COURS D'OPTIMISATION

Marc Dambrine & Grégory VIAL

4 octobre 2007

Table des matières

Chapitre 1

Rappels et compléments de calcul différentiel

1.1 Différentiabilité

1.1.1 Fonctions numériques d'une variable réelle

Soit I un intervalle ouvert de $\mathbb{R}$ et $f : I \to \mathbb{R}$ une application.

Définition 1.1 *f est dérivable sur I ssi pour tout $x \in I$, $\displaystyle\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ existe. On note alors cette limite $f'(x)$.*

Définition 1.2 *f est de classe $\mathscr{C}^1$ sur I – on note $f \in \mathscr{C}^1(I,\mathbb{R})$ – ssi f est dérivable sur I et l'application $x \mapsto f'(x)$ est continue sur I.*

Exemple 1.1 *Soit $f : \mathbb{R} \to \mathbb{R}$ définie par $f(x) = e^{2x} - x^2$. La fonction f est de classe $\mathscr{C}^1$ sur $\mathbb{R}$ car pour tout réel x, $f'(x) = 2e^{2x} - 2x$ qui définit une fonction continue sur $\mathbb{R}$.*

Exemple 1.2 *Soit $f : \mathbb{R}^+ \to \mathbb{R}$ définie par $f(x) = x^2 \sin \frac{1}{x}$ pour $x > 0$ et $f(0) = 0$. Alors f est dérivable sur $\mathbb{R}^+$:*

$$\forall x > 0,\ f'(x) = 2x \sin\frac{1}{x} - \sin\frac{1}{x^2} \quad \text{et } f'(0) = 0$$

Mais f n'est pas de classe $\mathscr{C}^1$ sur $\mathbb{R}^+$ car f' n'est pas continue en 0.

1.1.2 Fonctions vectorielles d'une variable réelle

Soit I un intervalle de $\mathbb{R}$ et $f : I \to \mathbb{R}^p$ une application. Pour tout $x \in I$, on note $f(x) = (f_1(x), \ldots, f_p(x))^\mathsf{T}$.

Définition 1.3 *f est de classe $\mathscr{C}^1$ sur I ssi pour tout $i \in \{1, \ldots, p\}$, f_i est de classe $\mathscr{C}^1$ sur I. On note alors, pour $x \in \mathbb{R}^n$,*

$$f'(x) = (f_1'(x), \ldots, f_p'(x))^\mathsf{T}$$

Exemple 1.3 *Soit $f : \mathbb{R}_+^* \to \mathbb{R}^2$ définie par*

$$\forall x \in \mathbb{R}_+^*, \ f(x) = (\sqrt{x} - x, e^x - \ln x)^\mathsf{T}$$

$f \in \mathscr{C}^1(\mathbb{R}_+^*)$ car

$$\forall x \in \mathbb{R}_+^*, \ f'(x) = \Big(\frac{1}{2\sqrt{x}} - 1, e^x - \frac{1}{x}\Big)^\mathsf{T}$$

qui est continue sur $\mathbb{R}_+^*$.

1.1.3 Fonctions vectorielles de plusieurs variables

Soit Ω un ouvert de $\mathbb{R}^n$ (pour simplifier, on peut considrer que Ω est un produit d'intervalles ouverts $I_1 \times \cdots \times I_n$) et $f : \Omega \to \mathbb{R}^p$ une application. On notera $x = (x_1, \ldots, x_n)^\mathsf{T}$ pour $x \in \Omega$.

Définition 1.4 *La fonction f est différentiable en $x \in \Omega$ ssi l'application*

$$t \in I_i \mapsto f(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n) \in \mathbb{R}^p$$

est dérivable en $t = x_i$. On note alors sa dérivée $\dfrac{\partial f}{\partial x_i}(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$ (appelée dérivée partielle de f selon la i^e direction).
En outre, la fonction f est dite de classe $\mathscr{C}^1$ sur Ω si elle est différentiable en chaque x de Ω, et chacune des ses dérivées partielles est continue sur Ω.

Exemple 1.4 *Soit $f : \mathbb{R}^2 \to \mathbb{R}^2$ définie par*

$$f(x_1, x_2) = (x_1^2 - 3x_2^3 + 2x_2 - 1, e^{2x_1} - x_2)^\mathsf{T}$$

Alors f est de classe $\mathscr{C}^1$ sur $\mathbb{R}^2$ et

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = (2x_1, 2e^{2x_1})^\mathsf{T} \quad et \quad \frac{\partial f}{\partial x_2}(x_1, x_2) = (-9x_2^2 + 2, -1)^\mathsf{T}$$

Définition 1.5 *Soit $f \in \mathscr{C}^1(\Omega, \mathbb{R}^p)$. On appelle matrice jacobienne de f en x la matrice $Jf(x)$ de taille $p \times n$, telle que*

$$[Jf(x)]_{ij} = \frac{\partial f_i}{\partial x_j}(x) \quad (1 \leq i \leq p, 1 \leq j \leq n)$$

Remarque 1.1 *On appelle aussi cette matrice la différentielle de f en x – on note alors $df(x)$ ou $f'(x)$.*

Exemple 1.5 *On reprend la fonction de l'exemple précédent. La matrice jacobienne est donnée par :*

$$Jf(x) = \begin{bmatrix} 2x_1 & -9x_2^2 + 2 \\ 2e^{2x_1} & -1 \end{bmatrix}$$

Exemple 1.6 *Soit $f : \mathbb{R}^3 \to \mathbb{R}^2$ l'application définie par*

$$f(x_1, x_2, x_3) = (e^{x_1} - x_3, x_2 + \sin x_3)^\mathsf{T}$$

Alors $f \in \mathscr{C}^1(\mathbb{R}^3, \mathbb{R}^2)$ et

$$Jf(x) = \begin{bmatrix} e^{x_1} & 0 & -1 \\ 0 & 1 & \cos x_3 \end{bmatrix}$$

On rappelle aussi le résultat de dérivation composée :

Théorème 1.1 *Soit $f : \mathbb{R}^n \to \mathbb{R}^p$ et $g : \mathbb{R}^p \to \mathbb{R}^q$ de classe $\mathscr{C}^1$. Alors $g \circ f : \mathbb{R}^n \to \mathbb{R}^q$ est de classe $\mathscr{C}^1$ et*

$$J[g \circ f](x) = Jg(f(x)) \cdot Jf(x)$$

Exemple 1.7 *Soit $f \in \mathscr{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ et $x, d \in \mathbb{R}^n$. On définit la fonction $\varphi : \mathbb{R} \to \mathbb{R}^p$ par*

$$\forall t \in \mathbb{R},\ \varphi(t) = f(x + td)$$

Alors $\varphi \in \mathscr{C}^1(\mathbb{R}, \mathbb{R}^p)$ et

$$\varphi = f \circ g \quad \textit{avec } g : t \in \mathbb{R} \mapsto x + td \in \mathbb{R}^n$$

Donc la formule de dérivation composée s'écrit :

$$\forall t \in \mathbb{R},\ \varphi'(t) = \underbrace{Jf(x + td)}_{p \times n} \cdot \underbrace{Jg(t)}_{n \times 1} = Jf(x + td) \cdot d \in \mathscr{M}_{p1}(\mathbb{R})$$

1.1.4 Cas particulier $p = 1$: gradient, hessienne

Dans toute la suite, on considère $f : \Omega \to \mathbb{R}$ une application.

Différentiation à l'ordre 1

Définition 1.6 *Si $f \in \mathscr{C}^1(\Omega, \mathbb{R})$, alors on appelle* gradient *de f en x le vecteur de $\mathbb{R}^n$*

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

Remarque 1.2 *Le gradient d'une application $f : \mathbb{R}^n \to \mathbb{R}$ est transposé de la jacobienne de f.*

Exemple 1.8 *Soit $f : \mathbb{R}^3 \to \mathbb{R}$ définie par $f(x) = \|x\|^2$. Si on écrit $x = (x_1, x_2, x_3)^\mathsf{T}$, alors*

$$f(x) = x_1^2 + x_2^2 + x_3^2$$

Alors le gradient est donné par $\nabla f(x) = (2x_1, 2x_2, 2x_3)^\mathsf{T} = 2x$.

Exemple 1.9 *Soit* $f : \Omega = (0, +\infty) \times \mathbb{R} \to \mathbb{R}$ *définie par*

$$f(x_1, x_2) = e^{x_1} + x_2 \ln x_1 + \cos x_2$$

Alors $f \in \mathscr{C}^1(\Omega, \mathbb{R})$ *et*

$$\nabla f(x) = \left(e^{x_1} + \frac{x_2}{x_1}, \ln x_1 - \sin x_2 \right)^\mathsf{T}$$

Exemple 1.10 *Soit* $A \in \mathscr{M}_n(\mathbb{R})$ *une matrice symétrique et* $b \in \mathbb{R}^n$*. Pour* $x \in \mathbb{R}^n$*, on pose*

$$f(x) = x^\mathsf{T} A x - b^\mathsf{T} x$$

Alors $f \in \mathscr{C}^1(\mathbb{R}^n, \mathbb{R})$ *et*

$$\forall x \in \mathbb{R}^n,\ \nabla f(x) = 2Ax - b$$

En effet $f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_j x_i - \sum_{i=1}^{n} b_i x_i$ *d'où*

$$\begin{aligned} \frac{\partial f}{\partial x_k}(x) &= \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i - b_k \\ &= (Ax)_k + (A^\mathsf{T} x)_k - b_k \end{aligned}$$

d'où le résultat puisque A est symétrique.

Remarquons que $f(x)$ est un nombre alors que $\nabla f(x)$ est un vecteur.

Différentiation à l'ordre 2

Définition 1.7 *f est de classe $\mathscr{C}^2$ sur Ω ssi $f \in \mathscr{C}^1(\Omega, \mathbb{R})$ et $\nabla f \in \mathscr{C}^1(\Omega, \mathbb{R}^n)$.*
On note alors Hf ou $\nabla^2 f$ la matrice jacobienne de ∇f ; elle est appelée hessienne *de f.*
Elle est donnée par

$$[Hf(x)]_{ij} = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

Exemple 1.11 *On reprend l'exemple* 1.8*. Alors* $f \in \mathscr{C}^2(\mathbb{R}^3, \mathbb{R})$ *et la hessienne est donnée par*

$$Hf(x) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Exemple 1.12 *La fonction de l'exemple* 1.9 *est de classe* $\mathscr{C}^2$ *sur* $\Omega = (0, +\infty[\times \mathbb{R}$ *et*

$$Hf(x) = \begin{pmatrix} e^{x_1} - \dfrac{x_2}{x_1^2} & \dfrac{1}{x_1} \\ \dfrac{1}{x_1} & -\cos x_2 \end{pmatrix}$$

On remarque que les deux matrices hessiennes calculées dans les exemples précédents sont des matrices symétriques. C'est en fait vrai en général :

Théorème 1.2 (Schwartz) *Soit $f \in \mathscr{C}^2(\Omega, \mathbb{R})$, alors*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

Exemple 1.13 *On revient sur l'exemple* (1.7) *: la dérivée seconde de l'application φ est donnée par*

$$\varphi''(t) = d^\mathsf{T} \nabla^2 f(x + td)d.$$

Formule de Taylor

On rappelle enfin la formule de Taylor à l'ordre 2 pour une fonction deux fois différentiable de $\mathbb{R}^n$ dans $\mathbb{R}$:

Théorème 1.3 *Soit $f \in \mathscr{C}^2(\Omega, \mathbb{R})$, alors pour tout $x \in \Omega$ et h suffisamment petit,*

$$f(x + h) = f(x) + \nabla f(x) \cdot h + \frac{1}{2} h^\mathsf{T} \nabla^2 f(x) h + o(\|h\|^2).$$

Rappel d'algèbre linéaire

Soit $A \in \mathscr{M}_n(\mathbb{R})$ une matrice *symétrique*.

Définition 1.8 *A est semi-définie positive ssi pour tout vecteur $x \in \mathbb{R}^n$, $x^\mathsf{T} A x \geq 0$.*
A est définie positive ssi A est semi-définie positive et $x^\mathsf{T} A x = 0 \Rightarrow x = 0$.

Exemple 1.14 *Soit*

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

Pour $x \in \mathbb{R}^2$,

$$x^\mathsf{T} A x = (x_1, x_2) \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1^2 - 2x_1 x_2 + 2x_2^2 = (x_1 - x_2)^2 + x_1^2 + x_2^2$$

ce qui prouve que A est définie positive.

Voici une caractérisation utile de la définition 1.8.

Proposition 1.4 *A est semi-définie positive ssi toutes les valeurs propres de A sont positives ou nulles.*
A est définie positive ssi toutes les valeurs propres de A sont strictement positives.

On rappelle qu'une matrice symétrique à coefficients réels est diagonalisable (en base orthonormale).

Exemple 1.15 *On reprend la matrice de l'exemple* 1.14. *Pour rechercher ses valeurs propres, on calcule le polynôme caractéristique et on le factorise :*

$$\chi_A(X) = \det(XI - A) = (X - 2)^2 - 1 = (X - 1)(X - 3)$$

Les deux valeurs propres sont donc 1 et 3 : elles sont strictement positives, donc A est définie positive en vertu du résultat 1.4.

1.2 Théorème de la fonction implicite

Soit $\Omega_1 \subset \mathbb{R}^n$ et $\Omega_2 \subset \mathbb{R}^p$ deux ouverts. On considère une application $f : \Omega_1 \times \Omega_2 \to \mathbb{R}^p$.

Notations : On désignera par (x, y) les éléments de $\Omega_1 \times \Omega_2$. La matrice $p \times p$

$$\left(\frac{\partial f_i}{\partial y_j}(x,y)\right), \ 1 \leq i \leq p \text{ et } 1 \leq j \leq p$$

est appelée jacobienne (ou différentielle) partielle par rapport à y et est notée $\partial_y f(x, y)$. On peut aussi définir de même la jacobienne partielle par rapport à x.
Dans le cas particulier d'une fonction numérique ($p = 1$), on parle de gradient partiel par rapport à y et on note

$$\nabla_y f(x,y) = \left[\frac{\partial f}{\partial y_1}(x,y), \cdots, \frac{\partial f}{\partial y_n}(x,y)\right]^\top$$

On rappelle maintenant un résultat essentiel du calcul différentiel :

Théorème 1.5 (fonction implicite) *On suppose que $f \in \mathscr{C}^1(\Omega_1 \times \Omega_2, \mathbb{R}^p)$. On suppose que $(a, b) \in \Omega_1 \times \Omega_2$ est tel que*

$$f(a,b) = 0 \quad \text{et} \quad \partial_y f(a,b) \text{ est inversible.}$$

Alors il existe des voisinages $V \subset \Omega_1$ de a et $W \subset \Omega_2$ de b tels que pour tout $x \in V$, l'équation (en y)

$$f(x,y) = 0$$

admette une et une seule solution $y = \varphi(x)$ dans W (en particulier $\varphi(a) = b$). De plus la fonction φ est de classe $\mathscr{C}^1(V, \mathbb{R}^p)$ et

$$\forall x \in V, \ \varphi'(x) = -\left[\partial_y f(x, \varphi(x))\right]^{-1} \partial_x f(x, \varphi(x))$$

DÉMONSTRATION : admise. ■

On peut remplacer l'expression "voisinage de" par "boule centrée en" dans l'énoncé précédent.

Exemple 1.16 *Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par*

$$f(x,y) = x + y^2$$

On calcule les différentielles partielles :

$$\partial_x f(x,y) = \frac{\partial f}{\partial x}(x,y) = 1 \quad \text{et} \quad \partial_y f(x,y) = \frac{\partial f}{\partial y}(x,y) = 2y$$

Considérons les deux cas suivants :

- $f(-1, 1) = 0$ *et* $\partial_y f(-1, 1) = 2 \neq 0$. *On peut donc appliquer le théorème* 1.5 *: il existe $\alpha > 0$ et une fonction $\varphi : (a - \alpha, a + \alpha) \to \mathbb{R}$ tels que*

$$\forall x \in (a-\alpha, a+\alpha), \ f(x, \varphi(x)) = 0 \text{ et } \varphi'(x) = \frac{1}{2\varphi(x)}$$

- $f(0, 0) = 0$ *et* $\partial_y f(0, 0) = 0$. *Donc le théorème* 1.5 *ne s'applique pas.*

On pourra vérifier que la fonction $\varphi(x) = \sqrt{x}$ convient dans le premier cas. Elle est candidate pour le deuxième, mais elle ne convient pas car n'est pas définie sur un voisinage de 0.

Chapitre 2

Généralités sur les problèmes d'optimisation

Dans toute la suite, on considère $f : A \subset \mathbb{R}^n \to \mathbb{R}$ où A est un sous-ensemble de $\mathbb{R}^n$.

2.1 Définitions

2.1.1 Optima absolus

Définition 2.1 *Si $x^* \in A$ est tel que*

$$\forall x \in A,\ f(x^*) \leq f(x)$$

alors on dit que f admet un minimum *(absolu) sur A en x^*. On note*

$$f(x^*) = \min_{x \in A} f(x) \quad \textit{et} \quad x^* = \operatorname*{argmin}_{x \in A} f(x)$$

Attention au vocabulaire : $f(x^*)$ est le minimum de f sur A ; f admet un minimum en x^* et x^* réalise le minimum de f sur A.

Exemple 2.1 *La fonction $f : [0,1] \to \mathbb{R}$ définie par*

$$\forall x \in \mathbb{R},\ f(x) = x^2 + 1$$

admet un minimum absolu sur $\mathbb{R}$:

$$\min_{x \in \mathbb{R}} f(x) = f(0) = 1 \quad \textit{et} \quad 0 = \operatorname*{argmin}_{x \in \mathbb{R}} f(x)$$

Exemple 2.2 *La fonction $\ln : (0,1) \to \mathbb{R}$ n'admet pas de minimum (car $\ln x \to -\infty$ quand $x \to 0$).*

Remarque 2.1 *Le minimum d'une fonction, s'il existe, est unique. Il peut cependant être atteint en plusieurs points différents.*

Exemple 2.3 *Soit $f : \mathbb{R} \to \mathbb{R}$ la fonction définie par*

$$\forall x \in \mathbb{R},\ f(x) = x^4 - 2x^2 + 1$$

Alors f admet 0 pour minimum absolu sur $\mathbb{R}$. Il est atteint en deux valeurs de x différentes : 1 et -1. En effet

$$\forall x \in \mathbb{R},\ f(x) = (x^2 - 1)^2$$

Exemple 2.4 *Soit $f : B \to \mathbb{R}$ donnée par $f(x, y) = 2 - x^2 - y^2$. B est la boule unité euclidienne fermée de $\mathbb{R}^2$:*

$$B = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

Alors f admet un minimum en tout point de la sphère unité ; ce minimum vaut 1.

Définition 2.2 *Si $x^* \in A$ est tel que*

$$\forall x \in A,\ f(x^*) \geq f(x)$$

alors on dit que f admet un maximum *(absolu) sur A en x^*. On note*

$$f(x^*) = \max_{x \in A} f(x) \quad \textit{et} \quad x^* = \operatorname*{argmax}_{x \in A} f(x)$$

Définition 2.3 *Si f admet en x^* un minimum ou un maximum, on dit qu'elle admet un* optimum *en x^*.*

Remarque 2.2 *Si f admet un minimum en x^*, alors $-f$ admet un maximum en x^*. C'est pourquoi, dans la suite, on ne parlera plus que de minimum, les énoncés concernant les maxima pourront être déduits facilement.*

2.1.2 Optima relatifs

Définition 2.4 *On dit que f admet un* minimum relatif *(ou* local*) sur A en x^* ssi il existe un voisinage V de x^* dans A tel que f admette un minimum absolu sur V en x^*.*

(on peut ici encore remplacer l'expression "voisinage de" par "petite boule centrée en").

Exemple 2.5 *Soit $f : \mathbb{R} \to \mathbb{R}$ la fonction définie par*

$$\forall x \in \mathbb{R},\ f(x) = x^3 - 3x + 1$$

Alors f admet en 1 un minimum relatif sur $\mathbb{R}$, mais pas un minimum absolu car $f(x)$ tend vers $-\infty$ quand $x \to -\infty$.

2.2 Optimisation sans contraintes – avec contraintes

Définition 2.5 *On appelle* problème de minimisation *:*

$$\text{“ Trouver } x^* \in \Omega \text{ tel que } f(x^*) = \min_{x \in A} f(x). \text{ ”}$$

Remarque 2.3 *Un problème de minimisation n'admet pas nécessairement de solution :*

$$x^* = \underset{x \in [1,+\infty)}{\operatorname{argmin}} \frac{1}{x} \quad \text{n'admet pas de solution.}$$

De plus s'il admet une solution, elle peut ne pas être unique :

$$x^* = \underset{x \in [0,4\pi]}{\operatorname{argmin}} \cos x \quad \text{admet } \pi \text{ et } 3\pi \text{ pour solutions.}$$

Vocabulaire : Si $A = \mathbb{R}^n$, on parle d'optimisation sans contraintes (ou libre), sinon il s'agit d'optimisation sous contrainte (ou liée).

$$A = \{x \in \mathbb{R}^n \mid g_1(x) = 0, \ldots, g_p(x) = 0\} \quad \text{contraintes d'égalités}$$

$$A = \{x \in \mathbb{R}^n \mid h_1(x) \leq 0, \ldots, h_q(x) \leq 0\} \quad \text{contraintes d'inégalités}$$

On peut aussi mélanger les deux types de contraintes :

$$A = \{x \in \mathbb{R}^n \mid g_1(x) = 0, \ldots, g_p(x) = 0 \text{ et } h_1(x) \leq 0, \ldots, h_q(x) \leq 0\}$$

On parle alors de contraintes égalités–inégalités. C'est le seul type d'optimisation liée que nous étudierons.

Définition 2.6 *On considère un problème d'optimisation avec contraintes de type inégalité et x^* une solution de ce problème :*

$$x^* = \underset{x \in A}{\operatorname{argmin}} f(x) \qquad (A = \{x \in \mathbb{R}^n \mid h_1(x) \leq 0, \ldots, h_q(x) \leq 0\})$$

Si $h_i(x^*) = 0$, on dit que la contrainte $h_i(x) \leq 0$ est serrée *en x^* (ou* saturée*, ou encore* active*). Dans le cas contraire ($h_i(x^*) < 0$), on dit que la contrainte* ne joue pas *(ou est* inactive*).*

2.3 Exemples de problèmes d'optimisation

Exemple 2.6 *Une entreprise produit un bien C à partir des matières premières A et B. Elle achète A au prix p_1 et B au prix p_2. On note f la fonction qui, à la quantité de matières premières, associe la quantité de bien produit (fonction de production). Le prix de vente de C est p.*

Quelles quantités de matières premières A et B l'entreprise doit-elle acheter pour que son profit soit maximum ?

Mathématiquement, ce problème s'écrit :

$$(x_1^*, x_2^*) = \underset{x_1, x_2 \in \mathbb{R}^+}{\operatorname{argmax}} \Big[p f(x_1, x_2) - p_1 x_1 - p_2 x_2 \Big]$$

Il s'agit d'un problème d'optimisation avec contraintes d'inégalités ($x_1 \geq 0$ et $x_2 \geq 0$).

Exemple 2.7 *On reprend l'exemple précédent, mais on suppose que le budget pour l'achat de matières premières est fixé à S.*

Le problème devient :

$$(x_1^*, x_2^*) = \underset{\substack{x_1, x_2 \in \mathbb{R}^+ \\ p_1x_1 + p_2x_2 = S}}{\operatorname{argmax}} \Big[pf(x_1, x_2) - p_1x_1 - p_2x_2 \Big] = \underset{\substack{x_1, x_2 \in \mathbb{R}^+ \\ p_1x_1 + p_2x_2 = S}}{\operatorname{argmax}} \Big[pf(x_1, x_2) - S \Big]$$

On est ici en présence d'un problème d'optimisation avec contraintes de type égalités–inégalités.

Exemple 2.8 *Dans l'exemple précédent, on suppose que le budget d'achat peut ne pas être totalement dépensé.*

Le problème d'optimisation devient

$$(x_1^*, x_2^*) = \underset{\substack{x_1, x_2 \in \mathbb{R}^+ \\ p_1x_1 + p_2x_2 \leq S}}{\operatorname{argmax}} \Big[pf(x_1, x_2) - p_1x_1 - p_2x_2 \Big]$$

Les contraintes sont toutes de type inégalité.

Remarque 2.4 *Comme le montrent les trois exemples précédents, les problèmes provenant de situations concrètes sont souvent des problèmes avec contraintes.*

Dans le dernier exemple, la contrainte

$$p_1x_2 + p_2x_2 \leq S$$

sera active. On le montrera mathématiquement, mais le contexte économique permet de le pressentir : le profit sera d'autant plus grand qu'il y aura d'avantage de matières premières.

Chapitre 3

Optimisation sans contrainte

On considère $f : \mathbb{R}^n \to \mathbb{R}$ et le problème de minimisation suivant :

$$\text{(P)} \qquad x^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x)$$

3.1 En dimension 1

On suppose ici $n = 1$.

3.1.1 Théorèmes d'existence et d'unicité

Théorème 3.1 *(i) Si* $\lim_{|x| \to +\infty} f(x) = +\infty$ *et f est continue sur* $\mathbb{R}$*, alors le problème* (P) *admet au moins une solution.*
(ii) Si, de plus, f est strictement convexe, il y a unicité.

DÉMONSTRATION : *(i)* Soit $a = f(0)$. Comme $\lim_{|x| \to +\infty} f(x) = +\infty$, il existe $R > 0$ tel que

$$|x| > R \Longrightarrow f(x) > f(0)$$

On en déduit

$$\min_{x \in \mathbb{R}} f(x) = \min_{|x| \leq R} f(x)$$

Or f est continue sur $I = [-R, R]$ et l'image d'un intervalle fermé borné par une application continue est un intervalle fermé borné :

$$f(I) = [m, M] \quad \text{avec } f(x^*) = m$$

ce qui prouve que

$$m = \min_{|x| \leq R} f(x)$$

d'où le résultat.
(ii) Notons x_1^* et x_2^* deux solutions et $x^* = (x_1^* + x_2^*)/2$. Alors, par stricte convexité,

$$f(x^*) < \frac{1}{2}\Big[f(x_1^*) + f(x_2^*)\Big] = f(x_1^*),$$

ce qui est une contradiction. ∎

Exemple 3.1 *Soit* $f(x) = x^6 - 3x^2 - x + 1$. *Alors* $f(x) \to +\infty$ *quand* $x \to \pm\infty$. *Donc* f *admet un minimum absolu sur* $\mathbb{R}$.

3.1.2 Condition nécessaire d'ordre 1

Théorème 3.2 *Si* $f \in \mathscr{C}^1(\mathbb{R}, \mathbb{R})$ *et admet en* x^* *un minimum relatif, alors*

$$f'(x) = 0$$

DÉMONSTRATION : Par définition de la dérivée :

$$f'(x^*) = \lim_{h \to 0} \frac{f(x^* + h) - f(x^*)}{h}$$

Or

$$\frac{f(x^* + h) - f(x^*)}{h} \quad \begin{cases} \geq 0 & \text{si } h \geq 0 \\ \leq 0 & \text{si } h \leq 0 \end{cases}$$

Par passage à la limite, $f'(x^*) \geq 0$ et $f'(x^*) \leq 0$ d'où $f'(x^*) = 0$. ■

Définition 3.1 *Un point* ξ *vérifiant* $f'(\xi) = 0$ *est appelé* point critique *de* f.

Exemple 3.2 *Soit* $f : \mathbb{R} \to \mathbb{R}$ *définie par*

$$\forall x \in \mathbb{R},\ f(x) = x^3 - 3x - 1$$

Les points candidats à être minimum de f *vérifient*

$$f'(x) = 3x^2 - 3 = 3(x^2 - 1) = 0 \Longrightarrow x = \pm 1$$

Une étude de la fonction f *permet de voir que* f *possède en* -1 *un maximum relatif et en* 1 *un minimum relatif.*

L'exemple précédent montre que la condition du théorème 3.2 n'est pas suffisante : elle ne permet pas de faire la différence entre maximum et minimum, elle ne distingue pas un optimum relatif d'un optimum absolu (c'est une condition locale). La situation est pire encore : les points critiques peuvent même ne pas être optimum relatif de f :

Exemple 3.3 *On considère la fonction* $x \mapsto x^3$. *Les points critiques sont les solutions de* $3x^2 = 0$, *donc il n'y en a qu'un :* 0, *qui ne correspond ni à un minimum relatif, ni à un maximum relatif de la fonction cube.*

3.1.3 Conditions d'ordre 2

On peut préciser le résultat vu au paragraphe précédent :

Théorème 3.3 *Soit* $f \in \mathscr{C}^2(\mathbb{R}, \mathbb{R})$. *Si* f *admet un minimum relatif en* x^*, *alors*

$$f'(x^*) = 0 \quad \text{et} \quad f''(x^*) \geq 0$$

DÉMONSTRATION : On sait déjà que $f'(x^*) = 0$. D'après la formule de Taylor, on peut écrire :

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^2 f''(x^*) + o(|x - x^*|^2)$$

D'où

$$\forall x \neq x^*,\ f''(x^*) = 2\frac{f(x) - f(x^*)}{(x - x^*)^2} + o(1)$$

Or

$$\forall x \neq x^*,\ \frac{f(x) - f(x^*)}{(x - x^*)^2} \geq 0$$

Le résultat s'en déduit par passage à la limite $x \to x^*$. ■

On verra dans les exemples suivants que la réciproque de ce théorème est fausse. On peut cependant donner une condition suffisante de minimum local :

Théorème 3.4 *Si $f \in \mathscr{C}^2(\mathbb{R}, \mathbb{R})$. Si f vérifie*

$$f'(x^*) = 0 \text{ et } f''(x^*) > 0$$

alors f admet un minimum relatif en x^*.

DÉMONSTRATION : Ici encore, on écrit la formule de Taylor

$$f''(x^*) = 2\frac{f(x) - f(x^*)}{(x - x^*)^2} + \varepsilon(x) \quad \text{avec } \lim_{x \to x^*} \varepsilon(x) = 0$$

Il existe $\eta > 0$ tel que

$$|x - x^*| \leq \eta \Longrightarrow |\varepsilon(x)| \leq f''(x^*)$$

On en déduit que

$$\forall x \neq x^*,\ \frac{f(x) - f(x^*)}{(x - x^*)^2} \geq 0 \Longrightarrow f(x^*) \leq f(x)$$

pour $|x - x^*| \leq \eta$, ce qui prouve que f admet en x^* un minimum relatif. ■

Remarque 3.1 *En appliquant ces deux théorèmes pour $-f$ on obtient des résultats similaires pour un maximum relatif : les conditions d'ordre 2 deviennent $f''(x^*) \leq 0$ et $f''(x^*) < 0$.*

Exemple 3.4 *On reprend les exemples du paragraphe précédent.*

- *Pour $f(x) = x^3 - 3x - 1$, on a*

  $$f''(-1) = -6 \text{ et } f''(1) = 6$$

  ce qui prouve que f admet un maximum relatif en -1 et un minimum relatif en 1.
- *Pour $f(x) = x^3$, on a $f''(0) = 0$ donc on ne peut pas conclure à l'aide du théorème 3.4.*

Remarque 3.2 *Les théorèmes précédents ne permettent pas de détecter un minimum absolu. Il faut par exemple faire une hypothèse de convexité pour obtenir un énoncé concluant à un minimum absolu.*

3.2 En dimension supérieure

3.2.1 Théorème d'existence

On peut énoncer le même résultat qu'en dimension 1 :

Théorème 3.5 *Si* $\lim_{\|x\|\to+\infty} f(x) = +\infty$ *et* f *est continue sur* $\mathbb{R}^n$*, alors le problème* (P) *admet au moins une solution. Il y a unicité si on ajoute une hypothse de stricte convexité.*

DÉMONSTRATION : La preuve est la même qu'en dimension 1, elle repose sur le fait que l'image d'un compact par une application continue est un compact. ■

Remarque 3.3 *L'énoncé précédent fait intervenir la norme de x, mais de laquelle s'agit-il ? En fait, peut importe car elles sont toutes équivalentes (* $\mathbb{R}^n$ *est de dimension finie n).*

Exemple 3.5 *Soit* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par*

$$\forall (x,y) \in \mathbb{R}^2,\ f(x,y) = e^{x^2} - x + y^4$$

Alors f *est continue et vérifie*

$$\lim_{\|(x,y)\|\to+\infty} f(x,y) = +\infty$$

En effet,

$$f(x,y) = [e^{x^2} - x] + [y^4] = h(x) + g(y) \quad \textit{avec} \quad \lim_{\pm\infty} h = \lim_{\pm\infty} g = +\infty$$

3.2.2 Condition nécessaire d'ordre 1

Théorème 3.6 *Si* $f \in \mathscr{C}^1(\mathbb{R}^n, \mathbb{R})$ *et admet en* x^* *un minimum relatif, alors*

$$\nabla f(x^*) = 0 = 0_{\mathbb{R}^n}$$

Cette égalité est appelée équation d'Euler.

DÉMONSTRATION : Soit $d \in \mathbb{R}^n$. On définit $\varphi : \mathbb{R} \to \mathbb{R}$ par

$$\forall t \in \mathbb{R},\ \varphi(t) = f(x^* + td)$$

Par la formule de dérivation composée, φ est de classe $\mathscr{C}^1$ sur $\mathbb{R}$ et

$$\forall t \in \mathbb{R},\ \varphi'(t) = \nabla f(x^* + td)^{\mathsf{T}} d$$

Comme x^* est un minimum relatif de f, 0 est un minimum relatif de φ. On en déduit :

$$\varphi'(0) = 0 \Longrightarrow \nabla f(x^*)^{\mathsf{T}} d = 0$$

Comme d est quelconque, on peut prendre $d = \nabla f(x^*)$, d'où le résultat. ■

Exemple 3.6 *Soit* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par*

$$f(x,y) = x^2 + y^2$$

Les points candidats à réaliser un minimum de f *vrifient*

$$\nabla f(x,y) = 2x = 0 \Longrightarrow x = 0$$

Donc seul $f(0)$ *est un probable minimum.*

3.2.3 Conditions d'ordre 2

Le théorème 3.3 devient

Théorème 3.7 *Soit $f \in \mathscr{C}^2(\mathbb{R}^n, \mathbb{R})$. Si f admet un minimum relatif en x^*, alors*

$$\nabla f(x^*) = 0 \quad \text{et} \quad \nabla^2 f(x^*) \text{ est semi-définie positive.}$$

DÉMONSTRATION : Soit φ définie comme dans la démonstration précédente. Alors

$$\forall t \in \mathbb{R},\ \varphi''(t) = d^{\mathsf{T}} \nabla^2 f(x^* + td)d$$

Par application du théorème 3.3, $d^{\mathsf{T}} \nabla^2 f(x^*)d \geq 0$. Comme d est quelconque, on en déduit le résultat. ■

De la même façon, voici le résultat correspondant au théorème 3.4

Théorème 3.8 *Si $f \in \mathscr{C}^2(\mathbb{R}^n, \mathbb{R})$. Si f vérifie*

$$\nabla f(x^*) = 0 \text{ et } \nabla^2 f(x^*) \text{ est définie positive.}$$

alors f admet un minimum relatif en x^*.

DÉMONSTRATION : Comme en dimension 1, on écrit la formule de Taylor

$$\forall x \in \mathbb{R}^n,\ f(x) = f(x^*) + \frac{1}{2}(x - x^*)^{\mathsf{T}} \nabla^2 f(x^*)(x - x^*) + o(\|x - x^*\|^2)$$

Soit $d \in \mathbb{R}^n$ et $\varepsilon > 0$, on pose $x = x^* + \varepsilon d$. On en déduit

$$\frac{f(x^* + \varepsilon d) - f(x^*)}{\|\varepsilon d\|^2} = \frac{(\varepsilon d)^{\mathsf{T}} \nabla^2 f(x^*)(\varepsilon d)}{2\|\varepsilon d\|^2} + \psi(\varepsilon)$$

avec $\psi(\varepsilon) \to 0$ quand $\varepsilon \to 0$. On peut récrire cette égalité :

$$f(x^* + \varepsilon d) - f(x^*) = \varepsilon^2 \Big(\frac{1}{2} d^{\mathsf{T}} \nabla^2 f(x^*)d + \|d\|^2 \psi(\varepsilon) \Big)$$

Par hypothèse, $d^{\mathsf{T}} \nabla^2 f(x^*)d > 0$ et, comme $\psi(\varepsilon) \to 0$, il existe $\varepsilon_0 > 0$ tel que

$$\varepsilon < \varepsilon_0 \Longrightarrow \Big| \|d\|^2 \psi(\varepsilon) \Big| \leq \frac{1}{2} d^{\mathsf{T}} \nabla^2 f(x^*)d$$

Cette dernière inégalité peut-être rendue vraie pour toute direction $\|d\| = 1$. Alors, pour $\varepsilon < \varepsilon_0$,

$$f(x) - f(x^*) = f(x^* + \varepsilon d) - f(x^*) \geq 0$$

d'où le résultat. ■

Exemple 3.7 *On reprend l'exemple précédent. La hessienne est constante :*

$$\forall x \in \mathbb{R}^2,\ \nabla^2 f(x) = 2I_2$$

Donc $\nabla^2 f(0)$ est définie positive, on peut donc conclure que f admet un minimum en 0 (ce qu'il est facile de voir par ailleurs).

Exemple 3.8 *Soit* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par*

$$\forall (x,y) \in \mathbb{R}^2,\ f(x,y) = x^2 + y^2 - xy$$

Les points critiques de f *sont les solutions de* $\nabla f(x,y) = 0$. *Or*

$$\nabla f(x,y) = (2x - y, 2y - x)^\mathsf{T}$$

D'où $x = y = 0$. *L'étude de la hessienne permet de conclure :*

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

qui est une matrice définie positive. Donc f *admet un minimum relatif en* $(0,0)$.

Remarque 3.4 *Comme en dimension* 1, *on ne peut pas conclure dans le cas où la hessienne est seulement semi-définie positive.*

Chapitre 4

Optimisation avec contraintes

Dans ce chapitre, on considère $f : \Omega \to \mathbb{R}^p$ où Ω est un ouvert.
Le problème qui nous intéresse est le suivant :

$$(Q) \qquad x^* = \underset{x \in K}{\operatorname{argmin}}\, f(x)$$

avec $K \subset \Omega$.

Théorème 4.1 *Soit x^* solution du problme (Q). Si x^* est à l'intérieur du domaine K, alors $\nabla f(x^*) = 0$.*

DÉMONSTRATION : La preuve est la même qu'en l'absence de contraintes, car on peut faire des petites variations dans toutes les directions autour de x^*. ■

Comme le montre l'exemple suivant, la condition d'annulation du gradient n'est pas nécessaire en présence de contraintes :

Exemple 4.1 *Soit le problème suivant :*

$$x^* = \underset{x \in [0,1]}{\operatorname{argmin}}(x^2 + x + 1)$$

Il est facile de voir que $x^* = 0$ en est l'unique solution. Cependant la dérivée en 0 vaut $1 \neq 0$.

Le problème provient du fait que le minimum est atteint au bord du domaine. Le but de ce chapitre est de trouver l'équivalent des conditions d'Euler dans le cas de l'optimisation avec contraintes. On se restreint aux contraintes de type égalité–inégalité.

4.1 Condition nécessaire d'ordre 1

4.1.1 Contraintes de type égalité

Dans ce paragraphe, on suppose que l'ensemble des contraintes K est donné par

$$K = \{x \in \mathbb{R}^n \mid g_1(x) = 0, \ldots, g_p(x) = 0\}$$

On notera $g(x)$ le vecteur $(g_1(x), \ldots, g_p(x))^\mathsf{T} \in \mathbb{R}^p$.

Définition 4.1 *On appelle* Lagrangien *du problème de minimisation* (Q) *l'application* $\mathscr{L}$ *de* $\Omega \times \mathbb{R}^p$ *valeurs dans* $\mathbb{R}$ *définie par*

$$\forall (x, \Lambda) \in \Omega \times \mathbb{R}^p, \ \mathscr{L}(x, \Lambda) = f(x) + \Lambda^{\mathsf{T}} g(x)$$

Ou encore

$$\forall x \in \Omega \quad \forall \lambda_1, \ldots, \lambda_p \in \mathbb{R}, \ \mathscr{L}(x, \lambda_1, \ldots, \lambda_p) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x)$$

Théorème 4.2 (extrema liés) *Soit* x^* *une solution de* (Q). *On suppose que les fonctions* f, g_1, $\ldots, g_p$ *sont de classe* $\mathscr{C}^1$ *et que les vecteurs*

$$\nabla g_1(x^*), \ldots, \nabla g_p(x^*) \in \mathbb{R}^n$$

sont linéairement indépendants. Alors

$$\exists \Lambda \in \mathbb{R}^p, \ \nabla_x \mathscr{L}(x^*, \Lambda) = 0$$

c'est-à-dire qu'il existe des réels $\lambda_1, \ldots, \lambda_p$ *tels que*

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

Cette équation est appelée équation d'Euler-Lagrange.

DÉMONSTRATION : Soit A la matrice

$$A = \left[\nabla g_1(x^*) \middle| \cdots \middle| \nabla g_p(x^*) \right]^{\mathsf{T}} = Jg(x^*) \in \mathscr{M}_{pn}(\mathbb{R})$$

L'hypothèse d'indépendance linéaire faite sur les gradients des fonctions g_i signifie que A est surjective. Elle admet donc un inverse à droite $B \in \mathscr{M}_{np}(\mathbb{R})$: $AB = I_p$.
On définit alors la fonction $G : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p$ par

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^p, \ G(x, y) = g(x^* + By + x)$$

On a $G(0,0) = g(x^*) = 0$ car $x^* \in K$. De plus G est de classe $\mathscr{C}^1$ et

$$\partial_y G(x, y) = Jg(x^* + By + x) \cdot B \Longrightarrow \partial_y G(0,0) = Jg(x^*) \cdot B = I_p$$

Le théorème de la fonction implicite s'applique donc : il existe un voisinage $V \subset \mathbb{R}^n$ de 0 et un voisinage $W \subset \mathbb{R}^p$ de 0 tels que, pour tout $x \in V$, l'équation $G(x, y) = 0$ admette une unique solution $y = \varphi(x) \in W$ (en particulier $\varphi(0) = 0$). On a donc $G(x, \varphi(x)) = 0$, d'où, après dérivation par rapport à x,

$$\partial_x G(x, \varphi(x)) + \partial_y G(x, \varphi(x)) \cdot J\varphi(x) = 0$$

Pour $x = 0$, on obtient :

$$\partial_x G(0,0) + \partial_y G(0,0) \cdot J\varphi(0) = 0$$

Or $\partial_x G(0,0) = Jg(0)$ et $\partial_y G(0,0) = I_p$ donc $J\varphi(0) = -Jg(x^*)$.
Si $x \in V$, alors $x^* + B\varphi(x) + x \in K$ et

$$f(x^*) \leq f(x^* + B\varphi(x) + x) \quad \forall x \in V$$

Donc $x = 0$ est un minimum relatif de $\tilde{f}$ définie par

$$\forall x \in \mathbb{R}^n,\ \tilde{f}(x) = f(x^* + B\varphi(x) + x)$$

Le théorème 3.6 s'applique à $\tilde{f}$:

$$0 = \nabla\tilde{f}(0) = \nabla f(x^*) + J\varphi(0)^\mathsf{T} B^\mathsf{T} \nabla f(x^*) = \nabla f(x^*) - Jg(x^*)^\mathsf{T} B^\mathsf{T} \nabla f(x^*)$$

En posant $\Lambda = -B^\mathsf{T}\nabla f(x^*)$, on obtient le résultat. ■

Remarque 4.1 *Les nombres $\lambda_1, \ldots, \lambda_p$ sont appelés* multiplicateurs de Lagrange*.*

Remarque 4.2 *Le théorème des extrema liés s'interprète simplement dans le cas où l'ensemble des contraintes est le cercle unité de $\mathbb{R}^2$. En effet, minimiser une fonction $f(x,y)$ sous la contrainte $x^2 + y^2 = 1$ revient à minimiser la fonction $\varphi(\theta) = f(\cos\theta, \sin\theta)$ pour θ dans $\mathbb{R}$. On est donc ramené un problème sans contrainte. La condition d'optimalité s'écrit*

$$\varphi'(\theta^*) = 0 \quad \Longleftrightarrow \quad -\sin\theta^* \frac{\partial f}{\partial x}(\cos\theta^*, \sin\theta^*) + \cos\theta^* \frac{\partial f}{\partial x}(\cos\theta^*, \sin\theta^*) = 0.$$

En interprétant cette dernière comme l'annulation d'un déterminant, elle montre que les vecteurs $\nabla f(\cos\theta^*, \sin\theta^*)$ et $(\cos\theta^*, \sin\theta^*)^\mathsf{T}$ sont colinéaires. Or ce dernier vecteur n'est autre que $\frac{1}{2}\nabla g(\cos\theta^*, \sin\theta^*)$, o $g(x,y) = x^2 + y^2 - 1$ définit la contrainte.
Ainsi, il existe un nombre réel μ tel que

$$\nabla f(\cos\theta^*, \sin\theta^*) = \mu\nabla g(\cos\theta^*, \sin\theta^*),$$

il suffit de poser $\lambda = -\mu$ pour retrouver le résultat annoncé.

Exemple 4.2 *Soit le problème de minimiser $x^2 - y$ sous la contrainte $x^2 + y^2 = 1$. Le Lagrangien s'écrit :*

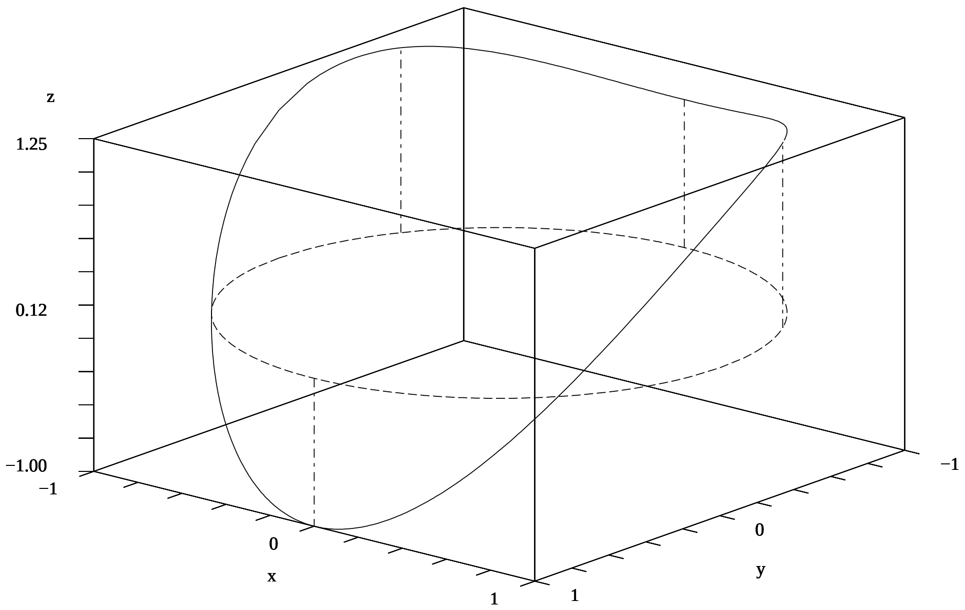
$$\mathscr{L}(x,y,\lambda) = x^2 - y + \lambda(x^2 + y^2 - 1)$$

si bien que la condition nécessaire d'ordre 1 s'écrit :

$$\begin{bmatrix} 2x \\ -1 \end{bmatrix} + \lambda \begin{bmatrix} 2x \\ 2y \end{bmatrix} = 0$$

Deux cas se présentent :
– *$x = 0$ d'où $y = \pm 1$.*
– *$x \neq 0$ et alors $\lambda = -1$ et $y = -1/2$. Comme $x^2 + y^2 = 1$, on en déduit $x = \pm\sqrt{3}/2$*
Il suffit maintenant de vérifier la condition d'indépendance linéaire, elle revient ici à la non nullité du gradient de la fonction définissant la contrainte, ce qui est aisé de vérifier.
En conclusion, les quatre points $(\pm\sqrt{3}/2, -1/2), (0, \pm 1)$ sont candidats. Il est facile de montrer que les deux premiers correspondent à un maximum absolu et que parmi les deux derniers, $(0,1)$ correspond à un minimum absolu et $(0,-1)$ à un minimum relatif (voir FIG 4.1*).*

FIG. 4.1 – Graphe de la fonction f de l'exemple 4.2



Remarque 4.3 *Il est crucial de vérifier la condition d'indépendance linéaire des gradients des fonctions définissant les contraintes. Considérons en effet la minimisation de* $x_1 + x_2^2$ *sous la contrainte* $x_1^3 - x_2^2 = 0$. *On voit facilement que* $(0,0)$ *est l'unique solution. Pourtant on n'a pas de condition d'ordre* 1 *:*

$$\nabla_x \mathscr{L}(x,\lambda) = \begin{bmatrix} 1 \\ 2x_2 \end{bmatrix} + \lambda \begin{bmatrix} 3x_1^2 \\ 2x_2 \end{bmatrix}$$

En $x = (0,0)$, *la condition d'ordre* 1 *fournit les deux égalités* $1 = 0$ *et* $0 = 0$, *ce qui est impossible.*

4.1.2 Contraintes de type inégalité

Dans ce paragraphe, on suppose que l'ensemble des contraintes K est donné par

$$K = \{x \in \mathbb{R}^n \mid h_1(x) \leq 0, \ldots, h_q(x) \leq 0\}$$

On notera $h(x)$ le vecteur $(h_1(x), \ldots, h_q(x))^\top \in \mathbb{R}^q$.

Définition 4.2 *On appelle* Lagrangien *du problème de minimisation* (Q) *la fonction* $\mathscr{L} : \Omega \times \mathbb{R}^q \to \mathbb{R}$ *définie par*

$$\forall (x, M) \in \Omega \times \mathbb{R}^q, \ \mathscr{L}(x, M) = f(x) + M^\mathsf{T} h(x)$$

Ou encore

$$\forall x \in \Omega \quad \forall \mu_1, \dots, \mu_q \in \mathbb{R}, \ \mathscr{L}(x, \mu_1, \dots, \mu_q) = f(x) + \sum_{i=1}^{q} \mu_i h_i(x)$$

Le résultat suivant fournit la condition nécessaire d'ordre 1 dans le cas des contraintes de type inégalités :

Théorème 4.3 (Kuhn et Tucker) *Soit x^* une solution de* (Q). *On suppose que les fonctions $f, h_1, \dots, h_q$ sont de classe $\mathscr{C}^1$ et que la condition suivante est vérifiée :*

$$(*) \qquad \exists d \in \mathbb{R}^n, \ h_i(x^*) = 0 \Longrightarrow d^\mathsf{T} \nabla h_i(x^*) < 0$$

Alors il existe des réels $\mu_1, \dots, \mu_q$ tels que

$$(i) \quad \mu_i \geq 0$$

$$(ii) \quad \mu_i h_i(x^*) = 0$$

$$(iii) \quad \nabla f(x^*) + \sum_{i=1}^{q} \mu_i \nabla h_i(x^*) = 0$$

DÉMONSTRATION : admise. ■

Remarque 4.4 *En dimension 1, le rsultat précédent est trs naturel : si f admet un minimum sur $[0, 1]$ en $x^* = 0$, alors $f'(x^*) \geq 0$. De même si $x^* = 1$, $f'(x^*) \leq 0$. C'est le signe des multiplicateurs de Kuhn-Tucker qui permet de retrouver ces conditions, associés aux contraintes $x \geq 0$ et $x \leq 1$.*

Remarque 4.5 *La condition (ii) signifie que si la contrainte n'est pas serrée (i.e. $h_i(x^*) < 0$), alors $\mu_i = 0$: elle n'intervient pas dans la condition d'ordre* 1. *On appelle (ii)* relation d'exclusion.
La condition $(*)$ est plus faible qu'une condition d'indépendance linéaire. On appelle $(*)$ hypothèse de qualification des contraintes.
Les coefficients $\mu_1, \dots, \mu_q$ sont appelés multiplicateurs de Kuhn et Tucker.
La condition (iii) s'écrit en termes de Lagrangien :

$$\nabla_x \mathscr{L}(x^*, M) = 0 \quad \textit{avec} \quad M = (\mu_1, \dots, \mu_n)$$

Exemple 4.3 *Soit $D = \{(x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 \leq 9\}$ et $f(x, y) = e^{x^2+y^2} + y^2 - 1$. Le Lagrangien est donné par*

$$\mathscr{L}(x, y, \lambda) = e^{x^2+y^2} + y^2 - 1 + \lambda(x^2 + y^2 - 9)$$

d'où

$$\nabla_{(x,y)} \mathscr{L}(x, y, \lambda) = \begin{bmatrix} 2xe^{x^2+y^2} + 2\lambda x \\ 2ye^{x^2+y^2} + 2y + 2\lambda y \end{bmatrix}$$

Donc la condition de Kuhn et Tucker s'écrit :

$$x(e^{x^2+y^2}+\lambda)=0 \quad et \quad y(e^{x^2+y^2}+1+\lambda)=0$$

Comme $\lambda \geq 0$, nécessairement $(x,y)=(0,0)$. Vérifions maintenant la condition de qualification des contraintes : elle est satisfaite car h et ∇h ne s'annulent pas simultanément.

Exemple 4.4 *Soit le problème de minimiser x sur l'ensemble*

$$K=\left\{(x,y)\in\mathbb{R}^2 \mid y\geq 0 \; et \; y\leq(1+x)^3\right\}$$

Le Lagrangien du problème est

$$\mathscr{L}(x,y,\mu_1,\mu_2)=x-\mu_1 y+\mu_2\Big[y-(1+x)^3\Big]$$

Donc

$$\nabla_{(x,y)}\mathscr{L}(x,y,\mu_1,\mu_2)=\begin{bmatrix} 1-3\mu_2(1+x)^2 \\ -\mu_1+\mu_2 \end{bmatrix}$$

La condition d'ordre 1 s'écrit :

$$1-3\mu_2(1+x)^2=0 \quad et \; \mu_1=\mu_2$$

De plus, les conditions d'exclusion fournissent :

$$\mu_1 y=0 \quad et \quad \mu_2\Big[y-(1+x)^3\Big]=0$$

- *Si $\mu_1=0$, alors $\mu_2=0$ d'où $1=0$, impossible ;*
- *Sinon, $y=0$ et alors $\mu_2(1+x)^3=0$. De même que dans le premier cas, on montre que $\mu_2\neq 0$, d'où $x=-1$.*

Le seul point candidat est $(-1,0)$.
Il reste à vérifier l'hypothèse de qualification des contraintes : si on note $h_1(x,y)=-y$ et $h_2(x,y)=y-(1+x)^3$ alors

$$\nabla h_1(-1,0)=\begin{bmatrix} 0 \\ -1 \end{bmatrix} \qquad et \qquad \nabla h_1(-1,0)=\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Et justement, elles ne sont pas vérifiées. Donc on ne peut pas appliquer le théorème de Kuhn et Tucker. Pourtant $(-1,0)$ est bien la solution recherchée.

Remarque 4.6 *Si on a affaire à un problème de maximisation, on considère $-f$ pour se ramener à un problème de minimisation, plutôt que de changer le signe des multiplicateurs.*

Exemple 4.5 *Soit $A\in\mathscr{M}_n(\mathbb{R})$ symétrique définie positive. On considère le problème*

$$x^*=\underset{\|x\|\leq 1}{\operatorname{argmax}}\, x^{\mathsf{T}}Ax$$

Le Lagrangien s'écrit $\mathscr{L}(x,\mu)=-x^{\mathsf{T}}Ax+\mu(\|x\|^2-1)$ d'où

$$\nabla_x\mathscr{L}(x^*,\mu)=-2Ax^*+2\mu x^*$$

La condition d'ordre 1 affirme donc que μ est valeur propre de A et x^* est vecteur propre associé. Ainsi

$$x^{*\mathsf{T}}Ax^* = \mu\|x\|^2$$

qui est maximum si μ est la plus grande valeur propre de A et x^* vecteur propre unitaire associé. La condition de qualification des contraintes est ici trivialement vérifiée.

Remarque 4.7 *En dimension 2, la condition de qualification des contraintes peut s'interpréter ainsi : les vecteurs gradients de contraintes serrées au point considéré sont situés dans un même demi-plan strict.*

4.1.3 Contraintes mixtes égalités–inégalités

Ici l'ensemble contrainte est donné par

$$K = \{x \in \mathbb{R}^n \mid g_1(x) = 0, \dots, g_p(x) = 0 \text{ et } h_1(x) \leq 0, \dots, h_q(x) \leq 0\}$$

Le Lagrangien s'écrit

$$\forall (x, \Lambda, \mathrm{M}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q,\ \mathscr{L}(x, \Lambda, \mathrm{M}) = f(x) + \Lambda^{\mathsf{T}} g(x) + \mathrm{M}^{\mathsf{T}} h(x)$$

La condition nécessaire d'ordre 1 est donnée par le théorème suivant :

Théorème 4.4 *Soit x^* une solution de* (Q). *On suppose que les fonctions f, $g_1, \dots, g_p$ et h_1, $\dots, h_q$ sont de classe $\mathscr{C}^1$. On fait, de plus, l'hypothèse de qualification des contraintes :*

$$(*) \quad \begin{array}{c} \exists d \in \mathbb{R}^n,\ d^{\mathsf{T}} \nabla g_i(x^*) = 0 \textit{ et } \big(h_i(x^*) = 0 \Longrightarrow d^{\mathsf{T}} \nabla h_i(x^*) < 0\big) \\ \textit{Les vecteurs } \nabla g_1(x^*), \dots, \nabla g_p(x^*) \textit{ sont linéairement indépendants.} \end{array}$$

Alors il existe des réels $\lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q$ tels que

$$\begin{array}{ll} (i) & \mu_i \geq 0 \\ (ii) & \mu_i h_i(x^*) = 0 \\ (iii) & g_i(x^*) = 0 \\ (iv) & \nabla f(x^*) + \displaystyle\sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{q} \mu_i \nabla h_i(x^*) = 0 \end{array}$$

DÉMONSTRATION : admise. ■

Exemple 4.6 *On se propose de minimiser x^2 sur l'ensemble contrainte :*

$$K = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 2 \textit{ et } x = y\}$$

On écrit le Lagrangien : $\mathscr{L}(x, y, \lambda, \mu) = x^2 + \lambda(x - y) + \mu(x^2 + y^2 - 2)$ si bien que

$$\nabla_x \mathscr{L}(x, y, \lambda, \mu) = \begin{pmatrix} 2x + \lambda + 2\mu x \\ -\lambda + 2\mu y \end{pmatrix}$$

La condition de Kuhn et Tucker s'écrit

$$2x^* + \lambda + 2\mu x^* = 0 \quad \textit{et} \quad 2\mu y^* = \lambda$$

On écrit aussi les relations de liaison et d'exclusion

$$x^* = y^* \quad \textit{et} \quad \mu(x^{*2} + y^{*2} - 2) = 0 \quad (\mu \geq 0)$$

On en déduit $x^*(1+2\mu) = 0$ d'où $x^* = 0$ puisque $\mu \geq 0$. Donc $y^* = 0$ (la condition de qualification des contraintes est vérifiée).

4.2 Condition suffisante d'ordre 2

Comme dans le cas de l'optimisation sans contrainte, on peut énoncer une condition suffisante faisant intervenir les dérivées d'ordre 2 du Lagrangien. On donne ici une version très faible du résultat, dans le cadre de contraintes de type "égalités".

Théorème 4.5 *On suppose que $f, g_1, \ldots, g_p$ sont de classe $\mathscr{C}^2$. Si x^* vérifie :*

$$\nabla_x \mathscr{L}(x, \Lambda) = 0 \quad \textit{et} \quad \nabla_x^2 \mathscr{L}(x, \Lambda) \textit{ est définie positive}$$

alors x^* est un minimum local de f sur K.

DÉMONSTRATION : admise. ∎

Chapitre 5

Algorithmes pour l'optimisation

5.1 Algorithmes de descente

5.1.1 Généralités

Méthodes de descente

Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ une application continue telle que le problème

$$x^* = \underset{x \in \Omega}{\operatorname{argmin}} f(x)$$

admette une unique solution.

Le but de ce paragraphe est de présenter une classe d'algorithmes – dits algorithmes de descente – pour déterminer numériquement x^*.

L'idée de la méthode consiste à construire une suite (x^k) telle que
- $\forall k \in \mathbb{N},\ x^k \in \Omega$;
- $f(x^{k+1}) \leq f(x^k)$.

On espère alors que la suite x^k converge vers x^* quand $k \to +\infty$. Par continuité de f,

$$f(x^k) \longrightarrow \min_{x \in \Omega} f(x)$$

5.1.2 Méthodes de gradient

Une méthode de descente peut s'écrire sous la forme :

$$x^{k+1} = x^k + d_k$$

La question est la suivante : x^k étant fixé, comment choisir la direction de descente d_k pour que $f(x^{k+1})$ soit inférieur à $f(x^k)$?

La formule de Taylor au premier ordre fournit une approximation à l'ordre 1 :

$$f(x^k + d_k) \simeq f(x^k) + \nabla f(x^k)^{\mathsf{T}} d_k$$

On veut que la différence $f(x^k) - f(x^k + d_k)$ soit maximum. Il s'agit donc de maximiser $-\nabla f(x^k)^{\mathsf{T}} d_k$. Or

$$|-\nabla f(x^k)^{\mathsf{T}} d_k| \leq \|\nabla f(x^k)\| \cdot \|d_k\| \quad \text{(inégalité de Cauchy-Schwartz)}$$

avec égalité pour $d_k = -\rho \nabla f(x^k)$ ($\rho > 0$). Cette remarque donne naissance aux méthodes dites de gradient :

$$x^{k+1} = x^k - \rho \nabla f(x^k)$$

5.1.3 Algorithme du gradient à pas fixe

Optimisation sans contrainte

On suppose ici $\Omega = \mathbb{R}^n$.
Afin de montrer la convergence de la méthode, on fait les hypothèses suivantes sur f :

$$\forall \omega \text{ borné} \quad \exists L_\omega > 0\ \forall x, y \in \omega,\ \|\nabla f(x) - \nabla f(y)\| \leq L_\omega \|x - y\| \quad (\nabla f \text{ est lipschitzien})$$

$$\exists \alpha > 0\ \forall x, y \in \mathbb{R}^n,\ (x-y)^\mathsf{T}(\nabla f(x) - \nabla f(y) \geq \alpha \|x - y\|^2 \qquad (f \text{ est } \alpha\text{−convexe})$$

Ces hypothèses impliquent que le problème de minimisation admet une unique solution.

On peut alors énoncer le théorème suivant :

Théorème 5.1 (convergence de la méthode du gradient) *On suppose que f vérifie les deux conditions ci-dessus, alors la suite construite par*

$$x^{k+1} = x^k - \rho \nabla f(x^k)$$

converge vers la solution x^* du problème de minimisation dès que ρ est suffisamment proche de 0.

DÉMONSTRATION : La suite x^k est bornée : en effet la suite $f(x^k)$ décroît, donc $f(x^k)$ appartient à la boule fermée $B(0, f(x^0))$ d'où $x^k \in B = f^{-1}(B(0, f(x^0)))$. Comme $B(0, f(x^0))$ est compacte et f continue, alors B est compact donc borné.
Soit $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ définie par

$$\forall x \in \mathbb{R}^n,\ \Phi(x) = x - \rho \nabla f(x)$$

Montrons que Φ est contractante sur B : soient $x, y \in B$,

$$\begin{aligned}
\|\Phi(x) - \Phi(y)\|^2 &= \|x - y - \rho(\nabla f(x) - \nabla f(y))\|^2 \\
&= \|x - y\|^2 + \rho^2 \|\nabla f(x) - \nabla f(y)\|^2 - 2\rho(x-y)^\mathsf{T}(\nabla f(x) - \nabla f(y)) \\
&\leq \|x - y\|^2 + \rho^2 L_B^2 \|x - y\|^2 - 2\alpha\rho \|x - y\|^2 \\
&= (1 - 2\alpha\rho + L_B^2 \rho^2) \|x - y\|^2
\end{aligned}$$

Une étude simple de la fonction $\psi : \rho \mapsto 1 - 2\alpha\rho + L_B^2\rho^2$ permet de dresser le tableau de variations suivant

x	$-\infty$	0	$\frac{\alpha}{L_B^2}$	$\frac{2\alpha}{L_B^2}$	$+\infty$
ψ	$+\infty \searrow$	$1 \searrow$	$\searrow 1 - \frac{\alpha}{L_B^2} \nearrow$	$\nearrow 1$	$\nearrow +\infty$

Donc Φ est strictement contractante pour $0 < \rho < \dfrac{2\alpha}{L_B^2}$. On conclut à l'aide du théorème du point fixe. ■

Remarque 5.1 *La convergence de l'algorithme est géométrique, de raison $\psi(\rho)$.*

Optimisation avec contraintes

Dans le cas de contraintes, l'algorithme précédent ne s'applique pas tel quel. En effet, même si $x^k \in \Omega$, il n'est pas sûr que $x^{k+1} \in \Omega$. Pour pallier cette difficulté, on modifie la méthode comme suit dans le cas où Ω est un convexe fermé :

$$x^{k+1} = \Pi_\Omega \Big[x^k - \rho \nabla f(x^k) \Big]$$

où Π_Ω est la projection sur Ω. Cette méthode converge sous les mêmes hypothèses :

Théorème 5.2 *Si f vérifie les hypothèses du paragraphe précédent et ρ est suffisamment proche de 0, alors la suite construite par*

$$x^{k+1} = \Pi_\Omega \Big[x^k - \rho \nabla f(x^k) \Big]$$

converge vers la solution x^* du problème de minimisation.

DÉMONSTRATION : La démonstration est identique à la précédente ; on utilise le fait que Π_Ω est contractante. ■

5.2 Méthode de Newton

Une autre façon d'aborder le problème consiste à travailler sur les équations d'Euler du problème (cas sans contraintes) ou les équations d'Euler-Lagrange (cas avec contraintes). On est donc ramené à résoudre une équation du type

$$F(x) = 0 \quad \text{où } F : \mathbb{R}^d \to \mathbb{R}^d$$

Nous allons ici décrire la méthode de Newton (ou méthode de Newton-Raphson) pour déterminer numériquement une racine de F.

Soit $x^0 \in \mathbb{R}^d$. On construit la suite (x^k) comme suit :

$$x^{k+1} = x^k - \Big[JF(x^k) \Big]^{-1} F(x^k)$$

(on suppose qu'il est licite d'écrire cette égalité).

Théorème 5.3 *Soit x^* solution de $F(x^*) = 0$. Si $F \in \mathscr{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ et $JF(x^*)$ est inversible, alors la suite x^k est bien définie et converge vers x^*.*

DÉMONSTRATION : Ici encore, on se ramène à un problème de point fixe : soit $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ définie par

$$\Phi(x) = x - \Big[JF(x) \Big]^{-1} F(x)$$

Montrons que Φ est contractante dans un voisinage de x^*. Puisque Φ est de classe $\mathscr{C}^1$, cela équivaut à montrer que

$$\| J\Phi(x^*) \| < 1$$

Or $JF(x)\Phi(x) = JF(x)x - F(x)$ donc

$$HF(x)\Phi(x) + JF(x)J\Phi(x) = HF(x)x + JF(x) - JF(x)$$

D'où $J\Phi(x^*) = \Big[JF(x^*)\Big]^{-2} HF(x^*)F(x^*) = 0$ donc $\|J\Phi(x^*)\| < 1$. ■

Remarque 5.2 *Comme $J\Phi(x^*) = 0$, la convergence est très rapide : on peut montrer qu'elle est quadratique, c'est-à-dire qu'à chaque étape le nombre de décimales exactes est doublé par rapport à l'étape précédente.*

5.3 Vers l'optimisation globale

Les méthodes décrites plut haut partagent toutes le même défaut de ne pas distinguer les minima locaux du minimum global. Des idées différentes ont été introduites, qui permettent de contourner plus ou moins cette difficulté. Les méthodes couramment utilisées dans cette optique sont les algorithmes *génétiques* ou *évolutionnaires*.

Annexe : méthode pratique de recherche d'extrema

Soit le problème d'optimisation (P) : "Trouver les minima de f sur K"

Écrire le Lagrangien du problème

Si l'ensemble contrainte est donné par (eventuellement $p = 0$ et/ou $q = 0$)

$$K = \{x \in \mathbb{R}^n \mid g_1(x) = 0, \ldots, g_p(x) = 0 \text{ et } h_1(x) \leq 0, \ldots, h_q(x) \leq 0\}$$

alors

$$\boxed{\mathscr{L}(x, \Lambda, M) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{i=1}^{q} \mu_i h_i(x)}$$

Écrire la CN d'ordre 1 et les relations de liaison et d'exclusion

$$\boxed{\begin{array}{c} \nabla_x \mathscr{L}(x^*, \Lambda, M) = 0 \Longleftrightarrow \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{q} \mu_i \nabla h_i(x^*) = 0 \\ g_i(x^*) = 0 \quad ; \quad \mu_i h_i(x^*) = 0 \quad \text{et} \quad \mu_i \geq 0 \end{array}}$$

Il s'agit d'un système de $n + p + q$ équations à $n + p + q$ inconnues.

Vérifier les conditions de qualification des contraintes

Pour chaque solution x^* du système précédent, on vérifie que

$$\boxed{\text{Les vecteurs } \nabla g_i(x^*) \text{ sont linéairement indépendants}}$$

$$\boxed{\exists d \in \mathbb{R}^n \quad \big(h_i(x^*) = 0 \Longrightarrow d^\top \nabla h_i(x^*) < 0\big) \text{ et } d^\top \nabla g_i(x^*) = 0}$$

N.B. les points qui ne satisfont pas la condition de qualification des contraintes sont de potentiels points de minimum...

Tester la condition suffisante d'ordre 2

$$\boxed{\nabla_x^2 \mathscr{L}(x^*, \Lambda) \text{ définie positive } \implies \text{ minimum}}$$

Dans le cas sans contraintes, on sait aussi que si la hessienne n'est pas semi-définie positive, alors on n'est pas en présence d'un minimum.

Bibliographie

[1] J.-M. ARNAUDIÈS, H. FRAYSSE. *Cours de mathématiques. 3*. Dunod, Paris 1989. Compléments d'analyse. [Complements of analysis].

[2] X. GOURDON. *Les Maths en tête, mathématiques pour M' : Analyse*. Ellipses, Paris 1994.

[3] J.-B. HIRIART-URRUTY. *L'optimisation*. Que sais-je ? PUF, Paris 1996.