Cours d'optimisation en dimension finie faible

Olivier LAFITTE

Contents

Ce cours présente des résultats fondamentaux d'optimisation, mais appliqués à la minimisation de fonctions de classe C^1 ou C^2 de deux ou trois variables réelles, sous contraintes égalité ou inégalité. Nous nous restreignons à une ou deux contraintes inégalité ou égalité, car, génériquement, en dimension 3, deux contraintes égalité conduisent à une variété de dimension 1, en dimension 2 deux contraintes égalité conduisent, génériquement, à un point. Nous étudierons aussi le cas quadratique en dimension quelconque.

Chapter 1

Le cas quadratique

1.1 Minimum d'une forme quadratique

Soit A une matrice quelconque de $\mathbb{R}^3 \times \mathbb{R}^3$, et $b \in \mathbb{R}^3$. On considère la fonction f, de $\mathbb{R}^3$ dans $\mathbb{R}$, donnée par

$$\begin{aligned} f(x_1, x_2, x_3) &= \tfrac{1}{2}(\textstyle\sum_{i,j} a_{ij}x_ix_j) - \sum b_ix_i = \tfrac{1}{2}(Ax, x) - (b, x) \\ &= \tfrac{1}{2}(a_{11}x_1^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{21}x_2x_1 + a_{22}x_2^2 + a_{23}x_2x_3 + a_{31}x_1x_3 + a_{32}x_2x_3 + a_{33}x_3^2) \\ &\quad -(b_1x_1 + b_2x_2 + b_3x_3) \end{aligned}$$

Lemme 1 *Soit $A^s = \frac{1}{2}(A + A^T)$, où A^T désigne la matrice transposée de A. Alors $\forall x, f(x) = \frac{1}{2}(A^s x, x) - (b, x)$.*

On remplace ainsi A par la matrice dont les composants extra-diagonaux sont $\tilde{a}_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$. Dans toute la suite, on prendra A symétrique puisqu'on se ramène à une matrice symétrique.

Lemme 2 *Il existe une base orthonormée de $\mathbb{R}^3$ telle que si (y_1, y_2, y_3) sont les coordonnées de x dans cette base, $\lambda_1, \lambda_2, \lambda_3$ sont les valeurs propres ordonnées associées et (c_1, c_2, c_3) les coordonnées de b dans cette base, alors $f(x) = \frac{1}{2}\lambda_1 y_1^2 - c_1y_1 + \frac{1}{2}\lambda_2 y_2^2 - c_2y_2 + \frac{1}{2}\lambda_3 y_3^2 - c_3y_3$.*

Comme A est symétrique, elle est diagonalisable dans une base orthonormée. Les colonnes d'une matrice de passage P sont donc des vecteurs propres de norme 1 associés à $\lambda_1, \lambda_2, \lambda_3$ (il y a donc 8 choix possibles pour P quand les trois valeurs propres sont distinctes). Comme les vecteurs sont de norme 1, on vérifie que $PP^T = I$ donc l'inverse de P est P^T. On a donc $AP = P\Delta$, où Δ est la matrice diagonale $Diag(\lambda_1, \lambda_2, \lambda_3)$. Prenant $x = Py$ (ou $y = P^Tx$, ce qui revient à dire que (y_1, y_2, y_3) sont les coordonnées de x dans la base (P_1, P_2, P_3), $x = y_1P_1 + y_2P_2 + y_3P_3$)

$$f(x) = \frac{1}{2}(APy, Py) - (b, Py) = \frac{1}{2}(P\Delta y, Py) - (P^Tb, y) = \frac{1}{2}(\Delta y, y) - (P^Tb, y) = g(y).$$

Une fois cette forme écrite, on a donc

$$g(y) = \frac{1}{2}(\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2) - (c_1 y_1 + c_2 y_2 + c_3 y_3).$$

Theorem 1 *La forme quadratique $f(x) = \frac{1}{2}(Ax, x) - (b, x)$ n'a pas de minimum fini si une des valeurs propres de A, matrice symétrique, est strictement négative.*
Cette forme quadratique a un unique minimum atteint en $x_0 = A^{-1}b$, de valeur $-\frac{1}{2}(A^{-1}b, b)$, quand toutes les valeurs propres de A sont strictement positives. Dans ce cas on a

$$f(x) = \frac{1}{2}(A(x - x_0), x - x_0) + f(x_0)$$

et les inégalités

$$\frac{1}{2}\lambda_{min}||x - x_0||^2 \leq f(x) - f(x_0) \leq \frac{1}{2}\lambda_{max}||x - x_0||^2$$

où λ_{min} et λ_{max} sont respectivement la plus petite et la plus grande valeur propre de la matrice A.
Lorsque le noyau de A n'est pas réduit à $\{0\}$, alors f n'admet de minimum que si b appartient à l'orthogonal du noyau de A, donc à l'image de A.

Démonstration. Le minimum de f est aussi le minimum de g (les points de minimum de f sont reliés à ceux de g par $x = Py$). Pour trouver le minimum de g, si il existe, on peut donc considérer successivement le minimum dans chacune des coordonnées y. On trouve les cas suivants

1. si une des valeurs propres est strictement négative, alors g a pour minimum $-\infty$ (il suffit de faire tendre la coordonnée correspondante vers l'infini)

2. si les trois valeurs propres sont strictement positives, en utilisant la forme réduite (valable pour tout $\lambda \neq 0$)

$$\frac{1}{2}\lambda x^2 - cx = \frac{1}{2}\lambda(x - \frac{c}{\lambda})^2 - \frac{c^2}{2\lambda}$$

   on vérifie que g se met sous la forme

$$g(x) = \frac{1}{2}\lambda_1(x - \frac{c_1}{\lambda_1})^2 - \frac{c_1^2}{2\lambda_1} + \frac{1}{2}\lambda_2(x - \frac{c_2}{\lambda_2})^2 - \frac{c_2^2}{2\lambda_2} + \frac{1}{2}\lambda_3(x - \frac{c_3}{\lambda_3})^2 - \frac{c_3^2}{2\lambda_3}$$

   ainsi g admet comme minimum la valeur $-\frac{c_1^2}{2\lambda_1} - \frac{c_2^2}{2\lambda_2} - \frac{c_3^2}{2\lambda_3}$, atteint au point $y = (\frac{c_1}{\lambda_1}, \frac{c_2}{\lambda_2}, \frac{c_3}{\lambda_3}) = \Delta^{-1}c$. Le point x correspondant est $x = Py = P\Delta^{-1}P^T b = (P\Delta P^{-1})^{-1}b = A^{-1}b$.

3. si une des valeurs propres est nulle, par exemple λ_3, on trouve que $g(x) = \frac{1}{2}(\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2) - (c_1 y_1 + c_2 y_2 + c_3 y_3)$. Si $c_3 \neq 0$ (c'est-à-dire si c n'appartient pas à l'image de la matrice Δ), alors le minimum de g est $-\infty$, si $c_3 = 0$ on se ramène au minimum d'une fonction de deux variables.

1.2 Minimum d'une forme quadratique sous contrainte affine

Nous pouvons aussi étudier la minimisation de f sous contrainte affine (premier cas de contrainte). Pour cela, on cherche à minimiser $f(x)$ sous la contrainte

$$V.x = K$$

où V est un vecteur non nul de l'espace. On note $C(x) = V.x - K$.
Nous démontrons le théorème suivant.

Theorem 2 • *minimum sous une contrainte égalité affine: On suppose V non nul et A symétrique définie positive. L'unique point de minimum de f sous la contrainte $V.x = K$ ($C(x) = 0$) est l'unique point x_* tel qu'il existe λ réel tel que*

$$Ax_* - b + \lambda V = 0, V.x_* = K$$

qui s'écrit aussi $f'(x_*) + \lambda C'(x_*) = 0, C(x_*) = 0$.
• *minimum sous deux contraintes égalité affines: On suppose V_1 et V_2 non colinéaires et A symétrique définie positive. L'unique point de minimum de f sous les contraintes $V_1.x = K_1, V_2.x = K_2$ est l'unique point x_{**} tel qu'il existe λ et μ tels que $Ax_{**} - b + \lambda V_1 + \mu V_2 = 0, V_1.x_{**} = K_1, V_2.x_{**} = K_2$, qui s'écrit encore $f'(x_{**}) + \lambda C_1'(x_{**}) + \mu C_2'(x_{**}) = 0$, x_{**} appartenant à l'ensemble des contraintes.*
• *Si V_1 et V_2 sont colinéaires, l'ensemble $\{x, V_1.x = K_1, V_2.x = K_2\}$ est soit vide soit est égal à $\{x, V_1.x = K_1\}$, donc on est ramené au cas d'une contrainte.*

Ce théorème s'appelle le théorème des multiplicateurs de Lagrange sous contraintes égalité, dans le cas très simplifié d'une fonction quadratique et de contraintes affines, mais toutes les idées du cas général sont présentes:
i) indépendance des contraintes (dans le sens où les vecteurs normaux aux hyperplans définissant les contraintes sont linéairement indépendants),
ii) relation sur les dérivées.
Dans ce théorème, le coefficient λ ou les coefficients λ et μ sont les multiplicateurs de Lagrange.

Il est bon de rappeler que si on a une contrainte affine, cet ensemble est un hyperplan, et le vecteur dérivé de la fonction affine caractérisant cette contrainte égalité est proportionnel au vecteur normal unitaire à l'hyperplan.

Preuve du théorème dans le cas d'une contrainte égalité Nous commencons par identifier un point satisfaisant la contrainte. On cherche par exemple $x_1 = \mu V$. La valeur de μ est alors donnée par $\mu(V.V) = K$. Comme V est non nulle, cela donne $\mu = \frac{K}{(V.V)}$. On pose alors $X = x - x_1$. On trouve alors

$$V.X = 0, f(x_1+X) = \frac{1}{2}(A(x_1+X), x_1+X) - (b, x_1+X) = f(x_1) + \frac{1}{2}(AX, X) + (Ax_1 - b, X).$$

Comme $f'(x) = Ax - b$ et $f''(x) = A$, cette égalité s'écrit

$$f(x_1 + X) = f(x_1) + (f'(x_1), X) + \frac{1}{2}(f''(x_1)X, X).$$

Comme V est non nul, au moins une de ses composantes est non nulle. Pour simplifier, nous supposons qu'il s'agit de V_3 et on introduit $\theta_1 = -\frac{V_1}{V_3}, \theta_2 = -\frac{V_2}{V_3}$. $V.X = 0$ est alors équivalent à $X_3 = \theta_1 X_1 + \theta_2 X_2$. On trouve alors $X = X_1 W_1 + X_2 W_2$ où les vecteurs W_1 et W_2 sont $(1, 0, \theta_1)$ et $(0, 1, \theta_2)$. On a alors

$$\begin{aligned} f(x_1 + X) \quad & = f(x_1) + \tfrac{1}{2}[(AW_1, W_1)X_1^2 + 2(AW_1, W_2)X_1X_2 + (AW_2, W_2)X_2^2] \\ & -(Ax_1 - b, W_1)X_1 + (Ax_1 - b, W_2)X_2 \\ & = f(x_1) + (f'(x_1), W_1)X_1 + (f'(x_1), W_2)X_2 \\ & +\tfrac{1}{2}(\tilde{A}X, X) \end{aligned}$$

où les éléments de la matrice $\tilde{A}$ sont (AW_i, W_j).
Le point critique de la fonction $(X_1, X_2) \to f(x_1 + X_1W_1 + X_2W_2)$ (qui est unique car la matrice $(AW_i, W_j)_{ij}$ est symétrique définie positive) est alors donné par les équations

$$(AW_1, W_1)X_1 + (AW_1, W_2)X_2 + (Ax_1 - b, W_1) = 0, (AW_1, W_2)X_1 + (AW_2, W_2)X_2 + (Ax_1 - b, W_2) = 0.$$

Le vecteur $A(X_1W_1 + X_2W_2) + Ax_1 - b$ est donc orthogonal à W_1 et à W_2. Il est donc colinéaire à V. Il existe donc λ tel que

$$A(X_1W_1 + X_2W_2 + x_1) - b + \lambda V = 0.$$

La méthode pour trouver $X_1W_1 + X_2W_2$ s'écrit alors directement (sans introduire les vecteurs W_1 et W_2). Pour cela, on écrit la solution $X = x_1 + X_1W_1 + X_2W_2$ en fonction de λ quelconque:

$$X = -x_1 + A^{-1}b - \lambda A^{-1}V$$

et on cherche λ de sorte que $V.X = 0$, soit

$$-V.x_1 + V.A^{-1}b + \lambda V.A^{-1}V = 0$$

puisque X doit vérifier la contrainte. Comme V est non nul et que A^{-1} est aussi symétrique définie positive $VA^{-1}V$ est non nul, ce qui permet de calculer λ.

On a donc vérifié le théorème, et il est donc beaucoup plus simple d'utiliser le théorème pour trouver le point de minimum.

Preuve du théorème dans le cas de deux contraintes égalité indépendantes
Les deux équations $V_1.x = K_1, V_2.x = K_2$ sont linéairement indépendantes puisque les vecteurs V_1 et V_2 ne sont pas colinéaires. Ainsi le système de deux équations à trois inconnues $V_1x = K_1, V_2x = k_2$ peut permettre de trouver deux des inconnues en fonction de la troisième (on étudie pour cela les sous-matrices 2;2 de la matrice comportant les deux lignes V_1 et V_2, et au moins un des sous-systèmes est inversible.). Dans ce cas on note W un vecteur orthogonal à V_1 et à V_2. Un choix possible de W est le produit vectoriel $V_1 \wedge V_2$. On sait alors qu'il existe x_2 tel que

$$V_1.x = k_1, V_2.x = k_2 \Leftrightarrow x = x_2 + tW.$$

On a alors

$$f(x) = f(x_2+tW) = \frac{1}{2}A(x_2+tW),(x_2+tW)-(b,x_2+tW) = \frac{1}{2}(AW,W)t^2+(Ax_2-b,W)t+f(x_2).$$

Le minimum de cette fonction est atteint pour t_* tel que $(AW,W)t_* + (Ax_2 - b, W) = 0$. Le vecteur $A(t_*W + x_2) - b$ est donc orthogonal à W. Il est donc combinaison linéaire de V_1 et V_2, ce qui revient donc à écrire l'égalité suivante pour $x = t_*W + x_2$:

$$Ax - b + \lambda V_1 + \mu V_2 = 0, x.V_1 = K_1, x.v_2 = K_2. \tag{1.1}$$

Remarque: Dans ce cas on a plus simplement $t = (AW,W)^{-1}(b - Ax_2, W)$, ce qui permet de trouver $x = x_2 + (AW,W)^{-1}(b - Ax_2, W)W$.

Preuve du deuxième item du théorème. La recherche du point de minimum ne se fait pas par l'intermédiaire de ces calculs, mais par une méthode directe comme dans le cas d'une contrainte. La démonstration ci-dessus a permis de trouver la **forme** (1.1) de x, que nous utilisons maintenant: On considère x de la forme (1.1). Alors $x = A^{-1}(b - \lambda V_1 - \mu V_2)$. On remplace cette valeur de x dans les deux égalités $V_1x = k_1; V_2x = k_2$; C'est un système de deux équations à deux inconnues λ et μ, dont la matrice est la matrice $(A^{-1}V_i, V_j)_{ij}$, qui est inversible. On détermine ainsi les deux multiplicateurs de Lagrange λ et μ.

Pour terminer cette partie, traitons le cas d'une forme quadratique en dimension 2 avec une contrainte égalité, soit $f(x_1,x_2) = \frac{1}{2}(A_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2) - (b_1x_1 + b_2x_2)$ sous la contrainte $x_1x_1 + c_2x_2 = k$.
L'espace des contraintes s'écrit $(x_1,x_2) = \frac{k}{c_1^2+c_2^2}(c_1,c_2) + t(-c_2,c_1)$ car le point $x_0 = \frac{k}{c_1^2+c_2^2}(c_1,c_2)$ appartient bien aux contraintes $((c_1,c_2) \neq (0,0))$, donc le point $x - x_0$ est orthogonal à (c_1,c_2), donc colinéaire à $(-c_2,c_1)$.
Il suffit alors de remplacer $(x_1,x_2) = x_0 + t(-c_2,c_1)$ dans l'expression $f(x) = \frac{1}{2}(Ax,x) - (b,x) = f(x_0) + t(Ax_0 - b, (-c_2,c_1)) + \frac{1}{2}t^2(A(-c_2,c_1),(-c_2,c_1))$ et de calculer le minimum en t.

Le cas où les deux contraintes sont linéairement dépendantes s'étudie en remarquant que, si les contraintes sont linéairement dépendantes, il existe un réel k non nul tel que $V_1 = kV_2$. Alors $V_1.x = K_1$ équivaut à $kV_2.x = K_2$, ou encore $V_2.x = \frac{K_2}{k}$. Dans le cas où $K_1 \neq \frac{K_2}{k}$, il n'y a aucune solution commune. Dans le cas ou $K_1 = \frac{K_2}{k}$, les deux égalités $V_1.x = K_1$ et $V_2.x = K_2$ sont équivalentes donc on s'est ramené à une seule contrainte.

Chapter 2

Etude des minima de fonctions quelconques définies sur $\mathbb{R}^3$

2.1 Minimum libre (sans contraintes)

On suppose dans ce chapitre, selon les cas, que f est de classe C^1 ou de classe C^2 si nécessaire. On commence par des résultats de représentation des fonctions par la fonction quadratique 'la plus proche'/

2.1.1 Formules de Taylor à l'ordre 1 et à l'ordre 2

On a, dans ce cas, les formules de Taylor suivantes, généralisation de la formule de Taylor en dimension 1, où $\vec{x} = (x, y, z)$ et $\vec{x}_0 = (x_0, y_0, z_0)$:

Proposition 1

$$f(\vec{x}) = f(\vec{x}_0) + \int_0^1 [(x-x_0)\partial_x f(\vec{x}_0+t(\vec{x}-\vec{x}_0)) + (y-y_0)\partial_y f(\vec{x}_0+t(\vec{x}-\vec{x}_0)) + (z-z_0)\partial_z f(\vec{x}_0+t(\vec{x}-\vec{x}_0))]dt \tag{2.1}$$

ainsi que

$$\begin{aligned} f(\vec{x}) = \quad & f(\vec{x}_0) + (x - x_0)\partial_x f(\vec{x}_0) + (y - y_0)\partial_y f(\vec{x}_0) + (z - z_0)\partial_z f(\vec{x}_0) \\ & + \int_0^1 (1-t)\big(Hessf(\vec{x}_0 + t(\vec{x} - \vec{x}_0))(\vec{x} - \vec{x}_0), (\vec{x} - \vec{x}_0)\big)dt \end{aligned} \tag{2.2}$$

où $Hessf$ est la matrice hessienne de f, supposée de classe C^2, des dérivées partielles seconde de f. C'est une matrice symétrique.

La preuve de ces deux résultats est une conséquence de la formule de Taylor appliquée à la fonction de la variable réelle t

$$\phi(t) = f(\vec{x}_0 + t(\vec{x} - \vec{x}_0)) = f(x_0 + t(x - x_0), y_0 + t(y - y_0), z_0 + t(z - z_0))$$

En effet, sa dérivée est égale à

$$\phi'(t) = (x-x_0)\partial_x f(\vec{x}_0+t(\vec{x}-\vec{x}_0))+(y-y_0)\partial_y f(\vec{x}_0+t(\vec{x}-\vec{x}_0))+(z-z_0)\partial_z f(\vec{x}_0+t(\vec{x}-\vec{x}_0))$$

et sa dérivée seconde s'obtient en dérivant $\phi'(t)$. Par exemple, la dérivée par rapport à t du terme $(y-y_0)\partial_y f(\vec{x}_0+t(\vec{x}-\vec{x}_0))$ est égale à

$$(y-y_0)[(x-x_0)\partial^2_{xy} f(\vec{x}_0+t(\vec{x}-\vec{x}_0))+(y-y_0)\partial^2_{yy} f(\vec{x}_0+t(\vec{x}-\vec{x}_0))+(z-z_0)\partial^2_{zy} f(\vec{x}_0+t(\vec{x}-\vec{x}_0))].$$

Il suffit alors d'écrire

$$\phi(1) = \phi(0) + \int_0^1 \phi'(t)dt$$

(formule de Taylor à l'ordre 1 sur $[0,1]$) et

$$\phi(1) = \phi(0) + t\phi'(0) + \int_0^1 (1-t)\phi''(t)dt$$

(formule de Taylor à l'ordre 2) pour obtenir respectivement (2.1) et (2.2).

2.1.2 Equation d'Euler pour le minimum d'une fonction de 3 variables. Condition de Legendre

On a alors les résultats

Theorem 3 *Si f a un minimum local au point X_0, alors $(\partial_x f, \partial_y f, \partial_z f)(X_0) = 0$. Cette quantité se note $\nabla f(X_0)$ ou $f'(X_0)$ par abus de langage. La réciproque est fausse. Si f admet un minimum local au point X_0, et si elle est de classe C^2, alors $\nabla f(X_0) = 0$ et la matrice hessienne en X_0 est une matrice positive (toutes ses valeurs propres sont positives).*

Preuve du théorème La démonstration du premier point se fait en considérant $f(x_0, y_0, z)$ et en remarquant que si f admet un minimum local en X_0, alors la fonction $z \to f(x_0, y_0, z)$ admet un minimum local en z_0, donc la dérivée par rapport à z en z_0 est nulle. On fait de même pour les deux autres composantes.

La démonstration du deuxième point utilise le résultat du premier point. On a alors, pour X dans un voisinage de X_0, par application de (2.1)

$$f(X) = f(X_0) + \int_0^1 (1-t)[Hessf(X_0+t(X-X_0))(X-X_0), (X-X_0)dt] \geq f(X_0)$$

On écrit alors $X = X_0 + \epsilon w$, où w est un vecteur de norme 1. Dans ce cas, l'inégalité $f(X) \geq f(X_0)$ implique

$$\epsilon^2 \int_0^1 (1-t)(Hessf(X_0 + t\epsilon w)w, w)dt \geq 0$$

En divisant par $\epsilon^2 \neq 0$ on trouve

$$\int_0^1 (1-t)(Hessf(X_0 + t\epsilon w)w, w)dt \geq 0$$

On utilise la continuité de la matrice hessienne, donc continuité uniforme, et en passant à la limite on trouve

$$\frac{1}{2}(Hessf(X_0)w, w) \geq 0.$$

Cette inégalité est valable quel que soit le vecteur w.

Or la matrice $Hessf(X_0)$ est symétrique donc elle est diagonalisable dans une base orthonormée. En appliquant l'inégalité ci-dessus à w vecteur propre quelconque de cette matrice, on trouve que toutes ses valeurs propres sont positives.

On déduit immédiatement le

Theorem 4 *On suppose que $\nabla f(X_0)(:= f'(X_0)) = 0$ et que la matrice hessienne $Hessf(X_0)$ a toutes ses valeurs propres strictement positives. Alors X_0 est un point de minimum local de la fonctionnelle.*

On utilise encore la formule de Taylor. On trouve

$$f(X) - f(X_0) = \int_0^1 (1-t)[Hessf(X_0 + t(X - X_0))(X - X_0), (X - X_0)dt].$$

Comme la matrice $Hessf(X_0)$ a toutes ses valeurs propres strictement positives, il existe $\alpha > 0$ et $\delta > 0$ tel que pour $||X - X_0|| \leq \delta$, toutes les valeurs propres de $Hessf(X)$ sont supérieures ou égales à α. Ainsi $(Hessf(X)w, w) \geq \alpha ||w||^2$. On en déduit que

$$f(X) - f(X_0) \geq \frac{1}{2}\alpha ||X - X_0||^2$$

pour $||X - X_0|| \leq \delta$, ainsi X_0 est un point de minimum local.

2.2 Minimum sous contraintes égalité

2.2.1 Les résultats principaux

Etudions le minimum sous contraintes égalité ou inégalité de cette fonction de trois variables. Nous traitons successivement

Theorem 5 *Soit f une fonction de classe C^1 sur un ouvert Ω de $\mathbb{R}^3$ et soit g une fonction de classe C^1 sur Ω. On suppose qu'il existe $x_* \in \Omega$ tel que $g(x_*) = 0$. Si g vérifie $\nabla_x g(x) \neq 0$ sur $\Omega' \subset \Omega$, alors un minimum local $x_0$de f sur $\bar{\Omega}'$ sous la contrainte $g = 0$ est solution de*

$$\exists \lambda \in \mathbb{R}, \nabla f(x_0) + \lambda \nabla g(x_0) = 0, g(x_0) = 0.$$

Autrement dit la recherche d'un point de minimum local sous contraintes se fait en formant l'équation $\nabla f(x) + \lambda \nabla g(x) = 0$, et en étudiant, pour toute solution $x(\lambda)$ de cette équation, l'équation $g(x(\lambda)) = 0$ qui est une équation à une inconnue réelle λ. **La condition $\nabla_x g(x_0) \neq (0,0,0)$ dans un voisinage de x_0 s'exprime par la phrase 'La contrainte g est régulière en x_0'.**

Preuve dans le cas d'une condition égalité régulière (gradient non nul) On écrit l'équation $g(x) = 0$ sur $\bar{\Omega}'$. Comme la contrainte est régulière, $\nabla g \neq (0,0,0)$. En traitant le problème localement, il existe une coordonnée pour laquelle la dérivée partielle est non nulle en ce point. Quitte à changer l'ordre des coordonnées, on peut supposer qu'il s'agit de x_3.

Dans ce cas, le théorème des fonctions implicites implique qu'il existe $\psi(x_1, x_2)$ telle que $g(x_1, x_2, x_3) = 0$ dans un voisinage de x_0 est équivalent à $x_3 = \psi(x_1, x_2)$. Dans ce voisinage, le problème de minimisation de f sous la contrainte $g = 0$ est équivalent à minimiser dans ce voisinage en (x_1, x_2) la fonction $F(x_1, x_2) = f(x_1, x_2, \psi(x_1, x_2))$.

On peut alors appliquer le théorème 3 (condition nécessaire: le gradient est nul en ce point). Ceci donne:

$$\begin{cases} \partial_1 f(x_0) + \partial_1 \psi(x_1^0, x_2^0) \partial_3 f(x_0) = 0 \\ \partial_2 f(x_0) + \partial_2 \psi(x_1^0, x_2^0) \partial_3 f(x_0) = 0. \end{cases}$$

Comme, d'autre part, l'identité $g(x_1, x_2, \psi(x_1, x_2)) = 0$ est satisfaite et comme $\partial_3 g(x_0) \neq 0$, on trouve

$$\partial_1 g(x_0) + \partial_1 \psi(x_1^0, x_2^0) \partial_3 g(x_0) = 0, \partial_2 g(x_0) + \partial_2 \psi(x_1^0, x_2^0) \partial_3 g(x_0) = 0.$$

On introduit alors

$$\lambda = -\frac{\partial_3 f(x_0)}{\partial_3 g(x_0)}.$$

On vérifie que les deux égalités que nous avons écrit ci-dessus impliquent

$$\begin{cases} \partial_1 f(x_0) + \lambda \partial_1 g(x_0) = 0 \\ \partial_2 f(x_0) + \lambda \partial_2 g(x_0) = 0. \end{cases}$$

Comme on a (bien sûr) $\partial_3 f(x_0) + \lambda \partial_3 g(x_0) = 0$ par définition de λ, Le théorème est démontré.

2.2.2 Minimum d'une fonction de trois variables dans le cas de deux contraintes inégalité

Theorem 6 *Soit f une fonction de classe C^1 sur un ouvert Ω de $\mathbb{R}^3$ et soit g_1, g_2 deux fonctions de classe C^1. Les conditions pour que x_0 soit un minimum local de f sous la contrainte $g_1 = g_2 = 0$, dans le cas où $(\nabla g_1(x_0), \nabla g_2(x_0))$*

forme une famille libre de $\mathbb{R}^3$ (on dit là encore que les contraintes sont régulières en x_0) sont

$$\exists(\lambda_1,\lambda_2)\in\mathbb{R}^2, \nabla_x f(x_0)+\lambda_1\nabla_x g_1(x_0)+\lambda_2\nabla_x g_2(x_0)=0, g_1(x_0)=g_2(x_0)=0.$$

Première partie de la preuve: résolution locale des deux contraintes Comme les vecteurs $(\nabla g_1(x_0),\nabla g_2(x_0))$ sont linéairement indépendants, une sous-matrice 2×2 extraite de la matrice formée des deux lignes est inversible. Pour fixer les idées, supposons que la matrice

$$\begin{pmatrix} \partial_1 g_1 & \partial_2 g_1 \\ \partial_1 g_2 & \partial_2 g_2 \end{pmatrix}(x_0)$$

soit inversible. Alors le théorème des fonctions implicites prouve qu'il existe deux fonctions $\psi_1(x_3),\psi_2(x_3)$ telles que $g_1(\psi_1(x_3),\psi_2(x_3),x_3)=g_2(\psi_1(x_3),\psi_2(x_3),x_3)=0$ et que, pour tout x_3 dans un voisinage de x_3^0, on obtienne ainsi l'unique solution du système $g_1=g_2=0$ dont les inconnues sont (x_1,x_2) et le paramètre est x_3. Si on veut le démontrer à partir du théorème des fonctions implicites scalaire, on commence par vérifier qu'un des éléments de la matrice au moins est non nul, ainsi on peut résoudre localement une des équations sous la forme $x_1=\psi(x_2,x_3)$par exemple. Il suffit de remplacer cette égalité dans l'autre équation pour résoudre à nouveau une équation de la forme $G(x_2,x_3)=0$. La condition $\partial_2 G(x_2^0,x_3^0)\neq 0$ provient du fait que le déterminant de la matrice est non nul (calcul laissé au lecteur).

Minimum de la fonction d'une variable obtenue en restreignant à l'espace des contraintes Ce résultat étant admis ou démontré, le minimum local de f sous les contraintes $g_1=g_2=0$ s'obtient en cherchant le minimum de $f(\psi_1(x_3),\psi_2(x_3),x_3)=0$. Ainsi on a l'équation (et les identités)

$$\left\{\begin{array}{l} \partial_1 f(x_0)\psi_1'(x_3^0)+\partial_2 f(x_0)\psi_2'(x_3^0)+\partial_3 f(x_0)=0 \\ \partial_1 g_1(x_0)\psi_1'(x_3^0)+\partial_2 g_1(x_0)\psi_2'(x_3^0)+\partial_3 g_1(x_0)=0 \\ \partial_1 g_2(x_0)\psi_1'(x_3^0)+\partial_2 g_2(x_0)\psi_2'(x_3^0)+\partial_3 g_2(x_0)=0 \end{array}\right.$$

On peut alors remplacer $\psi_1'(x_3^0),\psi_2'(x_3^0)$ obtenu par les deux dernières égalités dans la première égalité. On trouve (inversant la matrice 2×2)

$$\psi_1'(x_3^0)=\frac{1}{\partial_1 g_1\partial_2 g_2-\partial_1 g_2\partial_2 g_1}(-\partial_2 g_2\partial_3 g_1+\partial_2 g_1\partial_3 g_2)$$

et l'égalité similaire sur $\psi_2'(x_3^0)$. On remplace ces deux égalités dans la première égalité et on identifie explicitement λ et μ tels que

$$\lambda\partial_3 g_1(x_0)+\mu\partial_3 g_2(x_0)+\partial_3 f(x_0)=0.$$

On trouve en particulier les égalités (calculées à x_0):

$$\lambda=\frac{\partial_2 f\partial_1 g_2-\partial_1 f\partial_2 g_2}{\partial_1 g_1\partial_2 g_2-\partial_1 g_2\partial_2 g_1},\mu=\frac{\partial_1 f\partial_2 g_1-\partial_2 f\partial_1 g_1}{\partial_1 g_1\partial_2 g_2-\partial_1 g_2\partial_2 g_1}.$$

Notre résultat est donc démontré.
Nous nous intéressons dans la partie suivante à la résolution de problème sous contraintes inégalité

2.3 Contraintes inégalité

2.3.1 Cas quadratique

On commence par étudier le problème de minimisation de $\frac{1}{2}(Ax, x) - (b, x)$ où A est une matrice symétrique définie positive sous la contrainte $V.x < 0$ (nous avons considéré dans la partie précédente le cas $k \neq 0$ et nous avons démontré qu'on pouvait se ramener à l'hyperplan $V.X = 0$). Nous supposons $V \neq 0$ (dans le cas contraire, il n'y a pas de contrainte) et nous supposons $V_3 \neq 0$ (pour fixer les idées). Ainsi l'ensemble des contraintes s'écrit $V_3x_3 \leq -V_1x_1 - V_2x_2$.

Proposition 2 *Soit x_0 le minimum de $f(x) = \frac{1}{2}(Ax, x) - (b, x)$ sous la contrainte $V.x \leq 0$. Ce point x_0 existe et est unique, et il existe un réel $\lambda \geq 0$ tel que x est solution de $Ax - b + \lambda V = 0$, $V.x = 0$ dans le cas où $A^{-1}b$ vérifie $V.(A^{-1}b) > 0$, et est égal à $A^{-1}b$ dans le cas où $V.(A^{-1}b) \leq 0$. Dans le premier cas, on dit que la contrainte est active en x_0, dans le deuxième cas on dit que la contrainte est inactive en X_0.*

Démonstration: La fonction f admet un unique minimum sur $\mathbb{R}^3$, qui est $A^{-1}b$. Si ce point de minimum absolu appartient à l'ensemble des contraintes alors le minimum sous contrainte inégalité est égal au minimum absolu.
Si ce point de minimum n'appartient pas à l'espace des contraintes, alors on écrit le système de coordonnées généré par la base orthonormée e, f, g où $||V||e = V$ et où (f, g) est une base orthonormée du plan vectoriel orthogonal à V. Alors on a $x = y_1e + y_2f + y_3g$. La condition $V.x \leq 0$ est équivalente à $y_1 \leq 0$, et la fonction à minimiser s'écrit

$$\frac{1}{2}(k_{ij}y_iy_j) - (b, e)y_1 - (b, f)y_2 - (b, g)y_3,$$

où les coefficients k_{ij} sont de la forme (Ae, f) (par exemple).

On cherche donc le minimum de f sous la condition $y_1 \leq 0$, y_2, y_3 quelconques. Pour tout y_1, on calcule le minimum de la fonction en (y_2, y_3). On trouve des relations affines (que nous n'expliciterons pas) sur y_2, y_3 en fonction de y_1, ce qui permet d'aboutir au calcul du minimum d'une fonction de la forme $\frac{1}{2}ay_1^2 - cy_1$ pour $y_1 \leq 0$.
Il vient alors que si $\frac{c}{a} \leq 0$, le point obtenu est le minimum absolu de la fonction, ce qui a été exclu de notre calcul. Donc $\frac{c}{a} > 0$, et donc le minimum de la fonction sur $y_1 \leq 0$ est atteint en $y_1 = 0$.

On s'est donc ramené à trouver le minimum de f sur le plan vectoriel $y_1 = 0$. Comme on cherche le minimum de f sous la condition $V.x = 0$, il existe λ réel

tel que $f'(x) + \lambda V = 0$, ce qui donne l'égalité $Ax - b + \lambda V = 0$.
Si on avait $\lambda < 0$, alors on aurait $x = A^{-1}(b - \lambda V)$ et $0 \geq V.x = V.A^{-1}b - \lambda V.A^{-1}V \geq V.A^{-1}b$ donc le point $A^{-1}b$ appartiendrait à l'ensemble des contraintes, contradiction.

2.3.2 Minimisation de $f(x_1, x_2, x_3)$ sous une contrainte inégalité $g(x_1, x_2, x_3) \leq 0$

Proposition 3 *Soit x_0 un point de minimum local de f sous contrainte $g \leq 0$.*
i) Si $g(x_0) < 0$ (on dit que la contrainte est inactive en x_0), alors $\nabla f(x_0) = 0$,
ii) Si $g(x_0) = 0$ (on dit que la contrainte est active ou saturée en x_0) et si $\nabla g(x_0) \neq 0$, alors il existe $\lambda \geq 0$ tel que $\nabla f(x_0) + \lambda \nabla g(x_0) = 0$.

Preuve dans le premier cas (contrainte inactive) On considère x_0 un point de minimum local de f sous cette contrainte.
Si $g(x_0) < 0$, alors il existe un voisinage V_1 de x_0 pour lequel $g(x) < 0$. On sait que x_0 est un point de minimum local de f sous contraintes $g \leq 0$ donc il existe un voisinage W de x_0 tel que, sur ce voisinage $g(x) \leq 0$ et $f(x) \geq f(x_0)$. Sur $V_1 \cap W$, on a $f(x) \geq f(x_0)$ pour tout x, puisque la contrainte est satisfaite sur V_1. Ainsi $\nabla f(x_0) = 0$ et le premier item est démontré.

Preuve dans le deuxième cas (contrainte active) Si $g(x_0) = 0$, il faut pouvoir écrire localement la contrainte $g(x_1, x_2, x_3) \leq 0$. On suppose que $\nabla g(x_0) \neq 0$. Alors il existe une composante de ∇g non nulle. Pour simplifier l'écriture, nous supposons qu'il s'agit de $\partial_3 g(x_0)$. Le théorème des fonctions implicites permet de trouver une fonction $\psi(x_1, x_2)$ telle que, localement

$$g(x_1, x_2, \psi(x_1, x_2)) = 0, \psi(x_1^0, x_2^0) = x_3^0, \partial_3 g(x_1, x_2, \psi(x_1, x_2)) \neq 0.$$

La formule de Taylor avec reste intégral conduit à

$$g(x_1, x_2, x_3) = (x_3 - \psi(x_1, x_2)) \int_0^1 \partial_3 g(x_1, x_2, \psi(x_1, x_2) + t(x_3 - \psi(x_1, x_2)))dt.$$

On suppose dans un premier temps $\partial_3 g(x_1^0, x_2^0, x_3^0) > 0$. Alors, localement, $\partial_3 g(x_1, x_2, \psi(x_1, x_2) + t(x_3 - \psi(x_1, x_2))) > 0$, et donc $g(x_1, x_2, x_3) \leq 0 \Leftrightarrow x_3 \leq \psi(x_1, x_2)$.
On suppose donc que $f(x_1, x_2, \psi(x_1, x_2) + X_3)$ admet un minimum local en X_3 sous la contrainte $X_3 \leq 0$. Ceci indique que la dérivée partielle par rapport à X_3 de la fonction $X_3 \to f(x_1, x_2, \psi(x_1, x_2) + X_3)$ est négative au point $(x_1^0, x_2^0, X_3 = 0)$, et que, d'autre part, $(x_1, x_2) \to f(x_1, x_2, \psi(x_1, x_2))$ a sa dérivée nulle au point (x_1^0, x_2^0).
On a donc, au point x^0,

$$\partial_1 f + \partial_1 \psi \partial_3 f = 0, \partial_2 f + \partial_2 \psi \partial_3 f = 0$$

ainsi que

$$\partial_3 f \leq 0.$$

L'identité $g(x_1, x_2, \psi(x_1, x_2)) = 0$ implique les égalités $\partial_1 g + \partial_1\psi\partial_3 g = 0, \partial_2 g + \partial_2\psi\partial_3 g = 0$.
On introduit alors

$$\lambda = -\frac{\partial_3 f(x^0)}{\partial_3 g(x^0)}.$$

C'est le multiplicateur de Lagrange.
Comme dans le cas de la contrainte égalité, on trouve

$$\partial_1 f + \lambda\partial_1 g = 0, \partial_2 f + \lambda\partial_2 g = 0, \partial_3 f + \lambda\partial_3 g = 0$$

(la troisième égalité étant la définition de λ).
Comme $\partial_3 f \leq 0$, on en déduit

$$\lambda \geq 0.$$

Le multiplicateur de Lagrange est donc positif.
Le raisonnement est identique lorsque $\partial_3 g(x^0) < 0$. La proposition 3 est démontrée.

2.3.3 Minimisation de f sous les deux contraintes $g_1(x) \leq 0, g_2(x) \leq 0$

L'énoncé du théorème est identique au cas d'une contrainte, mais la démonstration est nécessaire pour comprendre les différences.

Proposition 4 *Soit x^0 un point de minimum local de f sous contrainte $g_1 \leq 0, g_2 \leq 0$. Suivant des hypothèses suffisantes mais non nécessaires sur g_1, g_2 au point x^0, il existe deux réels positifs λ, μ tels que*

$$\nabla f(x^0) + \lambda\nabla g_1(x^0) + \mu\nabla g_2(x^0) = 0$$

et tels que $\lambda g_1(x^0) = 0$, $\lambda g_2(x^0) = 0$.
Lorsqu'une contrainte est inactive ($g_1(x^0) < 0$) alors $\lambda = 0$ et tout se passe comme si on ne considérait pas cette contrainte.

Preuve Trois cas se présentent:
i) soit $g_1(x^0) < 0$ et $g_2(x^0) < 0$. Dans ce cas on dit que les deux contraintes sont inactives et l'équation à résoudre est $\nabla f(x^0) = 0$, puisque, localement, dans un voisinage de x^0, les contraintes n'interviennent pas et on se ramène à un minimum libre (et dans ce cas on peut choisir arbitrairement les multiplicateurs de Lagrange λ et μ).
ii) soit $g_1(x^0) = 0$ et $g_2(x^0) < 0$; Dans ce cas, seule la contrainte $g_1 = 0$ est active. Localement, au voisinage de x^0 on a, par continuité de g_2, $g_2(x) < 0$, donc il existe un voisinage V de x^0 où la contrainte g_2 est tout le temps vérifiée. Ainsi on se ramène au problème précédent (avec une seule contrainte active). Il

existe donc $\lambda \geq 0$ tel que $\nabla f(x^0)+\lambda\nabla g_1(x^0) = 0$. Ajoutant $\mu = 0$, donc le terme $0\nabla g_2(x^0) = 0$ dans cette équation, on trouve que l'égalité de la proposition 4 est vérifiée.
iii) soit $g_1(x^0) = g_2(x^0) = 0$. Le cas simple est celui où les vecteurs $\nabla g_1(x^0)$ et $\nabla g_2(x^0)$ sont linéairement indépendants. Alors il existe deux fonctions ψ_1 et ψ_2 telles que, localement, $g_1(\psi_1(x_3), \psi_2(x_3), x_3) = g_2(\psi_1(x_3), \psi_2(x_3), x_3) = 0$.
On introduit alors le changement de variable $(X_1, X_2, x_3) = (x_1 - \psi(x_3), x_2 - \psi(x_3), x_3)$. La formule de Taylor avec reste intégral conduit à

$$\begin{array}{ll} g_1(x_1, x_2, x_3) & = (x_1 - \psi_1(x_3)) \int_0^1 \partial_1 g_1(\psi_1(x_3) + t(x_1 - \psi_1(x_3)), x_2, x_3)dt \\ & +(x_2 - \psi_2(x_3)) \int_0^1 \partial_2 g_1(x_1, \psi_2(x_3) + t(x_1 - \psi_2(x_3)), x_2, x_3)dt \\ g_2(x_1, x_2, x_3) & = (x_1 - \psi_1(x_3)) \int_0^1 \partial_1 g_2(\psi_1(x_3) + t(x_1 - \psi_1(x_3)), x_2, x_3)dt \\ & +(x_2 - \psi_2(x_3)) \int_0^1 \partial_2 g_2(x_1, \psi_2(x_3) + t(x_1 - \psi_2(x_3)), x_2, x_3)dt. \end{array}$$

Les deux inégalités $g_1 \leq 0$ et $g_2 \leq 0$ peuvent se résumer en $A(x_1 - \psi_1(x_3)) + B(x_2 - \psi_2(x_3)) \leq 0, C(x_1 - \psi_1(x_3)) + D(x_2 - \psi_2(x_3)) \leq 0$, où A, B, C, D dépendent de (x_1, x_2, x_3). Lorsque $x = x^0$, les valeurs de A, B, C, D sont

$$\partial_1 g_1(x^0), \partial_2 g_1(x^0), \partial_1 g_2(x^0), \partial_2 g_2(x^0).$$

Le problème se ramène alors à minimiser $f(X_1 + \psi(x_3), X_2 + \psi(x_3), x_3)$ sous deux contraintes qui s'écrivent $AX_1 + BX_2 \leq 0, CX_1 + DX_2 \leq 0$. Nous notons alors $Y_1 = AX_1 + BX_2$, $Y_2 = CX_1 + DX_2$. On a alors $X_1 = \frac{DY_1-BY_2}{AD-BC}, X_2 = \frac{-CY_1+AY_2}{AD-BC}$, et le problème revient donc à minimiser $F(Y_1, Y_2, x_3) = f(\frac{DY_1-BY_2}{AD-BC} + \psi_1(x_3), \frac{-CY_1+AY_2}{AD-BC} + \psi_2(x_3), x_3)$ sous les contraintes $Y_1 \leq 0$ et $Y_2 \leq 0$. On en déduit donc que la dérivée de F par rapport à x_3 est nulle, et que les dérivées de F par rapport à Y_1, Y_2 sont négatives au point x^0.
La condition d'optimalité s'écrit

$$\partial_3 f + \psi_1' \partial_1 f + \psi_2' \partial_2 f = 0.$$

Les réels ψ_1', ψ_2' sont solution du système

$$\partial_1 g_1 \psi_1' + \partial_2 g_1 \psi_2' = -\partial_3 g_1, \partial_1 g_2 \psi_1' + \partial_2 g_2 \psi_2' = -\partial_3 g_2.$$

et on retrouve les mêmes multiplicateurs de Lagrange que dans le cas de deux contraintes égalité. Par exemple

$$\lambda = \frac{\partial_2 f \partial_1 g_2 - \partial_1 f \partial_2 g_2}{\partial_1 g_1 \partial_2 g_2 - \partial_1 g_2 \partial_2 g_1}.$$

D'autre part, la dérivée de F par rapport à Y_1, qui est négative, vaut

$$\frac{D\partial_1 f - C\partial_2 f}{AD - BC}$$

Remplacant par les valeurs de A, B, C, D calculées au point x^0, on obtient l'inégalité

$$\frac{\partial_2 g_2 \partial_1 f - \partial_2 f \partial_1 g_2}{\partial_1 g_1 \partial_2 g_2 - \partial_1 g_2 \partial_2 g_1} \leq 0$$

ce qui revient à $\lambda \geq 0$. Le théorème est démontré dans le cas où ∇g_1 et ∇g_2 sont linéairement indépendants au point x^0.

2.4 Conclusions

Dans cette introduction, on a pu donner des conditions nécessaires, dans des cas simples (dimension 3) pour obtenir les points de minimum local d'une fonction, ou sans contraintes, ou sous contraintes inégalité ou égalité.
Ceci conduit à l'écriture de l'équation d'Euler dans le cas de minimisation sans contraintes, et l'introduction de multiplicateurs de Lagrange dans le cas de minimisation sous contrainte.
La démonstration de ces résultats, que tout ingénieur doit savoir utiliser directement, s'apppuie sur le théorème des fonctions implicites, qui permet de résoudre localement les équations que donnent les contraintes égalité, et s'appliquent aussi aux contraintes inégalités, qui ne peuvent avoir un rôle que lorsqu'elles sont saturées, c'est à dire quand l'inégalité est en réalité une égalité. On a donc dans ce cas la compréhension de la représentation d'un hyperplan, ainsi que la représentation d'un demi-espace.