4 Conditions d'optimalité. Optimisation avec contraintes

Cette Section est consacrée à l'étude du problème d'optimisation dit avec contraintes :

$$\min \quad f(x)$$
$$\begin{cases} x \in \mathbb{R}^n, \\ g(x) \leq 0, \\ h(x) = 0 \end{cases}$$

avec $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^q$, $h : \mathbb{R}^n \to \mathbb{R}^p$ définies pour des entiers $p \geq 1$, $q \geq 1$. Dans cet énoncé, la contrainte d'inégalité sur la fonction vectorielle g doit être entendue composante par composante.

La résolution de ce problème fait intervenir des multiplicateurs de Lagrange. On introduit cette notion sur le cas simplifié d'une contrainte d'égalité.

4.1 Multiplicateurs de Lagrange. Le théorème des extrema liés

Pour simplifier l'exposé de la méthode on commence par le cas particulier où la contrainte d'égalité décrit l'intersection d'un nombre fini d'hyperplans. Plus précisément, on considère le problème :

$$\min \quad f(x)$$
$$\begin{cases} x \in \mathbb{R}^n, \\ h(x) = 0 \end{cases}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est supposée différentiable sur $\mathbb{R}^n$ et où $h : \mathbb{R}^n \to \mathbb{R}^p$ est définie par :

$$h(x) = \begin{pmatrix} \langle a_1, x \rangle \\ \vdots \\ \langle a_p, x \rangle \end{pmatrix}.$$

On peut supposer, quitte à la restreindre, que la famille $\{a_1, \cdots, a_p\}$ est libre. On pose $K = \{x \in \mathbb{R}^n, \ h(x) = 0\}$. Alors K est un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $n - p$. Des résultats de la Section 3, il résulte que si $x^* \in \mathbb{R}^n$ est un minimum local de f alors

$$\forall y \in K, \quad \langle \nabla f(x^*), y \rangle = 0$$

i.e que $\nabla f(x^*) \in K^\perp$ avec $K^\perp$ sev de $\mathbb{R}^n$ de dimension p. On vérifie immédiatement que $\mathrm{Vect}\{a_1, \cdots, a_p\} \subset K^\perp$ avec $\mathrm{dimVect}\{a_1, \cdots, a_p\} = p$, donc $\mathrm{Vect}\{a_1, \cdots, a_p\} = K^\perp$. On en déduit qu'il existe $(\lambda_1, \cdots, \lambda_p) \in \mathbb{R}^p$ t.q. $\nabla f(x^*) + \sum_{i=1}^p \lambda_i a_i = 0$. Les coefficients λ_i, $i \in [[1, p]]$, sont appelés multiplicateurs de Lagrange.

Dans le cas général où $h : \mathbb{R}^n \to \mathbb{R}^p$ est quelconque, on note $h_1, \cdots, h_p$ ses composantes et on pose :

$$K = \{x \in \mathbb{R}^n,\ h_1(x) = \cdots = h_p(x) = 0\}.$$

Le résultat suivant généralise le cas linéaire.

Théorème 4.1.1 (Extrema liés)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur $\mathbb{R}^n$ et soit $h : \mathbb{R}^n \to \mathbb{R}^p$ de classe $\mathcal{C}^1$. On pose :*

$$K = \{x \in \mathbb{R}^n,\ h_1(x) = \cdots = h_p(x) = 0\}.$$

et on suppose que f admet un minimum local en $x^* \in K$ tel que

$$\text{la famille } (\nabla h_1(x^*), \cdots, \nabla h_p(x^*)) \text{ est libre.} \tag{10}$$

Alors il existe $(\lambda_1, \cdots, \lambda_p) \in \mathbb{R}^p$ t.q. :

$$\nabla f(x^*) + \sum_{k=1}^{p} \lambda_k \nabla h_k(x^*) = 0. \tag{11}$$

Remarque 10 (Qualification des contraintes). La condition (10) est appelée condition de qualification des contraintes. Si elle n'est pas satisfaite, alors la conclusion du Théorème 4.1.1 tombe en défaut. En effet, si $f(x) = x$ et $h(x) = x^2$, alors $K = \{0\}$ et $f \equiv 0$ sur K. De plus : h est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et $h'(x) = 2x \Rightarrow h'(0) = 0$ donc $f'(0) + \lambda h'(0) = 1 \neq 0$, $\forall \lambda \in \mathbb{R}$. Donc on ne peut pas définir de multiplicateur de Lagrange dans ce cas.

Si on remplace la condition (11) par la condition élargie :

$$\exists (\lambda_0, \lambda_1, \cdots, \lambda_p) \in \mathbb{R}^{p+1} \text{ t.q. } \lambda_0 \nabla f(x^*) + \sum_{k=1}^{p} \lambda_k \nabla h_k(x^*) = 0. \tag{12}$$

alors on peut montrer ([10],[5]) que (10)$\Rightarrow \lambda_0 \neq 0$ et on est ramené à (11).

Démonstration. On ne restreint pas la généralité du raisonnement en supposant, pour simplifier les notations, que $n = 2$ et $p = 1$. On pose :

$$K = \{x \in \mathbb{R}^2,\ h(x) = 0\}.$$

On commence par ramener le problème à celui de la minimisation d'une fonction à une seule variable. En effet, on remarque que (10) équivaut à

$\nabla h(x^*) \neq 0$. On peut donc supposer, quitte à échanger les rôles de x_1 et x_2, que :

$$\frac{\partial h(x^*)}{\partial x_2} \neq 0.$$

Alors d'après le Théorème des fonctions implicites,

$$\exists \varepsilon > 0,\ \exists \varphi \in \mathcal{C}^1(\mathbb{R}, \mathbb{R}) :\ K \cap B(x^*, \varepsilon) = \{x \in B(x^*, \varepsilon),\ x_2 = \varphi(x_1)\}.$$

Autrement dit, dans un voisinage de x^*, K s'identifie au graphe d'une fonction φ de classe $\mathcal{C}^1$. Par hypothèse sur f, l'application $\tilde{f} : x_1 \mapsto f(x_1, \varphi(x_1))$ admet un minimum local en x_1^*. On en déduit alors, $\tilde{f}$ étant dérivable au voisinage de x^* :

$$\tilde{f}'(x_1^*) = 0 \iff \frac{\partial f}{\partial x_1}(x_1^*, \varphi(x_1^*)) + \varphi'(x_1^*)\frac{\partial f}{\partial x_2}(x_1^*, \varphi(x_1^*)) = 0 \tag{13}$$

Or on a aussi : $\forall x = (x_1, \varphi(x_1)) \in B(x^*, \varepsilon) \cap K$,

$$h(x_1, \varphi(x_1)) = 0 \Rightarrow \frac{\partial h}{\partial x_1}(x_1, \varphi(x_1)) + \varphi'(x_1)\frac{\partial h}{\partial x_2}(x_1, \varphi(x_1)) = 0 \tag{14}$$

$$\underset{\frac{\partial h(x^*)}{\partial x_2} \neq 0}{\Rightarrow} \varphi'(x_1^*) = -\frac{\frac{\partial h}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)},$$

ce qui montre que $\varphi'(x_1^*)$ est bien défini. De (13) et (14) on déduit de plus que :

$$\nabla h(x^*) \in \mathrm{Vect}\left\{\begin{pmatrix} 1 \\ \varphi'(x_1^*) \end{pmatrix}\right\}^\perp \text{ et } \nabla f(x^*) \in \mathrm{Vect}\left\{\begin{pmatrix} 1 \\ \varphi'(x_1^*) \end{pmatrix}\right\}^\perp \tag{15}$$

Dans $\mathbb{R}^2$, l'hyperplan Vect $\left\{\begin{pmatrix} 1 \\ \varphi'(x_1^*) \end{pmatrix}\right\}^\perp$ est une droite vectorielle, donc $\nabla h(x^*)$ et $\nabla f(x^*)$ sont colinéaires. □

Remarque 11. Par hypothèse $h \in \mathcal{C}^1$ donc on peut définir en tout $x \in K$ un vecteur unitaire normal, resp. tangentiel, à K en x noté $\vec{n}(x)$, resp. $\vec{\tau}(x)$. Dans la base orthonormée locale $(\vec{n}(x), \vec{\tau}(x))$, le vecteur $\nabla h(x)$ se décompose sous la forme :

$$\nabla h(x) = \frac{\partial h}{\partial n}(x)\vec{n}(x) + \frac{\partial h}{\partial \tau}(x)\vec{\tau}(x).$$

On a de même, f étant différentiable : $\forall x \in K$,

$$\nabla f(x) = \frac{\partial f}{\partial n}(x)\vec{n}(x) + \frac{\partial f}{\partial \tau}(x)\vec{\tau}(x).$$

Dans $K \cap B(x^*, \varepsilon)$:

$$\vec{\tau}(x^*) = \frac{1}{\sqrt{1+\varphi'(x_1)^2}} \begin{pmatrix} 1 \\ \varphi'(x_1) \end{pmatrix}$$

donc (15) entraîne :

$$\nabla h(x^*) = \frac{\partial h}{\partial n}(x^*)\vec{n}(x^*) \quad \text{et} \quad \nabla f(x^*) = \frac{\partial f}{\partial n}(x^*)\vec{n}(x^*)$$

ce qui s'interprète en écrivant que

$$\frac{\partial h}{\partial \tau}(x^*) := \nabla h(x^*) \cdot \vec{\tau}(x^*) = \frac{1}{\sqrt{1+\varphi'(x_1)^2}} \left(\frac{\partial h}{\partial x_1}(x^*) + \varphi'(x_1)\frac{\partial h}{\partial x_2}(x^*) \right) = 0$$

i.e. (14). On retrouve (13) de la même façon en écrivant que

$$\frac{\partial f}{\partial \tau}(x^*) = 0.$$

Exemple 14. Soit à résoudre :

$$\left\{ \begin{array}{ll} \inf & (x^4+y^4) \\ (x,y) \in \mathbb{R}^2, & \\ x^2+y^2=1 & \end{array} \right.$$

Comme $\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{pmatrix}$ est semi-définie positive, f est convexe sur $\mathbb{R}^2$ donc n'admet que des minima locaux sur $\mathbb{R}^2$. De plus $f \geq 0$ et $f(x,y) = 0$ est atteint en-dehors de la surface $x^2+y^2=1$. De plus,

$$x^2+y^2=1 \Rightarrow |x| \leq 1 \quad \text{et} \quad |y| \leq 1$$

donc

$$x^2+y^2=1 \Rightarrow x^4+y^4 \leq x^2+y^2=1$$

i.e. $0 \leq f \leq 1$ sur la surface $x^2+y^2=1$.

On a aussi, d'après l'inégalité de Cauchy-Schwartz : $\forall (x,y) \in \mathbb{R}^2$,

$$1 = x^2+y^2 \leq \sqrt{2}\sqrt{x^4+y^4} = \sqrt{2f(x,y)} \Rightarrow f(x,y) \geq \frac{1}{2}$$

Finalement : $\frac{1}{2} \leq f \leq 1$ sur la surface $x^2+y^2=1$.

La surface $x^2+y^2=1$ se paramétrise en :

$$x = \cos\theta, \quad y = \sin\theta, \quad \theta \in [0, 2\pi[$$

et on a alors :

$$f(x,y) = \frac{1}{4}(3+\cos(4\theta)) =: g(\theta)$$

avec g paire, périodique de période $\frac{\pi}{2}$, donc on est ramené au problème équivalent :

$$\min_{0\leq\theta\leq\frac{\pi}{4}} g(\theta).$$

L'étude des variations de g montre que

$$\min_{0\leq\theta\leq\frac{\pi}{4}} g(\theta) = g\left(\frac{\pi}{4}\right) = \frac{1}{2} = f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right).$$

On se propose de retrouver le résultat précédent par le Théorème 4.1.1 appliqué avec $f(x,y) = x^4+y^4$, $h(x,y) = x^2+y^2-1$. On a : $\forall (x,y) \in \mathbb{R}^2$,

$$\nabla f(x,y) = 4\begin{pmatrix} x^3 \\ y^3 \end{pmatrix}, \quad \nabla h(x,y) = 2\begin{pmatrix} x \\ y \end{pmatrix}.$$

Alors $\nabla h = 0 \iff x = y = 0$. On remarque que $(0,0)$ n'est pas solution. D'après le Théorème 4.1.1, on est ramené à chercher $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$ et $\lambda \in \mathbb{R}$ solution de

$$\begin{cases} x^3 = \lambda x, \\ y^3 = \lambda y, \\ x^2+y^2 = 1. \end{cases} \iff \begin{cases} x(x^2-\lambda) = 0, \\ y(y^2-\lambda) = 0, \\ x^2+y^2 = 1. \end{cases} \iff$$

$$\iff \begin{cases} x = 0, \\ \lambda = y^2, \\ \lambda = 1. \end{cases} \quad \text{ou} \quad \begin{cases} y = 0, \\ \lambda = x^2, \\ \lambda = 1. \end{cases} \quad \text{ou} \quad \begin{cases} \lambda = x^2, \\ \lambda = y^2, \\ 2\lambda = 1. \end{cases}$$

$$\Rightarrow \begin{cases} x = 0, \\ y = \pm 1, \\ f(x,y) = 1. \end{cases} \quad \text{ou} \quad \begin{cases} y = 0, \\ x = \pm 1, \\ f(x,y) = 1. \end{cases} \quad \text{ou} \quad \begin{cases} x = \pm\frac{1}{\sqrt{2}} \\ y = \pm\frac{1}{\sqrt{2}}, \\ f(x,y) = \frac{1}{2}. \end{cases}$$

On en déduit que les points extrémaux de f sur la surface $x^2+y^2 = 1$ sont $(0,\pm 1)$, $(\pm 1, 0)$ avec $f(0,\pm 1) = f(\pm 1, 0) = 1$ et $\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)$ avec $f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right) = \frac{1}{2}$. Comme $1 = \max_{x^2+y^2=1} f(x,y)$, on en déduit que $f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right) = \frac{1}{2}$ est l'unique solution.

Exemple 15 (Application à la démonstration du Théorème spectral)**.** Soit à résoudre :

$$\inf_{\|x\|=1} \langle Ax, x\rangle$$

lorsque A est une matrice symétrique réelle d'ordre n. D'après la théorie, l'application continue $x \mapsto \langle Ax, x\rangle$ atteint son minimum sur la surface $\|x\| = 1$ qui est un compact de $\mathbb{R}^n$, evn de dimension finie. On applique le Théorème 4.1.1 avec $f(x) = \langle Ax, x\rangle$ et $h(x) = \|x\|^2 - 1$. Alors on est ramené à chercher $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}$ solution de

$$\begin{cases} Ax = \lambda x, \\ \|x\|^2 = 1 \end{cases}$$

i.e. $\lambda \in \mathbb{R}$ valeur propre réelle de A et $x \in \mathbb{R}^n \setminus \{0\}$ vecteur propre associé. Pour une telle solution : $f(x) = \lambda$ et la valeur minimale cherchée est la plus petite valeur propre de A.

Théorème 4.1.2 (Théorème spectral)**.** *Toute matrice symétrique réelle d'ordre n est diagonalisable dans une base orthonormée de $\mathbb{R}^n$.*

Démonstration. On raisonne par récurrence sur la dimension n. Si $n = 1$, alors $A \in \mathbb{R}$ est un réel et $f(x) = Ax^2, \forall x \in \mathbb{R}$. On en déduit : $\inf_{|x|=1} Ax^2 = A$. On suppose que toute matrice symétrique réelle d'ordre n est diagonalisable dans une base orthonormée de $\mathbb{R}^n$ et on considère une matrice symétrique réelle A d'ordre $n+1$. On note $A' = (a_{ij})_{1\leq i,j\leq n}$ la sous-matrice de A formée des n premières lignes et des n premières colonnes de A. Par hypothèse sur A, A' est symétrique réelle d'ordre n, donc diagonalisable dans une bon de $\mathbb{R}^n$ par hypothèse de récurrence. On note $(e_i)_{1\leq i\leq n}$ une telle base avec $Ae_i = \lambda_i e_i$, $\forall i \in [[1, n]]$. Soit alors $e_{n+1} \in \text{Vect}\{(e_1, \cdots, e_n)\}^\perp$, $\|e_{n+1}\| = 1$. Alors $(e_1, \cdots, e_{n+1})$ est une bon de $\mathbb{R}^{n+1}$. On a :

$$\forall i \in [[1, n]], \quad \langle Ae_{n+1}, e_i\rangle \underset{A^T=A}{=} \langle e_{n+1}, Ae_i\rangle \underset{Ae_i=\lambda_i e_i}{=} 0.$$

i.e. : $Ae_{n+1} \in \text{Vect}\{(e_1, \cdots, e_n)\}^\perp$. On en déduit qu'il existe $\lambda_{n+1} \in \mathbb{R}$ t.q. $Ae_{n+1} = \lambda_{n+1}e_{n+1}$, i.e. que la bon $(e_1, \cdots, e_{n+1})$ diagonalise A, ce qui montre la propriété cherchée au rang $n + 1$. □

Exemple 16 (Inégalité arithmético-géométrique)**.** Soit l'application :

$$J : \mathbb{R}_+^n \setminus \{0\} \to \mathbb{R},\ x \mapsto \frac{\sqrt[n]{\prod_{i=1}^n x_i}}{\sum_{i=1}^n x_i}$$

On remarque que

$$\forall x \in \mathbb{R}_+^n \backslash \{0\}, \quad \forall \lambda > 0, \quad J(\lambda x) = J(x) = \frac{f(x)}{\|x\|_1}, \quad \text{où :} \quad f(x) := \sqrt[n]{\Pi_{i=1}^n x_i}.$$

On en déduit :

$$\min_{x \in \mathbb{R}_+^n \backslash \{0\}} J(x) = \min_{\|x\|_1 = n} J(x) = \frac{1}{n} \min_{\|x\|_1 = n} f(x).$$

resp. :

$$\max_{x \in \mathbb{R}_+^n \backslash \{0\}} J(x) = \max_{\|x\|_1 = n} J(x) = \frac{1}{n} \max_{\|x\|_1 = n} f(x).$$

On pose :

$$X = \{x \in \mathbb{R}_+^n, \ \|x\|_1 = n\}.$$

Alors :

$$X = \{x \in \mathbb{R}_+^n, \quad \frac{1}{n} \sum_{i=1}^n x_i = 1\}.$$

On remarque que : $X \subset [0,1]^n$ donc

$$\forall x \in X, \quad f(x) \leq 1.$$

On en déduit que X est fermé comme image réciproque par l'application continue $x \mapsto \frac{1}{n} \sum_{i=1}^n x_i$ du singleton $\{1\}$, et borné comme partie de la sphère de $\mathbb{R}^n$ centrée en 0, de rayon n pour la norme $\| \cdot \|_1$. Donc X est un compact de $\mathbb{R}^n$. Comme de plus f est continue sur $\mathbb{R}_+^n$, on en déduit que f atteint ses bornes sur X. f étant différentiable sur $(\mathbb{R}_+^*)^n$, on applique le Théorème 4.1.1 à $f|_{(\mathbb{R}_+^*)^n}$ avec $h(x) = \frac{1}{n} \sum_{i=1}^n x_i - 1$. On a : $\forall x \in (\mathbb{R}_+^*)^n$,

$$\frac{\partial h}{\partial x_i}(x) = \frac{1}{n}, \quad \frac{\partial f}{\partial x_i}(x) = \frac{1}{n x_i} \sqrt[n]{\Pi_{j=1}^n x_j}$$

D'après le Théorème 4.1.1 on est ramené à chercher $(x, \lambda) \in (\mathbb{R}_+^*)^n \times \mathbb{R}$ solution de

$$\begin{cases} \dfrac{1}{n x_i} \sqrt[n]{\Pi_{j=1}^n x_j} = \dfrac{\lambda}{n}, \quad 1 \leq i \leq n \\ \\ \sum_{i=1}^n x_i = n \end{cases} \iff \begin{cases} \sqrt[n]{\Pi_{j=1}^n x_j} = \lambda x_i, \quad 1 \leq i \leq n \\ \\ \sum_{i=1}^n x_i = n \end{cases}$$

$$\iff \begin{cases} \lambda = \lambda x_i, \quad 1 \leq i \leq n \\ \\ \sqrt[n]{\Pi_{j=1}^n x_j} = \lambda \\ \\ \sum_{i=1}^n x_i = n \end{cases} \iff \lambda = x_i = 1, \quad 1 \leq i \leq n$$

On a :

$$f(1,\cdots,1) = 1 = \max_X f \Rightarrow \max_{x\in\mathbb{R}^n_+\setminus\{0\}} J(x) = \frac{1}{n}$$

d'où on déduit : $\forall x \in \mathbb{R}^n_+ \setminus \{0\}$,

$$\sqrt[n]{\Pi_{i=1}^n x_i} \leq \frac{1}{n}\sum_{i=1}^n x_i.$$

4.2 Les Théorèmes de F.John et Karush-Kuhn-Tucker

On étudie le problème plus général des contraintes de type inégalités. La preuve du Théorème qui en résulte est plus ardue que celle du Théorème 4.1.1 et peut être consultée dans [5], [7], [10], [11].

On pose :

$$K = \{x \in \mathbb{R}^n, \quad h(x) = 0 \quad \text{et} \quad g(x) \leq 0\}. \tag{16}$$

où : $h : \mathbb{R}^n \to \mathbb{R}^p$ et $g : \mathbb{R}^n \to \mathbb{R}^q$ sont de classe $\mathcal{C}^1$. On introduit la notion de direction admissible.

Définition 4.2.1 (Direction admissible)**.** Soit K défini par (16). Pour tout $x \in K$, l'ensemble

$$K(x) = \{h \in \mathbb{R}^n,\ \exists (x_k)_{k\geq 0} \in K^{\mathbb{N}},\ \exists(\varepsilon_k)_{k\geq 0} \in (\mathbb{R}^*_+)^{\mathbb{N}},$$

$$\lim_{k\to+\infty} x_k = x,\ \lim_{k\to+\infty} \varepsilon_k = 0,\ \lim_{k\to+\infty} \frac{x_k - x}{\varepsilon_k} = h\}$$

est appelé le cône des directions admissibles de x.

Pour $x \in K$, l'ensemble $K(x)$ est l'ensemble des vecteurs tangents en x à une courbe contenue dans K et passant par x. En particulier, si K est une variété régulière, alors $K(x)$ coïncide avec l'espace tangent à K en $x \in K$. On considère le problème de minimisation :

$$\inf_{x\in K} f(x) \tag{17}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable. Soit $x^* \in K$ solution de (17) et soit $d \in K(x^*)$. Par définition de $K(x^*)$, il existe des suites $(x_k)_{k\geq 0} \in K^{\mathbb{N}}$ et $(\varepsilon_k)_{k\geq 0} \in (\mathbb{R}^*_+)^{\mathbb{N}}$ t.q.

$$\lim_{k\to+\infty} x_k = x^*, \quad \lim_{k\to+\infty} \varepsilon_k = 0 \quad \text{et} \quad \lim_{k\to+\infty} \frac{x_k - x^*}{\varepsilon_k} = d.$$

On a : $\forall k \geq 0$,

$$\begin{aligned} f(x_k) &= f(x^*) + \langle \nabla f(x^*), x_k - x^* \rangle + o(\|x_k - x^*\|) \\ &= f(x^*) + \varepsilon_k \langle \nabla f(x^*), \frac{x_k - x^*}{\varepsilon_k} \rangle + o\left(\varepsilon_k \left\| \frac{x_k - x^*}{\varepsilon_k} \right\| \right) \\ &= f(x^*) + \varepsilon_k \langle \nabla f(x^*), d \rangle + \|d\| o(\varepsilon_k) \\ &= f(x^*) + \varepsilon_k (\langle \nabla f(x^*), d \rangle + \|d\| o(1)) \end{aligned}$$

On en déduit :

$$f(x_k) \geq f(x^*) \iff \varepsilon_k (\langle \nabla f(x^*), d \rangle + \|d\| o(1)) \geq 0$$

$$\underset{\varepsilon_k > 0}{\iff} \langle \nabla f(x^*), d \rangle + \|d\| o(1) \geq 0 \iff \langle \nabla f(x^*), d \rangle \geq 0.$$

Finalement, on retrouve la condition d'optimalité du premier ordre du Théorème 3.0.1 :

$$\langle \nabla f(x^*), d \rangle \geq 0, \quad \forall d \in K(x^*).$$

En général, on ne connaît pas explicitement le cône $K(x^*)$.

Théorème 4.2.1 (F. John)**.** *Soit x^* un minimum local du problème (17). Alors il existe $(\lambda_1, \cdots \lambda_p) \in \mathbb{R}^p$, $(\mu_0, \mu_1, \cdots, \mu_q) \in \mathbb{R}_+^{q+1}$ t.q. :*

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla g_j(x^*) = 0.$$

avec

$$h(x^*) = 0, \quad g(x^*) \leq 0$$

$$\mu_j g_j(x^*) = 0, \quad \forall j \in [[1, q]], \quad \textit{(condition de complémentarité).}$$

Comme dans le Théorème 4.1.1, on montre que $\mu_0 \neq 0$ dès que les contraintes vérifient des condition dites de qualification.

Définition 4.2.2 (Contrainte active. Qualification des contraintes)**.** Soit K défini par (16) et soit $x \in K$.

1. L'ensemble $I(x) = \{i \in [[1, q]],\ g_i(x) = 0\}$ est appelé ensemble des contraintes actives en x.
2. On dit que les contraintes sont qualifiées en x s'il existe $d \in \mathbb{R}^n$ t.q.

$$\forall i \in [[1, p]], \quad \forall j \in I(x), \quad \langle \nabla h_i(x), d \rangle = 0 \quad \text{et} \quad \langle \nabla g_j(x), d \rangle < 0, \tag{18}$$

   et si les vecteurs $\nabla h_1(x), \cdots, \nabla h_p(x)$ sont linéairement indépendants.

La direction $d \in \mathbb{R}^n$ associée à $x \in K$ dans la Définition 4.2.2 peut être vue comme une direction entrante au sens où : $x + td \in K$ pour t suffisamment petit :

$$g_j(x+td) = g_j(x)+t\langle \nabla g_j(x), d\rangle+o(t) = t\langle \nabla g_j(x), d\rangle+o(t) < 0 \quad \text{quand } t \to 0^+$$

Remarque 12 (Autre qualification des contraintes). Une condition suffisante pour que (18) ait lieu est que les vecteurs $\nabla h_1(x), \cdots, \nabla h_p(x), \nabla g_1(x), \cdots, \nabla g_q(x)$ soient linéairement indépendants. En effet, si $r = |I(x)|$, on pose $I(x) = \{i_1, \cdots, i_r\}$ et on peut chercher $d \in \text{Vect}\{(\nabla h_1, \cdots, \nabla h_p, \nabla g_{j_1}, \cdots, \nabla g_{j_r})\}$ sous la forme :

$$d = \sum_{i=1}^{p} d_i \nabla h_i(x) + \sum_{j \in I(x)} d'_j \nabla g_j(x)$$

solution de :

$$\begin{aligned} \langle \nabla h_i(x), d\rangle &= 0, \quad \forall i \in [[1,p]], \\ \langle \nabla g_j(x), d\rangle &= -1, \quad \forall j \in I(x). \end{aligned}$$

Alors $(d_1, \cdots, d_p, d'_{j_1}, \cdots, d'_{j_r})$ est solution du système :

$$\underbrace{\begin{pmatrix} \langle \nabla h, \nabla h\rangle & \langle \nabla g, \nabla h\rangle \\ \langle \nabla h, \nabla g\rangle & \langle \nabla g, \nabla g\rangle \end{pmatrix}}_{A} \begin{pmatrix} d_1 \\ \vdots \\ d_p \\ d'_{j_1} \\ \vdots \\ d'_{j_r} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \end{pmatrix}$$

où

$$\begin{aligned} (\langle \nabla h, \nabla h\rangle)_{ij} &= \langle \nabla h_i, \nabla h_j\rangle, \quad \forall i,j \in [[1,p]] \\ (\langle \nabla g, \nabla h\rangle)_{ij} &= \langle \nabla g_i, \nabla h_j\rangle, \quad \forall i \in I(x), \quad \forall j \in [[1,p]] \\ (\langle \nabla g, \nabla\rangle)_{ij} &= \langle \nabla g_i, \nabla g_j\rangle, \quad \forall i,j \in I(x). \end{aligned}$$

La matrice A est la matrice de Gramm construite à partir de la famille libre $(\nabla h_1, \cdots, \nabla h_p, \nabla g_{j_1}, \cdots, \nabla g_{j_r})$ qui est libre par hypothèse, donc A est inversible et le problème admet une solution unique.

On peut énoncer le résultat principal.

Théorème 4.2.2 (Karush-Kuhn-Tucker)**.** *Soit x^* un minimum local du problème (17). On suppose que les contraintes sont qualifiées en x^*. Alors il existe $(\lambda_1, \cdots \lambda_p) \in \mathbb{R}^p$, $(\mu_1, \cdots, \mu_q) \in \mathbb{R}^q_+$ t.q. :*

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

avec

$$h(x^*) = 0, \quad g(x^*) \le 0$$

$$\mu_j g_j(x^*) = 0, \quad \forall j \in [[1, q]], \quad \textit{(condition de complémentarité).}$$

La condition de complémentarité traduit que si une contrainte est inactive, soit $g_j(x^*) < 0$, alors $\mu_j = 0$, i.e. le multiplicateur de Lagrange associé est nul.

Si de plus f est convexe, alors le Théorème 4.2.2 donne une condition et nécessaire et suffisante d'optimalité analogue au cas sans contrainte.

Exemple 17 (Application du Théorème 4.2.2)**.** Soit le problème de minimsation avec contraintes :

$$\inf_{x^2+y^2\ge 1} (x^4 + 3y^4) \tag{19}$$

On remarque que :

$$x^2 + y^2 = x^2 + \frac{1}{\sqrt{3}}\sqrt{3}y^2 \underset{Schwartz}{\le} \sqrt{1 + \frac{1}{3}}\sqrt{x^4 + 3y^4} = \frac{2}{\sqrt{3}}\sqrt{f(x, y)}$$

$$\Rightarrow f(x, y) \ge \frac{3}{4}\|(x, y)\|^2 \tag{20}$$

donc $\lim_{\|x,y)\|\to+\infty} f(x, y) = +\infty$, i.e. f est coercive. Comme de plus f est continue et que l'ensemble $K = \{(x, y) \in \mathbb{R}^2,\ g(x, y) := 1 - (x^2 + y^2) \le 0\}$ est fermé comme image réciproque du fermé $]-\infty, 0]$ par l'application continue g, on en déduit que le problème (19) admet une solution $(x^*, y^*) \in K$ qui minimise f sur K.

On se propose d'écrire les conditions d'optimalité du premier ordre. Soit (x, y) un minimiseur (global). D'après le Théorème 4.2.2, il existe $\mu \ge 0$ t.q. :

$$\nabla f(x, y) + \mu\nabla g(x, y) = 0 \iff \begin{cases} 4\begin{pmatrix} x^3 \\ 3y^3 \end{pmatrix} - 2\mu\begin{pmatrix} x \\ y \end{pmatrix} = 0, \\ \\ \mu(1 - x^2 - y^2) = 0 \end{cases}$$

$$\iff \begin{cases} x(2x^2 - \mu) = 0, \\ \\ y(6y^2 - \mu) = 0, \\ \\ \mu(x^2 + y^2 - 1) = 0 \end{cases} \iff \begin{cases} x = 0, \\ \\ y = 0, \\ \\ \mu = 0 \end{cases} \text{ou} \quad \begin{cases} x = 0, \\ \\ y^2 = \frac{1}{6}\mu, \\ \\ \mu > 0 \text{ et } x^2 + y^2 = 1 \end{cases}$$

$$\text{ou} \quad \begin{cases} x^2 = \frac{1}{2}\mu, \\ y = 0, \\ \mu > 0 \text{ et } x^2 + y^2 = 1 \end{cases} \qquad \text{ou} \quad \begin{cases} x^2 = \frac{1}{2}\mu, \\ y^2 = \frac{1}{6}\mu, \\ \mu > 0 \text{ et } x^2 + y^2 = 1 \end{cases}$$

$$\iff \begin{cases} x = 0, \\ y = 0, \\ \mu = 0 \end{cases} \quad \text{ou} \quad \begin{cases} x = 0, \\ y^2 = 1, \\ \mu = 6 \end{cases} \quad \text{ou} \quad \begin{cases} x^2 = 1, \\ y = 0, \\ \mu = 2 \end{cases} \quad \text{ou} \quad \begin{cases} x^2 = \frac{3}{4}, \\ y^2 = \frac{1}{4}, \\ \mu = \frac{3}{2} \end{cases}$$

$$\iff \begin{cases} x = 0, \\ y = 0, \\ \mu = 0 \end{cases} \quad \text{ou} \quad \begin{cases} x = 0 \\ y = \pm 1, \\ \mu = 6 \end{cases} \quad \text{ou} \quad \begin{cases} x = \pm 1, \\ y = 0, \\ \mu = 2 \end{cases} \quad \text{ou} \quad \begin{cases} x = \pm\frac{\sqrt{3}}{2}, \\ y = \pm\frac{1}{2}, \\ \mu = \frac{3}{2} \end{cases}$$

De (20) on déduit que $f \geq \frac{3}{4}$ sur K. On a $f(0,0) = 0$, $f(\pm 1, 0) = 1$, $f(0, \pm 1) = 3$ et $f\left(\pm\frac{\sqrt{3}}{2}, \pm\frac{1}{2}\right) = \frac{3}{4} = \min_K f$.

Bibliographie

[1] J.P. Demailly, Analyse Numérique et Equations Différentielles, P.U.Grenoble, Paris, 2006.

[2] H. Brezis, Analyse fonctionnelle : théorie et applications, Masson, Paris, 1983.

[3] P. Ciarlet, Introduction à l'analyse numérique matricielle et à l'optimisaion, Masson, Dunod, Paris.

[4] P. Lascaux, R. Théodor, Analyse numérique matricielle appliquée à l'art de l'ingénieur, Dunod, Paris.

[5] J.-B. Hiriart-Urruty, Convex Analysis and Minimization Algorithms I, Springer-Verlag, 1996.

[6] C. Zuily, H. Queffélec, Analyse pour l'Agrégation, 3ème Edition, Dunod, 2007.

[7] G. Allaire, Analyse Numérique et Optimisation, Editions de l'Ecole Polytchnique, 2005.

[8] G. Allaire, S.M. Kaber, Numerical Linear Algebra, Texts in Applied Mathematics, vol. 55, Springer, 2008.

[9] P. Lascaux, R. Théodor, Analyse numérique matricielle appliquée à l'art de l'ingénieur. Tome 1. Dunod, Paris, 1986.

[10] M. Bergounioux, Optimisation et contrôle des systèmes linéaires, Dunod, 2001.

[11] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, C. Sagastizabal, Optimisation numérique, coll. SMAI Mathématiques et Applications 27, Springer, 1997.