Optimisation Non Linéaire Sous Contraintes

A. Bazeniar

Enseignant chercheur
Centre universitaire Abdelhafid Boussouf
Mila, Algérie

Se reporter à des manuels de base et à certaines recherches

Septembre 2023

Chapitre 1

Conditions d'optimalité

1.1 Introduction

L'approche considérée ici pour l'obtention de ces conditions est basée sur les notions de descente et de direction admissible. l'étude de ces conditions a permet de développer les algorithmes de résolution et de vérifier la validités des résultats obtenues.

Dans ce chapitre, les fonctions sont toujours supposées différentiables à tout ordre. Si f est définie sur $S \subset \mathbb{R}^n$ et à valeurs réelles, sa différentielle en $x \in S$ est l'application linéaire notée

$$f(x+h) = f(x) + < \nabla f(x), h > + \|h\| \varepsilon(h), \quad \varepsilon(h) \to 0 \text{ lorsque } h \to 0. \tag{1.1}$$

Rappelons la formule de Taylor à l'ordre 2

$$f(x+h) = f(x) + < \nabla f(x), h > + \frac{1}{2} < \nabla^2 f(x), h > h + \|h^2\| \varepsilon(h), \quad \varepsilon(h) \to 0 \text{ lorsque } h \to 0. \tag{1.2}$$

1.2 Direction admissible

Définition 1.2.1. *(Direction admissible). Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ et $x \in S$. d est une direction admissible si*

$$\exists \lambda' > 0 \ \ \textit{tel que} \ \ x + \lambda d \in S, \ \ \textit{pour} \ \ 0 \leq \lambda \leq \lambda'.$$

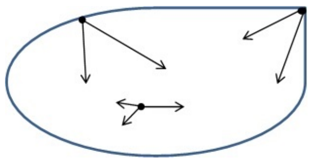

FIGURE 1.1 – Directions admissible

Remarque 1.2.1. 1. *Cette définition nous montre que tous les vecteurs d de $\mathbb{R}^n$ sont des directions admissibles, si le point $x + \lambda d$ est un point intérieur de S.*
2. *La nature de l'ensemble S intervient dans le formalisme d'équations ou d'inéquations de la condition d'optimalité, à savoir les points sur la frontière de S.*

1.2.1 Direction admissible à l'optimum

Théorème 1.2.1. *(condition nécessaire de Peano-Kantorovitch). Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ de classe C^1. Si x^* est un minimum local (globale) de f sur S, alors $\forall$ d direction admissible en x^*, on a*

$$\nabla^T f(x^*) d \geq 0. \tag{1.3}$$

Démonstration. Soit d une direction admissible en x^*. Considérons la fonction $g(\lambda) = f(x^* + \lambda d)$.
$\forall x \in S$ on a $f(x^*) \leq f(x)$. D'où $\lambda^* = 0$ est un minimum local de $g(\lambda)$ sur $[0, \lambda']$. On a

$$g(\lambda) = g(0) + \lambda g'(0) + o(\lambda),$$

et comme x^* est un minimum local, alors $g(\lambda) - g(0) \geq 0$.
Par conséquent $g'(\lambda) = [\nabla f(x^* + \lambda d]^T d \geq 0$,
on obtient $g'(0) = \nabla^T f(x^*) d \geq 0$. □

1.3 Existence et unicité de la solution

Définition 1.3.1. 1. *Une partie U de S est un ouvert de S si pour tout $x \in U$ il existe $\epsilon > 0$ tel que $B(x, \epsilon) \subset U$.*
2. *Une partie F de S est un fermé de S si et seulement si son complémentaire F^c dans S est ouvert.*

Remarque 1.3.1. *Dans* $\mathbb{R}$*, les intervalles ouverts sont des ouverts et les intervalles fermés sont des fermés. Plus généralement, toute boule ouverte est une partie ouverte et toute boule fermée est une partie fermée.*

Théorème 1.3.1. *(Existence : Théorème de Weierstrass).*

1. *Si* S *est un sous-ensemble fermé et borné de* $\mathbb{R}^n$ *et* f *continue, alors* f *admet au moins un point extremum.*
2. *Si* S *est un sous-ensemble fermé de* $\mathbb{R}^n$ *et si* f *est croissante à l'infini (*f *est coercive si* $f(x) \to +\infty$ *lorsque* $\| x \| \to +\infty$ *), alors* f *admet un point minimum globale sur* S*.*

Démonstration.

1. S est fermé et borné de $\mathbb{R}^n$ alors compact, et comme f est continue alors elle atteint ses bornes sur S. Donc x^* existe.
2. Soit l'ensemble $S_0 = \{x, x_0 \in S / f(x) \leq f(x_0)\}$, on voit que S_0 est un compact (fermé car f est continue, borné car f est coercive). Donc x^* existe.

□

Théorème 1.3.2. *(Unicité). Soit le problème de minimisation (PNL) avec* S *convexe et* $S \subset \mathbb{R}^n \to \mathbb{R}$ *convexe. Alors,*

1. *Tout minimum local de (PNL) est un minimum global de problème.*
2. *Si* f *est strictement convexe, il y a au plus un minimum de problème.*

Exemple 1.3.1. *Soit le problème (PNL) :*

$$\begin{cases} Min \ f(x) = x^2 \\ g: \quad x^4 - 1 \leq 0 \\ x \in \mathbb{R}^n, \end{cases} .$$

L'ensemble g *est fermé (car* $g^{(-1)}([-1,1])$ *est fermé) et* f *croissante à l'infinie. Comme* f *est continue, elle atteint ses bornes et elle admet un minimum global* x^**.*

1.4 Conditions necessaries d'optimalité

L'idée de cette section est basée sur la recherche d'une expression plus pratique de la condition (1.3).

1.4.1 Contraintes de type égalité

Soit l'ensemble S définie par,

$$S = \{x \in \mathbb{R}^n : h_1(x) = 0, \ldots, h_p(x) = 0\}.$$

- On dit que h réalise p contraintes et x est dit variable de décision.
- Cette ensemble des contraintes d'qualités est un sous-ensemble de $\mathbb{R}^n$ et qui représente une hypersurface.

1.4.1.1 Plan tangent

Définition 1.4.1. *On définit le plan tangent au point x^*, tout vecteur y qui est orthogonal avec les gradients de h_i, on écrit*

$$T = \{y, \nabla h(x^*)y = 0\}. \tag{1.4}$$

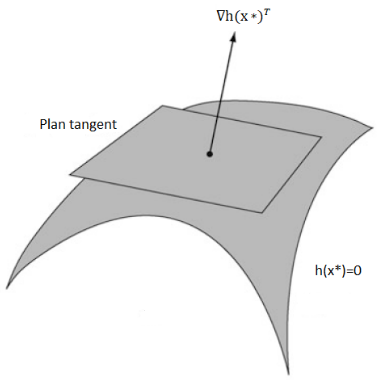

FIGURE 1.2 – Plan tangent.

Exemple 1.4.1. *Soit la contrainte $h(x_1, x_2) = x_1$. Le plan tangent est l'axe des x_2, car*

$$(1,0)^T(y_1, y_2) = 0 \Rightarrow y_1 = 0.$$

Définition 1.4.2. *On dit que x^* est un point régulier pour la contrainte $h(x) = 0$ si,*
- — $h(x^*) = 0$,
- — *Les vecteurs $\nabla h_i(x^*)$ sont linéairement indépendants.*

Théorème 1.4.1. *(CN1) Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ de classe C^1. On suppose que f admet en $x^* \in S$ un minimum local et que,*

$$\text{La famille} \quad (\nabla h_1(x^*), \dots, \nabla h_p(x^*)) \quad \text{est libre} \qquad (*)$$

Alors il existe des multiplicateurs de Lagrange $\lambda_1, \lambda_2, \dots, \lambda_p \in \mathbb{R}$ tel que,

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) = 0. \qquad (**)$$

$(**)$ Est appelée équation de Lagrange.
$(*)$ Condition de qualification des contraintes.

Remarque 1.4.1. — *La recherche des points stationnaires du problème se ramène à la résolution du système (si la solution existe) de l'équation* $(**)$.

— *Les gradients* ∇f *et* ∇h *sont orthogonaux aux courbes de niveau de* f *et* h *respectivement.*

— *Si* x^* *est extremum de* f *sous la contrainte* h *et les gradients* $\nabla f(x^*)$ *et* $\nabla h(x^*)$ *non nuls, les courbes de niveau de* f *et* h *sont tangentes en* x^*.

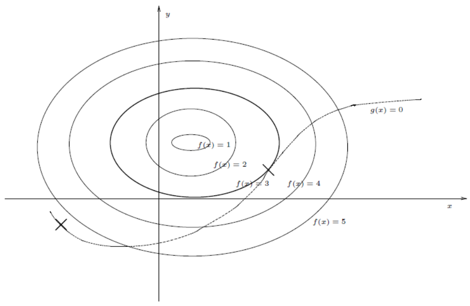

FIGURE 1.3 – Interprétation géométrique des multiplicateurs de Lagrange.

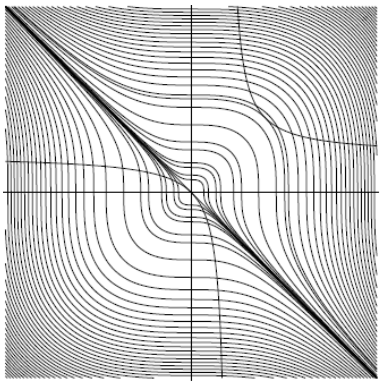

FIGURE 1.4 – Les courbes de niveau de f et h sont tangentes.

Démonstration. On se contente de l'ensemble $S = \{x \in \mathbb{R}^2 : h(x) = 0\}$. La généralisation sur $\mathbb{R}^n$ reste la même.
La condition $(*)$ indique que $\nabla h(x^*) = 0$. On suppose que $\partial h / \partial x_2 \neq 0$.
Ceci nous permet d'appliquer le théorème des fonctions implicite,
$h(x_1, x_2) = 0 \Rightarrow x_2 = \alpha(x_1)$, α une fonction continue et dérivable.

Maintenant, appliquant les équations d'Euler sur f et h, pour $i = \{1, \ldots, n-1\}$, on aura

$$\partial f / \partial x_1(x_1^*, \alpha(x_1^*)) + \partial f / \partial y(x_1^*, \alpha(x_i 1^*))\alpha'(x_1^*) = 0 \tag{1.5}$$

$$\partial h / \partial x_1(x_1^*, \alpha(x_1^*)) + \partial h / \partial x_2(x_1^*, \alpha(x_i^*))\alpha'(x_1^*) = 0 \tag{1.6}$$

De (1.5) et (1.6) on trouve,

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

Tel que,

$$\lambda = [\partial h/\partial x_2(x^*)]^{-1} \times \partial h/\partial x_2(x^*).$$

□

Exemple 1.4.2. $f(x,y,z) = (x-2)^2 + y^2 + z^2$ *sous la contrainte* $g : x^2 + 2y^2 + 3z^2 = 1$.
Le gradient est, $\nabla g(x,y,z) = (2x, 4y, 6z)^T$,
$\nabla g(x,y,z) \neq 0$ *car la contrainte g n'est pas active au point* $(0,0,0)$. *La contrainte est qualifiée.*
Les équations de Lagrange sont,

$$\begin{cases} x^2 + 2y^2 + 3z^2 = 1 \\ x - 2 = \lambda x \\ y = 2\lambda y \\ z = 3\lambda z \end{cases}$$

1.4.2 Interprétation de la valeur du multiplicateur λ (analyse de sensibilité)

Soit le problème (p), (f dif.)

$$\begin{cases} Min \;\; f(x) \\ s.c \;\; h(x) = b \\ x \in \mathbb{R}^n \end{cases}$$

Considérons la fonction $V(b)$, qui donne la valeur optimale du problème (p) pour une valeur b donnée . Si on varie la valeur de b, alors la valeur de la solution optimale varie. A partir de ça, on peut définir la fonction $x(b)$ comme,

$$V(b) = f(x(b)) \;\text{ et }\; h(x(b)) = b.$$

Maintenant, Mesurant la variation de $V(b)$,

$$\begin{aligned} V'(b) &= \frac{dV}{db}, \\ &= \frac{\partial f}{\partial x_1}\frac{dx_1}{db} + \cdots + \frac{\nabla f}{\nabla x_n}\frac{dx_n}{db}, \\ &= \lambda\left(\frac{\partial h}{\partial x_1}\frac{dx_1}{db} + \cdots + \frac{\nabla h}{\nabla x_n}\frac{dx_n}{db}\right), \\ &= \lambda\frac{db}{db}, \\ &= \lambda. \end{aligned}$$

Conséquence : le multiplicateur λ est sensible à la variation de la valeur de b et il donne l'information sur le taux de variation de la valeur de la fonction objective lorsque b augmente.

Calcul directe : On peut approcher la fonction V par son polynôme autour deb',

$$V(b) = V(b') + V'(b')(b - b') = V(b') + \lambda(b - b').$$

1.4.3 Contraintes de type inégalité

Soit l'ensemble des contraintes S définit par,

$$S = \{x \in \mathbb{R}^n : g_1(x) \leq 0, \dots, g_m(x) \leq 0\}.$$

Définition 1.4.3. *On dit qu'une contrainte g est active en x^* si,*

- $g(x^*) = 0$,
- *L'ensemble des indices des contraintes actives est,*

$$I(x^*) = \{i / g_i(x^*) = 0\}.$$

- *L'ensemble des vecteurs $E(x)$ qui active g est,*

$$E(x) = \{x \in \mathbb{R}^n, g_j(x) = 0, j \in I\}.$$

Théorème 1.4.2. *(CN1) Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ de classe C^1. On suppose que f admet en $x^* \in S$ un minimum local et qu'il existe une vecteur $v \in \mathbb{R}^n$ tel que,*

$$\forall j g_j(x^*) = 0 \Rightarrow < \nabla g_j(x^*), v > \leq 0 \qquad (*)$$

Alors il existe des multiplicateurs de Kuhn et Tucker $\mu_1, \mu_2, \dots, \mu_m \in \mathbb{R}$ tel que,

$$\nabla f(x^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(x^*) = 0. \qquad (**)$$

Avec

$$\begin{cases} \mu_j \geq 0, & \text{(Positivité)} \\ \mu_j g_j(x^*) = 0, & \text{(Relation d'exclusion).} \end{cases}$$

$(**)$ *Est appelée équation de Lagrange.*
$(*)$ *Condition de qualification des contraintes.*

Démonstration. 1. Les deux relations d'exclusion et $(**)$ sont une conséquence directe du théorème 1.4.1 (car il suffit de prendre $\mu_j = 0$ pour $j \notin I$).

2. Maintenant, on montre la positivité par absurde :
Supposons qu'il existe $\mu_k < 0, k \in I$.

D'autre part, on définit $y \in \mathbb{R}^n$ tel que :

$$\nabla g_j(x^*)y = 0, \quad j \in I,$$

$$\nabla g_k(x^*)y = -1, \ j \neq k.$$

Alors, y est une direction admissible,

$$\begin{aligned} \nabla f(x^*)^T y &= -\sum_{j=1}^{m} \mu_j \nabla g_j(x^*)^T y, \\ &= -\mu_k \nabla g_k(x^*)^T y, \\ &= \mu_k < 0. \end{aligned}$$

Ceci est impossible car x^* est un point minimum de f.

$\square$

Exemple 1.4.3. *Soit la fonction $f(x, y) = -xy$ à minimiser sur le domaine S*

$$x \geq 0, y \geq 0 \text{ et } x + y \leq 6.$$

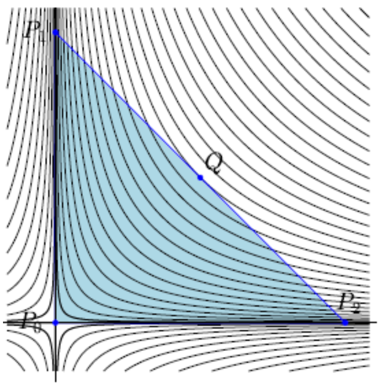

FIGURE 1.5 – Les lignes de niveau de $-xy$ sur le domaine S

1.4.4 Le théorème de Kuhn-Tucker avec lagrangien généralisé

Dans le cadre général du théorème de Kuhn et Tucker, la contrainte S est de la forme,

$$S = \{x \in \mathbb{R}^n, g_j(x) \leq 0, h_i(x) = 0\}.$$

On suppose que f admet en $x^* \in X$ un minimum local et qu'il existe une vecteur $v \in \mathbb{R}^n$ tel que,

$$< \nabla h_i(x^*), v >= 0,$$

$$\forall j \in I(x^*) < \nabla g_j(x^*), v > \leq 0.$$

Où $I(x^*)$ l'ensemble des indices des contraintes actives au point x^*.
Alors il existe des multiplicateurs $\lambda_1, \lambda_2, \ldots, \lambda_p \in \mathbb{R}$ et $\mu_1, \mu_2, \ldots, \mu_m \in \mathbb{R}$ tel que,

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(x^*) = 0.$$

Sous les conditions des deux théorèmes 2.3.1 et 2.3.2.

1.5 Conditions suffisantes d'optimalité

Rappelant les deux cas suivants :
Le Lagrangien associé au (PCE) est :

$$L(x, \lambda) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x).$$

Le Lagrangien associé au (PCI) est :

$$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x).$$

Théorème 1.5.1. *Soit $x^* \in \mathbb{R}^n$ un point régulier. Alors il existe λ^* tel que,*

- $\nabla L(x^*, \lambda^*) = 0$
- $d^T \nabla^2_{xx} L(x^*, \lambda^*) d \geq 0$, $d \neq 0$ *direction admissible*

Démonstration. Puisque x^* est minimum, on a $\nabla^2_{xx} f(x^*) \geq 0$ Et donc par les équations d'Euler,

$$d^T \nabla^2_{xx} f(x^*) d + \nabla f(x^*)^T d' \geq 0. \tag{1.7}$$

D'autre part,

$$d^T \nabla^2_{xx} h(x^*) d + \nabla h(x^*)^T d' = 0. \tag{1.8}$$

En multipliant par λ_i la relation 1.8 puis on fait la somme avec la relation 1.7, on aura

$$d^T (\nabla^2_{xx}(f(x^*) + \sum_{i=1}^{p} \nabla^2 h_i(x^*) \lambda_i) d + (\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*)^T) d' \geq 0.$$

Et comme $\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*))^T = 0$. L'inégalité est vérifiée.

□

Théorème 1.5.2. *(CS) pour (PCE)*
Soit $x^* \in \mathbb{R}^n$ *et* $\lambda^* \in \mathbb{R}^p$ *vérifiant,*

- $\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) = 0,$
- $h_i(x^*) = 0,$
- $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, \quad d \neq 0.$

Alors x^* *est un point minimum du problème.*

Théorème 1.5.3. *(CS) pour (PCI)*
Soit $x^* \in \mathbb{R}^n$ *et* $\mu^* \in \mathbb{R}^m$ *vérifiant,*

- $\nabla f(x^*) + \sum_{j=1}^{m} \mu_j^* \nabla g_i(x^*) = 0,$
- $g_j(x^*) \leq 0,$
- $\mu_j \geq 0,$
- $\mu_j g_j(x^*) = 0,$
- $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, \quad d \neq 0.$

Alors x^* *est un point minimum du problème.*