

PROBLÈMES D'OPTIMISATION AVEC CONTRAINTES D'ÉGALITÉS

1.1 Formulation du problème

Définition 1.1. Un problème de minimisation avec contraintes d'égalité peut être exprimé de la façon suivante :

$$\begin{cases} \min f(x), \\ h(x) = 0. \end{cases}$$

Où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction est appelée la fonction objectif et $h = (h_1, h_2, \dots, h_m)^T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction de classe $\mathcal{C}^1$ avec $m \leq n$ est appelée fonction contrainte.

Nous introduisons la définition suivante

Définition 1.2. Un point x^* satisfaisant aux contraintes $h_1(x^*) = 0, h_2(x^*) = 0, \dots, h_m(x^*) = 0$ est dit un point régulier des contraintes (ou point réalisable) si les vecteurs de gradient $\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_m(x^*)$ sont linéairement indépendants.

Remarque 1.1. 1. Soit $Dh(x^*)$ la matrice Jacobienne de la fonction h au point x^* donnée par :

$$Dh(x^*) = \left(\frac{\partial h_i}{\partial x_j}(x^*) \right) = \begin{bmatrix} \nabla h_1(x^*)^T \\ \nabla h_2(x^*)^T \\ \vdots \\ \nabla h_m(x^*)^T \end{bmatrix}.$$

Alors, x^* est régulier si et seulement si le $\text{rang}(Dh(x^*)) = m$.

2. L'ensemble des contraintes d'égalité $h_1(x^*) = 0, h_2(x^*) = 0, \dots, h_m(x^*) = 0, h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ décrit une surface définie par :

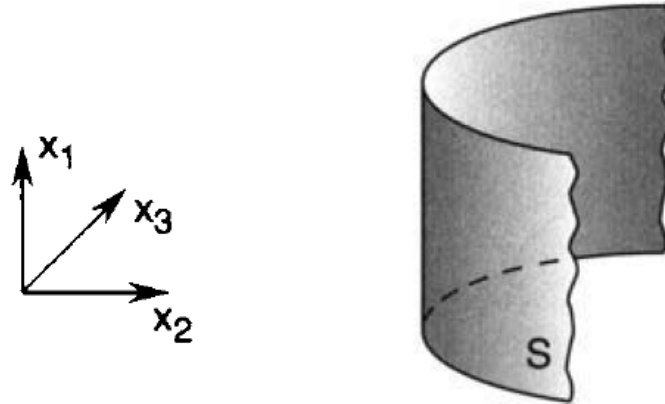
$$S = \{x \in \mathbb{R}^n : h_1(x^*) = 0, h_2(x^*) = 0, \dots, h_m(x^*) = 0\}.$$

3. Si les points de S sont réguliers, la dimension de la surface S est $n - m$.

Exemple 1.1. Soient $n = 3$ et $m = 1$. Soit $h_1(x_1, x_2, x_3) = x_2 - x_3^2 = 0$. Nous avons $\nabla h_1(x_1, x_2, x_3) = (0, 1, -2x_3)^T$. Donc, pour tout $x \in \mathbb{R}^3, \nabla h_1(x) \neq 0$. Dans ce cas,

$$\dim(S) = \dim\{x \in \mathbb{R}^3 : h_1(x) = 0\} = n - m = 2.$$

L'ensemble S est une surface dans $\mathbb{R}^3$. Voir Figure 1.1 pour une illustration graphique.



$$S = \{[x_1, x_2, x_3]^T : x_2 - x_3^2 = 0\}$$

FIGURE 1.1 – Surface dans $\mathbb{R}^3$.

Exemple 1.2. Soient $n = 3$ et $m = 2$. Par exemple, soit

$$\begin{aligned} h_1(x) &= x_1, \\ h_2(x) &= x_2 - x_3^2. \end{aligned}$$

Dans ce cas, $\nabla h_1(x) = (1, 0, 0)^T$ et $\nabla h_2(x) = (0, 1, -2x_3)^T$. Alors, les vecteurs $\nabla h_1(x)$ et $\nabla h_2(x)$ sont linéairement indépendants dans $\mathbb{R}^3$. Donc,

$$\dim(S) = \dim\{x \in \mathbb{R}^3 : h_1(x) = 0, h_2(x) = 0\} = n - m = 1.$$

L'ensemble des contraintes S est une courbe dans $\mathbb{R}^3$. Voir Figure 1.2 pour une illustration graphique.

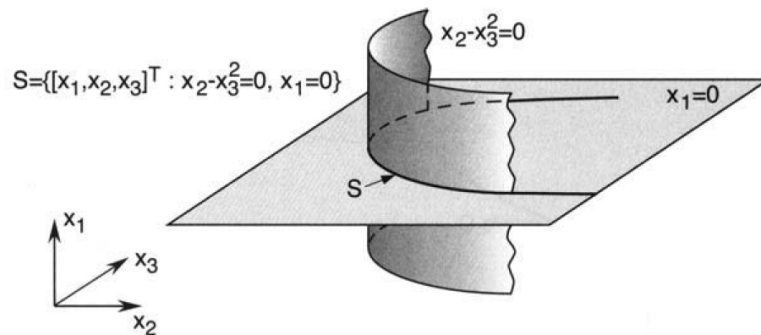


FIGURE 1.2 – Courbe dans $\mathbb{R}^3$.

Définition 1.3. Une courbe C sur une surface S est un ensemble de points $\{x(t) \in S : t \in]a, b[\}$ où $x :]a, b[\rightarrow S$ est une fonction continue.

Une illustration graphique de la définition d'une courbe est donnée dans Figure 1.3.

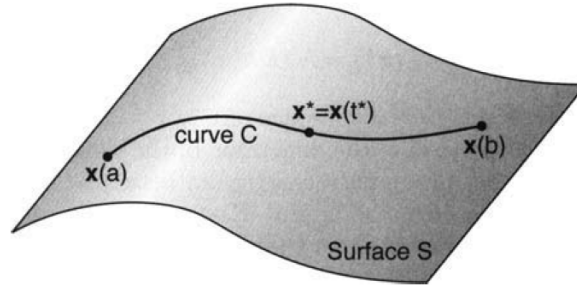


FIGURE 1.3 – Courbe sur une surface.

Définition 1.4. 1. La courbe $C = \{x(t) : t \in]a, b[\}$ est différentiable si

$$\dot{x}(t) = \frac{dx}{dt}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix} \text{ existe pour tout } t \in]a, b[.$$

Par conséquent, le vecteur $\dot{x}(t)$ est tangent à la courbe C à $x(t)$ (voir Figure 1.4).

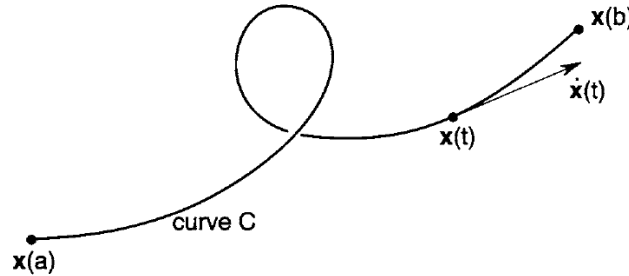


FIGURE 1.4 – Interprétation géométrique de la différentiabilité d'une courbe.

2. La courbe $C = \{x(t) : t \in]a, b[\}$ est de classe $\mathcal{C}^2$ si

$$\ddot{x}(t) = \frac{d^2x}{dt^2}(t) = \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \\ \vdots \\ \ddot{x}_n(t) \end{bmatrix}$$

existe pour tout $t \in]a, b[$.

1.2 Espaces tangents et normaux

Définition 1.5. (Espace tangent). Soit $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction de classe $\mathcal{C}^1$. L'espace tangent en un point x^* sur la surface $S = \{x \in \mathbb{R}^n : h(x) = 0\}$ est le sous-espace de $\mathbb{R}^n$ défini par :

$$T(x^*) = \{y \in \mathbb{R}^n : Dh(x^*)y = 0\}.$$

Remarque 1.2. 1. Notons que l'espace tangent $T(x^*)$ est le noyau de la matrice $Dh(x^*)$:

$$T(x^*) = \ker(Dh(x^*)).$$

2. Si x^* est un point régulier (point réalisable), alors $\dim(T(x^*)) = n - m$.

3. Nous définissons le plan tangent à x^* par l'ensemble suivant :

$$TP(x^*) = T(x^*) + x^* = \{x + x^* : x \in T(x^*)\}.$$

Figure 1.5 illustre la notion de plan tangent.

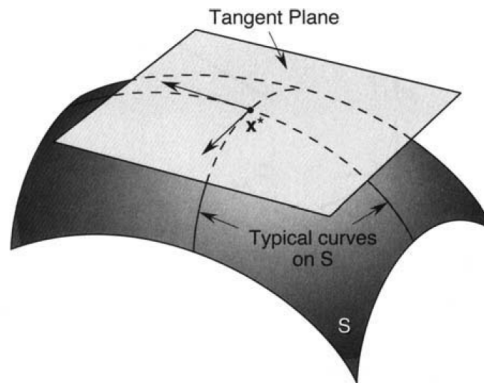


FIGURE 1.5 – Plan tangent à la surface S au point x^* .

Et Figure 1.6 illustre la relation entre le plan tangent et l'espace tangent dans $\mathbb{R}^2$ et $\mathbb{R}^3$.

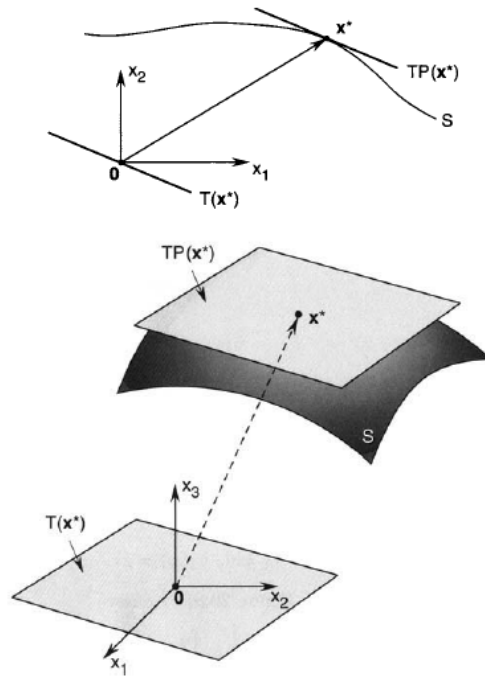


FIGURE 1.6 – Espaces tangent et plans dans $\mathbb{R}^2$ et $\mathbb{R}^3$.

Exemple 1.3. Soit $S = \{x \in \mathbb{R}^3 : h_1(x) = x_1 = 0, h_2(x) = x_1 - x_2 = 0\}$. Donc, S est l'axe x_3 dans $\mathbb{R}^3$ (voir Figure 1.7).

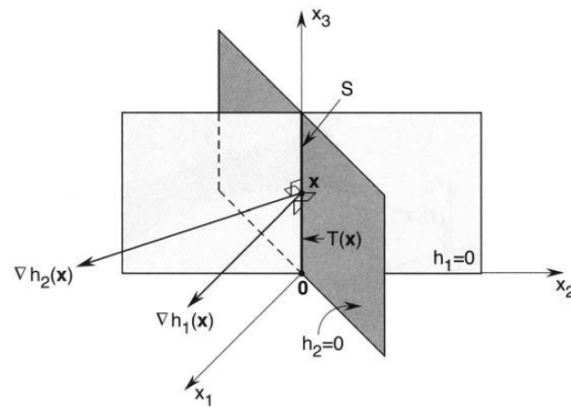


FIGURE 1.7 – La surface $S = \{x \in \mathbb{R}^3 : x_1 = 0, x_1 - x_2 = 0\}$.

Nous avons :

$$Dh(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \nabla h_2(x)^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}.$$

Pour tout $x \in S$, ∇h_1 et ∇h_2 sont linéairement indépendants, donc les points de S sont réguliers.

L'espace tangent au point $x \in S$ est donné par :

$$\begin{aligned} T(x) &= \left\{ y \in \mathbb{R}^3 : \nabla h_1(x)^T y = 0, \nabla h_2(x)^T y = 0 \right\} \\ &= \left\{ y \in \mathbb{R}^3 : \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = 0 \right\} \\ &= \left\{ [0, 0, \alpha]^T : \alpha \in \mathbb{R} \right\} \\ &= \text{Axe } x_3 \text{ dans } \mathbb{R}^3. \end{aligned}$$

Donc, $\dim(T(x)) = n - m = 1$

Théorème 1.1. Soient $x^* \in S$ un point régulier et $T(x^*)$ est l'espace tangent à x^* . Alors, $y \in T(x^*)$ si et seulement s'il existe une courbe différentiable dans S passant par x^* avec dérivée y en x^* .

Démonstration. 1. On suppose qu'il existe une courbe $\{x(t) : t \in]a, b[\}$ dans S tels que $x(t^*) = x^*$ et $\dot{x}(t^*) = y, t^* \in]a, b[$. Alors,

$$h(x(t)) = 0, \text{ pour tout } t \in]a, b[.$$

En dérivant $h(x(t))$ par rapport t , on obtient :

$$\frac{d}{dt}h(x(t)) = Dh(x(t))\dot{x}(t) = 0, \text{ pour tout } t \in]a, b[.$$

Par conséquent, à t^* nous obtenons :

$$Dh(x^*)y = 0,$$

donc, $y \in T(x^*)$.

2. Pour prouver l'implication inverse, nous devons utiliser le théorème de la fonction implicite. □

Nous introduisons maintenant la notion d'espace normal.

Définition 1.6. (Espace normal). L'espace normal $N(x^*)$ en un point x^* sur la surface $S = \{x \in \mathbb{R}^n : h(x) = 0\}$ est un sous-espace de $\mathbb{R}^n$ défini par :

$$\begin{aligned} N(x^*) &= \left\{ x \in \mathbb{R}^n : x = Dh(x^*)^T z : z \in \mathbb{R}^m \right\} \\ &= \left\{ x \in \mathbb{R}^n : x = z_1 \nabla h_1(x^*) + \dots + z_m \nabla h_m(x^*), z_1, \dots, z_m \in \mathbb{R} \right\}. \end{aligned}$$

De plus, $\dim(N(x^*)) = m$.

Définition 1.7. (Plan normal). Le plan normal en un point x^* est un sous-ensemble de $\mathbb{R}^n$ défini par :

$$NP(x^*) = x^* + N(x^*) = \{x^* + x \in \mathbb{R}^n : x \in N(x^*)\}.$$

Figure 1.8 illustre l'espace et le plan normaux dans $\mathbb{R}^3$ (c'est-à-dire $n = 3$ et $m = 1$).

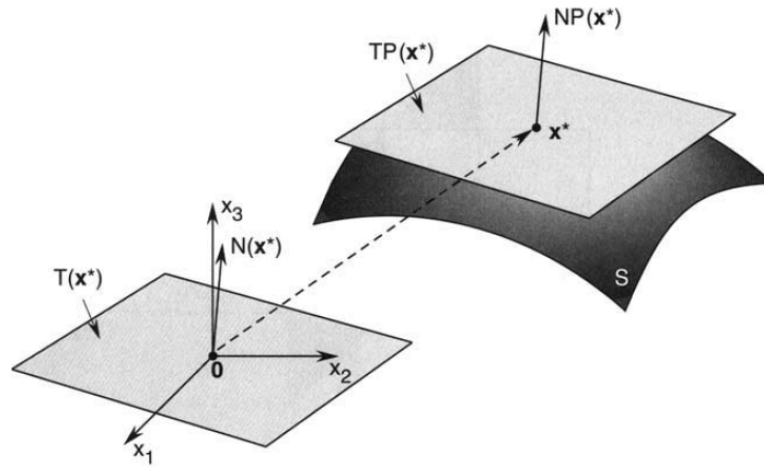


FIGURE 1.8 – Espace normal dans $\mathbb{R}^3$.

Nous montrons maintenant que l'espace tangent et l'espace normal sont des compléments orthogonaux l'un de l'autre.

Lemme 1.1. 1. Nous avons :

$$T(x^*) = N(x^*)^\perp \text{ et } T(x^*)^\perp = N(x^*).$$

2. De plus,

$$\mathbb{R}^n = T(x^*) \oplus N(x^*).$$

Démonstration. Par définition de $T(x^*)$, on peut écrire

$$T(x^*) = \{y \in \mathbb{R}^n : x^T y = 0 \text{ pour tout } x \in N(x^*)\}.$$

Par conséquent, par définition de $N(x^*)$, nous avons $T(x^*) = N(x^*)^\perp$. □

1.3 Condition de Lagrange

Dans cette section, nous présentons la condition nécessaire du premier ordre pour un problème d'optimisation avec contraintes d'égalité. Cette condition est connu par théorème de Lagrange.

Théorème 1.2. (Théorème de Lagrange pour $n = 2$ et $m = 1$). Soient $f, h : \mathbb{R}^2 \rightarrow \mathbb{R}$ deux fonctions différentiables et x^* est un point de minimum global (ou local) du problème suivant :

$$\begin{cases} \min f(x), \\ h(x) = 0. \end{cases}$$

Si $\nabla h(x^*) \neq 0$, alors $\nabla f(x^*)$ et $\nabla h(x^*)$ sont parallèles c'est à dire il existe $\lambda^* \in \mathbb{R}$ tel que

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0.$$

Démonstration. Soient $x^* \in \mathbb{R}^n$ tel que $h(x^*) = 0$ et $\nabla h(x^*) \neq 0$. L'ensemble de niveau de h au point x^* est donné par :

$$S = \{x \in \mathbb{R}^n : h(x) = h(x^*) = 0\}.$$

Donc, il existe une courbe $\{x(t) : t \in]a, b[\}$ tel que $x(t^*) = x^*$ et $x :]a, b[\rightarrow \mathbb{R}^2$ est de classe $\mathcal{C}^1$. Nous avons :

$$x(t^*) = x^*, \dot{x}(t^*) \neq 0, t^* \in]a, b[.$$

La fonction h est constante sur la courbe $\{x(t) : t \in]a, b[\}$. Alors, pour tout $t \in]a, b[$, on a :

$$h(x(t)) = 0.$$

Par conséquent, nous avons :

$$\frac{d}{dt}h(x(t)) = 0.$$

En appliquant la règle de chaine, nous obtenons :

$$\frac{d}{dt}h(x(t)) = \nabla h(x(t))^T \dot{x}(t) = 0.$$

D'où, $\nabla h(x^*)$ est orthogonal sur $\dot{x}(t^*)$.

On suppose que $x^* = x(t^*)$ est un point de minimum global de f sur l'ensemble

$$S = \{x \in \mathbb{R}^n : h(x) = 0\}.$$

Soit la fonction ϕ définie par :

$$\phi(t) = f(x(t)).$$

Grâce la condition nécessaire d'optimalité du premier ordre pour un problème d'optimisation sans contraintes, on obtient :

$$\frac{d\phi}{dt}(t^*) = 0.$$

Donc, on a :

$$0 = \frac{d\phi}{dt}(t^*) = \nabla f(x(t^*))^T \dot{x}(t^*) = \nabla f(x^*)^T \dot{x}(t^*).$$

D'où, $\nabla f(x^*)$ est orthogonal sur $\dot{x}(t^*)$. D'autre part, $\dot{x}(t^*)$ est un vecteur tangent de la courbe $\{x(t) : t \in]a, b[\}$ au point x^* , donc $\nabla f(x^*)$ est orthogonal sur la courbe $\{x(t) : t \in]a, b[\}$ au point x^* , voir Figure 1.9. Finalement, les deux vecteurs $\nabla f(x^*)$ et $\nabla h(x^*)$ sont parallèles, donc il existe $\lambda^* \in \mathbb{R}$

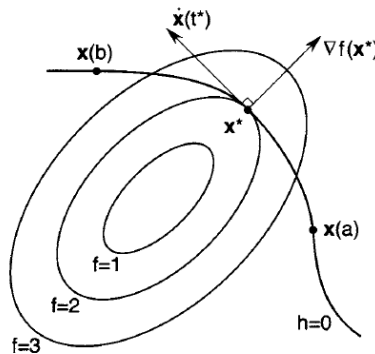


FIGURE 1.9 – Le gradient est orthogonal sur la courbe $\{x(t)\}$ au point de minimum de f .

tel que

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0.$$

□

- Remarque 1.3.** 1. Dans Théorème 1.2, λ^* est appelé le multiplicateur de Lagrange.
 2. Théorème 1.2 est valide pour un problème de maximisation, voir Figure 1.10.

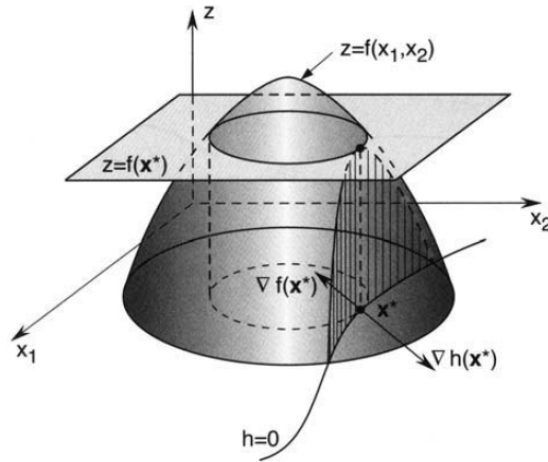


FIGURE 1.10 – Théorème de Lagrange pour un problème de maximisation avec $n = 2$ et $m = 1$.

3. La condition de Lagrange est nécessaire mais n'est pas suffisante. Voir Figure 1.11.

Nous avons maintenant généralisé le théorème de Lagrange pour des fonctions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ avec $m \leq n$.

Théorème 1.3. Soit $x^* \in \mathbb{R}^n$ est un point régulier du problème de minimisation avec contraintes d'égalité suivant :

$$\begin{cases} \min f(x), \\ h(x) = 0. \end{cases} \quad (1.1)$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ avec $m \leq n$ deux fonctions différentiables. Alors, si $x^* \in \mathbb{R}^n$ est un point de minimum local du problème (1.1), alors il existe $\lambda^* \in \mathbb{R}^m$ tel que

$$Df(x^*) + (\lambda^*)^T Dh(x^*) = 0^T.$$

Démonstration. Nous montrons que

$$\nabla f(x^*) = -Dh(x^*)^T \lambda^* \text{ pour tout } \lambda^* \in \mathbb{R}^m; \text{ c'est à dire } \nabla f(x^*) \in N(x^*).$$

Mais par Lemme ??, $N(x^*) = T(x^*)^\perp$. Il reste donc à montrer que $\nabla f(x^*) \in T(x^*)^\perp$. On suppose que $y \in T(x^*)$. Alors, par Théorème 1.1, Il existe une courbe différentiable $\{x(t) : t \in]a, b[\}$ tel que pour tout $t \in]a, b[$,

$$h(x(t)) = 0,$$

et il existe $t^* \in]a, b[$ satisfaisant

$$x(t^*) = x^*, \quad \dot{x}(t^*) = y.$$

On considère ϕ une fonction définie par :

$$\phi(t) = f(x(t)).$$

t^* est un point de minimum local de ϕ . D'après la condition nécessaire d'optimalité du premier ordre d'optimisation sans contraintes, nous avons :

$$\frac{d\phi}{dt}(t^*) = 0.$$

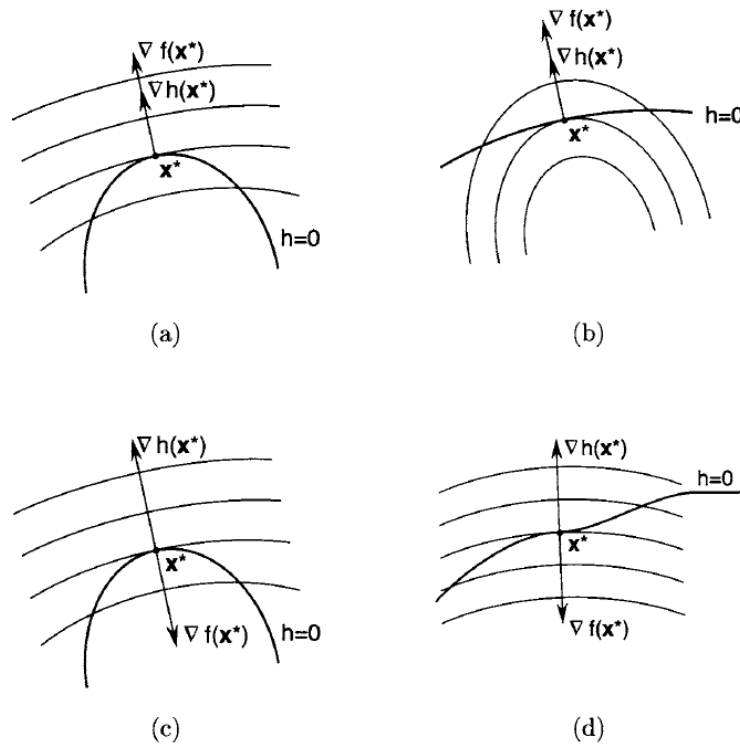


FIGURE 1.11 – (a) point de maximum (b) point de minimum (c) point de minimum (d) n'est pas un extrémum.

Avec la règle de chaîne, on obtient :

$$\frac{d\phi}{dt}(t^*) = Df(x^*) \dot{x}(t^*) = Df(x^*) y = \nabla f(x^*)^T y = 0.$$

Par conséquent, on a :

$$\nabla f(x^*) \in T(x^*)^\perp.$$

Ceci complète la preuve. □

Remarque 1.4. 1. Le théorème de Lagrange indique que si x^* est un extrémum local (ou global), alors le gradient de la fonction objective peut être exprimé comme une combinaison linéaire des gradients des contraintes.

2. Le vecteur λ^* est appelé vecteur multiplicateur de Lagrange.

3. La condition nécessaire, d'après Théorème 1.3, est donné par :

$$\nabla f(x^*) \in N(x^*).$$

Exemple 1.4. On considère le problème suivant :

$$\begin{cases} \min f(x), \\ h(x) = 0. \end{cases}$$

où $f(x) = x$ et

$$h(x) = \begin{cases} x^2 & \text{si } x < 0, \\ 0 & \text{si } 0 \leq x \leq 1, \\ (x-1)^2 & \text{si } x > 1. \end{cases}$$

L'ensemble réalisable est évidemment $[0, 1]$. Donc, $x^* = 0$ est un point de minimum local. Toutefois, $f'(x^*) = 1$ et $h'(x^*) = 0$. Par conséquent, x^* ne satisfait pas à la condition nécessaire dans le théorème de Lagrange. D'autre part, x^* n'est pas un point régulier, c'est pourquoi le théorème de Lagrange ne s'applique pas ici.

Définition 1.8. (Fonction de Lagrange). La fonction $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ définie par :

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x)$$

est appelée fonction de Lagrange (ou le Lagrangien).

Proposition 1.1. Si $x^* \in \mathbb{R}^n$ est un point d'extrémum local, alors il existe $\lambda^* \in \mathbb{R}^m$ tel que $D\mathcal{L}(x^*, \lambda^*) = 0$ c'est à dire

$$\begin{aligned} Df(x^*) + (\lambda^*)^T Dh(x^*) &= 0, \\ h(x^*) &= 0. \end{aligned}$$

Exemple 1.5. Étant donné une surface fixe du carton notée par A . Construire une boîte fermée de carton avec un volume maximum.

On note par x, y et z les dimensions de la boîte. Alors, le problème peut formulé comme suit :

$$\begin{aligned} \max \quad & xyz, \\ \text{s.t.} \quad & xy + xz + yz = \frac{A}{2}. \end{aligned}$$

On remarque que tous les points réalisables sont des points réguliers. Donc, la fonction de Lagrange est donnée par :

$$\mathcal{L}(x, y, z, \lambda) = xyz + \lambda \left(xy + xz + yz - \frac{A}{2} \right).$$

Donc, la condition de Lagrange est donnée par :

$$yz + \lambda(y + z) = 0, \tag{1.2}$$

$$xz + \lambda(x + z) = 0, \tag{1.3}$$

$$xy + \lambda(x + y) = 0, \tag{1.4}$$

$$xy + xz + yz = \frac{A}{2} \tag{1.5}$$

1. Si $\lambda = 0$, en sommant (1.2)-(1.4), on obtient $xy + xz + yz = 0$, donc contradiction avec (1.5).
2. Si $\lambda \neq 0$, en multipliant (1.2) par x et (1.3) par y , et avec la différence on obtient :

$$z\lambda(x - y) = 0.$$

Donc, $x = y$. En multipliant (1.3) par y et (1.4) par z et avec la différence, on obtient :

$$x\lambda(y - z) = 0.$$

D'où, $y = z$. D'après (1.5), nous avons :

$$x = y = z = \sqrt{A/6}.$$

Exemple 1.6. Calculer les extrémums de la fonction quadratique suivante :

$$f(x_1, x_2) = x_1^2 + x_2^2.$$

sur l'ellipse

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

La fonction de Lagrange est donnée par :

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

Nous avons :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_1} = 0, \\ \frac{\partial \mathcal{L}}{\partial x_2} = 0, \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 0. \end{cases} \Leftrightarrow \begin{cases} x_1 + \lambda x_1 = 0, \\ x_2 + 2\lambda x_2 = 0, \\ x_1^2 + 2x_2^2 = 1. \end{cases}$$

On remarque que tous les points réalisables sont des points réguliers. La première équation donne $x_1 = 0$ ou $\lambda = -1$.

1. Si $x_1 = 0$, avec la deuxième et troisième équations, on obtient :

$$\lambda = -1/2 \text{ et } x_2 = \pm 1/\sqrt{2}.$$

2. Si $\lambda = -1$, avec la deuxième et troisième équations, on obtient :

$$x_1 = \pm 1 \text{ et } x_2 = 0.$$

Les points qui sont vérifiés la condition de Lagrange :

$$x^{(1)} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \end{bmatrix}, x^{(2)} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \end{bmatrix}, x^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, x^{(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

D'autre part, on a :

$$f(x^{(1)}) = f(x^{(2)}) = 1/2$$

et

$$f(x^{(3)}) = f(x^{(4)}) = 1.$$

Donc, $x^{(1)}$ et $x^{(2)}$ points de minimum local et $x^{(3)}$ et $x^{(4)}$ points de maximum local, voir Figure 1.12.

Exemple 1.7. Soit le problème suivant :

$$\begin{cases} \max x^T Q x \\ x^T P x = 1 \end{cases}$$

où Q est une matrice d'ordre n , symétrique et semi-définie positive et P est une matrice d'ordre n , symétrique et définie positive. On pose

$$f(x) = x^T Q x \text{ et } h(x) = 1 - x^T P x.$$

On remarque que tous les points réalisables sont des points réguliers car P est définie positive. Soit la fonction de Lagrange :

$$\mathcal{L}(x, \lambda) = f(x) + \lambda h(x) = x^T Q x + \lambda(1 - x^T P x).$$

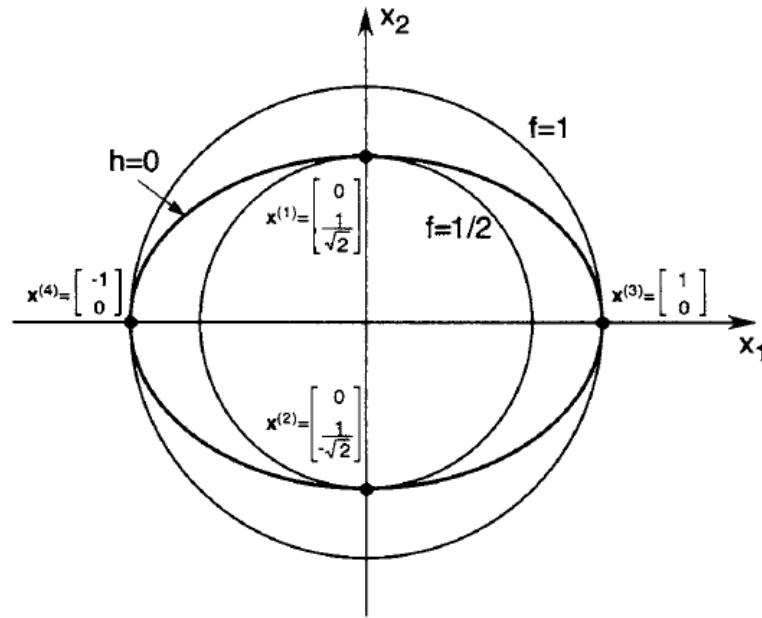


FIGURE 1.12 – Solution graphique.

Avec la condition de la Lagrange, on obtient :

$$\begin{aligned} Qx - \lambda Px &= 0, \\ 1 - x^T Px &= 0. \end{aligned}$$

D'après la première équation : on obtient :

$$P^{-1}Qx = \lambda x.$$

Alors, la solution du problème s'il existe, est un vecteur propre de la matrice $P^{-1}Q$ et le multiplicateur de Lagrange est la valeur propre associée. Soit (x^*, λ^*) une solution optimal, nous avons :

$$\begin{cases} P^{-1}Qx^* = \lambda^* x^*, \\ (x^*)^T Px^* = 1. \end{cases} \Rightarrow \lambda^* = (x^*)^T Qx^*.$$

Par conséquent, λ^* est la plus grande valeur propre de $P^{-1}Q$.

1.4 Conditions d'optimalité du second ordre

Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ deux fonctions de classe $\mathcal{C}^2$ avec $m \leq n$. Soit

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x) = f(x) + \lambda_1 h_1(x) + \dots + \lambda_m h_m(x)$$

la fonction de Lagrange. On note $\mathbf{L}(x, \lambda)$ la matrice hessienne de $\mathcal{L}(x, \lambda)$ par rapport à x :

$$\mathbf{L}(x, \lambda) = \mathbf{F}(x) + \lambda_1 \mathbf{H}_1(x) + \dots + \lambda_m \mathbf{H}_m(x)$$

où $\mathbf{F}(x)$ est la matrice hessienne de f au point x et $\mathbf{H}_k(x)$ est la matrice hessienne de h_k au point x , $k = 1, \dots, m$, donnée par :

$$\mathbf{H}_k(x) = \begin{bmatrix} \frac{\partial^2 h_k}{\partial x_1^2}(x) & \dots & \frac{\partial^2 h_k}{\partial x_n \partial x_1}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 h_k}{\partial x_1 \partial x_n}(x) & \dots & \frac{\partial^2 h_k}{\partial x_n^2}(x) \end{bmatrix}$$

Nous introduisons la notation $[\lambda \mathbf{H}(x)]$:

$$[\lambda \mathbf{H}(x)] = \lambda_1 \mathbf{H}_1(x) + \dots + \lambda_m \mathbf{H}_m(x).$$

Donc, nous avons :

$$\mathbf{L}(x, \lambda) = \mathbf{F}(x) + [\lambda \mathbf{H}(x)].$$

Théorème 1.4. (Conditions nécessaires du seconde ordre). Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ deux fonctions de classe $\mathcal{C}^2$ avec $m \leq n$ et x^* est un point régulier du problème de minimisation avec contraintes d'égalité suivant :

$$\begin{cases} \min f(x) \\ h(x) = 0 \end{cases} \quad (1.6)$$

Si x^* est un point de minimum local du problème (3.6), alors il existe $\lambda^* \in \mathbb{R}^m$ tels que

1. $Df(x^*) + (\lambda^*)^T Dh(x^*) = 0^T$.
2. Pour tout $y \in T(x^*)$, on a $y^T \mathbf{L}(x^*, \lambda^*) y \geq 0$.

Démonstration. D'après Théorème 1.3, il existe $\lambda^* \in \mathbb{R}^m$ tel que $Df(x^*) + (\lambda^*)^T Dh(x^*) = 0^T$. Il reste à prouver la deuxième partie du théorème. Soit $y \in T(x^*)$ c'est-à-dire que y appartient à l'espace tangent à $S = \{x \in \mathbb{R}^n : h(x) = 0\}$ au point x^* . Grâce Théorème 1.1, il existe une courbe $\{x(t) : t \in]a, b[\}$ sur S deux fois différentiable tels que

$$x(t^*) = x^*, \quad \dot{x}(t^*) = y \text{ où } t^* \in]a, b[.$$

Donc, t^* est un point de minimum local de la fonction ϕ définie par :

$$\phi(t) = f(x(t)).$$

D'après la condition nécessaire d'optimalité du seconde ordre du problème d'optimisation sans contraintes, nous obtenons :

$$\frac{d^2 \phi}{dt^2}(t^*) \geq 0.$$

En utilisant la formule

$$\frac{d}{dt} \left(y(t)^T z(t) \right) = z(t)^T \frac{dy}{dt}(t) + y(t)^T \frac{dz}{dt}(t)$$

et en appliquant la règle en chaîne, on obtient :

$$\begin{aligned} \frac{d^2 \phi}{dt^2}(t^*) &= \frac{d}{dt} [Df(x(t^*)) \dot{x}(t^*)] \\ &= \dot{x}(t^*)^T \mathbf{F}(x^*) \dot{x}(t^*) + Df(x^*) \ddot{x}(t^*) \\ &= y^T \mathbf{F}(x^*) y + Df(x^*) \ddot{x}(t^*) \geq 0. \end{aligned}$$

Puisque $h(x(t)) = 0$ pour tout $t \in]a, b[$, nous avons :

$$\frac{d^2}{dt^2} (\lambda^*)^T h(x(t)) = 0.$$

Donc,

$$\begin{aligned}
 \frac{d^2}{dt^2} (\lambda^*)^T h(x(t)) &= \frac{d}{dt} \left[(\lambda^*)^T \frac{d}{dt} h(x(t)) \right] \\
 &= \frac{d}{dt} \left[\sum_{k=1}^m \lambda_k^* \frac{d}{dt} h_k(x(t)) \right] \\
 &= \frac{d}{dt} \left[\sum_{k=1}^m \lambda_k^* D h_k(x(t)) \dot{x}(t) \right] \\
 &= \sum_{k=1}^m \lambda_k^* \frac{d}{dt} (D h_k(x(t)) \dot{x}(t)) \\
 &= \sum_{k=1}^m \lambda_k^* \left[\dot{x}(t)^T \mathbf{H}_k(x(t)) \dot{x}(t) + D h_k(x(t)) \ddot{x}(t) \right] \\
 &= \dot{x}^T(t) [\lambda^* \mathbf{H}(x(t))] \dot{x}(t) + (\lambda^*)^T D h(x(t)) \ddot{x}(t) \\
 &= 0.
 \end{aligned}$$

En particulier, pour $t = t^*$, on obtient :

$$y^T [\lambda^* H(x^*)] y + (\lambda^*)^T D h(x^*) \ddot{x}(t^*) = 0.$$

En ajoutant cette équation à l'inégalité

$$y^T \mathbf{F}(x^*) y + D f(x^*) \ddot{x}(t^*) \geq 0$$

on trouve

$$y^T (\mathbf{F}(x^*) + [\lambda^* \mathbf{H}(x^*)]) y + \underbrace{\left(D f(x^*) + (\lambda^*)^T D h(x^*) \right)}_{=0} \ddot{x}(t^*) \geq 0.$$

Donc,

$$y^T (\mathbf{F}(x^*) + [\lambda^* \mathbf{H}(x^*)]) y = y^T \mathbf{L}(x^*, \lambda^*) y \geq 0,$$

ce qui prouve le théorème. □

Théorème 1.5. (Conditions suffisantes d'optimalité du seconde ordre). Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ deux fonctions de classe $\mathcal{C}^2$ avec $m \leq n$.

1. On suppose qu'il existe $x^* \in \mathbb{R}^n$ et $\lambda^* \in \mathbb{R}^m$ tels que :

(a) $D f(x^*) + (\lambda^*)^T D h(x^*) = 0^T$.

(b) Pour tout $y \in T(x^*)$, $y \neq 0$, nous avons $y^T \mathbf{L}(x^*, \lambda^*) y > 0$.

Alors, x^* est un point de **minimum local strict** du problème de minimisation avec contraintes d'égalité suivant :

$$\begin{cases} \min f(x) \\ h(x) = 0 \end{cases}$$

2. On suppose qu'il existe $x^* \in \mathbb{R}^n$ et $\lambda^* \in \mathbb{R}^m$ tels que :

(a) $D f(x^*) + (\lambda^*)^T D h(x^*) = 0^T$.

(b) Pour tout $y \in T(x^*)$, $y \neq 0$, nous avons $y^T \mathbf{L}(x^*, \lambda^*) y < 0$.

Alors, x^* est un point de **maximum local strict** du problème de maximisation avec contraintes d'égalité suivant :

$$\begin{cases} \max f(x) \\ h(x) = 0 \end{cases}$$

Démonstration. En raisonnant par absurde. On suppose que x^* n'est pas un point de minimum local strict. Donc, il existe une suite (y_k) des points réguliers converge vers x^* telle que

$$\forall k \in \mathbb{N}, f(y_k) \leq f(x^*).$$

On peut écrire y_k sous la forme :

$$y_k = x^* + \delta_k w_k, \text{ où } w_k \in \mathbb{R}^n, \|w_k\| = 1 \text{ et } \delta_k > 0 \text{ pour tout } k \in \mathbb{N}.$$

Il est évident que

$$\delta_k \xrightarrow[k \rightarrow +\infty]{} 0 \text{ et } w_k \xrightarrow[k \rightarrow +\infty]{} w^*.$$

D'autre part, nous avons :

$$h(y_k) - h(x^*) = 0$$

En divisant sur δ_k et par passage à la limite $k \rightarrow +\infty$, on obtient :

$$\lim_{k \rightarrow +\infty} \frac{h(y_k) - h(x^*)}{\delta_k} = \lim_{k \rightarrow +\infty} \frac{h(x^* + \delta_k w_k) - h(x^*)}{\delta_k} = Dh(x^*) w^* = 0.$$

Donc, $w^* \in T(x^*)$. Avec la formule de Taylor, nous avons :

$$0 = h_j(y_k) = h_j(x^*) + \delta_k Dh_j(x^*) w_k + \frac{\delta_k^2}{2} w_k^T \mathbf{H}_j(\eta_j) w_k \quad (1.7)$$

et

$$0 \geq f(y_k) - f(x^*) = \delta_k Df(x^*) w_k + \frac{\delta_k^2}{2} w_k^T \mathbf{F}(\eta_0) w_k \quad (1.8)$$

où $\eta_j, j = 0, 1, \dots, m$ est un point dans le segment $]x^*, y_k[$. En multipliant (3.7) par λ_j^* et en sommant avec (1.8), on obtient :

$$0 \geq \delta_k \left(\underbrace{Df(x^*) + \sum_{j=1}^m \lambda_j^* Dh_j(x^*)}_{=0} \right) + \frac{\delta_k^2}{2} w_k^T \left[\mathbf{F}(\eta_0) + \sum_{j=1}^m \lambda_j^* \mathbf{H}_j(\eta_j) \right] w_k.$$

Donc,

$$w_k^T \left[\mathbf{F}(\eta_0) + \sum_{j=1}^m \lambda_j^* \mathbf{H}_j(\eta_j) \right] w_k \geq 0$$

Par passage à la limite $k \rightarrow +\infty$, on obtient :

$$(w^*)^T \mathbf{L}(x^*, \lambda^*) w^* \geq 0.$$

D'où, contradiction avec l'hypothèse (b) du théorème. □

Exemple 1.8. Résoudre le problème suivant :

$$\begin{cases} \max x^T Q x \\ x^T P x = 1 \end{cases}$$

où

$$Q = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{et} \quad P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

La fonction de Lagrange est donnée par :

$$\mathcal{L}(x, \lambda) = x^T Q x + \lambda (1 - x^T P x)$$

D'après la condition de Lagrange, on obtient :

$$P^{-1} Q x = \lambda x$$

où

$$P^{-1} Q = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Donc, λ^* est la plus grande valeur propre de la matrice $P^{-1} Q$, donc $\lambda^* = 2$. D'autre part, on pose $x = (x_1, x_2)^T$, on a :

$$\begin{cases} P^{-1} Q x = 2x, \\ x^T P x = 1 \end{cases} \Leftrightarrow \begin{cases} x_2 = 0, \\ 2x_1^2 + x_2^2 = 1. \end{cases} \Leftrightarrow x^* \in \left\{ \left(1/\sqrt{2}, 0 \right)^T, \left(-1/\sqrt{2}, 0 \right)^T \right\}.$$

Nous avons :

$$L(x^*, \lambda^*) = 2Q - 2\lambda^* P = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$$

L'espace tangent $T(x^*)$ de $\{x \in \mathbb{R}^2 : 1 - x^T P x = 0\}$ est donné par :

$$\begin{aligned} T(x^*) &= \{y \in \mathbb{R}^2 : (x^*)^T P y = 0\} \\ &= \{y \in \mathbb{R}^2 : [\sqrt{2} \ 0] y = 0\} \\ &= \{(0, a)^T : a \in \mathbb{R}\}. \end{aligned}$$

Pour tout $y \in T(x^*)$, $y \neq 0$, nous avons :

$$y^T \mathbf{L}(x^*, \lambda^*) y = \begin{bmatrix} 0 & a \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ a \end{bmatrix} = -2a^2 < 0.$$

D'où, x^* est un point de maximum local strict.

1.5 Minimisation quadratique avec contraintes d'égalité linéaire

Définition 1.9. Minimisation d'une fonction quadratique avec contraintes d'égalité linéaire est un problème d'optimisation avec contraintes donné par :

$$\begin{cases} \min \frac{1}{2} x^T Q x \\ Ax = b \end{cases} \quad (1.9)$$

où Q est une matrice d'ordre n , symétrique et définie positive. $A \in \mathcal{M}_{m,n}(\mathbb{R})$ avec $m \leq n$ et $\text{rg}(A) = m$.

Proposition 1.2. *Le problème (1.9) admet un point de minimum local strict x^* donné par :*

$$x^* = Q^{-1} A^T (A Q^{-1} A^T)^{-1} b.$$

Démonstration. La fonction quadratique est donnée par :

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^T Q x + \lambda (b - Ax).$$

Nous avons :

$$D\mathcal{L}(x, \lambda) = 0 \Leftrightarrow \begin{cases} x^T Q - \lambda^T A = 0^T, \\ Ax = b. \end{cases} \Leftrightarrow \begin{cases} x = Q^{-1} A^T \lambda, \\ Ax = A Q^{-1} A^T \lambda = b. \end{cases}$$

Donc, on a :

$$\lambda = (A Q^{-1} A^T)^{-1} b \quad \text{et} \quad x = Q^{-1} A^T (A Q^{-1} A^T)^{-1} b.$$

D'autre part, on a :

$$\mathbf{L}(x, \lambda) = Q.$$

D'où, x est un point de minimum local strict. □