Option Calcul Scientifique: Optimisation

Préparation Agrégation de Mathématiques
Université de Rennes 1
Isabelle Gruais

30 novembre 2021

1 Introduction

1.1 Le programme de l'Agrégation

— Optimisation et approximation
— Extrema des fonctions réelles de n variables réelles sans contraintes.
— Extrema des fonctions réelles de n variables réelles avec contraintes : multiplicateurs de Lagrange.
— Méthode des moindres carrés.

1.2 Vocabulaire

Soit $(V, \|\cdot\|)$ un espace vectoriel normé.

On considère le problème d'optimisation :

$$\inf_{x \in K} f(x) \tag{1}$$

avec $K \subset V$ et $f : K \to \mathbb{R}$ fonction coût ou critère.

Si $K = V$, resp. $K \neq V$, le problème (1) est dit sans contrainte, resp. avec contrainte.

Si $\dim K < +\infty$, resp. $\dim K = +\infty$, le problème (1) est dit d'optimisation en dimension finie, resp. en dimension infinie.

Sans perte de généralité, on peut se limiter au cas d'un problème de minimiation en remarquant que

$$\inf_{x \in K} f(x) \iff \sup_{x \in K}(-f(x))$$

On rappelle que si $x_0 \in V$ est solution de

$$x_0 \in K \quad \text{et} \quad f(x_0) = \inf_{x \in K} f(x)$$

alors on dit que la borne inf de (1) est atteinte en $x_0 \in K$ et on note

$$f(x_0) = \min_{x \in K} f(x).$$

Par abus de notation, on dit que x_0 est un minimum pour f sur K.

Si $V = \mathbb{R}$, on a le résultat fondamental :

Proposition 1.2.1. *Toute partie $X \subset \mathbb{R}$ de $\mathbb{R}$ non vide et minorée admet une borne inférieure et on a la caractérisation :*

$$m = \inf X \iff \begin{cases} \forall x \in X, \ m \leq x, \\ et \\ \forall \varepsilon > 0, \ \exists x_\varepsilon \in X, \ t.q. \ m \leq x_\varepsilon < m + \varepsilon. \end{cases}$$

On en déduit le critère équivalent :

Corollaire 1.2.2 (Suites minimisantes)**.** *Si $X \subset \mathbb{R}$ est une partie non vide minorée de $\mathbb{R}$, alors les propositions suivantes sont équivalentes :*

(i) $m = \inf X$

(ii) m est un minorant de X et il existe une suite $(x_n)_{n \geq 0}$ de points de X appelée minimisante qui converge vers m.

Le plan d'étude est le suivant :

(i) La théorie permet de déterminer si le problème de minimisation admet une solution.
Exemple 1. Une application continue $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ est majorée et minorée sur $[a, b]$ et atteint ses bornes sur ce compact.

(ii) Si oui, on identifie une telle solution grâce aux conditions nécessaires d'optimalité basées sur un développement de Taylor de f quand il existe : du premier ordre si f est de classe $\mathcal{C}^1$, du second ordre si f est de classe $\mathcal{C}^2$.
Exemple 2. Si $f(x) = x^2$, $\inf_{x \in [0,1]} f(x) = 0$ est atteint pour $x = 0$ solution de $f'(x) = 0$.
Exercice 1. Soit $p \in]0, +\infty[$ et soit $a = (a_n)_{n \geq 0} \in \ell^p$. Montrer que si $\frac{1}{p'} = 1 - \frac{1}{p}$, le problème

$$\sup_{u \in \ell^{p'}} \frac{|\sum_{n \geq 0} a_n u_n|}{\|u\|_{p'}}$$

est bien défini et admet une solution.

(iii) Si le problème de minimisation n'admet pas de solution, on cherche une suite minimisante, i.e. une suite $(x_n)_{n\geq 0} \in K^{\mathbb{N}}$ de K t.q. $f(x_n) \underset{n\to+\infty}{\to} \inf_K f$.

Exemple 3. La suite de terme général $u_n = 1 - \frac{1}{n}$, $n \geq 1$, est une suite minimisante pour le problème : $\sup]0,1[= 1$ qui n'a pas de solution dans $]0,1[$.

Exercice 2. Etudier $\sup_{u=(u_n)_{n\geq 0}\in c_0} |\sum_{n\geq 0} \frac{u_n}{2^n}|$.

(iv) Dans ce cours, on laisse de côté la question du choix de méthodes numériques pour calculer la solution du problème de minimisation quand la théorie ne permet pas de l'expliciter.

On termine par des exemples.

Exemple 4 (Dimension finie)**.** *Déterminer le volume maximal d'un parallélépipède rectangle de surface extérieure donnée égale à* 6*.*

On note a, b, c les longueurs des arêtes d'un parallélogramme rectangle donné. On est ramené à résoudre :

$$\sup_K V(a,b,c) \text{ avec } V(a,b,c) = abc,\ K = \{(a,b,c) \in (\mathbb{R}^+)^3,\ ab+bc+ca = 3\}$$

i.e. un problème d'optimisation avec contrainte.

Exemple 5. [Dimension infinie : problème dit de la Reine Didon] *Déterminer la portion d'aire maximale enclose par une courbe de longueur donnée L et le segment de ses extrémités données* $[a,b]$*.*

Soit $Y = \mathcal{C}^1([a,b])$ et soit

$$K = \{y \in Y,\ \int_a^b \sqrt{1+y'(x)^2}dx = L,\ y(a) = y(b) = 0\}.$$

D'après la théorie de l'intégration, on est ramené à résoudre :

$$\sup_{y\in K} \int_a^b y(x)dx.$$

On remarque que l'espace fonctionnel Y a été choisi pour que le problème d'optimisation soit bien défini.

1.3 Rappels de calcul différentiel

L'exemple 5 a montré que le cadre naturel de la théorie de l'optimisation est la classe $\mathcal{C}^1$ construite sur la notion de différentiabilité.

Définition 1.3.1 (Différentiabilité)**.** Soit E et F deux evns et soit $U \subset E$ un ouvert de E. Une application $f : U \to F$ est dite différentiable en $x_0 \in U$ s'il existe une application linéaire continue $df_{x_0} \in \mathcal{L}(E, F)$ appelée la différentielle de f en x_0 t.q.

$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - df_{x_0}(x - x_0)\|}{\|x - x_0\|} = 0$$

i.e. :

$$f(x) = f(x_0) + df_{x_0}(x - x_0) + o(\|x - x_0\|).$$

cette dernière écriture signifiant que f admet un DL d'ordre 1 au voisinage de x_0.

Remarque 1.

1. En dimension infinie, la différentiabilité d'une application dépend du choix de la norme utilisée, contrairement au cas de la dimension finie où toutes les normes sont équivalentes.
2. Par définition de la différentiabilité, la différenielle $df_{x_0} : E \to F$ en un point x_0 est continue sur E. Si en outre l'application résultante $df : U \to \mathcal{L}(E, F)$, $x \mapsto df_x$, est continue sur U , alors f est dite de classe $\mathcal{C}^1$ sur U.
3. Si on a montré que f est différentiable en x_0, le calcul pratique de $df_{x_0}(h)$ pour $h \in E$ donné s'effectue grâce à la formule

$$df_{x_0}(h) = \lim_{t \to 0} \frac{f(x_0 + th) - f(x_0)}{t}.$$

En toute rigueur, ce calcul donne la dérivée directionnelle de f en x_0, ou diffŕentielle au sens de Gâteaux de f en x_0 dans la direction h. Des définitions, il résulte immédiatement que si f est différentiable en x_0, alors elle admet une dérivéee directionnelle dans toutes les directions alors que la réciproque est fausse.

En résumé :
f de classe $\mathcal{C}^1$ en $x_0 \Rightarrow f$ est différentiable en $x_0 \Rightarrow f$ de classe $\mathcal{C}^0$ en x_0

$$\Downarrow$$

f dérivable en x_0 suivant toutes les directions

Exemple 6 (Contre-exemple)**.** L'application f définie par

$$f(x, y) = \begin{cases} \dfrac{x^3}{x + y} & \text{si} \quad x \neq -y, \\ 0 & \text{sinon} \end{cases}$$

admet une dérivée en $(0,0)$ suivant toutes les directions de $\mathbb{R}^2$ mais n'est pas continue en $(0,0)$

Soit $\lambda \in \mathbb{R}^*$. On a : $\forall (x,y) \in \mathbb{R}^2$, $x \neq -y$,

$$f(x,y) = \lambda \iff y = \frac{x^3}{\lambda} - x$$

donc

$$\lim_{(x,y)\to(0,0),\ y=\frac{x^3}{\lambda}-x} f(x,y) = \lambda \neq 0.$$

De plus, $\forall (x,y) \in \mathbb{R}^2$, $x \neq -y$,

$$\forall t \in \mathbb{R}^*, \quad f(tx,ty) = t^2 f(x,y) \Rightarrow \lim_{t\to 0} \frac{f(tx,ty)}{t} = 0 =: f'_{(0,0)}(x,y).$$

Remarque 2 (Différentiabilité d'ordre supérieur). Si V est un espace de Hilbert et si $f : V \to \mathbb{R}$ est supposée différentiable en $x_0 \in V$, d'après le Théorème de Riesz, il existe $\nabla f(x_0) \in V$, appelé le gradient de f en x_0, t.q. : $df_{x_0}(h) = \langle \nabla f(x_0), h\rangle_V$, $\forall h \in V$, conformément au DL : $\forall h \in V$,

$$f(x_0+h) = f(x_0) + df_{x_0}(h) + o(\|h\|_V) = f(x_0) + \langle \nabla f(x_0), h\rangle_V + o(\|h\|_V).$$

On dit que f est deux fois différentiable en $x_0 \in V$ si $df : U \subset V \to \mathcal{L}(V,\mathbb{R}) = V' \sim V$ est différentiable en x_0, i.e. s'il existe une application linéaire continue $L_{x_0} \in \mathcal{L}(V,V)$ t.q. :

$$df_{x_0+\xi} = df_{x_0} + L_{x_0}(\xi) + o(\|\xi\|) = df_{x_0} + d^2 f_{x_0}(\xi) + o(\|\xi\|)$$

en posant $d^2 f_{x_0} := L_{x_0}$, i.e. :

$$\nabla f(x_0+\xi) = \nabla f(x_0) + \nabla^2 f(x_0)\xi + o(\|\xi\|_V)$$

en utilisant la notation d'endomorphisme $\nabla^2 f(x_0)\xi := d^2 f_{x_0}(\xi)$ avec $\nabla^2 f(x_0) \in \mathcal{L}(V,V)$. On en déduit le DL à l'ordre 2 en x_0 :

$$f(x_0+h) = f(x_0) + \langle \nabla f(x_0), h\rangle_V + \frac{1}{2}\langle \nabla^2 f(x_0)h, h\rangle_V + o(\|h\|_V^2)$$

où $(h,\xi) \mapsto \langle \nabla^2 f(x_0)h, \xi\rangle_V$ est une forme bilinéaire continue sur $V \times V$.

En dimension finie, l'endomorphisme $\nabla^2 f(x_0) \in \mathcal{L}(V,V)$ s'identifie à la matrice hessienne de f en x_0.

1.4 Détour par la dimension finie

Le programme de l'Agrégation pour l'optimisation concerne essentiellement le cas de la dimension finie sur lequel on revient en détail. Dans la suite, on note $(e_1, \cdots, e_n)$ la base canonique de $\mathbb{R}^n$ et on munit $\mathbb{R}^n$ de sa structure euclidienne usuelle.

Définition 1.4.1 (Fonction de classe $\mathcal{C}^k$)**.** Soit $i \in [[1, n]]$ et soit $k \geq 2$. Soit U un ouvert de $\mathbb{R}^n$ et soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$.

1. On dit que f admet une dérivée partielle d'indice i en x_0 si f admet une dérivée directionnelle en x_0 dans la direction de e_i.
2. On dit que f est de classe $\mathcal{C}^k$ sur U si toutes ses dérivées partielles jusqu'à l'ordre k existent et sont continues sur U.

Dans la suite, on considère $f : U \subset \mathbb{R}^n \to \mathbb{R}$. Soit $x_0 \in K$.

(i) On suppose f différentiable en x_0. Alors :

$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0), h\rangle_V + \underset{h\to 0}{o}(\|h\|)$$

où $\nabla f(x_0)$ est le gradient de f en x_0, i.e. le vecteur de composantes $\dfrac{\partial f}{\partial x_1}(x_0), \cdots, \dfrac{\partial f}{\partial x_n}(x_0)$. La définition du gradient dépend du choix du produit scalaire et est une conséquence du théorème de représentation de Riesz appliqué à la différentielle de f en x_0. En dimension finie, le choix du produit scalaire canonique conduit aux formules ci-dessus pour le gradient et la matrice hessienne et peuvent alors être considérées comme génériques.

(ii) Si f est deux fois différentiable en x_0 alors

$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0), h\rangle + \frac{1}{2}\langle \nabla^2 f(x_0)h, h\rangle + \underset{h\to 0}{o}(\|h\|^2)$$

où $\nabla^2 f(x_0)$ est la matrice hessienne de f en x_0 formée par les dérivées partielles secondes de f en x_0 soit :

$$(\nabla^2 f(x_0))_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0), \quad i, j \in [[1, n]].$$

D'après le Théorème de Schwartz, si f est deux foix différentiable en x_0 alors la matrice hessienne est symétrique (réelle).

Remarque 3 (Contre-exemple de Peano). Si f n'est pas deux foix différentiable en x_0 alors la matrice hessienne n'est pas nécessairement symétrique, comme le montre le contre-exemple dû à Peano :

$$f(x,y) = \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2} & \text{si} \quad (x,y) \neq (0,0), \\ 0 & \text{sinon} \end{cases}$$

On en déduit :

$$\frac{|f(x,y)|}{\sqrt{x^2+y^2}} \leq \frac{|x||y|(x^2+y^2)}{(x^2+y^2)^{3/2}} = \frac{|x||y|}{(x^2+y^2)^{1/2}} \leq \frac{(x^2+y^2)}{2(x^2+y^2)^{1/2}} =$$

$$= \frac{1}{2}(x^2+y^2)^{1/2} \underset{(x,y)\to(0,0)}{\to} 0.$$

et donc f est différentiable en $(0,0)$ de différentielle $df_{(0,0)} = 0$. Des relations : $\forall (x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$,

$$f(x,y) = xy\left(1 - \frac{2y^2}{x^2+y^2}\right) = xy\left(\frac{2x^2}{x^2+y^2} - 1\right)$$

on déduit :$\forall (x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$,

$$\frac{\partial f}{\partial x}(x,y) = y\left(1 - \frac{2y^2}{x^2+y^2}\right) + \frac{4x^2y^3}{(x^2+y^2)^2},$$

$$\frac{\partial f}{\partial y}(x,y) = x\left(\frac{2x^2}{x^2+y^2} - 1\right) + \frac{4x^3y^2}{(x^2+y^2)^2}.$$

Il en résulte :

$$\lim_{y\to 0} \frac{1}{y}\left(\frac{\partial f}{\partial x}(0,y) - \frac{\partial f}{\partial x}(0,0)\right) = \lim_{y\to 0} \frac{1}{y}\frac{\partial f}{\partial x}(0,y) = \lim_{y\to 0} \frac{-y}{y} = -1.$$

i.e. :

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1.$$

De même :

$$\lim_{x\to 0} \frac{1}{x}\left(\frac{\partial f}{\partial y}(x,0) - \frac{\partial f}{\partial y}(0,0)\right) = \lim_{x\to 0} \frac{1}{x}\frac{\partial f}{\partial y}(x,0) = \lim_{x\to 0} \frac{x}{x} = 1.$$

i.e. :

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1 \neq \frac{\partial^2 f}{\partial y \partial x}(0,0),$$

donc la matrice hessienne de f en $(0,0)$ n'est pas symétrique.

On rappelle les formules de Taylor à l'ordre 2.

Proposition 1.4.1 (Formule de Taylor avec reste intégral)**.** *Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$ sur un ouvert U de $\mathbb{R}^n$ et soit $h \in \mathbb{R}^n$ t.q. $[x_0, x_0+h] \subset U$. Alors :*

$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0), h\rangle + \int_0^1 (1-t)\langle \nabla^2 f(x_0 + th)h, h\rangle dt$$

Proposition 1.4.2 (Formule de Taylor avec reste de Lagrange)**.** *Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$ sur un ouvert U de $\mathbb{R}^n$ et soit $h \in \mathbb{R}^n$ t.q. $[x_0, x_0 + h] \subset U$. On suppose que :*

$$M := \sup_{t\in[0,1]} \frac{|\langle \nabla^2 f(x_0 + th)h, h\rangle|}{\|h\|^2} < +\infty.$$

Alors :

$$|f(x_0 + h) - f(x_0) - \langle \nabla f(x_0), h\rangle| \leq \frac{M}{2}\|h\|^2.$$

2 Existence et unicité de la solution d'un problème d'optimisation

On peut retenir comme principe général que la compacité fournit des résultats d'existence et la convexité un cadre favorable pour l'unicité.

Dans la suite, on étudie les fonctions convexes d'abord dans le cas particulier de la dimension finie. Dans la Section 2.3, on considère le cas plus général des espaces de Hilbert.

2.1 Existence en dimension finie

Soit $f : K \subset \mathbb{R}^n \to \mathbb{R}$ continue. On considère le problème :

$$\min_{x\in K} f(x). \tag{2}$$

Sans hypothèse supplémentaire, ce problème n'a pas de solution. Pour le voir, il suffit de prendre $f : x \mapsto e^x$ et $K = \mathbb{R}$.

Théorème 2.1.1 (Existence en dimension finie)**.** *On suppose qu'il existe $x_0 \in \mathbb{R}^n$ t.q. $\{x \in \mathbb{R}^n,\ f(x) \leq f(x_0)\}$ soit borné. Alors le problème (2) admet une solution globale.*

Démonstration. On est ramené au problème :

$$\min_{x\in\tilde{K}} f(x) \quad \text{où} \quad \tilde{K} := \{x \in \mathbb{R}^n,\ f(x) \leq f(x_0)\}.$$

Par hypothèse, $\tilde{K}$ est borné. De plus, $\tilde{K} = f^{-1}(]-\infty, f(x_0)])$ est fermé comme image réciproque par f continue du fermé $]-\infty, f(x_0)]$ de $\mathbb{R}$. On en déduit que $\tilde{K}$ est compact et que f continue sur $\tilde{K}$ est bornée sur $\tilde{K}$ et y atteint ses bornes. En effet, par hypothèse, $m := \inf_{\tilde{K}} f > -\infty$ donc il existe une suite minimisante $(x_n)_{n\geq 0} \in \tilde{K}^{\mathbb{N}}$. Comme $\tilde{K}$ est compact, on peut en extraire une suite convergente $(x_{\varphi(n)})_{n\geq 0}$. Soit x^* sa limite. Par continuité de f :

$$\lim_{n\to+\infty} f(x_{\varphi(n)}) = f(x^*).$$

Par construction :

$$\lim_{n\to+\infty} f(x_{\varphi(n)}) = m$$

donc $f(x^*) = \inf_{x\in\tilde{K}} f(x) = \min_{x\in\tilde{K}} f(x)$. □

Remarque 4 (Deux remarques très utiles en pratique...). Il reste à voir comment le Théorème 2.1.1 s'applique au problème 2. On commence par rappeler que l'hypothèse de dimension finie est essentielle. Quand elle n'est pas vérifiée, il est facile de construire des contre-exemples.

1. Si K est compact, la continuité de f permet de conclure directement.
2. Si K est fermé et si f est coercive, i.e. si f vérifie :

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

   alors le Théorème 2.1.1 s'applique.

Remarque 5 (Semi-continuité inférieure). Le Théorème 2.1.1 s'applique encore si on suppose seulement que f est semi-continue inférieurement sur K, i.e. si : $\forall x_0 \in K$, $\forall \alpha < f(x_0)$, $f^{-1}(]\alpha, +\infty[) \in \mathcal{V}(x_0)$.

On montre que f est sci sur K ssi :

$$\forall \alpha \in \mathbb{R}, \quad f^{-1}(]-\infty, \alpha]) \quad \text{est fermé dans } \mathbb{R}^n,$$

ou, de façon équivalente, ssi : $\forall x_0 \in K$, $\forall \varepsilon > 0$, il existe un voisinage V de x_0 t.q. :

$$\forall x \in V, \quad f(x_0) - \varepsilon \leq f(x),$$

en abrégé : $f(x_0) \leq \liminf_{x\to x_0} f(x)$.

Démonstration. Il suffit de remarquer que si $x_n \underset{n\to+\infty}{\to} x^*$ alors $f(x^*) \leq \liminf_{n\to+\infty} f(x_n)$. □

Exemple 7. Soit $I \subset \mathbb{R}$ un sous-ensemble quelconque et soit $(f_j)_{j\in I}$ une famille de fonctions linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$. On pose :

$$f(x) = \sup_{j\in I} f_j(x), \quad \forall x \in \mathbb{R}^n.$$

Alors f est semi-continue inférieurement. En effet : soit $\alpha \in \mathbb{R}$. Par définition de f : $f^{-1}(]\alpha, +\infty[) = \cup_{j\in I} f_j^{-1}(]\alpha, +\infty[)$ est ouvert comme réunion d'ouverts.

Exemple 8. Soit le problème : $\min_{(x,y)\in K} f(x, y)$ avec

$$f(x, y) = x^4 + y^4 - x^2, \quad K = \{(x, y) \in \mathbb{R}^2,\ x + y \leq 4\}.$$

On a :

$$\lim_{\|(x,y)\|_2\to+\infty} f(x, y) = \lim_{\|(x,y)\|_2\to+\infty} (x^4 + y^4) = +\infty$$

i.e. f est coercive. Comme de plus, f est continue et K est fermé, on déduit du Théorème 2.1.1 que le problème admet une solution.

Exemple 9. On considère une subdivision régulière de $[0, 1]$, soit :

$$u_0 = 0 < u_1 < \cdots < u_{N+1} = 1,$$

définie par $u_i = ih$, $0 \leq i \leq N + 1$, $h = \frac{1}{N+1}$, où $N \geq 1$ est donné. Soit $(u_i, x_i)_{1\leq i\leq N}$ un nuage de points de $\mathbb{R}^2$. On pose $x_0 = 0$, $x_{N+1} = 1$. On pose $x = (x_i)_{1\leq i\leq N} \in \mathbb{R}^N$ et on note $f(x)$ la longueur de la courbe affine par morceaux passant par les points (u_i, x_i), $i \in [[0, N + 1]]$. On montre aisément que

$$f(x) = \sum_{i=0}^{N} \sqrt{(u_{i+i} - u_i)^2 + (x_{i+1} - x_i)^2} = h \sum_{i=0}^{N} \sqrt{1 + \frac{(x_{i+1} - x_i)^2}{h^2}}.$$

On s'intéresse au problème

$$\inf_{x\in\mathbb{R}^N} f(x). \tag{3}$$

On remarque que : $\forall k \in [[1, N + 1]]$,

$$f(x) \geq h \sum_{i=0}^{k-1} \frac{|x_{i+1} - x_i|}{h} = \sum_{i=0}^{k-1} |x_{i+1} - x_i| \geq |\sum_{i=0}^{k-1} (x_{i+1} - x_i)| = |x_k|,$$

i.e. : $f(x) \geq \|x\|_\infty$. Donc f est coercive. Comme de plus $\mathbb{R}^N$ est fermé, on déduit du Théorème 2.1.1 que le problème (3) admet une solution.

2.2 Unicité de l'optimum

La convexité fournit une condition suffisante d'unicité dans les problèmes d'optimum. On commence par rappeler les définitions.

Définition 2.2.1 (Ensembles convexes et fonctions convexes)**.** Un ensemble $K \subset \mathbb{R}^n$ est dit convexe s'il vérifie

$$\forall (x,y) \in K^2, \quad \forall t \in [0,1], \quad tx + (1-t)y \in K$$

Une fonction $f : K \subset \mathbb{R}^n \to \mathbb{R}$ définie sur un ensemble convexe $K \subset \mathbb{R}^n$ est dite convexe si :

$$\forall (x,y) \in K^2, \quad \forall t \in [0,1], \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Une fonction convexe $f : K \subset \mathbb{R}^n \to \mathbb{R}$ est dite strictement convexe si

$$\forall (x,y) \in K^2, \quad x \neq y \Rightarrow \forall t \in]0,1[, \quad f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

On rappelle qu'en dimension finie, toute fonction convexe a une régularité minimale dans le sens suivant.

Proposition 2.2.1. *Si $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ est une fonction convexe définie sur un ouvert de $\mathbb{R}^n$, alors f est continue sur Ω et lipschitzienne sur tout compact de Ω.*

Démonstration. Voir par exemple [5] pour la démonstration dans $\mathbb{R}^n$, et [6] pour le cas $n = 1$. $\square$

Corollaire 2.2.2. *Une fonction convexe $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ définie sur un ouvert de $\mathbb{R}^n$ est différentiable presque partout (au sens de la mesure de Lebesgue) sur Ω.*

Démonstration. C'est une conséquence de la propriété de Lipschitz et du Théorème de Rademacher. $\square$

On rappelle un résultat classique très utile en pratique mais nécessitant de la régularité.

Théorème 2.2.3 (Caractérisation des fonctions convexes régulières)**.**

1. *Si $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable, les propositions suivantes sont équivalentes.*
   (i) f est convexe sur $\mathbb{R}^n$.
   (ii) $\forall x, y \in \mathbb{R}^n$, $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$.
   (iii) $\forall x, y \in \mathbb{R}^n$, $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$.

2. *Si $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable, les propositions suivantes sont équivalentes.*
(i) f est strictement convexe sur $\mathbb{R}^n$.
(ii) $\forall x, y \in \mathbb{R}^n$, $x \neq y \Rightarrow f(y) > f(x) + \langle \nabla f(x), y - x \rangle$.
(iii) $\forall x, y \in \mathbb{R}^n$, $x \neq y \Rightarrow \langle \nabla f(y) - \nabla f(x), y - x \rangle > 0$.
3. *Si $f : \mathbb{R}^n \to \mathbb{R}$ est deux fois différentiable, les propositions suivantes sont équivalentes.*
(i) f est convexe sur $\mathbb{R}^n$.
(ii) $\forall x \in \mathbb{R}^n$, la matrice hessienne $\nabla^2 f(x)$ est semi-définie positive.

Démonstration. 1. $(i) \Rightarrow (ii)$: On pose :

$$\forall t \in [0,1], \quad \varphi(t) = f((1-t)x + ty).$$

Par hypothèse sur f, φ est dérivable sur $[0,1]$, de dérivée donnée par

$$\forall t \in [0,1], \quad \varphi'(t) = \langle \nabla f((1-t)x + ty), y - x \rangle.$$

De plus, la définition de la convexité entraîne directement :

$$\forall t \in [0,1], \quad \varphi(t) = \varphi(t + (1-t)0) \leq t\varphi(1) + (1-t)\varphi(0). \tag{4}$$

On en déduit :

$$\forall t \in]0,1], \quad \varphi(1) \geq \varphi(0) + \frac{1}{t}(\varphi(t) - \varphi(0))$$

puis, quand $t \to 0^+$:

$$\varphi(1) \geq \varphi(0) + \varphi'(0) \tag{5}$$

i.e. (ii) :

$$f(y) \geq f(0) + \langle \nabla f(x), y - x \rangle.$$

$(ii) \Rightarrow (iii)$ Par hypothèse :

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \text{et} \quad f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$$

d'où, par linéarité du produit scalaire :

$$\langle \nabla f(x), y - x \rangle \leq f(y) - f(x) \leq \langle \nabla f(y), y - x \rangle$$

$$\Rightarrow \langle \nabla f(x), y - x \rangle \leq \langle \nabla f(y), y - x \rangle$$

i.e. (iii).

$(iii) \Rightarrow (i)$ Soit $t, s \in]0,1[$, $t \neq s$. Par hypothèse sur f, φ est continue sur $[0,1]$ et dérivable sur $]0,1[$. On a :

$$\langle \nabla f((1-t)x+ty) - \nabla f((1-s)x+sy), (t-s)(y-x)\langle \underset{(iii)}{\geq} 0$$

i.e. :

$$(t-s)(\varphi'(t) - \varphi'(s)) \geq 0.$$

Cela donne, en divisant par $|t-s|^2 > 0$:

$$\frac{\varphi'(t) - \varphi'(s)}{t-s} > 0$$

i.e. φ' est croissante sur $]0,1[$.
Soit $t \in]0,1[$. On a :

$$\begin{aligned} f((1-t)x+ty) - (1-t)f(x) - tf(y) &= \varphi(t) - (1-t)\varphi(0) - t\varphi(1) \\ &= (1-t)(\varphi(t) - \varphi(0)) + t(\varphi(t) - \varphi(1)) \end{aligned}$$

Du Théorème des valeurs intermédiaires on déduit qu'il existe $u \in]0,t[$ et $s \in]t,1[$ t.q. :

$$\varphi(t) - \varphi(0) = t\varphi'(u), \quad \varphi(t) - \varphi(1) = (t-1)\varphi'(s).$$

On en déduit :

$$\begin{aligned} f((1-t)x+ty) - (1-t)f(x) - tf(y) &= \underbrace{(1-t)t}_{\geq 0}\underbrace{(\varphi'(u) - \varphi'(s))}_{\leq 0} \\ &\leq 0 \end{aligned}$$

2. $(i) \Rightarrow (ii)$. On suppose f strictement convexe. Soit $s, t \in]0,1[$ et soit $x, y \in \mathbb{R}^n$, $x \neq y$. On a

$$\varphi(t) < (1-t)\varphi(0) + t\varphi(1) \underset{t>0}{\Rightarrow} \frac{\varphi(t) - \varphi(0)}{t} < \varphi(1) - \varphi(0).$$

De même :

$$\varphi(st) < (1-s)\varphi(0) + s\varphi(t) \underset{s>0}{\Rightarrow} \frac{\varphi(st) - \varphi(0)}{st} < \frac{\varphi(t) - \varphi(0)}{t} < \varphi(1) - \varphi(0).$$

On en déduit, quand $s \to 0^+$:

$$\varphi'(0) \leq \frac{\varphi(t) - \varphi(0)}{t} < \varphi(1) - \varphi(0)$$

i.e. (ii).
$(ii) \Rightarrow (iii)$ Soit $x \neq y$. Par hypothèse :

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle \quad \text{et} \quad f(x) > f(y) + \langle \nabla f(y), x - y \rangle$$

i.e. :

$$\langle \nabla f(x), y - x \rangle < f(y) - f(x) < \langle \nabla f(y), y - x \rangle$$

donc

$$\langle \nabla f(x), y - x \rangle < \langle \nabla f(y), y - x \rangle$$

i.e. (iii).
$(iii) \Rightarrow (i)$ Soit $t \in]0, 1[$. Par le même raisonnement que dans 1., il existe $u \in]0, t[$ et $s \in]t, 1[$ t.q., en posant $w = (1 - u)x + uy$, $z = (1 - s)x + sy$:

$$(1-t)f(x)+tf(y) = f((1-t)x+ty)+t(1-t)\langle \nabla f(w)-\nabla f(z), x-y\rangle =$$

$$= f((1-t)x+ty)+\frac{t(1-t)}{s-u}\langle \nabla f(w)-\nabla f(z), w-z\rangle \underset{(iii)}{>} f((1-t)x+ty).$$

3. $(i) \Rightarrow (ii)$ D'après le Théorème de Schwartz, f étant deux fois différentiable sur $\mathbb{R}^n$, sa matrice hessienne est symétrique en tout point de $\mathbb{R}^n$. Par hypothèse, φ est deux fois dérivable, de dérivée seconde donnée par : $\forall t \in [0, 1]$,

$$\varphi''(t) = \langle \nabla^2 f((1-t)x+ty)(y-x), y-x\rangle \Rightarrow \langle \nabla^2 f(x)(y-x), y-x\rangle = \varphi''(0)$$

avec

$$\varphi''(0) = \lim_{t \to 0^+} \frac{1}{t}(\varphi'(t) - \varphi'(0)).$$

On remarque que, par hypothèse :

$$\forall t \in]0, 1], \quad \frac{1}{t}(\varphi'(t)-\varphi'(0)) = \frac{1}{t}\langle \nabla f((1-t)x+ty)-\nabla f(x), y-x\rangle \geq 0$$

donc, en passant à la limite quand $t \to 0^+$: $\varphi''(0) \geq 0$, i.e. (ii).
$(ii) \Rightarrow (i)$ Soit $x, y \in \mathbb{R}^n$. Par hypothèse,

$$\forall t \in [0, 1], \quad \varphi''(t) = \langle \nabla^2 f((1 - t)x + ty)(y - x), y - x\rangle \geq 0,$$

donc φ' est croissante sur $[0, 1]$. On en déduit, φ étant de classe $\mathcal{C}^1$ sur $[0, 1]$:

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

i.e. :

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

De 1., on déduit que f est convexe, i.e. (i).

$\square$

Exemple 10 (Convexité d'une fonction quadratique)**.** Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique réelle d'ordre n, $b \in \mathbb{R}^n$ un vecteur et soit $c \in \mathbb{R}$. On considère la fonction :

$$f : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c$$

En utilisant la symétrie de A, on obtient directement :

$$\forall x, h \in \mathbb{R}^n, \quad f(x+h) = f(x) + \langle Ax - b, h\rangle + \frac{1}{2}\langle Ah, h\rangle$$

ce qui conduit à :

$$\frac{1}{\|h\|}\,|f(x+h) - f(x) - \langle Ax - b, h\rangle| = \frac{\langle Ah, h\rangle}{2\|h\|} \leq \frac{\|A\|}{2}\|h\| \underset{h\to 0}{\to} 0$$

i.e. :

$$\forall x, h \in \mathbb{R}^n, \quad df_x(h) = \langle Ax - b, h\rangle \Rightarrow \nabla f(x) = Ax - b.$$

Il en résulte que ∇f est différentiable sur $\mathbb{R}^n$, de différentielle définie par : $\nabla^2 f(x) = A$, $\forall x \in \mathbb{R}^n$.

En particulier, on déduit de ce calcul et du Théorème 2.2.3 que f est convexe, resp. strictement convexe, ssi A est semi-définie positive, resp. strictement définie positive.

Théorème 2.2.4. *On conidère le problème (2) avec f convexe et K convexe, éventuellement de dimension infinie. Alors*

1. *Tout mininum local est global.*
2. *Si f est strictement convexe alors il y a au plus un minimum.*

Démonstration. 1. On suppose que le problème (2) admet une solution x^* et que cette solution est un minimum local. Il existe $r > 0$ t.q.

$$\forall x \in K, \quad \|x - x^*\| < r \Rightarrow f(x^*) \leq f(x).$$

Soit $x \in K$, $\|x - x^*\| \geq r$. On pose

$$y = x^* + \frac{r}{2}\frac{x - x^*}{\|x - x^*\|}.$$

Par construction :
$$\|y - x^*\| = \frac{r}{2} < r.$$
De plus :
$$y = \left(1 - \frac{r}{2\|x - x^*\|}\right) x^* + \frac{r}{2\|x - x^*\|} x$$
avec
$$0 < \frac{r}{2\|x - x^*\|} \leq \frac{1}{2} < 1$$
donc $y \in K$ par convexité de K. On en déduit, par hypothèse sur x^* : $f(x^*) \leq f(y)$. De plus, par convexité de f sur K :
$$f(y) \leq \left(1 - \frac{r}{2\|x - x^*\|}\right) f(x^*) + \frac{r}{2\|x - x^*\|} f(x).$$
On en déduit :
$$f(x^*) \leq \left(1 - \frac{r}{2\|x - x^*\|}\right) f(x^*) + \frac{r}{2\|x - x^*\|} f(x) \iff$$
$$\iff \frac{r}{2\|x - x^*\|} f(x^*) \leq \frac{r}{2\|x - x^*\|} f(x) \underset{r>0}{\iff} f(x^*) \leq f(x).$$
Donc x^* est un minimum global sur K.

2. On suppose que f est strictement convexe sur K et qu'il y a deux minima $x_1^* \neq x_2^*$. De 1., on déduit que $f(x_1^*) = f(x_2^*) = \min_K f$. Soit $t \in]0,1[$. Par hypothèse : $x_1^* \neq x_2^* \Rightarrow$
$$f(tx_1^* + (1-t)x_2^*) < tf(x_1^*) + (1-t)f(x_2^*) = \min_K f$$
avec $tx_1^* + (1-t)x_2^* \in K$ par convexité de K. Contradiction.

$\square$

2.3 Existence en dimension infinie

Dans ce paragraphe, on énonce un résultat d'existence en dimension infinie dans le cas où f vérifie une hypothèse de convexité forte. Dans le cas général, il est beaucoup plus difficile d'établir un résultat d'existence en dimension infinie. A titre d'exemple, on peut considérer l'espace de Hilbert de dimension infinie défini par :
$$\ell^2(\mathbb{N}) = \{u = (u_n)_{n\geq 0}, \sum_{n\geq 0} |u_n|^2 < +\infty\}.$$

muni du produit scalaire : $(x, y) \mapsto \langle x, y\rangle = \sum_{n\geq 0} x_n y_n$. On considère la fonctionnelle f définie par :

$$\forall x \in \ell^2(\mathbb{N}), \quad f(x) = (\|x\|^2 - 1)^2 + \sum_{n\geq 0} \frac{x_n^2}{n+1}$$

et on s'intéresse au problème de minimisation :

$$\inf_{x\in\ell^2(\mathbb{N})} f(x).$$

On remarque immédiatement que f est coercive. En effet :

$$\forall x \in \ell^2(\mathbb{N}), \quad f(x) \geq (\|x\|^2 - 1)^2 \underset{\|x\|\to+\infty}{\to} +\infty.$$

Pourtant f n'admet pas de minimum sur $\ell^2(\mathbb{N})$. En effet : $f \geq 0$ et

$$f(x) = 0 \iff \|x\| = 1 \quad \text{et} \quad \sum_{n\geq 0} \frac{x_n^2}{n+1} = 0 \Rightarrow x_n = 0, \quad \forall n \geq 0$$

ce qui contredit $\|x\| = 1$, donc $f > 0$ sur $\ell^2(\mathbb{N})$.

Soit $x^{(n)} \in \ell^2(\mathbb{N})$ la série définie par :

$$x_k^{(n)} = \delta_{k,n} = \begin{cases} 1 & \text{si} \quad k = n, \\ 0 & \text{sinon} \end{cases}$$

On a :

$$\forall n \geq 0, \quad f(x^{(n)}) = \frac{1}{n+1} \underset{n\to+\infty}{\to} 0$$

i.e. $(x^{(n)})_{n\geq 0}$ est une suite minimisante de $\ell^2(\mathbb{N})$.

On remarque que la suite $(x^{(n)})_{n\geq 0}$ est bornée mais non compacte dans $\ell^2(\mathbb{N})$.

Dans ce qui suit, on étudie un cas d'existence en dimension infinie.

Définition 2.3.1 (Fonction α-ellitique)**.** Soit V un espace de Hilbert, $\langle\cdot,\cdot\rangle$ un produit scalaire sur V. Soit $K \subset V$ un convexe. Une fonction $f : K \to \mathbb{R}$ est dite fortement convexe ou uniformément convexe ou α-convexe ou α-elliptique s'il existe $\alpha > 0$ t.q. : $\forall (x, y) \in K^2$, $\forall t \in [0, 1]$,

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) - \frac{\alpha}{2} t(1 - t)\|x - y\|^2.$$

On vérifie immédiatement que l'uniforme convexité entraîne la stricte convexité et donc la convexité. La convexité correspond au cas $\alpha = 0$.

Exemple 11 (Liens entre les variantes de la convexité)**.** On donne des exemples élémentaires qui seront repris en détails dans la suite. En particulier, on étudiera la convexité des formes quadratiques en dimension finie.

1. Les fonctions affines de $\mathbb{R}$ dans $\mathbb{R}$ sont convexes mais non strictement convexes.
2. Une fonction α-elliptique est strictement convexe donc convexe.
3. La fonction $f : x \mapsto -\ln x$ est strictement convexe sur $]0,+\infty[$ mais non elliptique. En effet, f est de classe $\mathcal{C}^\infty$ sur $]0,+\infty[$ et on a $f''(x) = \frac{1}{x^2} > 0$, $\forall x > 0$. On raisonne par l'absurde en supposant qu'il existe $\alpha > 0$ t.q. $\forall t \in [0,1]$, $\forall x, y \in]0,+\infty[$,
$$-\ln(tx+(1-t)y) \leq -t\ln(x) - (1-t)\ln(y) - \frac{\alpha}{2}t(1-t)|x-y|^2.$$
Soit $t \in]0,1[$ et soit $y > 0$. Alors : $\forall x > y$,
$$-\frac{\ln(tx+(1-t)y)}{|x-y|^2} \leq \frac{-t\ln(x)-(1-t)\ln(y)}{|x-y|^2} - \frac{\alpha}{2}t(1-t).$$
avec
$$\frac{\ln(tx+(1-t)y)}{|x-y|^2} = \frac{\ln(x)+\ln(t)+\frac{(1-t)}{t}\frac{y}{x}+o\left(\frac{y}{x}\right)}{|x|^2\left(1-\frac{2y}{x}+o\left(\frac{y}{x}\right)\right)} \underset{x\to+\infty}{\sim} \frac{\ln(x)}{x^2}$$
donc
$$\lim_{x\to+\infty} -\frac{\ln(tx+(1-t)y)}{|x-y|^2} = 0.$$
De même :
$$\frac{-t\ln(x)-(1-t)\ln(y)}{|x-y|^2} = -\frac{t\ln(x)\left(1+\frac{(1-t)}{t}\frac{\ln(y)}{\ln(x)}+o\left(\frac{1}{\ln(x)}\right)\right)}{|x|^2\left(1-\frac{2y}{x}+o\left(\frac{y}{x}\right)\right)}$$
$$\underset{x\to+\infty}{\sim} -\frac{t\ln(x)}{x^2}$$
donc
$$\lim_{x\to+\infty} \frac{-t\ln(x)-(1-t)\ln(y)}{|x-y|^2} = 0.$$
On en déduit, quand $x \to +\infty$:
$$0 \leq -\frac{\alpha}{2}t(1-t)$$
en contradiction avec $-\frac{\alpha}{2}t(1-t) < 0$.

4. On vérifie aisément que $x \mapsto x^2$ est ellitique de rapport $\alpha = 2$. En effet :

$$(tx + (1-t)y)^2 = tx^2 + (1-t)^2y^2 - t(1-t)(x-y)^2.$$

La proposition ci-dessous précise les liens entre convexité et uniforme convexité.

Proposition 2.3.1. *Soit $(V, \langle \cdot, \cdot \rangle)$ un espace de Hilbert et soit $f : V \to \mathbb{R}$ une application.*

1. *La fonction f est α-elliptique ssi $f - \frac{\alpha}{2}\| \cdot \|^2$ est convexe.*
2. *On suppose f continue. Alors f est α-elliptique ssi*

$$\forall (x,y) \in V^2, \quad f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}(f(x) + f(y)) - \frac{\alpha}{8}\|x-y\|^2.$$

Démonstration. 1. $\Rightarrow$ On suppose f α-elliptique et on pose $\varphi = f - \frac{\alpha}{2}\|\|^2$. Alors : $\forall t \in [0,1]$, $\forall x, y \in V$,

$$\varphi(tx + (1-t)y) = f(tx + (1-t)y) - \frac{\alpha}{2}\|tx + (1-t)y\|^2$$

$$\leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2 - \frac{\alpha}{2}\|tx + (1-t)y\|^2$$

avec

$$\begin{aligned} t(1-t)\|x-y\|^2 + \|tx + (1-t)y\|^2 &= t(1-t)(\|x\|^2 + \|y\|^2 - 2\langle x, y \rangle) + \\ &+ t^2\|x\|^2 + (1-t)^2\|y\|^2 + 2t(1-t)\langle x, y \rangle \\ &= t(1-t)(\|x\|^2 + \|y\|^2) + t^2\|x\|^2 + (1-t)^2\|y\|^2 \\ &= t\|x\|^2 + (1-t)\|y\|^2 \end{aligned}$$

d'où on déduit :

$$\begin{aligned} \varphi(tx + (1-t)y) &\leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}(t\|x\|^2 + (1-t)\|y\|^2) \\ &= t\varphi(x) + (1-t)\varphi(y). \end{aligned}$$

$\Leftarrow$ Inversement, on suppose φ convexe. Alors : $\forall t \in [0,1]$, $\forall x, y \in V$,

$$f(tx + (1-t)y) = \varphi(tx + (1-t)y) + \frac{\alpha}{2}\|tx + (1-t)y\|^2$$

$$\leq t\varphi(x) + (1-t)\varphi(y) + \frac{\alpha}{2}(t^2\|x\|^2 + (1-t)^2\|y\|^2 + 2t(1-t)\langle x, y \rangle)$$

$$= tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2,$$

i.e. f est α-elliptique.

2. Il suffit de montrer que $f : V \to \mathbb{R}$ continue est convexe ssi

$$\forall (x,y) \in V^2, \quad f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y)$$

Pour cela on raisonne par récurrence sur n en montrant :

$$(\mathcal{P}_n) : \quad \forall (x,y) \in V^2, \quad \forall k \in \{0, \cdots, 2^n\},$$
$$f\left(\frac{k}{2^n}x + \left(1 - \frac{k}{2^n}\right)y\right) \leq \frac{k}{2^n}f(x) + \left(1 - \frac{k}{2^n}\right)f(y).$$

On voit médiatement que $(\mathcal{P}_0)$ est vraie. On suppose $(\mathcal{P}_n)$ vraie. Soit alors $k \in \{0, \cdots, 2^{n+1}\}$.
Si $2^n < k \leq 2^{n+1}$, alors :

$$f\left(\frac{k}{2^{n+1}}x + \left(1 - \frac{k}{2^{n+1}}\right)y\right) = f\left(\frac{k-2^n}{2^{n+1}}x + \left(1 - \frac{k-2^n}{2^{n+1}}\right)y + \frac{x-y}{2}\right)$$
$$= f\left(\frac{k-2^n}{2^n}\frac{x}{2} + \left(1 - \frac{k-2^n}{2^n}\right)\frac{y}{2} + \frac{x}{2}\right)$$

On en déduit, par hypothèse sur f :

$$f\left(\frac{k}{2^{n+1}}x + \left(1 - \frac{k}{2^{n+1}}\right)y\right) \leq \frac{1}{2}f\left(\frac{k-2^n}{2^n}x + \left(1 - \frac{k-2^n}{2^n}\right)y\right) + \frac{1}{2}f(x)$$
$$\underset{(\mathcal{P}_n)}{\leq} \frac{k-2^n}{2^{n+1}}f(x) + \frac{1}{2}\left(1 - \frac{k-2^n}{2^n}\right)f(y) + \frac{1}{2}f(x) =$$
$$= \frac{k}{2^{n+1}}f(x) + \left(1 - \frac{k}{2^{n+1}}\right)f(y)$$

On suppose que $k \in \{0, \cdots 2^n\}$. Alors :

$$f\left(\frac{k}{2^{n+1}}x + \left(1 - \frac{k}{2^{n+1}}\right)y\right) = f\left(\frac{k}{2^n}\frac{x}{2} + \left(1 - \frac{k}{2^n}\right)\frac{y}{2} + \frac{y}{2}\right)$$
$$\leq \frac{1}{2}f\left(\frac{k}{2^n}x + \left(1 - \frac{k}{2^n}\right)y\right) + \frac{1}{2}f(y)$$
$$\underset{(\mathcal{P}_n)}{\leq} \frac{k}{2^{n+1}}f(x) + \frac{1}{2}\left(1 - \frac{k}{2^n}\right)f(y) + \frac{1}{2}f(y)$$
$$= \frac{k}{2^{n+1}}f(x) + \left(1 - \frac{k}{2^{n+1}}\right)f(y)$$

Finalement dans tous les cas $(\mathcal{P}_{n+1})$ est vraie. Par récurrence sur $n \geq 0$, on en déduit que $(\mathcal{P}_n)$ est vraie $\forall n \geq 0$.
Soit $t \in [0,1]$ et soit $x, y \in V$. On a : $\forall n \geq 0$,

$$[0,1] = \cup_{k=0}^{2^n-1} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right].$$

Il existe donc une suite $(k_n)_{n\geq 0} \in \mathbb{N}^*$ t.q. :

$$\forall n \geq 0, \quad 0 \leq k_n \leq 2^n - 1 \quad \text{et} \quad \frac{k_n}{2^n} \leq t \leq \frac{k_n+1}{2^n} \Rightarrow t = \lim_{n\to+\infty} \frac{k_n}{2^n}.$$

Par continuité de f :

$$f(tx + (1-t)y) = \lim_{n\to+\infty} f\left(\frac{k_n}{2^n}x + \left(1 - \frac{k_n}{2^n}\right)y\right)$$

avec $\forall n \geq 0$,

$$f\left(\frac{k_n}{2^n}x + \left(1 - \frac{k_n}{2^n}\right)y\right) \leq \frac{k_n}{2^n}f(x) + \left(1 - \frac{k_n}{2^n}\right)f(y)$$

$$\underset{n\to+\infty}{\rightarrow} tf(x) + (1-t)f(y).$$

On en déduit, quand $n \to +\infty$:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Donc f est convexe sur V.
On suppose maintenant que f est α-elliptique. On en déduit le résultat en prenant $t = 1 - t = \frac{1}{2}$ dans la définition.
Inversement, on suppose que :

$$\forall (x,y) \in V^2, \quad f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y) - \frac{\alpha}{8}\|x-y\|^2.$$

On pose :

$$\forall x \in V, \quad \varphi(x) = f(x) - \frac{\alpha}{2}\|x\|^2.$$

Alors : $\forall (x,y) \in V^2$,

$$\varphi\left(\frac{x+y}{2}\right) = f\left(\frac{x+y}{2}\right) - \frac{\alpha}{8}\|x+y\|^2$$

$$\leq \frac{1}{2}f(x) + \frac{1}{2}f(y) - \frac{\alpha}{8}\|x-y\|^2 - \frac{\alpha}{8}\|x+y\|^2 =$$

$$= \frac{1}{2}f(x) + \frac{1}{2}f(y) - \frac{\alpha}{4}(\|x\|^2 + \|y\|^2) = \frac{1}{2}\varphi(x) + \frac{1}{2}\varphi(y).$$

Du résultat préliminaire, on déduit que φ est convexe sur V, i.e., d'après 1., que f est α-elliptique sur V.

□

Dans le cas où f est régulière, il existe des caractérisations de l'uniforme convexité qui peuvent être vues comme des conséquences du Théorème 2.2.3.

Corollaire 2.3.2 (Caractérisation des fonctions uniformément convexes)**.**

1. *Si $f : V \to \mathbb{R}$ est différentiable, alors les propositions suivantes sont équivalentes.*
   (i) f est α-elliptique.
   (ii) $\forall (x, y) \in V^2$, $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}\|x - y\|^2$.
   (iii) $\forall (x, y) \in V^2$, $\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq \frac{\alpha}{2}\|x - y\|^2$.
2. *Si $f : V \to \mathbb{R}$ est deux fois différentiable, alors les propositions suivantes sont équivalentes.*
   (i) f est α-elliptique.
   (ii) $\forall (x, h) \in V^2$, $\langle \nabla^2 f(x)h, h\rangle \geq \alpha\|h\|^2$.

Démonstration. 1. De la Proposition 2.3.1, on déduit que f est α-elliptique ssi φ est convexe ce qui, d'après le Théorème 2.2.3, se caractérise par les deux conditions équivalentes : $\forall (x, y) \in V^2$,

$$\varphi(y) \geq \varphi(x) + \langle \nabla\varphi(x), y - x\rangle$$

$$\langle \nabla\varphi(y) - \nabla\varphi(x), y - x\rangle \geq 0$$

où

$$\forall x \in V, \quad \nabla\varphi(x) = \nabla f(x) - \alpha x.$$

On conclut en remarquant que : $\forall (x, y) \in V^2$,

$$\varphi(y) \geq \varphi(x) + \langle \nabla\varphi(x), y - x\rangle$$

$$\iff f(y) - \frac{\alpha}{2}\|y\|^2 \geq f(x) - \frac{\alpha}{2}\|x\|^2 + \langle \nabla f(x) - \alpha x, y - x\rangle$$

$$\iff f(y) \geq f(x) + \frac{\alpha}{2}(\|y\|^2 - \|x\|^2 - 2\langle x, y - x\rangle) =$$

$$= f(x) + \frac{\alpha}{2}(\|x\|^2 + \|y\|^2 - 2\langle x, y\rangle) = f(x) + \frac{\alpha}{2}\|x - y\|^2$$

et que de même : $\forall (x, y) \in V^2$,

$$\langle \nabla\varphi(y) - \nabla\varphi(x), y - x\rangle \geq 0 \iff$$

$$\langle \nabla f(y) - \nabla f(x) - \alpha(y - x), y - x\rangle \geq 0 \iff \langle \nabla f(y) - \nabla f(x), y - x\rangle \geq \alpha\|x - y\|^2.$$

2. De même, en utilisant la Proposition 2.3.1 et le Théorème 2.2.3, f α-elliptique ssi

$$\forall (x,h) \in V^2, \quad \langle \nabla^2 \varphi(x) h, h\rangle \geq 0$$

où $\nabla^2\varphi(x) = \nabla^2 f(x) - \alpha I$, ce qui donne : $\forall (x,h) \in V^2$,

$$\langle \nabla^2 \varphi(x) h, h\rangle \geq 0 \iff \langle \nabla^2 f(x) h - \alpha h, h\rangle \geq 0 \iff \langle \nabla^2 f(x) h, h\rangle \geq \alpha \|h\|^2$$

□

Exemple 12 (Uniforme convexité d'une fonction quadratique)**.** Soit $A \in \mathcal{M}_n(\mathbb{R})$ symétrique d'ordre n, soit $b \in \mathbb{R}^n$ et soit $c \in \mathbb{R}$. Soit f la fonction quadratique définie par :

$$\forall x \in \mathbb{R}^n, \quad f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c.$$

On a vu dans l'Exemple 10 que f est convexe, resp. strictement convexe, ssi A est semi-definie positive, resp. définie positive et que de plus $\nabla^2 f(x) = A$, $\forall x \in \mathbb{R}^n$. La matrice A étant symétrique, elle est diagonalisable dans une base orthonormée de $\mathbb{R}^n$, i.e. A se décompose sous la forme : $A = U^T D U$ où $U \in \mathcal{O}_n(\mathbb{R})$ est orthogonale et où D est diagonale, formée par les valeurs propres $\lambda_1^2 \leq \cdots \leq \lambda_n^2$ de A, avec la convention $\lambda_i \geq 0$, $\forall i \in [[1, n]]$. On en déduit : $\forall h \in \mathbb{R}^n$,

$$\langle Ah, h\rangle = \langle U^T D U h, h\rangle = \langle DUh, Uh\rangle = \sum_{i=1}^{n} \lambda_i^2 (Uh)_i^2 \geq \lambda_1^2 \|Uh\|^2$$

$$\underset{U \in \mathcal{O}_n(\mathbb{R})}{=} \lambda_1^2 \|h\|^2$$

d'où on déduit que f est λ_1^2-elliptique dès que $\lambda_1 > 0$, i.e. dès que A est définie positive. Si $h \in \mathbb{R}^n \setminus \{0\}$ vérifie $Ah = \lambda_1^2 h$, alors $\langle Ah, h\rangle = \lambda_1^2 \|h\|^2$ et donc

$$\lambda_1^2 = \min_{h \neq 0} \frac{\langle Ah, h\rangle}{\|h\|^2}$$

i.e., λ_1^2 est la meilleure constante d'ellipticité de A.

Remarque 6. Si $f : V \to \mathbb{R}$ est α-elliptique et différentiable, alors f est coercive. En effet : soit $(x, y) \in V^2$. D'après le Corollaire 2.3.2 :

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}\|y - x\|^2 =$$

$$= f(x) + \|x-y\| \left(\left\langle \nabla f(x), \frac{y-x}{\|x-y\|} \right\rangle + \frac{\alpha}{2}\|y-x\| \right)$$

avec, pour $x \in \mathbb{R}^n$ fixé et $\|y\| \to +\infty$:

$$\|y-x\| = \|y\| \left(1 - \frac{\langle x, y\rangle}{\|y\|} + o\left(\frac{1}{\|y\|}\right)\right) \underset{\|y\|\to+\infty}{\sim} \|y\|$$

$$\left| \left\langle \nabla f(x), \frac{y-x}{\|y-x\|} \right\rangle \right| \leq \|\nabla f(x)\|$$

donc

$$f(y) \geq f(x) + \frac{\alpha}{2}\|y\|^2(1+o(1)) \underset{\|y\|\to+\infty}{\to} +\infty$$

i.e. : $\lim_{\|y\|\to+\infty} f(y) = +\infty$.

On est maintenant en mesure d'énoncer le résultat d'existence en dimension infinie annoncé dans l'introduction.

Théorème 2.3.3. *Soit K un convexe fermé non vide d'un espace de Hilbert V et soit f une fonction convexe continue sur K. Alors il existe un unique minimum $x^* \in K$ de f sur K et on a*

$$\forall x \in K, \quad \|x - x^*\|^2 \leq \frac{4}{\alpha}(f(x) - f(x^*)).$$

En particulier, toute suite minimisante de f sur K converge vers x^*.

Démonstration. La démonstration repose sur le résultat technique suivant :

Lemme 2.3.4. *Soit f une fonction α-convexe sur K. Alors, il existe deux constantes $\alpha_1 > 0$, $\alpha_2 \in \mathbb{R}$ t.q. :*

$$\forall x \in K, \quad f(x) \geq \alpha_1\|x\|^2 + \alpha_2.$$

Démonstration du Lemme. C'est une conséquence du Théorème de Hahn-Banach géométrique appliqué à l'épigraphe de f noté $\mathcal{E}(f)$ et défini par :

$$\mathcal{E}(f) = \{(\lambda, x) \in \mathbb{R} \times K,\ f(x) \leq \lambda\}.$$

En effet, f étant convexe et continue, $\mathcal{E}(f)$ est fermé comme image réciproque du fermé $[0{+}\infty[$ de $\mathbb{R}$ par l'application continue $(\lambda, x) \mapsto \lambda - f(x)$ et convexe par suite de la convexité de f. Soit $(\lambda_0, x_0) \in \mathbb{R} \times K$ t.q. $\lambda_0 < f(x_0)$. Alors $(\lambda_0, x_0) \notin \mathcal{E}(f)$ et d'après le Théorème de Hahn-Banach de séparation des convexes compacts et des convexes fermés, il existe une forme linéaire continue qui sépare strictement $\mathcal{E}(f)$ et le singleton $\{(\lambda_0, x_0)\}$ dans $\mathbb{R}\times V$ au sens suivant : il existe une forme linéaire continue $L \in V'$ et il existe $(\beta, \delta) \in \mathbb{R}^2$ t.q. :(6)

$$\forall (\lambda, x) \in \mathcal{E}(f), \quad \beta\lambda + L(x) > \delta > \beta\lambda_0 + L(x_0). \tag{6}$$

En considérant $\lambda \to +\infty$ dans (6), on voit que $\beta \geq 0$ et en prenant $x = x_0$ dans (6) on voit que $\beta > 0$. En remarquant que $(f(x), x) \in \mathcal{E}(f)$, $\forall x \in K$, on en déduit que

$$\forall x \in K, \quad \beta f(x) + L(x) > \delta.$$

De plus, l'uniforme convexité de f entraîne alors, compte tenu de la linéarité de L : $\forall (x, y) \in K^2$,

$$\frac{\beta}{2}(f(x) + f(y)) + \frac{1}{2}(L(x) + L(y)) - \frac{\alpha\beta}{8}\|x - y\|^2$$

$$\underset{\beta>0}{\geq} \beta f\left(\frac{x+y}{2}\right) + L\left(\frac{x+y}{2}\right) > \delta.$$

Comme L est continue, il en résulte :

$$\frac{\beta}{2}(f(x) + f(y)) + \frac{1}{2}\|L\|\|x\| + \frac{1}{2}L(y) > \delta + \frac{\alpha\beta}{8}\|x - y\|^2$$

$$= \delta + \frac{\alpha\beta}{8}(\|x\|^2 - 2\langle x, y\rangle + \|y\|^2) \geq \delta + \frac{\alpha\beta}{8}(\|x\|^2 - 2\|x\|\|y\| + \|y\|^2)$$

i.e. :

$$f(x) > \frac{\alpha}{4}\|x\|^2 - \left(\frac{\alpha}{4}\|y\| + \frac{\|L\|}{\beta}\right)\|x\| + C \geq \frac{\alpha}{8}\|x\|^2 + C'$$

où $C, C' \in \mathbb{R}$ sont des constantes indépendantes de x. □

Du Lemme 2.3.4, on déduit que f est coercive, donc, K étant fermé par hypothèse, que le problème $\inf_K f$ admet une solution dans K.

Soit $(x_n)_{n\geq 0}$ une suite minimisante. On a : $\forall n, m \geq 0$,

$$f\left(\frac{x_m + x_n}{2}\right) \leq \frac{1}{2}f(x_m) + \frac{1}{2}f(x_n) - \frac{\alpha}{8}\|x_m - x_n\|^2. \tag{7}$$

On en déduit, dans un premier temps :

$$\inf_K f \leq \liminf_{m,n\to+\infty} f\left(\frac{x_m + x_n}{2}\right) \leq \limsup_{m,n\to+\infty} f\left(\frac{x_m + x_n}{2}\right) \leq$$

$$\leq \lim_{m\to+\infty} \frac{1}{2}f(x_m) + \lim_{n\to+\infty} \frac{1}{2}f(x_n) = \inf_K f$$

i.e. :

$$f\left(\frac{x_m+x_n}{2}\right) \underset{m,n\to\infty}{\to} \inf_K f.$$

On obtient, en reportant dans (7) :

$$\inf_K f + \liminf_{m,n\to+\infty} \|x_m - x_n\|^2 \leq \inf_K f + \limsup_{m,n\to+\infty} \|x_m - x_n\|^2 \leq \inf_K f$$

$$\Rightarrow \liminf_{m,n\to+\infty} \|x_m - x_n\|^2 \leq \limsup_{m,n\to+\infty} \|x_m - x_n\|^2 \leq 0$$

donc :

$$\lim_{m,n\to+\infty} \|x_m - x_n\|^2 = 0$$

i.e. la suite $(x_n)_{n\geq 0}$ est de Cauchy dans V complet, donc convergente vers $x^* \in V$. Comme K est fermé, il en résulte que $x_n \underset{n\to+\infty}{\to} x^* \in K$, et par continuité de f : $f(x^*) = \inf_K f$.

On a d'autre part : $\forall n, p \geq 0$,

$$\frac{\alpha}{8}\|x_{n+p} - x_n\|^2 + f\left(\frac{x_{n+p}+x_n}{2}\right) \leq \frac{1}{2}f(x_{n+p}) + \frac{1}{2}f(x_n).$$

On en déduit, quand $p \to +\infty$: $\forall n \geq 0$,

$$\frac{\alpha}{8}\|x^* - x_n\|^2 + \inf_K f \leq \frac{1}{2}\inf_K f + \frac{1}{2}f(x_n) \iff$$

$$\iff \frac{\alpha}{4}\|x^* - x_n\|^2 \leq f(x_n) - \inf_K f$$

□

Remarque 7. Le Théorème 2.3.3 reste vrai si on suppose f semi-continue inférieure. En effet, le même raisoement montre que $x_n \underset{n\to+\infty}{\to} x^* \in K$ avec

$$\inf_K f \leq f(x^*) \leq \liminf_{n\to+\infty} f(x_n) = \lim_{n\to+\infty} f(x_n) = \inf_K f \Rightarrow f(x^*) = \inf_K f.$$

3 Conditions d'optimalité. Optimisation sans contrainte

Conformément au programme de l'Agrégation, on se limite au cas de la dimension finie, laissant en remarques des prolongements pour le cas de la dimension infinie.

Théorème 3.0.1 (Inéquation d'Euler)**.** *Soit $K \subset V$ un sous-ensemble convexe d'un espace de Hilbert V. Soit $f : K \to \mathbb{R}$ différentiable sur K. Si x est un minimum local de f sur K alors x est solution de l'inéquation d'Euler :*

$$\forall y \in K, \quad \langle \nabla f(x), y - x \rangle \geq 0.$$

Si de plus f est convexe alors x est un minimum global de f sur K.

Démonstration. Soit $y \in K$. Par hypothèse sur f, l'application

$$\varphi : [0,1] \to \mathbb{R}, \quad t \mapsto f((1-t)x + ty)$$

est dérivable sur $[0,1]$, de dérivée donnée par :

$$\forall t \in [0,1], \quad \varphi'(t) = \langle \nabla f((1-t)x + ty), y - x \rangle.$$

De plus φ admet un minimum local en $t = 0$, i.e. il existe $r \in]0,1[$ t.q. :

$$\forall t \in]0, r[, \quad \varphi(t) \geq \varphi(0) \Rightarrow \frac{1}{t}(\varphi(t) - \varphi(0)) \geq 0.$$

On en déduit :

$$\lim_{t \to 0} \frac{1}{t}(\varphi(t) - \varphi(0)) = \varphi'(0) \geq 0$$

i.e. $\langle \nabla f(x), y - x \rangle \geq 0$.

On suppose que de plus f est convexe. Alors, d'après le Théorème 2.2.3 :

$$\forall y \in K, \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \geq f(x)$$

i.e. x est un minimum global. $\square$

3.1 Conditions d'optimalité. Optimisation sans contrainte

Sot $f : \mathbb{R}^n \to \mathbb{R}$. On considère le problème :

$$\min_{x \in \mathbb{R}^n} f(x). \tag{8}$$

On se propose d'étudier la généralisation au cas de la dimension $n \geq 2$ de la condition suffisante classique en dimension 1.

Théorème 3.1.1 (Conditions nécessaires)**.** *Soit $x^* \in \mathbb{R}^n$ solution du problème (8).*

1. *Si f est différentiable en x^*, alors $\nabla f(x^*) = 0$ et x^* est appelé point stationnaire ou critique.*
2. *Si f est deux fois différentiable en x^*, alors la matrice $\nabla^2 f(x^*)$ est semi-définie positive.*

Démonstration. 1. On suppose f différentiable en x^* solution de (8). Soit $x \in \mathbb{R}^n$. Alors l'application

$$\varphi : \mathbb{R} \to \mathbb{R}, \quad t \mapsto f((1-t)x^* + tx) \tag{9}$$

admet un minimum en $t = 0$ et est dérivable en $t = 0$, de dérivée définie par : $\varphi'(0) = \langle \nabla f(x^*), x - x^* \rangle$. On a :

$$\forall t \in]0, \infty[, \quad \frac{\varphi(t) - \varphi(0)}{t} \geq 0$$

donc

$$\varphi'(0) = \lim_{t \to 0^+} \frac{\varphi(t) - \varphi(0)}{t} \geq 0.$$

On a de même : On a :

$$\forall t \in] - \infty, 0[, \quad \frac{\varphi(t) - \varphi(0)}{t} \leq 0$$

donc

$$\varphi'(0) = \lim_{t \to 0^-} \frac{\varphi(t) - \varphi(0)}{t} \leq 0,$$

ce qui entraîne que $\varphi'(0) = 0$, i.e. $\langle \nabla f(x^*), x - x^* \rangle = 0$. Ceci étant vrai $\forall x \in \mathbb{R}^n$, il en résulte que $\nabla f(x^*) = 0$.

2. On suppose f deux fois différentiable en x^*. Soit $x \in \mathbb{R}^n$. Alors φ définie par (13) admet une dérivée seconde en $t = 0$ définie par :

$$\varphi''(0) = \langle \nabla^2 f(x^*)(x - x^*), x - x^* \rangle.$$

De plus :

$$\begin{aligned} \varphi(t) &= \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(0) + o(t^2) \\ &\underset{\varphi'(0)=0}{=} \varphi(0) + \frac{t^2}{2}\varphi''(0) + o(t^2) \geq \varphi(0) \end{aligned}$$

Si $\varphi''(0) \neq 0$, alors $\varphi(t) \sim \varphi(0) + \frac{t^2}{2}\varphi''(0)$ et $\varphi''(0) + o(1) > 0$, donc $\varphi''(0) > 0$. On en déduit que $\varphi''(0) \geq 0$ en général, i.e.

$$\langle \nabla^2 f(x^*)(x - x^*), x - x^* \rangle \geq 0.$$

Ceci étant vrai $\forall x \in \mathbb{R}^n$, il en résulte que $\nabla^2 f(x^*)$ est semi-définie positive.

$\square$

Remarque 8. 1. Si f est deux fois différentiable en x^*, la hessienne en x^* n'est pas en général définie positive. Par exemple, si $f(x) = x^4$, alors f atteint son minimum absolu sur $\mathbb{R}$ en $x = 0$ et $f''(0) = 0$.

2. Le Théorème 3.1.1 fournit des conditions nécessaires mais non suffisantes. En effet, si $f(x) = x^3$, alors $f'(0) = f''(0) = 0$ et f n'a pas de minimum sur $\mathbb{R}$.

Théorème 3.1.2 (Conditions suffisantes)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois différentiable en x^* t.q. $\nabla f(x^*) = 0$. On suppose en outre que :*
(i) soit $\nabla^2 f(x^*)$ est définie positive ;
(ii) soit f est deux fois différentiable dans un voisinage de x^*, de hessienne semi-définie positive dans ce voisinage.
Alors x^* est un minimum local pour f.

Démonstration. (i) Soit $x \in \mathbb{R}^n$, $x \neq x^*$, et soit

$$\varphi : \mathbb{R} \to \mathbb{R}, \quad \varphi(t) = f((1-t)x^* + tx).$$

Alors φ est deux fois dérivable en 0 et on a :

$$\varphi'(0) = \langle \nabla f(x^*), x - x^* \rangle = 0$$

$$\varphi''(0) = \langle \nabla^2 f(x^*)(x - x^*), x - x^* \rangle > 0$$

On en déduit :

$$\varphi(t) = \varphi(0) + \frac{t^2}{2}\varphi''(0) + o(t^2)$$

i.e. :

$$\frac{\varphi(t) - \varphi(0)}{t^2} \underset{t\to 0}{\sim} \frac{1}{2}\varphi''(0) > 0$$

Donc il existe $\eta > 0$ t.q.

$$\varphi(t) > \varphi(0), \quad \forall t \in]-\eta, \eta[=: I\eta$$

i.e. :

$$f((1-t)x^* + tx) \geq f(x^*), \quad \forall t \in I_\eta.$$

(ii) Par hypothèse, f est deux fois différentiablke dans un voisinage V_{x^*} de x^*. Soit $r > 0$ t.q. $B(x^*, r) \subset V_{x^*}$ et soit $x \in B(x^*, r) \setminus \{x^*\}$: $0 < \|x - x^*\| < r$. Avec les mêmes notations que dans (i), φ est deux fois différentiable dans un voisinage I_x de 0 défini par :

$$t \in I_x \iff |t|\|x - x^*\| < r \iff t \in \left] -\frac{r}{\|x - x^*\|}, \frac{r}{\|x - x^*\|} \right[=: I_x.$$

Soit $t \in I_x$. De la formule d'Euler-Mac Laurin, on déduit qu'il existe $c_t \in I_x$ t.q.

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(c_t) = \varphi(0) + \frac{t^2}{2}\varphi''(c_t) \geq \varphi(0).$$

$\square$

On obtient un résultat optimal dans le cas convexe.

Théorème 3.1.3 (Cas convexe. Condition nécessaire et suffisante.)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe et différentiable sur $\mathbb{R}^n$. Une CNS pour que $x^* \in \mathbb{R}^n$ soit un minimum local (et donc global) de f est que x^* soit un point critique de f i.e. vérifie :*

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

Démonstration. $\Rightarrow$ Soit $x^* \in \mathbb{R}^n$ un point critique de f sur $\mathbb{R}^n$. D'après la caractérisation des fonctions convexes du Théorème 2.2.3 :

$$\forall x \in \mathbb{R}^n, \quad f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*)$$

i.e., x^* est un minimum global de f sur $\mathbb{R}^n$.

$\Leftarrow$ Inversement on suppose que f admet un minimum local en x^*. Alors $\nabla f(x^*) = 0$ d'après le Théorème 3.1.1. $\square$

Dans la suite, on étudie en détails deux problèmes fondamentaux en mathématiques appliquées : la minimisation d'une fonctionnelle quadratique et la méthode des moindres carrés.

3.2 Minimisation d'une fonctionnelle quadratique. Optimisation sans contrainte

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée réelle symétrique, soit $b \in \mathbb{R}^n$ et soit $c \in \mathbb{R}$. On définit la fonctionnelle quadratique (somme d'une forme quadratique et d'une fonction affine)

$$f : \mathbb{R}^n \to R, \quad x \mapsto \frac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle + c$$

et on considère le problème de minimisation :

$$\min_{x \in \mathbb{R}^n} f(x).$$

Dans l'Exemple 10, on a montré que f est deux fois différentiable sur $\mathbb{R}^n$, de différentielle définie par le gradient : $\nabla f(x) = Ax - b$, de matrice hessienne constante $\nabla^2 f(x) = A$. En particulier, f est convexe ssi A est semi-définie positive.

On en déduit que si A est semi-définie positive, alors f admet un minimum local (et donc global) s'il existe $x^* \in \mathbb{R}^n$ solution de $Ax^* = b$, ce qui est réalisé ssi $b \in \mathrm{Im}(A)$ avec, en dimension finie : $\mathrm{Im}(A) = \mathrm{Ker}(A^T)^\perp = \mathrm{Ker}(A)^\perp$.

Dans le cas général où A est symétrique réelle, A et digonalisable dans une bon, i.e. il existe $U \in \mathcal{O}_n(\mathbb{R})$ orthogonale t.q. $A = U^T D U$ avec D diagonale formée des valeurs propres de A, soit :

$$\lambda_1 \leq \cdots \leq \lambda_n.$$

On distingue plusieurs cas suivant le signe de la plus petite valeur propre λ_1.

(i) Si $\lambda_1 < 0$, soit $u_1 \in \mathbb{R}^n$ t.q. $Au_1 = \lambda_1 u_1$ et $\|u_1\| = 1$. On a : $\forall t \in \mathbb{R}$,

$$f(tu_1) = \frac{t^2}{2}\lambda_1 - t\langle b, u_1\rangle + c = \frac{\lambda_1}{2}\left(t - \frac{\langle b, u_1\rangle}{\lambda_1}\right)^2 + \frac{\langle b, u_1\rangle^2}{2\lambda_1} + c$$

avec $\lim_{t\to+\infty}\left|t - \frac{\langle b,u_1\rangle}{\lambda_1}\right| = +\infty \Rightarrow$

$$\lim_{t\to+\infty} f(tu_1) = -\infty$$

donc $\inf_{\mathbb{R}^n} f = -\infty$ et le problème (8) n'a pas de solution.

(ii) Si $\lambda_1 = 0$ et $b \notin \mathrm{Ker}(A)^\perp$, alors il existe $u \in \mathrm{Ker}(A)$ t.q. $\langle b, u\rangle \neq 0$. On a : $\forall t \in \mathbb{R}$,

$$f(tu) = -t\langle b, u\rangle + c.$$

On peut supposer, quitte à remplacer u par $-u$, que $\langle b, u\rangle > 0$. Alors

$$\lim_{t\to+\infty} f(tu) = -\infty \Rightarrow \inf_{x\in\mathbb{R}^n} f(x) = -\infty$$

et le problème (8) n'a pas de solution. Autrement dit, A étant semi-définie positive, f est convexe non minorée sur $\mathbb{R}^n$. En particulier, l'équation $Ax = b$ n'a pas de solution i.e. f n'a pas de point critique, car $b \notin (\mathrm{Ker}(A)^\perp$.

(iii) Si $\lambda_1 = 0$ et $b \in (\mathrm{Ker}(A))^\perp$, alors on remarque que $(\mathrm{Ker}(A))^\perp = \mathrm{Im}(A^T) = \mathrm{Im}(A)$ i.e. l'équation des points critiques $Ax = b$ est bien posée et admet une infinité de solutions. Plus précisément, $b \in \mathrm{Ker}(A)^\perp = \mathrm{Im}(A) \Rightarrow \exists x^* \in \mathbb{R}^n$ t.q. $b = Ax^*$ et alors

$$Ax = b \iff A(x - x^*) = 0 \iff x \in x^* + \mathrm{Ker}(A).$$

Soit alors $x \in \mathbb{R}^n$ et soit $x = x' + x''$ la décomposition de x dans la somme directe $\mathbb{R}^n = \mathrm{Ker}(A) \oplus \mathrm{Ker}(A)^\perp$. On a

$$\langle Ax, x\rangle = \langle Ax'', x\rangle = \langle x'' Ax\rangle = \langle Ax'', x''\rangle$$

et

$$b \in \mathrm{Ker}(A)^\perp \Rightarrow \langle b, x\rangle = \langle b, x''\rangle.$$

Donc $f(x) = f(x'')$ et on en déduit que $\inf_{x\in\mathbb{R}^n} f(x) = \inf_{x\in(\mathrm{Ker}(A))^\perp} f(x)$. Or $\lambda_1 = 0 \Rightarrow \nabla^2 f(x) = A$ est semi-définie positive, $\forall x \in \mathbb{R}^n$, donc f est convexe et tout minimum local est un minimum global. De plus, l'équation $\nabla f(x) = 0 \iff Ax = b$ admet une unique solution x^* dans $(\mathrm{Ker}(A))^\perp$. On en déduit que les solutions du problème (8) sont les vecteurs de la forme $x^* + x$, $x \in \mathrm{Ker}(A)$ où x^* est l'unique solution de

$$Ax^* = b \quad \text{et} \quad x^* \in (\mathrm{Ker}(A))^\perp.$$

Le calcul montre directement que :

$$\inf_{x\in\mathbb{R}^n} f(x) = f(x^*) = -\frac{1}{2}\langle b, x^*\rangle + c.$$

(iv) Si $\lambda_1 > 0$, alors $\mathrm{Im}(A) = \mathbb{R}^n$ et l'équation $Ax = b$ admet une unique solution x^*. De plus A est définie positive donc f est (strictement) convexe et $x^* = A^{-1}b$ est l'unique solution du problème (8). On a

$$\inf_{xin\mathbb{R}n} f(x) = f(x^*) = -\frac{1}{2}\langle b, A^{-1}b\rangle + c.$$

3.3 La méthode des moindres carrés

On pourra se reporter à [8] pour des compléments sur cette méthode.

Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$ une matrice de taille $m \times n$ (m lignes et n colonnes) avec $m > n$. Soit $b \in \mathbb{R}^m$. On considère le problème :

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|^2.$$

Le problème à n inconnues et m équations $Ax = b$ est mal posé en général. Il revient à minimiser la distance entre b et $\mathrm{Im}(A) = \mathrm{Ker}(A^T)^\perp$, i.e., puisque $\mathrm{Ker}(A^T)^\perp$ est fermé en dimension finie, à chercher le projeté de b sur $\mathrm{Ker}(A^T)^\perp$.

Soit l'application :

$$f : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto f(x) = \frac{1}{2}\|Ax - b\|^2.$$

Un calcul direct montre que

$$\forall x \in \mathbb{R}^n, \quad f(x) = \frac{1}{2}\|Ax\|^2 - \langle Ax, b\rangle + \frac{1}{2}\|b\|^2$$

$$= \frac{1}{2}\langle A^T Ax, x\rangle - \langle A^T b, x\rangle + \frac{1}{2}\|b\|^2$$

et on est ainsi ramené au problème de la Section 3.2 avec $A^T A$ symétrique semi-définie positive ($\lambda_1 \geq 0$ avec les notations de la Section 3.2.)

Soit $A^T Ax = 0$. Alors :

$$\langle A^T Ax, x\rangle = \|Ax\|^2 = 0 \Rightarrow Ax = 0$$

donc $\mathrm{Ker}(A^T A) \subset \mathrm{Ker}(A)$. Comme il est immédiat que $\mathrm{Ker}(A) \subset \mathrm{Ker}(A^T A)$, on en déduit que $\mathrm{Ker}(A^T A) = \mathrm{Ker}(A)$. De plus :

$$A^T b \in \mathrm{Im}(A^T) = \mathrm{Ker}(A)^\perp = \mathrm{Ker}(A^T A)^\perp$$

D'après le Théorème du rang en dimension finie : $n = \dim(\mathrm{Ker}(A)) + \dim(\mathrm{Im}(A))$. Si $\dim(\mathrm{Im}(A)) < n$, alors $\mathrm{Ker}(A^T A) = \mathrm{Ker}(A) \neq \{0\}$ et l'ensemble des solutions du problème (8) est $x^* + \mathrm{Ker}(A)$, où x^* est l'unique solution de

$$A^T Ax^* = A^T b, \quad x^* \in \mathrm{Ker}(A)^T.$$

Si $\dim(\mathrm{Im}(A)) = n$, alors $\mathrm{Ker}(A^T A) = \mathrm{Ker}(A) = \{0\}$, i.e. $A^T A$ est carrée d'ordre n, injective donc inversible et le problème (8) admet pour unique solution la solution de $A^T Ax^* = A^T b$.

Remarque 9 (Pseudo-inverse). Si A est injective alors $A^T A$ est inversible et on peut définir la pseudo-inverse ou inverse généralisée de A en posant $A^\dagger := (A^T A)^{-1} A^T$.

Avec cette définition :

$$A^T Ax^* = A^T b \iff x^* = A^\dagger b.$$

On vérifie immédiatement que $A^\dagger$ est l'inverse à gauche de A.

Dans le cas général où A n'est pas nécessairement injective, $A^T A$ est symétrique donc se décompose sous la forme $A^T A = U^T \Sigma^2 U$ où $U \in \mathcal{O}_n(\mathbb{R})$ est orthogonale d'ordre n, Σ est diagonale, de valeurs propres $\sigma_i \geq 0$. On en déduit ([9], 1.2.7) : $(AU^T)^T AU^T = \Sigma^2$, i.e., en notant $c_j \in \mathbb{R}^m$ le $j$ème vecteur colonne de AU^T, $j \in [[1, n]]$, $c_i^T c_j = \sigma_j^2 \delta_{ij}$, $\forall i, j \in [[1, n]]$. Alors : $\|c_j\| = \sigma_j$, $\forall j \in [[1, n]]$. On pose

$$v_j = \sigma_j^{-1} c_j \quad \text{si} \quad \sigma_j \neq 0,$$

et on complète la famille (v_j) en une base orthonormée encore notée (v_j) de $\mathbb{R}^m$. Soit $V \in \mathcal{O}_m(\mathbb{R})$ la matrice orthogonale de vecteurs colonnes les v_j, $j \in [[1, m]]$ et soit $\Sigma' := \begin{pmatrix} \Sigma \\ 0 \end{pmatrix}$ de taille $m \times n$. On a : $\forall i \in [[1, m]]$, $\forall j \in [[1, n]]$,

$$(V\Sigma')_{ij} = \sum_{k=1}^{m} V_{ik}\Sigma'_{kj} = \sum_{k=1}^{n} V_{ik}\Sigma_{kj} = V_{ij}\sigma_j = \sigma_j(v_j)_i = (\sigma_j v_j)_i = (c_j)_i$$

i.e. : $V\Sigma' = AU^T$. Les éléments diagonaux de Σ sont appelés les valeurs propres singulières de A et $A = V\Sigma' U$ est la décomposition en valeurs propres singulières de A. Si A est injective, alors Σ est inversible et on a

$$A^\dagger = (U^T\Sigma^2 U)^{-1}U^T\Sigma'^T V^T = U^T\Sigma^{-2}UU^T\Sigma'^T V^T = U^T \begin{pmatrix} \Sigma^{-1} & 0 \end{pmatrix} V^T$$

Soit $A = QR$ la décomposition QR de A avec

$$R = \begin{pmatrix} R' \\ 0 \end{pmatrix} \quad \text{et } R' \text{ inversible si } A \text{ est injective}$$

On a $A^TA = R^TR = R'^TR' \Rightarrow$:

$$A^\dagger = (R^TR)^{-1}(QR)^T = R'^{-1}R'^{-T}R^TQ^T = \begin{pmatrix} R'^{-1} & 0 \end{pmatrix} Q^T$$

On vérifie directement que l'opération de pseudo-inversion est involutive et commute avec la transposition et la conjugaison.

Exemple 13 (Régression linéaire)**.** On pourra se reporter à [10], Chapitre 2, pour la régression linéaire sans contraintes et à [10], Chapitre 3, pour la régression linéaire avec contraintes.

On considère un nuage de points $(M_i)_{i\in[[1,m]]}$, $M_i = (t_i, x_i)$, $\forall i \in [[1, m]]$, résultant en pratique de mesures. Conformément à une modélisation théorique préalable, on peut s'attendre à ce qu'ils se concentrent autour d'une droite restant à déterminer. On est alors ramené au problème de minimisation :

$$\min_{(\alpha,\beta)\in\mathbb{R}^2} \underbrace{\sum_{i=1}^{m} |x_i - \alpha t_i - \beta|^2}_{f(\alpha,\beta)},$$

où $f(\alpha, \beta)$ est la somme des distances des points $(M_i)_{i\in[[1,m]]}$ à la droite du plan $x = \alpha t + \beta$. En posant :

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad A = \begin{pmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix}$$

on réécrit le problème sous la forme :

$$\min_{(\alpha,\beta)\in\mathbb{R}^2} \left\| x - A \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\|^2 \iff \min_{(\alpha,\beta)\in\mathbb{R}^2} \left\| A \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - x \right\|^2$$

où $\|\cdot\|$ désgne la norme de $\mathbb{R}^m$. On est donc ramené à résoudre un problème de moindres carrés de matrice $A = (\; t \quad u \;)$ où u désigne le vecteur de $\mathbb{R}^m$ de composantes $u_i = 1$, $i \in [[1, m]]$. De l'égalité :

$$A^T A = \left(\begin{array}{c} t^T \\ u^T \end{array} \right)(\; t \quad u \;) = \begin{pmatrix} \|t\|^2 & t^T u \\ t^T u & \|u\|^2 \end{pmatrix}$$

il résulte que

$$\mathrm{d}et(A^T A) = \|t\|^2\|u\|^2 - (t^T u)^2 \geq 0$$

avec

$$\mathrm{d}et(A^T A) = 0 \iff \|t\|\|u\| = |t^T u|$$

ce qui est réalisé ssi les vecteurs t et u sont colinéaires. On en déduit : si $t_1 = \cdots = t_n$, i.e. alors les points M_i sont alignés sur la droite $t = t_1$. Sinon, $\mathrm{d}et(A^T A) > 0$, alors $\min_{(\alpha,\beta)} f(\alpha, \beta) = f(\alpha^*, \beta^*)$ avec (α^*, β^*) unique solution de

$$A^T A \begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix} = A^T x$$

donné par les formules :

$$\alpha^* = \frac{m t^T x - (t^T u)(t^T x)}{m\|t\|^2 - (t^T u)^2}, \quad \beta^* = \frac{\|t\|^2 (u^T x) - (t^T u)(t^T x)}{m\|t\|^2 - (t^T u)^2}.$$

4 Conditions d'optimalité. Optimisation avec contraintes

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

$$\begin{array}{cc} \min & f(x) \\ \left\{ \begin{array}{l} x \in \mathbb{R}^n, \\ g(x) \leq 0, \\ h(x) = 0 \end{array} \right. & \end{array}$$

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 :

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

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 $\dim\mathrm{Vect}\{a_1, \cdots, a_p\} = p$, 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 :

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

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. $\square$

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 :

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

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

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

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

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

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

$$\left\{\begin{array}{l} Ax = \lambda x, \\ \|x\|^2 = 1 \end{array}\right.$$

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 $e_{n+1} \in \mathrm{Vect}\{(e_1, \cdots, e_n)\}^\perp$, $\|e_{n+1}\| = 1$. 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. : $Ae_{n+1} \in \mathrm{Vect}\{(e_1, \cdots, e_n)\}^\perp$. 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 :

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

On remarque que

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

On en déduit :

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

resp. :

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

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 : $\forall x \in (\mathbb{R}_+^*)^n$,

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

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

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

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

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 : $\forall x \in \mathbb{R}_+^n \setminus \{0\}$,

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

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 :

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

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

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

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

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 $d \in \mathrm{Vect}\{(\nabla h_1, \cdots, \nabla h_p, \nabla g_{j_1}, \cdots, \nabla g_{j_r})\}$ 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 :

$$\langle \nabla h_i(x), d \rangle = 0, \quad \forall i \in [[1, p]],$$

$$\langle \nabla g_j(x), d \rangle = -1, \quad \forall j \in I(x).$$

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

$$(\langle \nabla h, \nabla h \rangle)_{ij} = \langle \nabla h_i, \nabla h_j \rangle, \quad \forall i, j \in [[1, p]]$$

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

$$(\langle \nabla g, \nabla \rangle)_{ij} = \langle \nabla g_i, \nabla g_j \rangle, \quad \forall i, j \in I(x).$$

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

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

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

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 :

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

On remarque que :

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

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

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 $K = \{(x,y) \in \mathbb{R}^2,\ g(x,y) := 1 - (x^2 + y^2) \leq 0\}$ 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 $\mu \geq 0$ t.q. :

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

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

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

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

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

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.