exercice 2 : Approximation de fonction

Dans cet exercice, nous cherchons un polynôme h qui approxime la fonction g au sens des moindres carrés. Cela signifie que h doit minimiser la grandeur suivante :

$$f = \sum_{i=1}^{m} [g(x_i) - h(x_i)]^2$$

Résolution générale du problème

Cet exercice se prête bien à un traitement matriciel. On pose :

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_m) \end{bmatrix} = \begin{bmatrix} x_1^0 & \cdots & x_1^n \\ \vdots & & \vdots \\ x_m^0 & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = Xa \quad \text{et} \quad G = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_m) \end{bmatrix}$$

On a alors :

$$\begin{aligned}
f =& (H-G)^T(H-G) \\
=& H^TH + G^TG - 2G^TH (\text{car } G^TG \text{ est un scalaire}) \\
=& a^TX^TXa + G^TG - 2G^TXa
\end{aligned}$$

Par définition, X^TX est une matrice symétrique ; et en posant $Q = 2X^TX$ et $b = -2X^TG$, on se ramène à la fonction quadratique de l'exercice prédécent. Donc :

$$\frac{\partial f}{\partial a}(a) = a^TQ + b^T$$

La condition du premier ordre nous donne :

$$a^T Q + b^T = 0 \implies a = -b^T Q^{-1} = -Q^{-T} b$$

La matrice $X^T X$ est diagonalisable. En effet, $Q = X^T X$:

$$Q = \begin{bmatrix} x_1^0 & \cdots & x_m^0 \\ \vdots & & \vdots \\ x_1^n & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} x_1^0 & \cdots & x_1^n \\ \vdots & & \vdots \\ x_m^0 & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}$$

Donc $q_{ij} = \sum\limits_{k=1}^{m} x_k^{i+j-2}$: Q est bien symétrique. Par ailleurs, $\forall v \in \mathbb{R}^{n+1}$, $v^T Q v = v^T X^T X v = (Xv)^T (Xv) = ||Xv|| \geq 0$.

Donc $X^T X$ est semi-définie positive ; donc Q est semi-définie positive (car $Q = 2X^T X$). L'optimum a trouvé est donc bien un mimimum au sens des moindres carrés.

Application numérique

Dans l'application :

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \text{et} \quad G = \begin{bmatrix} 8 \\ 25 \\ 52 \end{bmatrix}$$

On obtient :

$$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \quad \text{et} \quad X^T G = \begin{bmatrix} 85 \\ 214 \end{bmatrix}$$

Nous cherchons $a = -Q^{-T} b$, donc Q^{-1} et $-b$:

$$Q^{-1} = (2X^T X)^{-1} = \frac{1}{12} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = Q^{-T} \quad \text{et} \quad -b = 2X^T G = \begin{bmatrix} 170 \\ 428 \end{bmatrix}$$

Par conséquent :

$$a = (2X^T X)^{-1} (2X^T G) = \begin{bmatrix} -\frac{47}{3} \\ 22 \end{bmatrix}$$

Donc, $h(x) = 22x - \frac{47}{3}$.

Nous pouvons vérifier la validité de cette approximation car $h(1) = 6,33$, $h(2) = 28,33$ et $h(3) = 50,33$ pour $g(1) = 8$, $g(2) = 25$ et $g(3) = 52$.

Exercice 3 : Production

Remarques d'ordre général pour l'optimisation

Dans cet exercice, nous cherchons à maximiser une quantité. Notre problème peut donc être modélisé de la façon suivante :

$$\max_{x_i \geq 0} qh(x_1 \dots x_n) - \sum_{i=1}^{n} p_i x_i$$

Nous négligerons les contraintes de positivité ; et nous les vérifierons *a posteriori*. En effet, chaque x_i est nécessairement positif car il représente une quantité de matière première.

La condition du premier ordre pour trouver un optimum est d'annuler le gradient.

$$\forall i \in \{1, \dots, n\}, \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = q\frac{\partial h(x_1 \dots x_n)}{\partial x_i} - p_i$$

La condition du second ordre est d'avoir un Hessien semi-défini négatif (resp. défini négatif) afin d'avoir un maximum (resp. strict). Ici,

$$\frac{\partial^2 f}{\partial x^2} = q \begin{bmatrix} \frac{\partial^2 h}{\partial x^2} & \dots & \frac{\partial^2 h}{\partial x_1 x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 h}{\partial x_n x_1} & \dots & \frac{\partial^2 h}{\partial x_n^2} \end{bmatrix}$$

Application 1

Soit $h(x_1, x_2) = x_1^{\frac{1}{3}} x_2^{\frac{1}{3}}$. La condition du premier ordre nous donne :

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{q}{3} x_1^{-\frac{2}{3}} x_2^{\frac{1}{3}} - p_1 \\ \frac{q}{3} x_1^{\frac{1}{3}} x_2^{-\frac{2}{3}} - p_2 \end{bmatrix} = 0$$

Il nous faut résoudre le système :

$$\begin{cases} \frac{q}{3} x_1^{-\frac{2}{3}} x_2^{\frac{1}{3}} - p_1 = 0 \\ \frac{q}{3} x_1^{\frac{1}{3}} x_2^{-\frac{2}{3}} - p_2 = 0 \end{cases} \iff \begin{cases} x_2^{\frac{1}{3}} = \frac{3p_1 x_1^{\frac{2}{3}}}{q} \\ x_1^{\frac{1}{3}} = \frac{3p_2 x_2^{\frac{2}{3}}}{q} \end{cases} \iff \begin{cases} x_2 = \frac{27 p_1^3 x_1^2}{q^3} & (1) \\ x_1 = \frac{27 p_2^3 x_2^2}{q^3} & (2) \end{cases}$$

On injecte (1) dans (2) :

$$\begin{aligned} x_1 =& 27 p_2^3 (27 p_1^3 x_1^2 \frac{1}{q^3})^2 \frac{1}{q^3} \\ =& 27 p_2^3 (27^2 p_1^6 x_1^4 \frac{1}{q^6}) \frac{1}{q^3} \\ =& 27^3 p_1^6 p_2^3 x_1^4 \frac{1}{q^{18}} \end{aligned}$$

$$x_1^3 = \frac{q^{18}}{27^3 p_1^6 p_2^3} \iff x_1 = \frac{q^3}{27 p_1^2 p_2}$$

Le raisonnement est identique pour x_2 ; et avec $p_1 = p_2 = 1$, $x_1 = x_2 = \frac{q^3}{27}$.

La condition du second ordre nous donne :

$$\frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} \frac{q}{9}(-2x_1^{-\frac{5}{3}} x_2^{\frac{1}{3}}) & \frac{q}{9}(x_1^{-\frac{2}{3}} x_2^{-\frac{2}{3}}) \\ \frac{q}{9}(x_1^{-\frac{2}{3}} x_2^{-\frac{2}{3}}) & \frac{q}{9}(-2x_1^{\frac{1}{3}} x_2^{-\frac{5}{3}}) \end{bmatrix}$$

Au point $\hat{x}$ considéré, le Hessien est :

$$\frac{\partial^2 f}{\partial x^2}(\hat{x}) = \frac{9}{q^3}\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

Ici, le déterminant est positif (il vaut 3) et la trace est négative (elle vaut -4) ; donc le produit des valeurs propres est positif et leur somme est négative. Par conséquent les deux valeurs propres sont du même signe et de somme négative : les deux valeurs propres sont négatives.

Si l'on veut s'en convaincre par le calcul, le polynôme caractéristique vaut :

$$P(\lambda) = \begin{vmatrix} -2-\lambda & 1 \\ 1 & -2-\lambda \end{vmatrix} = (2+\lambda)^2 - 1 = (\lambda+1)(\lambda+3)$$

Application 2

Soit $h(x_1, x_2) = x_1^2 + x_2^2$. La condition du premier ordre nous donne :

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 2qx_1 - p_1 \\ 2qx_2 - p_2 \end{bmatrix} = 0$$

Nous avons donc $x_1 = \frac{p_1}{2q}$ et $x_2 = \frac{p_2}{2q}$.

La condition du second ordre nous donne :

$$\frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} 2q & 0 \\ 0 & 2q \end{bmatrix}$$

Or, le Hessien est défini positif. Nous avons trouvé un minimum alors que nous cherchions un maximum.

Dans cette application, le bon sens peut nous permettre de trouver directement ce résultat. En effet, on peut produire indifféremment le produit final avec du produit 1 ou du produit 2. Or le produit 2 coûte moins cher : on ne va donc utiliser que du produit 2.

De plus, avec $p_2 = 1$ et $q = 1$, on a un bénéfice de $x_2^2 - x_2$. Or cette quantité est décroissante et bornée sur $[0, 1]$ tandis qu'elle est croissante sur $[1, +\infty[$ et $\lim\limits_{x_2 \to \infty} x_2^2 - x_2 = \infty$.

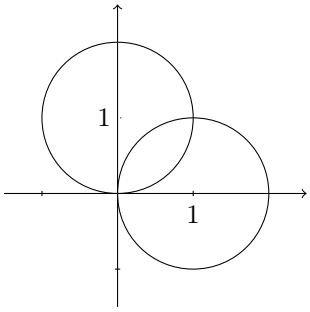
Plus on augmente la production, plus on augmente le bénéfice. Le maximum est atteint en $x_2 = +\infty$. Il n'y a donc pas de maximum réel au problème. Cela traduit graphiquement le fait que l'on cherche le maximum d'une fonction convexe (coercive) : il est en $+\infty$. On ne peut trouver des maxima dans $\mathbb{R}$ que pour des fonctions (localement) concaves.

TD 2 : Qualification

Exercice 1 : directions et cônes admissibles

Premier domaine

$$\Omega = \{(x, y) \in \mathbb{R}^2 : (x-1)^2 + y^2 - 1 = 0 \text{ et } x^2 + (y-1)^2 - 1 = 0\}$$



(Approche graphique) Qu'est-ce qu'une direction admissible ? Considérons un domaine dans $\mathbb{R}^n$ et un point donné $\hat{x}$ de ce domaine, une direction est admissible si l'on peut se déplacer dans cette direction à partir du point $\hat{x}$ sans sortir du domaine. L'ensemble des directions admissibles en un point est noté $\mathcal{D}_{\hat{x}}(\Omega)$.

Pour notre exercice, nous pouvons remarquer que $\Omega = \{(0,0)(1,1)\}$. Ainsi au voisinage de $(0,0)$, le domaine est vide. Il est donc impossible de se déplacer sur le domaine dans une direction autre que le vecteur nul. Ainsi $\mathcal{D}_{(0,0)}(\Omega) = \{(0,0)\}$. En effet, le vecteur nul est toujours une direction admissible (il est toujours possible de ne pas se déplacer).

(Approche par la définition) Bien entendu, une direction admissible est définie formellement.

Définition 1 (Direction admissible)**.** *Soit un domaine $\Omega \subseteq \mathbb{R}^n$ et un point $\hat{x}$ de ce domaine. Le vecteur $v \in \mathbb{R}^n$ est une direction admissible si et seulement si $\exists \varphi \in \mathcal{C}^1(\mathbb{R}^+, \mathbb{R}^n)$ telle que :*

- $\varphi(0) = \hat{x}$
- $\varphi'(0) = v$
- $\exists \eta > 0, \forall 0 < t < \eta, \varphi(t) \in \Omega$

Si l'on se considère au point $(0,0)$, par définition $\varphi(0) = (0,0)$. Comme φ est continue et que, pour t petit, $\varphi(t) \in \Omega$, la seule fonction φ respectant ces deux propriétés est $\varphi(t) = (0,0)$. On peut s'en convaincre en raisonnant par l'absurde :

Supposons qu'il existe une direction admissible $v \neq (0,0)$. La formule de Taylor appliquée au φ correspondant nous indique que $\varphi(t) = \varphi(0) + \varphi'(0).t + o(t)$. Or $\varphi(t) \in \Omega$. Donc $\varphi(t) = (0,0)$; par conséquent $vt + o(t) = (0,0)$. On sait que $\lim\limits_{t \to 0} \frac{o(t)}{t} = (0,0)$. Or $\frac{o(t)}{t} = v$ et $v \neq (0,0)$ par hypothèse. C'est absurde : il n'existe donc pas de direction admissible non nulle. Un raisonnement identique peut être tenu si l'on se considère au point $(1,1)$

(Approche par la régularité) La notion de régularité ne s'applique qu'aux contraintes égalités. On dit que les contraintes sont régulières en un point $\hat{x}$ si et seulement si leurs gradients en ce point forment une famille libre. Si l'on considère des contraintes régulières alors $\mathcal{D}_{\hat{x}}(\Omega) = \ker(\nabla g(\hat{x}))$ où g est le vecteur des contraintes égalités g_i. Dans notre exercice, nous avons deux contraintes : $g_1(x,y) = (x-1)^2 + y^2 - 1$ et $g_2(x,y) = x^2 + (y-1)^2 - 1$. Les gradients des contraintes sont :

$$\nabla g_1(x,y) = \begin{bmatrix} 2(x-1) \\ 2y \end{bmatrix} \quad \text{et} \quad \nabla g_2(x,y) = \begin{bmatrix} 2x \\ 2(y-1) \end{bmatrix}$$

et au point considéré :

$$\nabla g_1(0,0) = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \quad \text{et} \quad \nabla g_2(0,0) = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

Comme il y a indépendance linéaire entre ces deux vecteurs, les contraintes sont régulières ; donc :

$$\mathcal{D}_{(0,0)}(\Omega) = \ker \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} = \{(0,0)\}$$

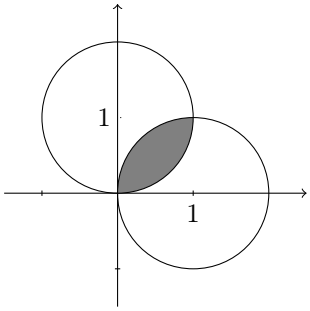
Deuxième domaine

$$\Omega = \{(x,y,z) \in \mathbb{R}^3 : (x-1)^2 + y^2 - 1 = 0 \text{ et } x^2 + (y-1)^2 - 1 = 0\}$$

Si $\Omega \subset \mathbb{R}^3$ alors $\Omega = \{(0,0,z),(1,1,z)\}$. En se plaçant au point $\hat{x} = (0,0,z)$ ou $\hat{x} = (1,1,z)$, on ne peut se déplacer sans sortir du domaine que le long de l'axe z. Par conséquent $\mathcal{D}_{\hat{x}}(\Omega) = \{(0,0,z) : z \in \mathbb{R}\}$

Troisième domaine

$$\Omega = \{(x,y) \in \mathbb{R}^2 : (x-1)^2 + y^2 - 1 \leq 0 \text{ et } x^2 + (y-1)^2 - 1 \leq 0\}$$



(Approche graphique) En se plaçant en $\hat{x} = (0,0)$, on voit que l'on peut se déplacer dans les directions $(v_1, v_2) \in \mathbb{R}^2$ si $v_1 > 0$ et $v_2 > 0$. Qu'en est-il des axes ? Le cône des directions admissibles est l'enveloppe convexe de $\mathcal{D}_{\hat{x}}(\Omega)$ donc les demi-axes sont inclus. En d'autres termes, le cône d'admissibilité étant un fermé, les tangentes à $\mathcal{D}_{\hat{x}}(\Omega)$ y sont incluses. Nous allons voir formellement que $\mathcal{C}_{(0,0)}(\Omega) = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 \geq 0, v_2 \geq 0\}$.

(Approche par la définition) Nous cherchons les $(v_1, v_2) \in \mathbb{R}^2$ tels que $\varphi(0) = (0,0)$, $\varphi'(0) = (v_1, v_2)$ et $\varphi(t) \in \Omega$ pour tout t petit et positif. Nous allons donc chercher à écrire une fonction φ qui convient (satisfait les propriétés de la Définition **??**) pour les différentes valeurs de (v_1, v_2), c'est-à-dire pour v_i strictement positif, pour $v_1 = 1$ et $v_2 = 0$, pour $v_1 = 0$ et $v_2 = 1$ et pour un des v_i négatif.

Pour $v_1 > 0$ et $v_2 > 0$, nous pouvons poser $\varphi(t) = (v_1, v_2)t$. En effet, φ est continue ; $\varphi(0) = (0,0)$, $\varphi'(t) = (v_1, v_2)$ (donc $\varphi'(0) = (v_1, v_2)$) et pour t petit, $\varphi(t) \in \Omega$.

Pour $v_1 = 1$ et $v_2 = 0$, nous pouvons prendre par exemple $\varphi(t) = (t, t^2)$. En effet, φ est continue ; $\varphi(0) = (0,0)$ et $\varphi'(t) = (1, 2t)$ (donc $\varphi'(0) = (1,0)$). Il nous reste à prouver que pour t petit, $\varphi(t)$ reste dans le domaine. Il est trivial de le vérifier pour la premier contrainte. Quant à la seconde, on a (avec $x = t$ et $y = t^2$:

$$t^2 + (t^2 - 1)^2 - 1 = t^2 + t^4 - 2t^2 + 1 - 1 = t^4 - t^2$$

Cette quantité est négative pour $0 < t < 1$: la contrainte est donc bien vérifiée. Par conséquent, $\varphi(t) \in \Omega$. Un raisonnement identique pour $v_1 = 0$ et $v_2 = 1$ où $\varphi(t) = (t^2, t)$ peut être tenu.

Pour $v_1 < 0$ ou $v_2 < 0$, raisonnons par l'absurde en supposant qu'il existe une fonction φ admissible. On peut alors écrire pour φ et t tendant vers 0 le développement de Taylor à l'ordre 1 :

$$\begin{aligned}\varphi(t) =& \varphi(0) + \frac{\partial \varphi}{\partial t}(0).t + o(t)\\ \varphi(t) =& \varphi(0) + \varphi'(0).t + o(t)\end{aligned}$$

Or $\varphi'(0) = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$; et par définition du reste de Taylor : $\lim\limits_{t \to 0} \frac{||o(t)||}{t} = 0$.

Donc $\exists \eta > 0 : \forall t < \eta$, $\frac{||o(t)||}{t} < \min(|v_1|, |v_2|)$; donc $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.t + o(t)$ a une composante négative ; donc $\forall\ \varphi \in \mathcal{C}^1(\mathbb{R}^+, \mathbb{R}^n)$, $\varphi(t) \notin \Omega$.

Nous pouvons en déduire que $\mathcal{C}_{(0,0)}(\Omega) = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 \geq 0 \text{ et } v_2 \geq 0\}$.

(Approche par la qualification) Les contraintes sont qualifiées si elles sont convexes et que l'intérieur du domaine est non vide. Il existe d'autres propriétés, vues en cours, permettant de conclure à la qualification des contraintes. Ici, le Hessien des contraintes est le suivant :

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \geq 0$$

Le Hessien H est défini positif : les contraintes sont convexes ; et il est trivial de montrer que l'intérieur du domaine est non vide. Les contraintes sont donc qualifiées.

Si les contraintes h_i sont qualifiées, nous savons que :

$$\mathcal{C}_{\hat{x}}(\Omega) = G_{\hat{x}}(\Omega) = \{v \in \mathbb{R}^n : \forall i \in \mathcal{A}(\hat{x}), \frac{\partial h_i}{\partial x}(\hat{x}).v \leq 0\}$$

(Rappel) : $\mathcal{A}(\hat{x})$ est l'ensemble des contraintes actives (ou saturées) au point considéré. Une contrainte h_i est saturée en $\hat{x}$ si $h_i(\hat{x}) = 0$.

Dans notre exercice, les deux contraintes sont saturées en $(0, 0)$. Nous cherchons donc : $\{(v_1, v_2) \in \mathbb{R}^2 : \frac{\partial h_i}{\partial x}(\hat{x}).(v_1, v_2) \leq 0\}$

$$\begin{cases} \begin{bmatrix} -2 & 0 \end{bmatrix} . \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \leq 0 \\ \begin{bmatrix} 0 & -2 \end{bmatrix} . \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \leq 0 \end{cases} \implies \begin{cases} -2v_1 \leq 0 \\ -2v_2 \leq 0 \end{cases}$$

Par conséquent, $v_1 \geq 0$ et $v_2 \geq 0$; et :

$$\mathcal{C}_{(0,0)}(\Omega) = G_{(0,0)}(\Omega) = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 \geq 0 \text{ et } v_2 \geq 0\}$$

Exercice 2

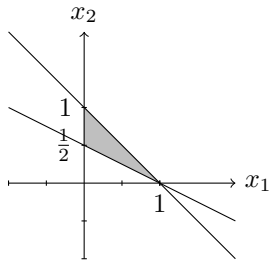
Soit le problème d'optimisation suivant :

$$(P) : \min x \in \Omega_i(-x_1)$$

$$\Omega_1 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 - 1 \leq 0, -x_1 - 2x_2 + 1 \leq 0, -x_1 \leq 0, -x_2 \leq 0\}$$

$$\Omega_2 = \{(x_1, x_2) \in \mathbb{R}^2 : (x_1 + x_2 - 1)(x_1 + 2x_2 - 1) \leq 0, -x_1 \leq 0, -x_2 \leq 0\}$$

Approche graphique



Nous pouvons remarquer que les deux ensembles Ω_1 et Ω_2 correspondent graphiquement à un même domaine. Il s'agit donc d'un domaine identique exprimé de deux manières différentes. Nous allons voir que la manière d'écrire les contraintes va avoir une influence sur la résolution du problème.

Ici, nous pouvons voir que la solution du problème (P) est $\hat{x} = (1, 0)$.

Cônes admissibles pour Ω_1

Déterminons $G_{(1,0)}(\Omega_1)$. En ce point, il y a trois contraintes saturées :

$$\begin{aligned} h_1 &: x_1 + x_2 - 1 \leq 0 \\ h_2 &: -x_1 - 2x_2 + 1 \leq 0 \\ h_3 &: -x_2 \leq 0 \end{aligned}$$

Calculons les gradients des contraintes :

$$\frac{\partial h_1}{\partial x}(\hat{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \frac{\partial h_2}{\partial x}(\hat{x}) = \begin{bmatrix} -1 \\ -2 \end{bmatrix} \qquad \frac{\partial h_3}{\partial x}(\hat{x}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Ainsi :

$$G_{(1,0)}(\Omega_1) = \{\begin{bmatrix} 1 \\ 1 \end{bmatrix}.v \leq 0\} \cap \{\begin{bmatrix} -1 \\ -2 \end{bmatrix}.v \leq 0\} \cap \{\begin{bmatrix} 0 \\ -1 \end{bmatrix}.v \leq 0\}$$

Nous pouvons remarquer que chaque sous-ensemble correspond à l'expression d'une des contraintes saturées. Comme $v_1 + v_2 \leq 0$ et $-v_1 - 2v_2 \leq 0$, on a nécessairement $-v_2 \leq 0$. La troisième contrainte est toujours vérifiée. Donc :

$$G_{(1,0)}(\Omega_1) = \{\begin{bmatrix} 1 \\ 1 \end{bmatrix}.v \leq 0\} \cap \{\begin{bmatrix} -1 \\ -2 \end{bmatrix}.v \leq 0\}$$

Nous vérifions bien que $G_{(1,0)}(\Omega_1) = \mathcal{C}_{(1,0)}(\Omega_1)$. Les contraintes sont donc qualifiées, confirmant ainsi le théorème de qualification (voir cours).

Cône admissible pour Ω_2

Déterminons $G_{(1,0)}(\Omega_2)$. En ce point, il y a deux contraintes saturées :

$$\begin{aligned} h_1 &: (x_1 + x_2 - 1)(x_1 + 2x_2 - 1) \leq 0 \\ h_2 &: -x_2 \leq 0 \end{aligned}$$

Calculons les gradients des contraintes :

$$\frac{\partial h_1}{\partial x} = \begin{bmatrix} 2x_1 + 3x_2 - 2 \\ 3x_1 + 4x_2 - 3 \end{bmatrix} \implies \frac{\partial h_1}{\partial x}(\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \frac{\partial h_2}{\partial x}(\hat{x}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Ainsi :

$$\mathcal{G}_{(1,0)}(\Omega_2) = \{\begin{bmatrix} 0 \\ 0 \end{bmatrix}.v \leq 0\} \cap \{\begin{bmatrix} 0 \\ -1 \end{bmatrix}.v \leq 0\} = \{(0,0)\}$$

Nous pouvons remarquer que l'ensemble $G_{(1,0)}(\Omega_2)$ n'est pas égal au cône des directions admissibles (identique à $\mathcal{D}_{(1,0)}(\Omega_1)$) : les contraintes ne sont pas qualifiées.

Interprétation

Pour un même problème, la manière d'écrire les contraintes va influer sur leur qualification. Ceci est important car les contraintes doivent être qualifiées pour que l'on puisse appliquer les théorèmes vus en cours. Cette application illustre le fait que la qualification des contraintes dépend de leur écriture et n'est pas une propriété intrinsèque de l'ensemble des points décrits par Ω.

En d'autres termes, bien poser le problème c'est l'avoir déjà résolu.

Exercice 3

Considérons la fonction paramétrée :

$$f_m(x) = mx_1^3 + \frac{3}{2}x_1^2 + x_2^2 + 3x_3^2 - 2x_2$$

Optimisation sans contrainte

Dans cet exercice, nous chercherons à optimiser la fonction sans contrainte puis nous vérifierons *a posteriori* que l'optimum se trouve dans le domaine. Cherchons donc une solution $\hat{x} \in \mathbb{R}^3$ en satisfaisant les conditions du premier et du second ordre.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 3mx_1^2 + 3x_1 \\ 2x_2 - 2 \\ 6x_3 \end{bmatrix} = 0$$

Nous avons deux cas, $m = 0$ et $m \neq 0$. Si $m = 0$ alors $x_1 = 0$, $x_2 = 1$ et $x_3 = 0$. Si $m \neq 0$ alors $x_1 = 0$ ou $x_1 = -\frac{1}{m}$, $x_2 = 1$ et $x_3 = 0$.

$$\frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} 6mx_1 + 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

Si $m = 0$ alors le Hessien est défini positif et $(0, 1, 0)$ est un minimum global de f_m. Si $m \neq 0$ alors le Hessien est défini positif pour $(0, 1, 0)$ mais pas pour $(-\frac{1}{m}, 1, 0)$. Par conséquent, la solution $\hat{x}$ est $(0, 1, 0)$.

Satisfaction des contraintes

Soit $\Omega = \{x \in \mathbb{R}^3 : x_2 > 0\}$. La solution $\hat{x}$ trouvée est bien dans le domaine Ω.

Soit $\Omega = \{x \in \mathbb{R}^3 : x_2 > 1\}$. La solution $\hat{x}$ trouvée est sur la frontière du domaine mais n'en fait pas partie. Il n'existe pas de solution au problème car – quel que soit le point considéré – on peut toujours trouver un point du domaine "plus optimal".

Soit $\Omega = \{x \in \mathbb{R}^3 : x_2 \geq 2\}$. La solution $\hat{x}$ trouvée n'appartient pas au domaine. Cependant, il est possible de trouver un point sur la frontière du domaine (lui appartenant) qui minimise pourtant la fonction. Optimiser une fonction et vérifier *a posteriori* que les contraintes sont satisfaites ne suffisent pas pour trouver cet optimum : les contraintes doivent être intégrées à la résolution du problème.

TD 3 : Problèmes avec contraintes égalités

Exercice 1

Modélisation

Notre problème est le suivant :

$$(P) : \min f(x_1, x_2) = 2x_1 + 3x_2 - 1 \quad \text{avec} \quad (x_1, x_2) \in \mathbb{R}^2$$
$$\text{sous la contrainte} \quad g(x_1, x_2) = x_1^2 + \frac{3}{2}x_2^2 - 6 = 0$$

Il nous faut dans un premier temps vérifier la régularité des contraintes (puisqu'il s'agit de contraintes égalités) afin de pouvoir utiliser les théorèmes vus en cours. Les contraintes égalités sont *régulières* si leurs gradients forment une famille libre. Ici, nous n'avons qu'une seule contrainte : son gradient est un seul vecteur qui est toujours libre (s'il est non nul).

Ici, $\frac{\partial g}{\partial x}(x) = \begin{bmatrix} 2x_1 \\ 3x_2 \end{bmatrix}$.

$\frac{\partial g}{\partial x}(x)$ s'annule en $(0, 0)$ mais ce point ne fait pas partie du domaine : les contraintes sont donc régulières sur tout Ω.

Nous pouvons alors poser le Lagrangien :

$$\mathcal{L}(x_1, x_2, \lambda) = 2x_1 + 3x_2 - 1 + \lambda(x_1^2 + \frac{3}{2}x_2^2 - 6)$$

Condition du premier ordre

En un point $\hat{x}$ solution du problème, il existe un vecteur $\hat{\lambda}$ (dont les composantes sont les multiplicateurs de Lagrange) tel que le gradient du Lagrangien en $(\hat{x}, \hat{\lambda})$ est nul. Si les contraintes sont régulières, le vecteur $\hat{\lambda}$ est unique. Dans notre exercice, nous n'avons qu'une contrainte : $\hat{\lambda}$ est un scalaire.

$$\frac{\partial \mathcal{L}}{\partial x} = \begin{bmatrix} 2 + 2\lambda x_1 \\ 3 + 3\lambda x_2 \end{bmatrix} = 0$$

Par conséquent, $x_1 = x_2 = -\frac{1}{\lambda}$

Pour trouver λ, injectons le résultat précédent dans la contrainte g :

$$\begin{aligned}\frac{1}{\lambda^2} + \frac{3}{2\lambda^2} - 6 =& 0\\ \frac{5}{2\lambda^2} =& 6\\ \lambda^2 =& \frac{5}{12}\\ \hat{\lambda} =& \pm\sqrt{\frac{5}{12}}\end{aligned}$$

Nous avons donc deux candidats à l'optimalité $\hat{x} = (\sqrt{\frac{12}{5}}, \sqrt{\frac{12}{5}})$ et $\hat{x}' = (-\sqrt{\frac{12}{5}}, -\sqrt{\frac{12}{5}})$.

La condition du premier ordre ne suffit pas pour conclure car les contraintes ne sont pas affines.

Condition du second ordre

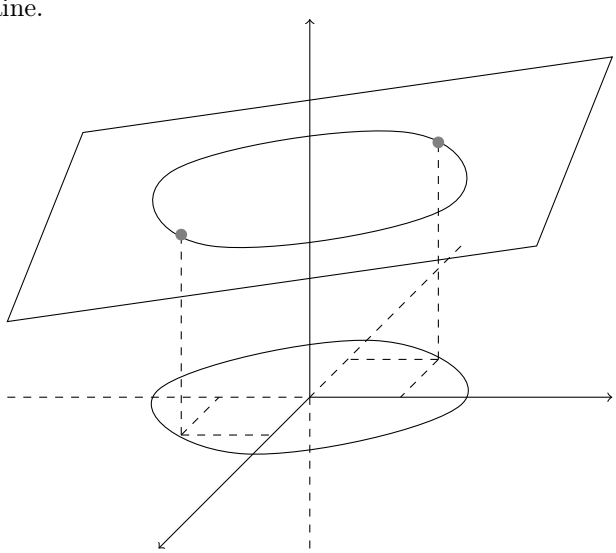
Au deuxième ordre, le Hessien doit être défini positif dans le cône admissible $\mathcal{D}_{\hat{x}}$. Plus formellement, $\forall\ v \in \mathcal{D}_{\hat{x}}$, $v^T \frac{\partial^2 \mathcal{L}}{\partial x^2}(\hat{x}) v \geq 0$.

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(\hat{x}) = \begin{bmatrix} 2\lambda & 0 \\ 0 & 3\lambda \end{bmatrix}$$

Le Hessien est déjà diagonal. Nous pouvons remarquer que si $\hat{\lambda}$ est négatif, $\forall\ v \in \mathcal{D}_{\hat{x}}$, $v^T \frac{\partial^2 \mathcal{L}}{\partial x^2} v \leq 0$; tandis que pour $\hat{\lambda}$ positif, $\forall\ v \in \mathcal{D}_{\hat{x}'}$, $v^T \frac{\partial^2 \mathcal{L}}{\partial x^2} v \geq 0$.

Ainsi, seule la solution $\hat{x}'$ vérifie la condition du second ordre. Nous sommes en présence d'un minimum ; et $(x_1, x_2) = (-\sqrt{\frac{12}{5}}, -\sqrt{\frac{12}{5}})$.

Comme l'illuste la figure, l'autre point correspond au maximum sur le domaine.



Exercice 2

Il nous faut trouver le pavé de dimension x, y, z dont le volume est maximum et de surface donnée c. ‘

Modélisation

Nous pouvons modéliser notre problème de la manière suivante :

$$\begin{aligned}(P) : \min f(x,y,z) = -xyz \quad &\text{avec} \quad (x,y,z) \in \mathbb{R}^3 \\ \text{sous la contrainte} \quad &g(x,y,z) = (xy+yz+xz) - \frac{c}{2} = 0\end{aligned}$$

Nous devons y ajouter trois autres contraintes : x, y, z doivent être strictement positif ; donc $(x,y,z) \in \mathbb{R}^{+3}$. Ce domaine est ouvert et nous ne pouvons optimiser dessus. Il nous faut raisonner sur un fermé et vérifier *a posteriori* que la solution se trouve dans le domaine (et non pas sur la frontière).

Comme dans l'exercice précédent, il nous faut vérifier la régularité des contraintes. Comme il n'y a qu'une seule contrainte, le gradient des contraintes est un unique vecteur (non nul sur Ω), toujours libre. Les contraintes sont régulières. Nous pouvons poser le Lagrangien. Posons $X = (x,y,z)$

$$\mathcal{L}(X,\lambda) = -xyz + \lambda(xy+yz+xz) - \lambda\frac{c}{2}$$

Condition du premier ordre

$$\frac{\partial \mathcal{L}}{\partial X} = \begin{bmatrix} -yz + \lambda(y+z) \\ -xz + \lambda(x+z) \\ -yx + \lambda(x+y) \end{bmatrix} = 0$$

Il nous faut résoudre le système :

$$\begin{cases} -yz + \lambda(y+z) = 0 & (1) \\ -xz + \lambda(x+z) = 0 & (2) \\ -yx + \lambda(x+y) = 0 & (3) \\ xy + xz + zx - \frac{c}{2} = 0 & (4) \end{cases}$$

On écrit $(1) - (2)$:

$$\begin{aligned} -yz + \lambda(y+z) + xz - \lambda(x+z) &= 0 \\ xz - yz + \lambda y - \lambda x &= 0 \\ z(x-y) + \lambda(y-x) &= 0 \\ (z-\lambda)(x-y) &= 0 \end{aligned}$$

Nous avons deux solutions : $z = \lambda$ ou $x = y$. Dans le premier cas, remplaçons z dans (1) :

$$-\lambda y + \lambda y - \lambda^2 = 0 \quad \Longrightarrow \quad \lambda = 0 \quad \text{et} \quad z = 0$$

La dimension z ne peut pas être nulle ; donc $x = y$. Remplaçons y dans (3) :

$$x^2 - 2\lambda x = 0 \implies x = y = 2\lambda$$

Remplaçons dans (2) :

$$-2\lambda z + 2\lambda^2 + \lambda z = 0 \implies \lambda z = 2\lambda^2 \implies z = 2\lambda = x = y$$

Notre solution est un cube. Déterminons maintenant λ en l'injectant dans la contrainte :

$$(4\lambda^2 + 4\lambda^2 + 4\lambda^2) - \frac{c}{2} = 0 \implies 24\lambda^2 = c \implies \lambda = \pm\sqrt{\frac{c}{24}} = \pm\frac{1}{2}\sqrt{\frac{c}{6}}$$

Comme x, y, z doivent être positifs, λ doit être positif ; donc $\lambda = \frac{1}{2}\sqrt{\frac{c}{6}}$ et $x = y = z = 2\lambda = \sqrt{\frac{c}{6}}$

La condition du premier ordre ne suffit pas pour conclure car les contraintes ne sont pas affines.

Condition du second ordre

Le Hessien doit être défini positif dans toutes les directions admissibles à partir du point solution $\hat{X}$. Ici,

$$\frac{\partial^2 \mathcal{L}}{\partial X^2}(\hat{X}) = -\begin{bmatrix} 0 & \lambda & 0 \\ \lambda & 0 & \lambda \\ \lambda & \lambda & 0 \end{bmatrix} = -\lambda \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Les valeurs propres de $\frac{\partial^2 \mathcal{L}}{\partial X^2}(\hat{X})$ sont $\{\lambda, \lambda, -2\lambda\}$. On ne peut conclure à la positivité du Hessien sur $\mathbb{R}^3$. Il faut donc regarder plus précisement l'ensemble des directions admissibles en $\hat{X}$. Soit $v \in \mathcal{D}_{\hat{X}}(\Omega)$. Comme les contraintes sont régulières, $v \in \ker \nabla g(\hat{X})$. Ici,

$$\frac{\partial g}{\partial X}(\hat{X}) = 4\lambda \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Donc $v = (v_1, v_2, v_3) \in \mathcal{D}_{\hat{X}}(\Omega)$ si et seulement si $v_1 + v_2 + v_3 = 0$; et nous pouvons examiner $v^T \frac{\partial^2 \mathcal{L}}{\partial X^2}(\hat{X})v$:

$$v^T \frac{\partial^2 \mathcal{L}}{\partial X^2}(\hat{X})v = -2\lambda \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$
\begin{aligned}
v^T \frac{\partial^2 \mathcal{L}}{\partial X^2}(\hat{X})v =& -\lambda(v_1(v_2+v_3)+v_2(v_1+v_3)+v_3(v_1+v_2)) \\
=& -\lambda(-(v_2+v_3)^2+v_2(-v_2)+v_3(-v_3)) \\
=& -\lambda(-(v_2+v_3)^2-v_2^2-v_3^2) \\
=& \lambda((v_2+v_3)^2+v_2^2+v_3^2)
\end{aligned}
$$

Comme $\lambda > 0$ et $v_1+v_2+v_3 = 0$, $v^T \frac{\partial^2 \mathcal{L}}{\partial x^2} v \geq 0$. La condition du second ordre est vérifiée et nous avons bien trouvé un minimum en $\hat{X}$.

Exercice 3

Dans cet exercice, nous désirons trouver une loi de probabilité P qui maximise l'entropie pour une variable aléatoire X (prenant les valeurs $x_1 \dots x_n$ avec une probabilité $p_1 \dots p_n$) et dont la valeur moyenne est un m donné.

Modélisation

Nous pouvons modéliser notre problème de la manière suivante :

$$
\begin{aligned}
(Pb) : \max_{P \in \mathbb{R}^{+n}} f(P) = -\sum_{i=1}^{n} p_i \log_2(p_i) \\
\text{sous la contrainte} \quad g_1(P) = \sum_{i=1}^{n} p_i - 1 = 0 \\
\text{sous la contrainte} \quad g_2(P) = \sum_{i=1}^{n} p_i x_i - m = 0
\end{aligned}
$$

On néglige les contraintes de positivité. Nous les vérifierons *a posteriori*. Afin de vérifier la régularité des contraintes, calculons le gradient des contraintes :

$$
\frac{\partial g}{\partial P} = -\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}
$$

Si l'on fait l'hypothèse qu'au moins deux x_i sont distincts, les colonnes du gradient forment une famille libre : les contraintes sont régulières et nous pouvons poser le Lagrangien :

$$
\mathcal{L}(P, \lambda) = -\sum_{i=1}^{n} p_i \log_2(p_i) + \lambda_1(\sum_{i=1}^{n} p_i - 1) + \lambda_2(\sum_{i=1}^{n} p_i x_i - m)
$$

Condition du premier ordre

$$\frac{\partial \mathcal{L}}{\partial P} = \begin{bmatrix} \log_2(p_1) + \frac{1}{\ln 2} + \lambda_1 + \lambda_2 x_1 \\ \vdots \\ \log_2(p_n) + \frac{1}{\ln 2} + \lambda_1 + \lambda_2 x_n \end{bmatrix} = 0$$

Comme $\log_2(p_i) = \frac{\ln(p_i)}{\ln 2}$, nous obtenons n équations de la forme suivante :

$$\begin{aligned} &\log_2(p_i) + \frac{1}{\ln 2} + \lambda_1 + \lambda_2 x_i = 0 \\ &\frac{\ln(p_i)}{\ln 2} = -\frac{1}{\ln 2} - \lambda_1 - \lambda_2 x_i \\ &\ln(p_i) = (\ln 2)(-\frac{1}{\ln 2} - \lambda_1 - \lambda_2 x_i) \\ &p_i = 2^{(-\frac{1}{\ln 2} - \lambda_1 - \lambda_2 x_i)} \end{aligned}$$

Ce système ne peut être résolu analytiquement de façon simple. Nous vous proposons alors de retirer une contrainte : celle sur la valeur moyenne prise par la variable X.

Dans ce cas, le Lagrangien est :

$$\mathcal{L}(P, \lambda) = -\sum_{i=1}^{n} p_i \log_2(p_i) + \lambda(\sum_{i=1}^{n} p_i - 1)$$

La condition du premier ordre nous donne :

$$p_i = 2^{(-\frac{1}{\ln 2} - \lambda)}$$

En intégrant ce résultat dans la contrainte, nous obtenons :

$$\sum_{i=1}^{n} 2^{(-\frac{1}{\ln 2} - \lambda)} = 1$$

Comme $2^{(-\frac{1}{\ln 2} - \lambda)}$ est un réel fixé alors $2^{(-\frac{1}{\ln 2} - \lambda)} = \frac{1}{n}$. Par conséquent, $\forall\ i$, $p_i = \frac{1}{n}$. C'est la loi uniforme qui maximise l'entropie.

Nous pouvons alors remarquer que la contrainte est affine, auxquel cas la condition du premier ordre est nécessaire et suffisante.

TD4 : Problèmes avec contraintes inégalité

Exercice 1 : Programmation quadratique

Soit $Q \in \mathbb{R}^{n\times n}$ un matrice définie positive, a et b dans $\mathbb{R}^n$ avec $a \neq 0$ et $c \in \mathbb{R}$.

On veut résoudre

$$(P) : \min_{h(x)\leq 0} f(x), \quad avec \quad \left\{ \begin{array}{rcl} f(x) & = & \frac{1}{2}x^TQx - b^Tx \\ h(x) & = & a^Tx - c \end{array} \right. \quad x \in \mathbb{R}^n$$

La contrainte h est affine, elle est donc *qualifiée*.

On calcule le Lagrangien et ses dérivées partielles. soit $\mu \in \mathbb{R}$

$$\mathcal{L}(x,\mu) = \frac{1}{2}x^TQx - b^Tx + \mu(a^Tx - c) \tag{1}$$

$$\frac{\partial \mathcal{L}}{\partial x}(x,\mu) = \frac{1}{2}x^T(Q+Q^T) - b^T + \mu a^T \tag{2}$$

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(x,\mu) = \frac{1}{2}(Q+Q^T) \tag{3}$$

$\frac{\partial^2 \mathcal{L}}{\partial x^2}$ est une matrice définie positive. Si une solution est trouvée avec la condition du premier ordre pour $\mu \geq 0$ alors cette solution est bien un minimum.

$$\left\{ \begin{array}{rcl} \frac{\partial \mathcal{L}}{\partial x}(\hat{x},\hat{\mu}) & = & 0 \\ \hat{\mu}h(\hat{x}) & = & 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{rcl} \frac{1}{2}\hat{x}^T(Q+Q^T) - b^T + \hat{\mu}a^T & = & 0 \\ \hat{\mu}(a^T\hat{x} - c) & = & 0 \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{rcl} \frac{Q+Q^T}{2}\hat{x} - b + \hat{\mu}a & = & 0 \\ \hat{\mu}(a^T\hat{x} - c) & = & 0 \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{rcl} \hat{x} & = & (\frac{Q+Q^T}{2})^{-1}(b - \hat{\mu}a) \\ \hat{\mu}(a^T\hat{x} - c) & = & 0 \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{rcl} \hat{x} & = & 2(Q+Q^T)^{-1}(b - \hat{\mu}a) \\ \hat{\mu}(a^T\hat{x} - c) & = & 0 \end{array} \right.$$

1er cas : $\hat{\mu} = 0$

Dans ce cas : $\boxed{\hat{x} = 2(Q+Q^T)^{-1}b}$

2ème cas : $\hat{\mu} \neq 0$

On a alors

$$\begin{array}{rcl} a^T\hat{x} - c = 0 & \Rightarrow & 2a^T(Q+Q^T)^{-1}(b-\hat{\mu}a)) - c = 0 \\ & \Rightarrow & \hat{\mu}(2a^T(Q+Q^T)^{-1}a) = 2a^T(Q+Q^T)^{-1}b - c \end{array}$$

On a donc

$$\left\{ \begin{array}{rcl} \hat{\mu} & = & \frac{2a^T(Q+Q^T)^{-1}b-c}{2a^T(Q+Q^T)^{-1}a} \\ \hat{x} & = & 2(Q+Q^T)^{-1}(b - \frac{2a^T(Q+Q^T)^{-1}b-c}{2a^T(Q+Q^T)^{-1}a}a) \end{array} \right.$$

Pour que l'on ait un minimum, il faut que μ soit positif, c'est à dire que $2a^T(Q+Q^T)^{-1}b \geq c$.

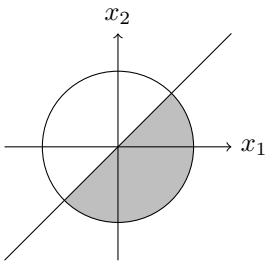
Dans ce cas, on a $\boxed{\hat{x} = 2(Q+Q^T)^{-1}(b - \dfrac{2a^T(Q+Q^T)^{-1}b - c}{2a^T(Q+Q^T)^{-1}a}a)}$

Interprétation : c est l'ordonnée à l'origine de la contrainte. Il définit donc la position de l'hyperplan de la contrainte par rapport au minimum global. Quand c devient plus petit que la quantité mentionnée plus haut, le minimum global sort du domaine et la contrainte commence à participer à tirer le minimum dans le domaine ($\mu \neq 0$). A la limite, on peut remarquer qu'on a à la fois $h(x) = 0$ et $\mu = 0$.

Exercice 2

On veut résoudre le problème suivant :

$$(P) : \min_{\substack{h_1(x_1,x_2)\leq 0 \\ h_2(x_1,x_2)\leq 0}} x_1 + x_2, \quad avec \quad \left\{ \begin{array}{rcl} h_1(x_1,x_2) & = & x_2 - x_1 \\ h_2(x_1,x_2) & = & x_1^2 + x_2^2 - 1 \end{array} \right. \quad (x_1,x_2) \in \mathbb{R}^2$$



Qualification des contraintes

Il faut tout d'abord vérifier que les contraintes sont bien qualifiées. Pour cela, calculons $\frac{\partial h_i}{\partial x}$.

On a $\frac{\partial h_1}{\partial x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ et $\frac{\partial h_2}{\partial x} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$

La famille des $\frac{\partial h_i}{\partial x}$ est donc libre si et seulment si on a $x_1 \neq -\frac{1}{2}$ et $x_2 \neq \frac{1}{2}$. Ce couple particulier n'appartient pas au domaine souhaité pour le problème. Les contraintes sont donc qualifiées.

Lagrangien : condition du 1er ordre

Ecrivons le Lagrangien et ses dérivées partielles. Soient μ_1 et μ_2 dans $\mathbb{R}^+$.

$$\begin{array}{rcl} \mathcal{L}(x_1, x_2, \mu_1, \mu_2) & = & x_1 + x_2 + \mu_1(x_2 - x_1) + \mu_2(x_1^2 + x_2^2 - 1) \\ & = & \mu_2 x_1^2 + (-\mu_1 + 1)x_1 + \mu_2 x_2^2 + (\mu_1 + 1)x_2 - \mu_2 \end{array}$$

On a donc

$$\begin{array}{rcl} \frac{\partial \mathcal{L}}{\partial x_1}(x_1, x_2, \mu_1, \mu_2) & = & 1 - \mu_1 + 2\mu_2 x_1 \\ \frac{\partial \mathcal{L}}{\partial x_2}(x_1, x_2, \mu_1, \mu_2) & = & 1 + \mu_1 + 2\mu_2 x_2 \end{array}$$

$$\left\{ \begin{array}{rcl} \frac{\partial \mathcal{L}}{\partial x_1}(\hat{x}_1, \hat{x}_2, \hat{\mu}_1, \hat{\mu}_2) & = & 0 \\ \frac{\partial \mathcal{L}}{\partial x_2}(\hat{x}_1, \hat{x}_2, \hat{\mu}_1, \hat{\mu}_2) & = & 0 \\ \hat{\mu}_1(\hat{x}_2 - \hat{x}_1) & = & 0 \\ \hat{\mu}_2(\hat{x}_1^2 + \hat{x}_2^2 - 1) & = & 0 \end{array} \right.$$

ce qui est équivalent à

$$\left\{ \begin{array}{rcl} 1 - \hat{\mu}_1 + 2\hat{\mu}_2\hat{x}_1 & = & 0 \\ 1 + \hat{\mu}_1 + 2\hat{\mu}_2\hat{x}_2 & = & 0 \\ \hat{\mu}_1(\hat{x}_2 - \hat{x}_1) & = & 0 \\ \hat{\mu}_2(\hat{x}_1^2 + \hat{x}_2^2 - 1) & = & 0 \end{array} \right.$$

1er cas : $\hat{\mu}_1 = \hat{\mu}_2 = 0$

On a alors $1 - \hat{\mu}_1 + 2\hat{\mu}_2\hat{x}_1 = 0 \Leftrightarrow 1 = 0$. **Impossible !**

2eme cas : $\hat{\mu}_1 = 0$ et h_2 saturée

$$\left\{ \begin{array}{rcl} 1 + 2\hat{\mu}_2\hat{x}_1 & = & 0 \\ 1 + 2\hat{\mu}_2\hat{x}_2 & = & 0 \\ \hat{x}_1^2 + \hat{x}_2^2 - 1 & = & 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} \hat{x}_1 & = & -\frac{1}{2\hat{\mu}_2} \\ \hat{x}_2 & = & -\frac{1}{2\hat{\mu}_2} \\ \hat{\mu}_2(\frac{1}{2\hat{\mu}_2^2} - 1) & = & 0 \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{rcl} \hat{\mu}_2 & = & \frac{\sqrt{2}}{2} \\ \hat{x}_1 & = & -\frac{\sqrt{2}}{2} \\ \hat{x}_2 & = & -\frac{\sqrt{2}}{2} \end{array} \right.$$

3eme cas : $\hat{x}_2 = 0$ et h_1 saturée

$$\left\{ \begin{array}{rcl} 1 - \hat{x}_1 & = & 0 \\ 1 + \hat{x}_1 & = & 0 \\ \hat{x}_2 - \hat{x}_1 & = & 0 \end{array} \right.$$

On a donc $1 = 0$, ce qui est **Impossible !**

4eme cas : h_1 et h_2 saturées

Les deux contraintes sont saturées, on a donc $\hat{x}_1 = \hat{x}_2$ et $\hat{x}_1^2 + \hat{x}_2^2 = 1$.

On a alors $(\hat{x}_1, \hat{x}_2) = \pm(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$.

En soustrayant les deux premières équations issues de la dérivation du Lagrangien, on trouve $\hat{\mu}_1 = 0$ et $\hat{\mu}_2 = -\frac{1}{2\hat{x}_2}$. On cherche $\hat{\mu}_2 \geq 0$, cela nous permet d'éliminer un des deux cas, on a au final : $(\hat{x}_1, \hat{x}_2, \hat{\mu}_1, \hat{\mu}_2) = (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2})$.
On peut remarquer que $\mu_1 = 0$, on a bien retrouvé le résultat du second cas, ce qui illustre le fait que le minimum est trouvé pile sur la contrainte sans que cette dernière ne participe à le tirer dans un sens ou un autre.

f est affine et les contraintes sont convexes, la condition d'optimalité du premier ordre suffit donc à conclure que l'unique point trouvé est bien un minimum. Pour s'entraîner, on peut tout de même aller vérifier la condition du second ordre.

Lagrangien : Condition du 2eme ordre et solution

On voit donc que la seule solution possible est

$$\begin{cases} \mu_1 & = & 0 \\ \mu_2 & = & \frac{\sqrt{2}}{2} \\ \hat{x_1} & = & -\frac{\sqrt{2}}{2} \\ \hat{x_2} & = & -\frac{\sqrt{2}}{2} \end{cases}$$

On a $\frac{\partial^2 \mathcal{L}}{\partial x^2}(\hat{x}, \hat{\mu}) = \begin{bmatrix} 2\hat{\mu}_2 & 0 \\ 0 & 2\hat{\mu}_2 \end{bmatrix}$. $\mu_2 > 0$ donc $\frac{\partial^2 \mathcal{L}}{\partial x^2}$ est définie positive.
Le minimum est donc bien atteint en $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$

Exercice 3

On remarque qu'il s'agit là de la généralisation de l'exercice 2 à des fonctions affines et deux contraintes affine et quadratique quelconques.

Soient a un vecteur de $\mathbb{R}^n \setminus \{0\}$, Q une matrice de$\mathbb{R}^{n \times n}$ symétrique définie positive, et b et c deux réels avec $c > \frac{1}{2} a^T Q a$.
On considère les fonctions h_1 et h_2 de $\mathbb{R}^n$ dans $\mathbb{R}$ définies par

$$h_1(x) = a^T x$$
$$h_2(x) = \frac{1}{2} x^T Q x - c$$

On veut résoudre

$$(P) : \min_{x \in \Omega} (a^T x - b), \text{ avec } \Omega = \{x \in \mathbb{R}^n : h_1(x) \leq 0 \text{ et } h_2(x) \leq 0\}$$

Qualification des contraintes

$\frac{\partial^2 h_1}{\partial x^2} =\geq 0$ donc h_1 est convexe (0 désigne ici la matrice nulle).
$\frac{\partial^2 h_2}{\partial x^2} = Q \geq 0$ donc h_2 est convexe.
On a $h_1(-a) < 0$ et $h_2(-a) < 0$ donc le domaine est non-vide : les contraintes sont qualifiées sur Ω.

Lagrangien et condition du 1er ordre

Soient μ_1 et μ_2 dans $\mathbb{R}$.

$$\mathcal{L}(x, \mu_1, \mu_2) = a^T x - b + \mu_1 a^T x + \frac{1}{2}\mu_2 x^T Q x - \mu_2 c$$

$$\frac{\partial \mathcal{L}}{\partial x}(x, \mu_1, \mu_2) = a^T + \mu_1 a^T + \mu_2 x^T Q$$

Avec la condition du 1er ordre, on a le système suivant :

$$\left\{ \begin{array}{lcl} a^T + \mu_1 a^T + \hat{\mu}_2 \hat{x}^T Q & = & 0 \\ \hat{\mu}_1 a^T \hat{x} & = & 0 \\ \hat{\mu}_2 (\frac{1}{2}\hat{x}^T Q \hat{x} - c) & = & 0 \end{array} \right.$$

1er cas : $\hat{\mu}_1 = 0$ et $\hat{\mu}_2 = 0$

Dans ce cas, on obtient $a^T = 0$ avec la première équation du système, ce qui est en contradiction avec notre hypothèse sur a. Ce cas est donc **impossible !**

2eme cas : h_1 est active et $\hat{\mu}_2 = 0$

La première équation nous donne $\hat{\mu}_1 + 1 = 0$. Cela implique que $\mu_1 > 0$, ce cas est donc **impossible !**.

3eme cas : $\mu_1 = 0$ et h_2 est active

On a

$$\left\{ \begin{array}{lcl} a^T + \hat{\mu}_2 \hat{x}^T Q & = & 0 \\ \frac{1}{2}\hat{x}^T Q \hat{x} - c & = & 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{lcl} \hat{x} & = & -\frac{1}{\hat{\mu}_2} Q^{-1} a \\ \hat{x}^T Q \hat{x} & = & 2c \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{lcl} \frac{1}{2\hat{\mu}_2^2} a^T Q^{-T} Q Q^{-1} & = & c \\ \hat{x} & = & -\frac{1}{\hat{\mu}_2} Q^{-1} a \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{lcl} \hat{\mu}_2 & = & \sqrt{\frac{a^T Q^{-T} a}{2c}} \\ \hat{x} & = & -\sqrt{\frac{2c}{a^T Q^{-T} a}} Q^{-1} a \end{array} \right.$$

4eme cas : $h1$ et h_2 sont actives

La première équation nous indique que $(1 + \hat{\mu}_1)a = -\hat{\mu}_2 Q x$ et donc que $\hat{x} = -\frac{1+\hat{\mu}_1}{\hat{\mu}_2} Q^{-1} a$.

Or la seconde équation précise que $a^T x = 0$, on a donc : $-\frac{1+\hat{\mu}_1}{\hat{\mu}_2} a^T Q^{-1} a = 0$. Or $a^T Q a \neq 0$ car $a \neq 0$ et Q définie positive. Donc $\hat{\mu}_1 = -1$, ce qui est **impossible !**.

Comme à l'exercice précédent, il est immédiat de vérifier que f est affine et que les contraintes sont convexes : le théorème d'optimalité au premier ordre suffit à conclure que le point $\hat{x} = -\sqrt{\frac{2c}{a^T Q^{-T} a}} Q^{-1} a$ (qui existe toujours) est bien un minimum. On peut s'entraîner à vérifier la condition du second ordre.

Lagrangien et condition du 2eme ordre

On a $\frac{\partial^2 \mathcal{L}}{\partial x^2}(\hat{x}, \hat{\mu}) = Q \geq 0$ car Q est définie positive. La solution trouvée est donc le minimum global sous contraintes.

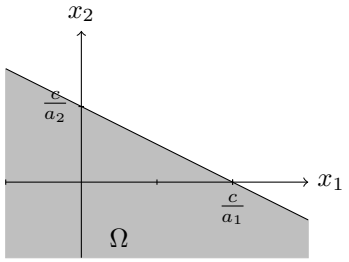
TD5 : Dualité

Exercice 1

Soient a_1, a_2 et c dans $\mathbb{R}$ avec $(a_1, a_2) \neq (0, 0)$. On veut résoudre :

$$(P) : \min_{f(x_1,x_2)=(x_1^2+x_2^2)} a_1x_1 + a_2x_2 - c \leq 0$$

On pose $\Omega = \{(x_1, x_2) : h(x_1, x_2) = a_1x_1 + a_2x_2 - c \leq 0\}$.



Problème primal - problème dual

Le problème primal est

$$\inf_{x\in\mathbb{R}^2} \sup_{\mu\geq 0} \mathcal{L}(x, \mu)$$

Le problème dual est

$$\sup_{\mu\geq 0} \inf_{x\in\mathbb{R}^2} \mathcal{L}(x, \mu)$$

Qualification des contraintes

La fonction h est affine et donc qualifiée.

Lagrangien - fonction duale

Soit $\mu \in \mathbb{R}$, $(x_1, x_2) \in \mathbb{R}^2$

$$\mathcal{L}(x_1, x_2, \mu) = x_1^2 + x_2^2 + \mu(a_1x_1 + a_2x_2 - c)$$

On note $\psi(\mu)$ la fonction duale.

$$\psi(\mu) = \inf_{x_1,x_2\in\mathbb{R}^2} \mathcal{L}(x_1, x_2, \mu)$$

$$\frac{\partial \mathcal{L}}{\partial x}(\hat{x_1}, \hat{x_2}, \mu) = 0 \quad \Rightarrow \quad \left\{ \begin{array}{lcl} \frac{\partial \mathcal{L}}{\partial x_1}(\hat{x_1}, \hat{x_2}, \mu) & = & 0 \\ \frac{\partial \mathcal{L}}{\partial x_2}(\hat{x_1}, \hat{x_2}, \mu) & = & 0 \end{array} \right.$$
$$\Rightarrow \left\{ \begin{array}{lcl} 2\hat{x_1} + a_1\mu & = & 0 \\ 2\hat{x_2} + a_1\mu & = & 0 \end{array} \right.$$
$$\Rightarrow \left\{ \begin{array}{lcl} \hat{x_1} & = & -\frac{a_1}{2}\mu \\ \hat{x_2} & = & -\frac{a_2}{2}\mu \end{array} \right.$$

f et h sont convexes. $(\hat{x_1}, \hat{x_2})$ est donc un minimum.
On a donc

$$\psi(\mu) = \mathcal{L}(\hat{x_1}, \hat{x_2}, \mu) = -\frac{\mu^2}{4}(a_1^2 + a_2^2) - \mu c$$

Résoudre le problème dual revient à résoudre

$$\sup_{\mu \geq 0} psi(\mu)$$

On calcule donc $psi'(\mu)$.

$$psi'(\mu) = -\frac{\mu}{2}(a_1^2 + a_2^2) - c$$

1er cas : $c \geq 0$

Dans ce cas, $\forall \mu \geq 0$, $psi'(\mu) \leq 0$. $psi(\mu)$ est donc décroissante. Le maximum est alors atteint pour $\mu = 0$, c'est à dire $\boxed{\hat{x_1} = \hat{x_2} = 0}$

2eme cas : $c < 0$

Dans ce cas, on a

$$\begin{array}{lcl} psi'(\mu) = 0 & \Rightarrow & \frac{\mu}{2}(a_1^2 + a_2^2) = -C \\ & \Rightarrow & \mu = \frac{-2c}{a_1^2 + a_2^2} \end{array}$$

μ	0		$\frac{-2c}{a_1^2+a_2^2}$		$+\infty$
$psi'(\mu)$		$+$	0	$-$	
$\psi(\mu)$		$\nearrow$		$\searrow$	

On a alors

$$\boxed{\hat{x_1} = \frac{a_2 c}{a_1^2 + a_2^2}, \hat{x_2} = \frac{a_1 c}{a_1^2 + a_2^2}}$$

Exercice 2 : Programmation quadratique

Cet exercice est l'équivalent de l'exercice 1 du TD4.

Soit Q une matrice de $\mathbb{R}^{n \times n}$ symétrique définie positive, a et b dans $\mathbb{R}^n$ avec $a \neq 0$ et $c \in \mathbb{R}$. On veut résoudre

$$(P) : \min_{h(x) \leq 0} f(x) \text{ avec } \left\{ \begin{array}{lcl} f(x) & = & \frac{1}{2}x^T Q x - b^T x \\ h(x) & = & a^T x - c \end{array} \right. , x \in \mathbb{R}^n$$

Qualification des contraintes

h est affine donc qualifiée.

Lagrangien - fonction duale

f et h sont convexes. On écrit le Lagrangien et la fonction duale ψ qui lui est associée. Soit $\mu \in \mathbb{R}$.

$$\mathcal{L}(x,\mu) = \frac{1}{2}x^TQx - b^Tx + \mu(a^Tx - c)$$

$$\psi(\mu) = \inf_{x\in\mathbb{R}^n} \mathcal{L}(x,\mu)$$

.

$$\begin{array}{rcl} \frac{\partial \mathcal{L}}{\partial x}(\hat{x},\mu) = 0 & \Rightarrow & \frac{1}{2}\hat{x}^T(Q+Q^T) - b^T + \mu a^T = 0 \\ & \Rightarrow & Q\hat{x} - b + \mu a = 0 \\ & \Rightarrow & \hat{x} = Q^{-1}(b - \mu a) \end{array}$$

On a donc :

$$\psi(\mu) = \mu^2[\frac{1}{2}a^TQ^{-1}a - a^TQ^{-1}a] + \mu[-\frac{1}{2}a^TQ^{-1}b - \frac{1}{2}b^TQ^{-1}a + b^TQ^{-1}a + a^TQ^{-1}b - c] + \frac{1}{2}b^TQ^{-1}b - b^TQ^{-1}b$$

c'est à dire :

$$\psi(\mu) = \mu^2[-\frac{1}{2}a^TQ^{-1}a] + \mu[a^TQ^{-1}b - c] - \frac{1}{2}b^TQ^{-1}b$$

On a donc :

$$\psi'(\mu) = -\mu a^TQ^{-1}a + a^TQ^{-1}b - c$$

$$\psi'(\hat{\mu}) = 0 \Leftrightarrow \hat{\mu} = \frac{a^TQ^{-1}b - c}{a^TQ^{-1}a}$$

1er cas : $a^TQ^{-1}b - c \leq 0$

Dans ce cas, $\forall \mu \geq 0$, $\psi'(\mu) \leq 0$. $\psi(\mu)$ est donc décroissante. Le maximum est alors atteint pour $\mu = 0$, c'est à dire $\boxed{\hat{x} = Q^{-1}b}$.

2eme cas : $a^TQ^{-1}b - c > 0$

μ	0		$\frac{a^TQ^{-1}b-c}{a^TQ^{-1}a}$		$+\infty$
$\psi'(\mu)$		$+$	0	$-$	
$\psi(\mu)$		$\nearrow$		$\searrow$	

Le minimum est atteint pour

$$\boxed{\hat{x} = Q^{-1}(b - \frac{a^T Q^{-1} b - c}{a^T Q^{-1} a} a)}$$

On retrouve bien le résultat de l'exercice 1, TD4.

Exercice 3

Soit $c \in \mathbb{R}$ tel que $c < 0$. On définit les fonctions f, $h \colon \mathbb{R}^n \to \mathbb{R}$ par

$$\begin{array}{rcl} f(x) & = & \sum_{i=1}^n x_i^4 \\ h(x) & = & (\sum_{i=1}^n x_i) - c \end{array} \quad , x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$$

On veut résoudre

$$(P) : \min_{x \in \Omega} f(x) \text{ avec } \Omega = \{x \in \mathbb{R}^n : h(x) \leq 0\}$$

Question 1 : f convexe

On a

$$\frac{\partial f}{\partial x}(x) = \begin{bmatrix} \vdots \\ 4x_i^3 \\ \vdots \end{bmatrix}$$

et

$$\frac{\partial^2 f}{\partial x^2}(x) = \begin{bmatrix} 12x_1^2 & & 0 \\ & \ddots & \\ 0 & & 12x_n^2 \end{bmatrix}$$

$\frac{\partial^2 f}{\partial x^2} \geq 0$ donc f est convexe.

Question 2 : Problème dual

Soit $\mu \in \mathbb{R}$. On a

$$\mathcal{L}(x, \mu) = \sum_{i=1}^n x_i^4 + \mu(\sum_{i=1}^n x_i) - \mu c$$

.

$$\begin{array}{rcl} \frac{\partial \mathcal{L}}{\partial x_i}(\hat{x}, \mu) = 0 & \Rightarrow & 4\hat{x}_i^3 + \mu = 0 \\ & \Rightarrow & \hat{x}_i = -\left(\frac{\mu}{4}\right)^{\frac{1}{3}} \end{array}$$

f est convexe donc $\hat{x} = (-\left(\frac{\mu}{4}\right)^{\frac{1}{3}}, \ldots, -\left(\frac{\mu}{4}\right)^{\frac{1}{3}})^T$ est bien un minimum.
On écrit alors la fonction duale :

$$
\begin{aligned}
\psi(\mu) = \mathcal{L}(\hat{x}, \mu) &= \textstyle\sum_{i=1}^{n}((\frac{\mu}{4})^{\frac{1}{3}}) - \mu \sum_{i=1}^{n}((\frac{\mu}{4})^{\frac{1}{3}}) - \mu c \\
&= n\left(\frac{\mu}{4}\right)^{\frac{4}{3}} - n\frac{\mu^{\frac{4}{3}}}{4^{\frac{1}{3}}} - \mu c \\
&= n\mu^{\frac{4}{3}}(\frac{1}{4^{\frac{4}{3}}} - \frac{1}{4^{\frac{1}{3}}}) - \mu c \\
&= n\mu^{\frac{4}{3}}(\frac{4^{\frac{1}{3}} - 4^{\frac{4}{3}}}{4^{\frac{5}{3}}}) - \mu c \\
&= \tau\mu^{\frac{4}{3}} - \mu c
\end{aligned}
$$

On cherche $\inf\limits_{\mu \geq 0} \psi(\mu)$.

$$
\begin{aligned}
\psi'(\mu) = 0 \quad &\Rightarrow \quad \tfrac{4}{3}\tau\mu^{\frac{1}{3}} - c = 0 \\
&\Rightarrow \quad \tfrac{4}{3}\tau\mu^{\frac{1}{3}} = c \\
&\Rightarrow \quad \mu^{\frac{1}{3}} = \tfrac{3c}{4\tau} \\
&\Rightarrow \quad \mu = \left(\tfrac{3c}{4\tau}\right)^3
\end{aligned}
$$

τ et c sont négatifs donc on a bien un μ positif.
On a donc

$$\hat{x}_i = -\left(\frac{\mu}{4}\right)^{\frac{1}{3}} = -\frac{3c}{4\tau}\frac{1}{4^{\frac{1}{3}}} = -\frac{3c}{4^{\frac{4}{3}}\tau} = \frac{c}{n}$$

et $\boxed{\hat{x} = (\frac{c}{n}, \ldots, \frac{c}{n})^T}$