COURS OPTIMISATION

Cours à l'ISFA, en M1SAF

Ionel Sorin CIUPERCA

Table des matières

Chapitre 1

Introduction

1.1 Motivation

Voir cours en amphi

1.2 Le problème général d'optimisation

On se donne :

1. Une fonction $J : I\!R^n \mapsto I\!R$ (fonction coût)
2. Un ensemble $U \subset I\!R^n$ (ensemble des contraintes)

On cherche à minimiser J sur U, c'est à dire, on cherche $x^* \in U$ tel que

$$J(x^*) = \min_{x \in U} J(x)$$

ou équivalent

$$J(x^*) \leq J(x), \quad \forall\, x \in U.$$

Chapitre 2

Quelques rappels de calcul différentiel, analyse convexe et extremum

2.1 Rappel calcul différentiel

2.1.1 Quelques Notations

1. Pour tout $n \in I\!N^*$, $I\!R^n$ désigne l'espace **euclidien** $I\!R \times I\!R \times \cdots I\!R$ ("produit n fois"). En général un vecteur $x \in I\!R^n$ sera noté $x = (x_1, x_2, \cdots x_n)^T$ (vecteur colonne).
2. On note $e_1, e_2, \cdots e_n$ les éléments de la **base canonique** de $I\!R^n$, où e_i est le vecteur de $I\!R^n$ donné par :

$$(e_i)_j = \delta_{ij} = \begin{cases} 0 & \text{si} \quad j \neq i \\ 1 & \text{si} \quad j = i \end{cases}, \qquad \forall\, i, j = 1, 2 \cdots n \tag{2.1}$$

   (symboles de **Kronecker**).
3. Pour tous $x, y \in I\!R^n$ on note par $< x, y > \in I\!R$ le **produit scalaire** de x et y, qui est donné par

$$< x, y >= \sum_{i=1}^{n} x_i\, y_i.$$

4. Pour tout $x \in I\!R^n$ on note par $\|x\| \geq 0$ la **norme euclidienne** de x, donnée par

$$\|x\| = \sqrt{< x, x >} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

   Pour tous $x \in I\!R^n$ et $r > 0$ on notera par $B(x, r)$ la **boule ouverte** du centre x et rayon r, donnée par

$$B(x, r) = \{y \in I\!R^n, \quad \|y - x\| < r\}.$$

5. Si $a, b \in I\!R^n$ on note $[a, b]$ le sous-ensemble de $I\!R^n$ donné par

$$[a, b] = \{a + t(b - a) \equiv (1 - t)a + tb,\ t \in [0, 1]\}.$$

L'ensemble $[a, b]$ est aussi appellé **le segment** reliant a à b.
Remarques :
· $[a, b] = [b, a]$ (Exo !)
· Si $a, b \in I\!R$ avec $a < b$ alors on retrouve le fait que $[a, b]$ désigne l'intervalle des nombres $x \in I\!R$ tels que $a \leq x \leq b$.

6. On a

$$< Bx,\ y > = < x, By > \quad \forall\, x \in I\!R^n,\ y \in I\!R^m,\ B \in \mathcal{M}_{m,n}(I\!R)$$

7. Rappellons aussi l'inégalité de Cauchy-Schwarz :

$$| < x\ ,\ y > | \leq \|x\| \cdot \|y\| \quad \forall\, x, y \in I\!R^n.$$

2.1.2 Rappel gradient et hessienne

Soit $\Omega \subset I\!R^n$ un **ouvert** et $f : \Omega \to I\!R$.

1. On dit que f est de classe C^m sur Ω ($f \in C^m(\Omega)$) si toutes les dérivées partielles jusqu'à l'ordre m existent et sont continues.
2. Pour tout $x \in \Omega$ et tout $i \in \{1, 2, \cdots, n\}$ on note (quand $\exists$)

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \mapsto 0} \frac{1}{t} [f(x + te_i) - f(x)].$$

(c'est la **dérivée partielle** de f en x de direction x_i.)

3. Pour tout $x \in \Omega$ on note (quand $\exists$)

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots \frac{\partial f}{\partial x_n} \right)^T \in I\!R^n, \quad \forall\, x \in \Omega$$

(le **gradient** de f en x).
On note aussi

$$J_f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots \frac{\partial f}{\partial x_n} \right) \in I\!R^n, \quad \forall\, x \in \Omega$$

(la **Jacobienne** de f en x). On a

$$\nabla f = (J_f)^T$$

4. Pour tous $x \in \Omega$ et $h \in I\!R^n$ on note (quand $\exists$)

$$\frac{\partial f}{\partial h}(x) = \lim_{t \mapsto 0} \frac{1}{t} [f(x + th) - f(x)] = g'(0).$$

(c'est la **dérivée directionnelle** de f en x de direction h) où on a noté $g(t) = f(x+th)$.
Remarques :
i) $\frac{\partial f}{\partial 0}(x) = 0$

ii) $\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x)$
Nous rappellons aussi la formule :

$$\frac{\partial f}{\partial h}(x) = < \nabla f(x), h >, \quad \forall\, x \in \Omega \quad \forall\, h \in I\!R^n.$$

5. Pour $x \in \Omega$ on note (quand $\exists$) $\nabla^2 f(x) =$ la matrice carrée $\in \mathcal{M}_n(I\!R)$ donnée par

$$\left(\nabla^2 f(x)\right)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \quad \forall\, i,j = 1,2,\cdots n.$$

($\nabla^2 f(x)$ s'appelle aussi **la matrice hessienne** de f en x).
Remarque : Si $f \in C^2(\Omega)$ alors $\nabla^2 f(x)$ est une matrice **symmétrique** $\forall x \in \Omega$ (c'est le Théorème de Schwarz).

Proposition 2.1. *(Gradient de la composée) Supposons qu'on deux ouverts $\Omega \subset I\!R^n$ et $U \subset I\!R$ et deux fonctions $f : \Omega \mapsto I\!R$ et $g : U \mapsto I\!R$ avec en plus $f(\Omega) \subset U$ (on peut alors définir $g \circ f : \Omega \mapsto I\!R$). Supposons que f, g sont de classe C^1. Alors $g \circ f$ est aussi de classe C^1 avec en plus*

$$\nabla(g \circ f)(x) = g'(f(x))\nabla f(x) \quad \forall\, x \in \Omega.$$

Preuve très facile !

Proposition 2.2. *(lien entre ∇ et ∇^2)*
a) *La i - ème ligne de $\nabla^2 f(x)$ est la Jacobienne du i - ème élément de ∇f.*
b) *On a*

$$\nabla^2 f(x)h = \nabla < \nabla f(x), h >, \quad \forall\, x \in \Omega,\ \forall\, h \in I\!R^n.$$

Démonstration. **a)** évidente
b) On a :

$$\frac{\partial}{\partial x_i} < \nabla f(x), h > = \frac{\partial}{\partial x_i}\left(\sum_{j=1}^n \frac{\partial f}{\partial x_j}(x) h_j\right) = \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i x_j}(x) h_j = \left(\nabla^2 f(x) h\right)_i.$$

□

Quelques exemples importantes :

1. Si $f : I\!R^n \to I\!R$ est une fonction **constante** alors $\nabla f = \nabla^2 f = 0$.

2. Soit $f : I\!R^n \to I\!R$ définie par

$$f(x) = < a\,,\, x > \quad \forall\, x \in I\!R^n,$$

où $a \in I\!R^n$ est un vecteur donné (c'est à dire, f est une fonction **linéaire**). Alors on calcule facilement : $\frac{\partial f}{\partial x_k} = a_k$, donc

$$\nabla f = a$$

(le gradient est constant).
Ceci nous donne

$$\nabla^2 f = 0.$$

3. Soit $f : I\!R^n \to I\!R$ donnée par

$$f(x) = < Ax\,,\, x > \quad \forall\, x \in I\!R^n,$$

où $A \in \mathcal{M}_n(I\!R)$ est un matrice carrée, réelle, de taille n (c'est à dire, f est la fonction **quadratique** associée à la matrice A). Alors pour un $p \in \{1, 2, \cdots n\}$ fixé, on peut écrire

$$f(x) = \sum_{i,j=1}^{n} A_{ij}x_i x_j = A_{pp}x_p^2 + \sum_{j=1, j\neq p}^{n} A_{pj}x_p x_j + \sum_{i=1, i\neq p}^{n} A_{ip}x_i x_p + \sum_{i,j=1, i\neq p, j\neq p}^{n} A_{ij}x_i x_j$$

ce qui nous donne

$$\frac{\partial f}{\partial x_p} = 2A_{pp}x_p + \sum_{j=1, j\neq p}^{n} A_{pj}x_j + \sum_{i=1, i\neq p}^{n} A_{ip}x_i = \sum_{j=1}^{n} A_{pj}x_j + \sum_{i=1}^{n} A_{ip}x_i = (Ax)_p + (A^T x)_p.$$

Nous avons donc obtenu :

$$\nabla f(x) = (A + A^T)x, \quad \forall\, x \in I\!R^n.$$

On peut aussi écrire

$$\frac{\partial f}{\partial x_i}(x) = \sum_{k=1}^{n} (A + A^T)_{ik}x_k, \quad \forall\, i = 1, \cdots, n.$$

On a alors immédiatement :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = (A + A^T)_{ij}, \quad \forall\, i, j = 1, \cdots, n.$$

c'est à dire

$$\nabla^2 f(x) = A + A^T, \quad \forall\, x \in I\!R^n$$

(donc la hessienne de f est constante).
Remarque : En particulier, si A est **symmétrique** (c'est à dire $A = A^T$) alors

$$\nabla\ < Ax\,,\, x > = 2Ax, \quad \forall\, x \in I\!R^n.$$

$$\nabla^2\ < Ax\,,\, x > = 2A, \quad \forall\, x \in I\!R^n.$$

2.1.3 Rappel formules de Taylor

Proposition 2.3. *(sans preuve)*
Soit $\Omega \subset \mathbb{R}^n$ ouvert, $f : \Omega \mapsto \mathbb{R}, a \in \Omega$ et $h \in \mathbb{R}^n$ tels que $[a, a+h] \subset \Omega$. Alors :

1. *Si $f \in C^1(\Omega)$ alors*
   i) $f(a+h) = f(a) + \int_0^1 < \nabla f(a+th), h > \, dt$
   (formule de Taylor à l'ordre 1 avec reste intégral).
   ii) $f(a+h) = f(a) + < \nabla f(a+\theta h), h > \quad avec \quad 0 < \theta < 1$
   (formule de Taylor - Maclaurin à l'ordre 1)
   iii) $f(a+h) = f(a) + < \nabla f(a), h > + o(\|h\|)$
   (formule de Taylor - Young à l'ordre 1)
2. *Si $f \in C^2(\Omega)$ alors*
   i) $f(a+h) = f(a) + < \nabla f(a), h > + \int_0^1 (1-t) < \nabla^2 f(a+th)h, h > \, dt$
   (formule de Taylor à l'ordre 2 avec reste intégral).
   ii) $f(a+h) = f(a) + < \nabla f(a), h > + \frac{1}{2} < \nabla^2 f(a+\theta h)h, h > \quad avec \quad 0 < \theta < 1$
   (formule de Taylor - Maclaurin à l'ordre 2)
   iii) $f(a+h) = f(a) + < \nabla f(a), h > + \frac{1}{2} < \nabla^2 f(a)h, h > + o(\|h\|^2)$
   (formule de Taylor - Young à l'ordre 2).

Remarque : Dans la proposition précédente la notation $o(\|h\|^k)$ pour $k \in \mathbb{N}^*$ signifie une expression qui tend vers 0 plus vite que $\|h\|^k$ (c'est à dire, si on la divise par $\|h\|^k$, le résultat tend vers 0 quand $\|h\|$ tend vers 0).

2.2 Convexité

2.2.1 Rappel fonctions convexes

Définition 2.4. Un ensemble $U \subset \mathbb{R}^n$ est dit **convexe** si $\forall x, y \in U$ on a $[x, y] \subset U$ (quelque soit deux points dans U, tout le segment qui les unit est dans U).

Définition 2.5. Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \to \mathbb{R}$ une fonction.

1. On dit que f est **convexe** sur U si
$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y), \quad \forall\, x, y \in U, \ \ \forall\, t \in [0,1]$$
2. On dit que f est **strictement convexe** sur U si
$$f(tx + (1-t)y) < tf(x) + (1-t)f(y), \quad \forall\, x, y \in U \text{ avec } x \neq y, \ \ \forall\, t \in]0,1[.$$
3. On dit que f est **concave** (respectivement **strictement concave**) si $-f$ est convexe (respectivement strictement convexe).

Remarque : Il est facile de voir que toute fonction strictement convexe est convexe, mais que la réciproque n'est pas vraie en général.
Par exemple une application affine $f(x) = Ax + b$ est convexe (et aussi concave) mais elle n'est pas strictement convexe (ni strictement concave).
On montre facilement le résultat utile suivant :

Proposition 2.6. *Soit $U \subset \mathbb{R}^n$ un ensemble convexe, $p \in \mathbb{N}^*$, $f_1, f_2, \cdots, f_p : U \to \mathbb{R}$ des fonctions convexes et $\gamma_1, \gamma_2, \cdots, \gamma_n$ des constantes strictement positives.*
Posons $f = \gamma_1 f_1 + \gamma_2 f_2 + \cdots \gamma_p f_p$. Alors on a :
a) *La fonction f est convexe (donc toute combinaison linéaire avec des coefficients strictement positifs de fonctions convexes est convexe).*
a) *Si au moins une des fonctions $f_1, \cdots, f_p$ est strictement convexe alors f est strictement convexe.*

Démonstration. Laissée en exercice ! □

Il est en général difficile de vérifier la convexité d'une fonction en utilisant uniquement la définition (essayez avec $f(x) = x^2$ ou avec $f(x) = x^4$!) La proposition suivant donne des critères de convexité plus faciles à utiliser pour montrer la convexité ou la convexité stricte d'une fonction.

Proposition 2.7. *Soit $\Omega \subset \mathbb{R}^n$ ouvert, $U \subset \Omega$ avec U convexe et $f : \Omega \mapsto \mathbb{R}$ une fonction. Alors on a :*
Partie I *(caractérisation de la convexité avec "∇").*
Supposons que f est de classe C^1. Alors

1. *f est convexe sur U si et seulement si :*

$$f(y) \geq f(x) + < \nabla f(x), y - x >, \quad \forall\, x, y \in U$$

2. *f est strictement convexe sur U si et seulement si :*

$$f(y) > f(x) + < \nabla f(x), y - x >, \quad \forall\, x, y \in U \text{ avec } x \neq y.$$

3. *f est convexe sur U si et seulement si ∇f est* **monotone sur** *U, c'est à dire*

$$< \nabla f(y) - \nabla f(x)\,,\, y - x > \geq 0 \quad \forall\, x, y \in U.$$

4. *Si ∇f est strictement monotone sur U (c'est à dire :*

$$< \nabla f(y) - \nabla f(x)\,,\, y - x > > 0 \quad \forall\, x, y \in U \quad \text{avec} \quad x \neq y \quad)$$

alors f est strictement convexe sur U.

Partie II *(caractérisation de la convexité avec "∇^2").*
Supposons que f est de classe C^2. Alors

1. *f est convexe sur U si et seulement si :*

$$< \nabla^2 f(x)(y-x), y-x > \geq 0, \quad \forall\, x, y \in U$$

2. *Si*

$$< \nabla^2 f(x)(y-x), y-x > > 0, \quad \forall\, x, y \in U \text{ avec } x \neq y$$

alors f est strictement convexe sur U.

Remarques :

1. Dans le cas particulier où $\Omega = U = I\!R^n$ alors les deux inégalités de la **Partie II** de la proposition précédente peuvent s'écrire :

$$< \nabla^2 f(x)h, h > \geq 0, \quad \forall\, x, h \in I\!R^n$$

et respectivement

$$< \nabla^2 f(x)h, h > > 0, \quad \forall\, x, h \in I\!R^n, \text{ avec } h \neq 0.$$

2. Dans le cas particulier $n = 1$ et Ω un intervalle ouvert dans $I\!R$, on a $\nabla^2 f(x) = f''(x)$, donc $< \nabla^2 f(x)(y-x), y-x >= f''(x)(y-x)^2$. Alors les deux inégalités de la **Partie II** de la proposition précédente peuvent s'écrire :

$$f''(x) \geq 0, \quad \forall\, x \in U$$

et respectivement

$$f''(x) > 0, \quad \forall\, x \in U.$$

On rétrouve un résultat bien connu !

Exemple : Soit $f : I\!R \to I\!R$ donnée par $f(x) = x^2$. Comme $\nabla f(x) = f'(x) = 2x$ on a $\forall\, x, y \in I\!R$:

$$f(y) - f(x) - < \nabla f(x), y-x >= y^2 - x^2 - 2x(y-x) = y^2 + x^2 - 2xy = (y-x)^2 > 0 \quad \text{si } x \neq y$$

et ceci montre la stricte convexité de cette fonction, en utilisant la **Partie I2)** de la proposition précédente.
On peut aussi utiliser la Partie I4) :

$$< \nabla f(y) - \nabla f(x),\ y - x > = 2(y-x)^2 > 0 \ \text{ si } \ y \neq x$$

donc f est strictement convexe.
C'est encore plus facile si on utilise la **Partie II** :

$$f''(x) = 2 > 0$$

donc f est strictement convexe.

2.2.2 Fonctions elliptiques, fonctions coercives

Définition 2.8. Soit $f : I\!R^n \to I\!R$. On dit que f est une fonction **elliptique** si elle est de classe C^1 et si $\exists \alpha > 0$ telle que

$$< \nabla f(x) - \nabla f(y)\,,\, x - y > \geq \alpha \|x - y\|^2, \quad \forall\, x, y \in I\!R^n.$$

Définition 2.9. Soit $\Omega \subset I\!R^n$ un ensemble **non borné** et $f : \Omega \to I\!R$. On dit que f est **coercive** sur Ω si on a

$$\lim_{x \in \Omega, \|x\| \mapsto +\infty} f(x) = +\infty.$$

Proposition 2.10. *Soit $f : I\!R^n \to I\!R$ une fonction elliptique. Alors elle est strictement convexe et coercive. Elle vérifie en plus l'inégalité :*

$$f(y) \geq f(x) + < \nabla f(x), y - x > + \frac{\alpha}{2}\|y - x\|^2, \quad \forall x, y \in I\!R^n. \tag{2.2}$$

Démonstration. Montrons d'abord l'inégalité (2.2).
On utilise la formule de Taylor à l'ordre 1 avec reste intégral. On a

$$f(y) = f(x) + \int_0^1 < \nabla f(x + t(y - x)), y - x > \, dt$$

ce qui nous donne

$$f(y) - f(x) - < \nabla f(x), y - x > = \int_0^1 < \nabla f(x + t(y - x)) - \nabla f(x)\,,\, t(y - x) > \frac{1}{t}\, dt.$$

En utilisant le fait que f est elliptique on déduit :

$$< \nabla f(x + t(y - x)) - \nabla f(x)\,,\, t(y - x) > \geq \alpha \|t(y - x)\|^2$$

ce qui nous donne (car $t > 0$) :

$$f(y) - f(x) - < \nabla f(x), y - x > \geq \alpha \|y - x\|^2 \int_0^1 t\, dt = \frac{\alpha}{2}\|y - x\|^2$$

ce qui prove (2.2).
Montrons la stricte convexité de f : de la définition de l'ellipticité on déduit :

$$< \nabla f(x) - \nabla f(y)\,,\, x - y > > 0, \quad \forall\, x, y \in I\!R^n \quad \text{avec} \quad x \neq y.$$

ce qui nous donne la convexité stricte de f, comme conséquence de la Partie I4 de la Proposition 2.7 (autre méthode : utiliser (2.2) et la Partie I2 de la Proposition 2.7).
Montrons la coercivité de f.
En prennant $x = 0$ en (2.2) on obtient

$$f(y) \geq f(0) + < \nabla f(0), y > + \frac{\alpha}{2}\|y\|^2. \tag{2.3}$$

Nous utilisons l'inégalité de Cauchy-Schwarz

$$| < \nabla f(0), y > | \leq \|\nabla f(0)\| \, \|y\|$$

qui nous donne

$$< \nabla f(0), y > \geq -\|\nabla f(0)\| \, \|y\|, \quad \forall \, y \in I\!R^n.$$

En utilisant cette dernière inégalité en (2.3) on déduit

$$f(y) \geq f(0) - \|\nabla f(0)\| \, \|y\| + \frac{\alpha}{2}\|y\|^2$$

ce qui nous donne

$$f(y) \mapsto +\infty \quad \text{si} \quad \|y\| \mapsto +\infty.$$

La preuve de la proposition est alors terminée.

□

Proposition 2.11. *(Caractérisation de l'ellipticité avec "∇^2")*
Soit f une fonction de classe C^2. Alors f est elliptique si et seulement si $\exists \beta > 0$ tel que

$$< \nabla^2 f(x)h, h > \geq \beta \|h\|^2, \quad \forall x, h \in I\!R^n. \tag{2.4}$$

Démonstration. **a)** Supposons que f est elliptique et montrons (2.4). Soit $h \in I\!R^n$ fixé et notons $g : I\!R^n \mapsto I\!R$ la fonction donnée par $g(x) = < \nabla f(x), h >, \ \forall \, x \in I\!R^n$. Nous avons en utilisant aussi la Proposition 2.2 :

$$< \nabla^2 f(x)h, h > = < \nabla g(x), h > = \frac{\partial g}{\partial h}(x) = \lim_{t \mapsto 0} \frac{< \nabla f(x + th), h > \, - \, < \nabla f(x), h >}{t}$$

En utilisant la bilinéarité du produit scalaire et ensuite le fait que f est elliptique, on obtient :

$$< \nabla^2 f(x)h, h > = \lim_{t \mapsto 0} \frac{< \nabla f(x + th) - \nabla f(x) \, , \, th >}{t^2} \geq \alpha \frac{\|th\|^2}{t^2} = \alpha \|h\|^2$$

ce qui nous donne (2.4) avec $\beta = \alpha$.
b) Supposons maintenant que (2.4) est satisfaite et montrons que f est elliptique. Soient $x, y \in I\!R^n$ fixées arbitraires, et considérons la fonction $g_1 : I\!R^n \mapsto I\!R$ donnée par $g_1(z) = < \nabla f(z) \, , \, x - y >, \ \forall \, z \in I\!R^n$. Alors

$$< \nabla f(x) - \nabla f(y) \, , \, x - y > = \, g_1(x) - g_1(y) = < \nabla g_1(y + \theta(x - y)) \, , \, x - y >$$

avec $\theta \in]0, 1[$ (on a utilisé l'une des formules de Taylor).
D'autre part, nous avons

$$\nabla g_1(z) = \nabla^2 f(z)(x - y)$$

et ceci nous permet de déduire, en utilisant aussi (2.4) :

$$< \nabla f(x) - \nabla f(y) \, , \, x - y > = < \nabla^2 f(y + \theta(x - y))(x - y) \, , \, x - y > \, \geq \beta \|x - y\|^2.$$

On a donc obtenu l'ellipticité de f avec $\alpha = \beta$. □

On a aussi le résultat suivant :

Proposition 2.12. *Soit $p \in \mathbb{N}^*$, $f_1, f_2, \cdots, f_p : \mathbb{R}^n \to \mathbb{R}$ des fonctions de classe C^1 et convexes et $\gamma_1, \gamma_2, \cdots, \gamma_n$ des constantes strictement positives.*
Si au moins une des fonctions $f_1, \cdots, f_p$ est elliptique alors $f \equiv \gamma_1 f_1 + \gamma_2 f_2 + \cdots \gamma_p f_p$ est elliptique.

Démonstration. Soit $i \in \{1, 2, \cdots p\}$ tel que f_i soit elliptique. Alors il existe $\alpha > 0$ tel que

$$< \nabla f_i(y) - \nabla f_i(x)\,,\, y - x > \geq \alpha \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n.$$

D'autre part, comme tous les f_k sont convexes on a

$$< \nabla f_k(y) - \nabla f_k(x)\,,\, y - x > \geq 0 \quad \forall x, y \in \mathbb{R}^n, \quad \forall k \neq i.$$

En multipliant la première inégalité par γ_i et les autres par γ_k et en sommant en k, on obtient immédiatement le résultat. □

2.2.3 Exemples des fonctions elliptiques

1. **Si $n = 1$:**
   Toute fonction $f : \mathbb{R} \to \mathbb{R}$ de classe C^2 et satisfaisant :

   $$\exists\, \alpha > 0, \quad f''(x) \geq \alpha, \quad \forall\, x \in \mathbb{R}$$

   est une fonction elliptique (c'est une conséquence immédiate de la Proposition 2.11).
   Exemples :
   i) $f(x) = ax^2 + bx + c$ avec $a > 0$
   ii) $f(x) = x^2 + \sin(x)$ (car $f''(x) = 2 - \sin(x) \geq 1$, $\forall\, x \in \mathbb{R}$).
2. **Le cas général ($n \in \mathbb{N}^*$)**
   Soit $f : \mathbb{R}^n \to \mathbb{R}$ donnée par

   $$f(x) = \frac{1}{2} < Ax\,,\, x > - < b, x > + c, \quad \forall\, x \in \mathbb{R}^n \tag{2.5}$$

   avec $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée réelle et **symmétrique** de taille n, avec $b \in \mathbb{R}^n$ un vecteur et $c \in \mathbb{R}$ un scalaire *(on appelle encore, par abus de language, fonction (ou forme)* **quadratique** *une fonction de ce type)*. On calcule facilement :

   $$\nabla f(x) = Ax - b$$

   $$\nabla^2 f(x) = A$$

   (donc la hessienne de f est constante).
   Comme A est symmétrique alors on sait que toutes les valeurs propres de A sont réelles et que

   $$< Ah\,,\, h > \geq \lambda_{min} \|h\|^2, \quad \forall\, h \in \mathbb{R}^n$$

où $\lambda_{min} \in I\!R$ est la plus petite valeur propre de A.
Rémarquons que l'inégalité précédente devient égalité si h est un vecteur propre associé à la valeur propre λ_{min}.
Rappellons aussi que A est une matrice **définie positive** si et seulement si $\lambda_{min} > 0$.
(*par définition une matrice carrée réelle $B \in \mathcal{M}_n(I\!R)$ est définie positive si*
$< Bx, x > > 0, \quad \forall\, x \in I\!R^n,\ x \neq 0$).
En utilisant la Proposition 2.11 on déduit que f est une fonction elliptique si et seulement si A est une matrice définie positive.
(Voir des exemples "concrètes" en TD).

2.3 Conditions nécéssaires de minimum

Définition 2.13. Soit $U \subset I\!R^n,\ u^* \in U$ et $f : U \to I\!R$.

1. On dit que u^* est un point de minumum **absolu** (ou **global**) de f sur U si
$$f(u) \geq f(u^*), \quad \forall\, u \in U.$$
2. On dit que u^* est un point de minumum **relatif** (ou **local**) de f sur U si $\exists V$ voisinage de u^* dans $I\!R^n$, tel que
$$f(u) \geq f(u^*), \quad \forall\, u \in U \cap V.$$
3. On dit que u^* est un point de maximum absolu (respectivement relatif) de f sur U si u^* est un point de minumum absolu (respectivement relatif) de $-f$ sur U.
4. On dit que u^* est un point d'**extremum** absolu (respectivement relatif) de f sur U si u^* est : soit un point de minumum absolu (respectivement relatif) de f sur U, soit un point de maximum absolu (respectivement relatif) de f sur U.

Dans toute la suite du cours on parlera uniquement de la minimisation d'une fonction f ; pour la maximisation, faire la minimisation de la fonction $-f$.
Remarques :
- Un point de minimum absolu est clairement un point de minimum relatif.
- Si on dit simplement *minimum* on comprends *minimum absolu.*

On va commencer par le lemme suivant :

Lemme 2.14. *Soit $U \subset I\!R^n,\ a \in I\!R^n$ et u^* un élément appartenant à* **l'intérieur** *de U ($u^* \in \overset{\circ}{U}$). Alors les deux assertions suivantes sont équivalentes :*

1.
$$< a\,,\, u - u^* > \ \geq 0, \quad \forall\, u \in U. \tag{2.6}$$

2.
$$a = 0. \tag{2.7}$$

Démonstration. L'implication (2.7) $\Rightarrow$ (2.6) est évidente. Pour montrer l'implication inverse, soit $w \in I\!R^n$ arbitraire, avec $w \neq 0$. Alors il existe $\theta_0 > 0$ telle que

$$u^* + \theta w \in U, \quad \forall\, \theta \in [-\theta_0, \theta_0]$$

(il suffit de prendre $\theta_0 = \frac{r}{2\|w\|}$ où $r > 0$ est tel que $B(u^*, r) \subset U$). Alors en prennant $u^* + \theta w$ à la place de u dans (2.6) on déduit

$$< a\,,\, \theta w > \;\geq 0, \quad \forall\, \theta \in [-\theta_0, \theta_0].$$

En prennant d'abord $\theta = \theta_0$ et ensuite $\theta = -\theta_0$ dans l'inégalité précédente, on déduit

$$< a\,,\, w > \;= 0 \quad \text{ et ceci quelque soit } \; w \in I\!R^n, w \neq 0. \tag{2.8}$$

Comme l'égalité précédente est évidemment valable aussi pour $w = 0$ alors elle est valable pour tout $w \in I\!R^n$. Ceci donne immédiatement (2.7) (il suffit de prendre $w = a$ dans (2.8)). □

On utilisera dans la suite la définition suivante :

Définition 2.15. Soit $U \subset I\!R^n$ un ensemble et $u^* \in U$. On dit que $w \in I\!R^n$ est une **direction admissible** pour u^* en U s'il existe $t_0 > 0$ tel que $u^* + tw \in U \quad \forall\, t \in [0, t_0]$.

Exemples :

1. Si $u^* \in \overset{\circ}{U}$ alors tout vecteur $w \in I\!R^n$ est une direction admissible pour u^* en U.
2. Si U est convexe alors pour tout $v \in U$ le vecteur $v - u^*$ est une direction admissible pour u^* en U. (*Preuve :* prendre $t_0 = 1$ dans la définition précédente).

Lemme 2.16. *Soit $\Omega \subset I\!R^n$ un ouvert, $U \subset \Omega$, $f : \Omega \to I\!R$ une fonction de classe C^1 et $u^* \in U$ un point de minimum relatif de f sur U. Soit $w \in I\!R^n$ une direction admissible pour u^* en U. Alors*

$$< \nabla f(u^*)\,,\, w > \geq 0.$$

Démonstration. Soit V un voisinage de u^* en $I\!R^n$ tel que

$$f(u) \geq f(u^*), \quad \forall u \in U \cap V. \tag{2.9}$$

Soit $t_1 > 0$ tel que

$$u^* + tw \in V, \quad \forall\, t \in [-t_1, t_1]$$

(pour $w \neq 0$ il suffit de prendre $t_1 = \frac{r}{2\|w\|}$ où $r > 0$ est tel que $B(u^*, r) \subset V$; pour $w = 0$ tout t_1 convient).
D'autre part on a

$$u^* + tw \in U, \quad \forall\, t \in [0, t_0]$$

où t_0 est comme dans la Définition 2.15.
On déduit alors de (2.9)

$$f(u^* + tw) - f(u^*) \geq 0, \quad \forall\, t \in [0, \min\{t_0, t_1\}].$$

En divisant par t et en passant à la limite pour $t \to 0,\ t > 0$, on obtient :

$$\frac{\partial f}{\partial w}(u^*) \geq 0$$

ce qui est exactement l'inégalité demandée. □

Nous pouvons alors énoncer :

Théorème 2.17. *Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $U \subset \Omega$ un ensemble convexe et $f : \Omega \to \mathbb{R}$ une fonction de classe C^1. Soit $u^* \in U$ un point de minimum relatif de f sur U. Alors*

1.

$$< \nabla f(u^*)\,,\, u - u^* > \ \geq 0, \quad \forall\, u \in U \tag{2.10}$$

(c'est **l'inéquation d'Euler***).*

2. Si en plus u^* est dans l'intérieur de U ($u^* \in \overset{\circ}{U}$) alors la condition (2.10) est équivalente au

$$\nabla f(u^*) = 0 \tag{2.11}$$

(c'est **l'équation d'Euler***).*

Démonstration. **1.** C'est une conséquence immédiate du Lemme 2.16 et du fait que $u - u^*$ est une direction admissible pour u^* en U (car U est convexe).
2. Cette partie est une conséquence immédiate du Lemme 2.14 avec $a = \nabla f(u^*)$.
□

Remarques :

1. Ces relations respectives ((2.10) et (2.11)) donnent seulement des conditions **nécéssaires** de minimum relatif, qui ne sont pas en général suffisantes.
2. Si U est un ensemble **ouvert** (par exemple si U est l'espace entier, $U = \mathbb{R}^n$) alors u^* est forcément dans l'intérieur de U et donc l'équation d'Euler (2.11) peut être utilisée comme condition nécessaire de minimum relatif.

Le résultat suivant nous donne aussi des conditions suffisantes de minimum :

Théorème 2.18. *Soit $\Omega \subset \mathbb{R}^n$ ouvert, $U \subset \Omega$ un ensemble convexe, $f : \Omega \to \mathbb{R}$ une fonction de classe C^1 et* **convexe** *et $u^* \in U$. Alors les trois assertions suivantes sont équivalentes :*

1. u^* est un point de minimum absolu de f sur U
2. u^* est un point de minimum relatif de f sur U
3. $< \nabla f(u^*)\,,\, u - u^* > \ \geq 0, \quad \forall\, u \in U$.

Démonstration. 1. $\Rightarrow$ 2. est évidente et 2. $\Rightarrow$ 3. est une conséquence immédiate du Théorème 2.17.
Montrons 3. $\Rightarrow$ 1. Grâce à la convexité de f et à la Partie I1) de la Proposition 2.7 nous avons

$$f(u) - f(u^*) \geq \, < \nabla f(u^*) \, , \, u - u^* > \quad \forall \, u \in U.$$

De 3. nous obtenons alors 1. □

Remarque : Dans le cas où $u^* \in \overset{\circ}{U}$ alors l'assertion 3. du Théorème 2.18 peut être remplacée (grâce au Lemme 2.14) par l'équation d'Euler :

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

2.4 Existence et unicité d'un point de minimum

Théorème 2.19. *(Existence)*
Soit $U \subset I\!R^n$ un ensemble non-vide et fermé et $f : U \to I\!R$ une fonction continue. On suppose
· Soit U est borné
· Soit U est non bornée et f est une fonction coercive sur U.
Alors il existe **au moins** *un point de minimum de f sur U (c'est à dire, $\exists u^* \in U$ tel que $f(u^*) \leq f(u), \quad \forall \, u \in U$)*

Démonstration. On distingue deux cas :
Cas 1) L'ensemble U est borné.
Alors comme U est aussi fermé, U est compact. Comme f est continue, le Théorème de Weierstrass nous assure que f est bornée sur U et elle atteint ses bornes. Donc il existe au moins un point de minimum absolu de f sur U.
Cas 2) L'ensemble U est non borné.
Soit $a \in U$ et considérons l'ensemble

$$E = \{x \in U, \ f(x) \leq f(a)\} \qquad \text{Remarque : } a \in E.$$

Il est facile de montrer :

1. E **est fermé**
   (car $E = f^{-1}(]-\infty, \ f(a)])$, donc E est l'image inverse d'un intervalle fermé par une fonction continue)
2. E **est borné**
   (supposons le contraire : alors il existe une suite $x_k \in E$ avec $\|x_k\| \to +\infty$ pour $k \to +\infty$. Comme f est coercive sur U, ceci entrainne $f(x_k) \mapsto +\infty$ ce qui est absurde, car $f(x_k) \leq f(a), \ \ \forall \, k \in I\!N.$)

On déduit alors que E est un ensemble compact dans $I\!R^n$. Du Théorème de Weierstrass, $\exists u^* \in E$ tel que

$$f(u^*) \leq f(u), \quad \forall \, u \in E.$$

Mais d'autre part, on a

$$f(u^*) < f(u), \quad \forall\, u \in U - E$$

(car $f(u^*) \leq f(a) < f(u), \quad \forall\, u \in U - E$).
Ceci prouve que u^* est un point de minumum absolu de f sur U, ce qui finit la preuve. □

Théorème 2.20. *(Unicité)*
Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \to \mathbb{R}$ une fonction strictement convexe. Alors il existe **au plus** *un point de minimum de f sur U.*

Démonstration. On va raisonner par absurd. Soient $u_1, u_2 \in U$ avec $u_1 \neq u_2$ deux points de minimum de f sur U. Nous avons donc :

$$f(u_1) = f(u_2) \leq f(u), \quad \forall\, u \in U. \tag{2.12}$$

Comme f est strictement convexe, on a

$$f\left(\frac{1}{2}u_1 + \frac{1}{2}u_2\right) < \frac{1}{2}f(u_1) + \frac{1}{2}f(u_2) = f(u_1)$$

(car $f(u_1) = f(u_2)$) et ceci contredit (2.12). □

Corollaire 2.21. *(Existence et unicité)*
Soit $U \subset \mathbb{R}^n$ un ensemble fermé et convexe et $f : \mathbb{R}^n \to \mathbb{R}$ une fonction elliptique. Alors **il existe** *un* **unique** *point de minimum de f sur U.*

Démonstration. **Existence :** Remarquons d'abord que f est continue, car elle est C^1. On a alors deux situations :

1. Si U est borné alors l'existence est immédiate du Thórème 2.19.
2. Si U est non borné alors comme conséquence de la Proposition 2.10 f est coercive sur $\mathbb{R}^n$ donc sur U ; alors l'existence résulte encore du Thórème 2.19.

Unicité : Toujours de la Proposition 2.10 on déduit que f est strictement convexe. Alors l'unicité est une conséquence immédiate du Thórème 2.20.

□

Chapitre 3

Optimisation sans contraintes

On considère ici le cas particulier où l'ensemble de contraintes U est l'espace entier $\mathbb{R}^n$. On va donc considérer le problème de minimisation :

$$\min_{u \in \mathbb{R}^n} J(u) \tag{3.1}$$

où $J : \mathbb{R}^n \mapsto \mathbb{R}$ est une fonction donnée.
C'est un problème de **minimisation sans contraintes**.
Le problème (3.1) peut s'écrire

$$\text{Trouver } u^* \in \mathbb{R}^n \quad \text{tel que} \quad J(u^*) \leq J(u), \quad \forall\, u \in \mathbb{R}^n. \tag{3.2}$$

On supposera que u^* existe (éventuellement qu'il est unique) et on se propose de trouver une **approximation numérique** de u^*, en construisant une suite $\{u^{(k)}\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ telle que

$$u^{(k)} \mapsto u^* \quad \text{pour} \quad k \mapsto +\infty.$$

Remarque : Si $J \in C^1$ alors u^* satisfait nécessairement l'équation d'Euler

$$\nabla J(u^*) = 0$$

(vu en Chapitre 2).
Rappel : Si en plus J est convexe alors l'équation d'Euler est aussi une condition **suffisante** de minimum. Si en plus J est strictement convexe, alors u^* est unique.
Alors une méthode numérique qui peut être envisagée c'est de résoudre numériquement l'équation d'Euler, qui est en fait un système de n équations avec n inconnues, du type

$$\text{Trouver} \quad u \in \mathbb{R}^n \quad \text{tel que} \quad F(u) = 0$$

où $F : \mathbb{R}^n \to \mathbb{R}^n$ est une fonction donnée (ici $F = \nabla J$).
On peut utiliser la méthode numérique de **Newton-Raphson** (vue déjà en L3 en dimension 1 ?)
En particulier si J est donnée par

$$J(x) = \frac{1}{2} < Ax\,,\, x > \; - \; < b\,,\, x > \; + c$$

avec A matrice symmétrique, b vecteur et c scalaire (c'est à dire, J est quadratique) alors nous avons $\nabla J(u) = Au - b$, et l'équation d'Euler devient un **système algébrique linéaire**

$$Au = b.$$

On peut utiliser alors toute méthode numérique connue (vue en L3).

On se propose ici de donner d'autres méthodes numérique qui sont spécifiques à l'optimisation. Ces méthodes consitent à construire la suite approximante $\{u^{(k)}\}$ par **reccurence**, c'est à dire : on se donne un point initial $u^{(0)} \in I\!R^n$ et on construit $u^{(k+1)}$ comme une fonction de $u^{(k)}$. L'expression générale de $u^{(k+1)}$ sera

$$u^{(k+1)} = u^{(k)} + \rho_k d^{(k)} \tag{3.3}$$

avec $d^{(k)} \in I\!R^n$ des vecteurs qu'on appelle **directions de descente** et $\rho_k \in I\!R$ des scalaires qu'on appelle **facteurs de descente**. Ces noms viennent du fait qu'on cherchera toujours à avoir (entre autres)

$$J(u^{(k+1)}) < J(u^{(k)}), \quad \forall\, k \in I\!N \tag{3.4}$$

(la suite $\{J(u^{(k)})\}$ doit être strictement décroissante).

L'algorithme associé est en général de la forme :

pas 1. *Faire $k = 0$, choisir $u^{(0)} \in I\!R^n$ (par exemple $u^{(0)} = 0$) , choisir $k_{max} \in I\!N$ assez grand (par exemple $k_{max} = 1000$).*

pas 2. *Tant que ((“test arrêt” est faux) et ($k < k_{max}$)) faire $u^{(k+1)} = u^{(k)} + \rho_k d^{(k)}$ et $k = k + 1$*

pas 3. *Si (“test arrêt” est vrai) alors $u^{(k)}$ est une approximation du point de minimum recherché. Sinon, la méthode n'a pas convergé.*

Comme “test arrêt” on choisit le plus souvent : $\nabla J(u^{(k)} = 0$ (en pratique le test sera : $\|\nabla J(u^{(k)}\| \leq \epsilon$ où ϵ est un nombre strictement positiv et petit, qui représente un niveau de tolérance admis, à fixer au début (par exemple $\epsilon = 10^{-6}$)).

Il y a un grand nombre de méthodes numériques de minimisation, suivant le choix qui est fait pour $d^{(k)}$ et ρ_k.

3.1 Méthodes de relaxation

3.1.1 Description de la méthode

Cette méthode consiste à faire le choix suivant pour les directions de descente :

$$d^{(0)} = e_1, \quad d^{(1)} = e_2, \ \cdots \ d^{(n-1)} = e_n$$

ensuite on recommence

$$d^{(n)} = e_1, \quad d^{(n+1)} = e_2, \ \cdots \ d^{(2n-1)} = e_n$$

et ainsi de suite ...
(rappellons que $\{e_1, e_2, \cdots e_n\}$ sont les vecteurs de la base canonique de $\mathbb{R}^n$).
Donc en général on a :

$$d^{(k)} = e_l$$

si et seulement si l est le rest de la division de $k+1$ par n.

On prend ensuite les facteurs $\rho_k \in \mathbb{R}$ tels que

$$J(u^{(k)} + \rho_k d^{(k)}) = \min_{\rho \in \mathbb{R}} J(u^{(k)} + \rho d^{(k)}), \quad \forall\, k \in \mathbb{N} \tag{3.5}$$

(en supposant que ces minimum existent).
Finalement on pose

$$u^{(k+1)} = u^{(k)} + \rho_k d^{(k)}.$$

C'est la méthode (ou l'algorithme) de **relaxation**.

On peut écrire cet algorithme de la manière équivalente suivante :
On suppose connu le vecteur $u^{(k)} = (u_1^{(k)}, u_2^{(k)}, \cdots u_n^{(k)})^T$ et on calcule
$u^{(k+1)} = (u_1^{(k+1)}, u_2^{(k+1)}, \cdots u_n^{(k+1)})^T$ en n pas successifs par les formules suivantes :

$$J(u_1^{(k+1)}, u_2^{(k)}, \cdots u_n^{(k)}) = \min_{y \in \mathbb{R}} J(y, u_2^{(k)}, \cdots u_n^{(k)})$$

$$J(u_1^{(k+1)}, u_2^{(k+1)}, u_3^{(k)} \cdots u_n^{(k)}) = \min_{y \in \mathbb{R}} J(u_1^{(k+1)}, y, u_3^{(k)} \cdots u_n^{(k)})$$

$$\cdots$$

$$J(u_1^{(k+1)}, \cdots u_n^{(k+1)}) = \min_{y \in \mathbb{R}} J(u_1^{(k+1)}, \cdots u_{n-1}^{(k+1)}, y)$$

Ci-dessus on a utilisé le fait que (par exemple)

$$\min_{\rho \in \mathbb{R}} J(u_1^{(k)} + \rho, u_2^{(k)}, \cdots u_n^{(k)}) = \min_{y \in \mathbb{R}} J(y, u_2^{(k)}, \cdots u_n^{(k)}),$$

égalité obtenue en faisant le changement de variable $y = u_1^{(k)} + \rho$.

Le but dans la suite est de démontrer que l'algorithme de relaxation est bien défini, au moins dans le cas d'une fonction elliptique. Pour cela montrons d'abord le résultat suivant :

Proposition 3.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ et $a, b \in \mathbb{R}^n$. Définissons la fonction $g : \mathbb{R} \mapsto \mathbb{R}$, $g(t) = f(a + tb)$, $\forall\, t \in \mathbb{R}$. Alors on a :*

1. *Si f est convexe alors g est convexe.*
2. *Si f est strictement convexe et $b \neq 0$ alors g est strictement convexe.*
3. *Si f est elliptique et $b \neq 0$ alors g est elliptique.*

Démonstration. 1. Soient $t_1, t_1 \in I\!R,\ \theta \in [0,1]$. Alors

$$g(\theta t_1 + (1-\theta)t_2) = f(a + [\theta t_1 + (1-\theta)t_2]b) = f(\theta(a+t_1 b) + (1-\theta)(a+t_2 b))$$

(on a utilise l'égalité : $a = \theta a + (1-\theta)a$).
Avec la convexité de f on obtient

$$g(\theta t_1 + (1-\theta)t_2) \leq \theta f(a+t_1 b) + (1-\theta) f(a+t_2 b) = \theta g(t_1) + (1-\theta) g(t_2)$$

ce qui donne le résultat.

2. C'est pareil qu'en 1. avec $t_1 \neq t_2$ et $\theta \in]0,1[$. Comme $b \neq 0$ alors $a + t_1 b \neq a + t_2 b$ et donc on peut remplacer dans la démonstration précédente "$\leq$" par "$<$".
3. Soient $t_1,\ t_2 \in I\!R$. On a

$$[g'(t_1) - g'(t_2)]\,(t_1 - t_2) = [< \nabla f(a+t_1 b)\,,\, b > - < \nabla f(a+t_1 b)\,,\, b >]\,(t_1 - t_2) =$$

$$= < \nabla f(a+t_1 b) - < \nabla f(a+t_2 b)\,,\, (t_1 - t_2)b > \ \geq \alpha \|b\|^2 (t_1 - t_2)^2$$

ce qui montre le résultat, car $\alpha \|b\|^2 > 0$.

□

Corollaire 3.2. *Soit $f : I\!R^n \to I\!R$ et $a_1, a_2, \cdots a_{i-1}, a_{i+1}, \cdots a_n \in I\!R, i \in \{1, 2, \cdots n\}$ données. Soit $g : I\!R \mapsto I\!R$ définie par*

$$g(t) = f(a_1, a_2, \cdots a_{i-1}, t, a_{i+1}, \cdots a_n)$$

(fonction partielle par rapport à la i-ème variable).
Alors on a :

1. *Si f est convexe alors g est convexe*
2. *Si f est strictement convexe alors g est strictement convexe*
3. *Si f est elliptique alors g est elliptique.*

Démonstration. Il suffit d'appliquer la Proposition 3.1 avec $a = (a_1, a_2, \cdots a_{i-1}, 0, a_{i+1}, \cdots a_n)$, $b = e_i \neq 0$. □

Le résultat théorique principal de cette section est la suivant :

Théorème 3.3. *Soit $J : I\!R^n \to I\!R$ une fonction eliptique. Alors la méthode de relaxation est bien définie et elle converge (c'est à dire, pour tout $u^{(0)} \in I\!R^n$ la suite $u^{(k)}$ construit par cet algorithme converge vers l'unique point de minimum de J).*

Démonstration. Du Corollaire 3.2 on déduit que la fonction

$$y \in I\!R \to J(u_1^{(k+1)}, \cdots u_{i-1}^{(k+1)}, y, u_{i+1}^k, \cdots u_n^k) \in I\!R$$

est une fonction elliptique (car f est elliptique). On déduit alors qu'il existe un unique point de minimum de cette fonction, ce qui montre que l'algorithme de relaxation est bien défini. La convergence de $u^{(k)}$ est admise. □

Remarque : Cette méthode de relaxation est interessante si ces minimisations sur $I\!R$ peuvent ce faire de manière exacte.

3.1.2 Cas particulier des fonctions quadratiques

Dans ce paragraph nous supposons que $J : I\!R^n \to I\!R$ est donnée par

$$J(x) = \frac{1}{2} < Ax\,,\, x > - < b\,,\, x > + c \tag{3.6}$$

avec $A \in \mathcal{M}_n(I\!R)$ matrice SDP (c'est à dire matrice symmétrique et définie positive), $b \in I\!R^n$, $c \in I\!R$ (donc c'est une forme quadratique associée à une matrice SDP).
On a besoin de calculer $y^* \in I\!R$ tel que

$$\min_{y \in I\!R} g(y) = g(y^*)$$

où la fonction $g : I\!R \mapsto I\!R$ est définie par

$$g(y) = J(u_1^{(k+1)}, \cdots u_{i-1}^{(k+1)}, y, u_{i+1}^{(k)}, \cdots u_n^{(k)})$$

avec $u_1^{(k+1)}, \cdots u_{i-1}^{(k+1)}, u_{i+1}^{(k)}, \cdots u_n^{(k)} \in I\!R$ données.
Pour faciliter l'écriture, faisons les notations suivantes :

$$v_1 = u_1^{(k+1)},\ \ v_2 = u_2^{(k+1)}, \cdots v_{i-1} = u_{i-1}^{(k+1)},\ \ v_{i+1} = u_{i+1}^{(k)}\ , \cdots v_n = u_n^{(k)}$$

Comme J est elliptique alors g est aussi elliptique, donc y^* existe et est unique ; y^* est l'unique solution de l'équation d'Euler

$$g'(y^*) = 0.$$

Mais $g'(y)$ n'est autre que $\frac{\partial J}{\partial x_i}(v_1, v_2, \cdots v_{i-1}, y, v_{i+1}, \cdots v_n)$ ce qui nous donne

$$g'(y) = \sum_{j=1, j\neq i}^{n} A_{ij} v_j + A_{ii} y - b_i.$$

Donc on doit résoudre en $y^* \in I\!R$:

$$\sum_{j=1, j\neq i}^{n} A_{ij} v_j + A_{ii} y^* - b_i = 0$$

ce qui nous donne

$$y^* = \frac{1}{A_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{n} A_{ij} v_j \right)$$

Remarque : On a $A_{ii} > 0$ car $A_{ii} =< Ae_i\,,\, e_i > > 0$ (A est SDP).

On a donc montré :

Proposition 3.4. *Dans le cas où J est donnée par* (3.6) *(J quadratique avec A une matrice SDP) la méthode de relaxation s'écrit :*

$$u_1^{(k+1)} = \frac{1}{A_{11}} \left(b_1 - \sum_{j>1}^{n} A_{1j} u_j^{(k)} \right)$$

$$u_2^{(k+1)} = \frac{1}{A_{22}} \left(b_2 - A_{21} u_1^{(k+1)} - \sum_{j>2}^{n} A_{1j} u_j^{(k)} \right)$$

$$\cdots$$

$$u_i^{(k+1)} = \frac{1}{A_{ii}} \left(b_i - \sum_{j<i}^{n} A_{ij} u_j^{(k+1)} - \sum_{j>i}^{n} A_{ij} u_j^{(k)} \right)$$

$$\cdots$$

$$u_n^{(k+1)} = \frac{1}{A_{nn}} \left(b_n - \sum_{j<n}^{n} A_{nj} u_j^{(k+1)} \right)$$

En plus la méthode converge (c'est une conséquence du Théorème 3.3).

Remarque : L'algorithme ci-dessus n'est autre que l'algorithme de Gauss-Seidel (vu en L3) pour la résolution numérique du système linéaire

$$Ax = b.$$

Ceci s'explique par le fait que le point de minimum u^* de J satisfait l'équation d'Euler $\nabla J(u^*) = 0$ qui devient dans le cas J quadratique : $Au^* = b$.

3.2 Méthodes de gradient

Idée : prendre comme direction de descente :

$$d^{(k)} = -\nabla J(u^{(k)})$$

Justification : $J(u^{(k)})$ est orthogonal à la ligne du niveau de J dans le point $u^{(k)}$ et la fonction J diminue dans la direction $-\nabla J(u^{(k)})$. Cette afirmation se justifie par le calcul suivant, où on utilise un développement de Taylor :

$$J\left(u^{(k)} - \rho \nabla J(u^{(k)})\right) = J(u^{(k)}) + < \nabla J(u^{(k)}) \, , \, -\rho \nabla J(u^{(k)}) > + o(\rho)$$

(on suppose que J est de classe C^1). Nous avons alors :

$$J\left(u^{(k)} - \rho \nabla J(u^{(k)})\right) - J(u^{(k)}) = -\rho \|\nabla J(u^{(k)})\|^2 + o(\rho).$$

Le membre de droite de l'égalité précédente est < 0 si $\rho > 0$ avec ρ "assez petit" et $\nabla J(u^{(k)}) \neq 0$.

Les **méthodes de gradient** sont définis alors par les relations :

$$u^{(k+1)} = u^{(k)} - \rho_k \nabla J(u^{(k)}) \tag{3.7}$$

où les facteurs de descente $\rho_k \in I\!R$ sont à choisir.
Il y a plusieurs méthodes de gradient suivant le choix que nous faisons pour ρ_k.

3.2.1 Méthodes de gradient à pas optimal

La métode est la suivante : $u^{(k+1)}$ est donnée par (3.7) où $\rho_k \in I\!R$ est tel que

$$J\left(u^{(k)} - \rho_k \nabla J(u^{(k)})\right) = \min_{\rho \in I\!R} J\left(u^{(k)} - \rho \nabla J(u^{(k)})\right) \tag{3.8}$$

(en supposant qu'un tel minimum existe).
Remarque : Nous faisons l'itération (3.7) dans le cas où $\nabla J(u^{(k)}) \neq 0$, donc le problème de minimisation (3.8) a un sens.

On a le résultat suivant :

Théorème 3.5. *Soit $J : I\!R^n \mapsto I\!R$ une fonction elliptique. Alors la méthode de gradient à pas optimal donnée par* (3.7) *et* (3.8) *est bien définie et converge.*

Démonstration. Comme conséquence de la Proposition 3.1 la fonction

$$\rho \in I\!R \mapsto J\left(u^{(k)} - \rho \nabla J(u^{(k)}\right) \in I\!R$$

est une fonction elliptique, donc il existe un unique point de minimum de cette fonction. Ceci montre que la méthode de gradient à pas optimal est bien définie.
La convergence est admise ! □

Cas particulier : fonctions quadratiques.
On suppose ici J comme dans le paragraph 3.1.2 (donc J quadratique associée à une matrice SDP).
Nous devons calculer $\rho_k \in I\!R$ qui minimise la fonction $g : I\!R \mapsto I\!R$ donnée par

$$g(\rho) = J\left(u^{(k)} - \rho \nabla J(u^{(k)}\right).$$

Alors ρ_k satisfait nécessairement

$$g'(\rho_k) = 0.$$

Un calcul simple nous donne

$$g'(\rho) = - < \nabla J\left(u^{(k)} - \rho \nabla J(u^{(k)}\right) , \nabla J(u^{(k)}) >$$

c'est à dire, comme $\nabla J(u^{(k)}) = Au^{(k)} - b$

$$g'(\rho) = - < A\left(u^{(k)} - \rho \nabla J(u^{(k)})\right) - b , Au^{(k)} - b >= -\|Au^{(k)} - b\|^2 + \rho < A\left(Au^{(k)} - b\right) , Au^{(k)} - b >$$

On obtient alors

$$\rho_k = \frac{\|Au^{(k)} - b\|^2}{< A\left(Au^{(k)} - b\right) , Au^{(k)} - b >} \tag{3.9}$$

Rémarquons qu'on a $< A\left(Au^{(k)} - b\right) , Au^{(k)} - b > \ > 0$ (car A est SDP et $Au^{(k)} - b = \nabla J(u^{(k)}) \neq 0$).

Donc la méthode de gradient à pas optimal dans le cas J quadratique est :

$$u^{(k+1)} = u^{(k)} - \rho_k \left(Au^{(k)} - b\right)$$

avec ρ_k données par (3.9) (valable uniquement pour $Au^{(k)} - b \neq 0$).
Remarque : Comme $g'(\rho_k) = 0$, on déduit immédiatement

$$< \nabla J(u^{(k+1)}) \, , \, \nabla J(u^{(k)}) > = 0. \tag{3.10}$$

3.2.2 Autres méthodes du type gradient

Si J n'est pas quadratique, il peut être difficile de trouver ρ_k solution de (3.8). Quand on ne peut pas calculer ρ_k de manière exacte, il faut utiliser à chaque pas k une méthode numérique pour approcher ρ_k, ce qui alourdit les calculs. On peut se satisfaire uniquement d'une estimation "assez large" pour ρ_k. Le résultat suivant nous donne un intervalle tel que si tous les ρ_k se trouvent dans cet intervalle, alors la méthode du gradient converge :

Théorème 3.6. *Soit $J : \mathbb{R}^n \to \mathbb{R}$ une fonction elliptique (c'est à dire, $J \in C^1$ et $\exists \alpha > 0$ telle que*

$$< \nabla J(x) - \nabla J(y) \, , \, x - y > \; \geq \alpha \, \|x - y\|^2, \quad \forall \, x, y \in \mathbb{R}^n \quad)$$

et telle que ∇J soit lipschitzienne (c'est à dire : $\exists M > 0$ tel que

$$\|\nabla J(x) - \nabla J(y)\| \leq M \|x - y\|, \quad \forall \, x, y \in \mathbb{R}^n \quad).$$

Supposons que la suite $\{\rho_k\}_{k \in \mathbb{N}} \subset \mathbb{R}$ satisfait la propriété suivante : il existe a_1, a_2 avec $0 < a_1 \leq a_2 < \frac{2\alpha}{M^2}$ tels que

$$a_1 \leq \rho_k \leq a_2, \quad \forall \, k \in \mathbb{N}. \tag{3.11}$$

Alors la méthode générale de gradient (3.7) converge et la convergence est au moins géométrique, c'est à dire : $\exists \beta \in [0, 1[$ tel que

$$\|u^{(k)} - u^*\| \leq \beta^k \|u^{(0)} - u^*\|, \quad \forall \, k \in \mathbb{N}$$

où u^* est l'unique point de minimum de J sur $\mathbb{R}^n$.

Démonstration. Nous savons que u^* est l'unique solution de l'équation

$$\nabla J(u^*) = 0.$$

Notre but est de montrer que $u^{(k)} - u^* \to 0$ pour $k \to +\infty$. Nous avons :

$$u^{(k+1)} - u^* = u^{(k)} - \rho_k \nabla J(u^{(k)}) - u^* = u^{(k)} - u^* - \rho_k \left[\nabla J(u^{(k)}) - \nabla J(u^*)\right].$$

Utilisons la formule : $\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2 < x, y >$ pour tous $x, y \in \mathbb{R}^n$. Nous avons

$$\|u^{(k+1)} - u^*\|^2 = \|u^{(k)} - u^*\|^2 + \rho_k^2 \|\nabla J(u^{(k)}) - \nabla J(u^*)\|^2 - 2\rho_k < u^{(k)} - u^* \, , \, \nabla J(u^{(k)}) - \nabla J(u^*) > .$$

En utilisant l'ellipticité de J et le fait que ∇J est lipschitzienne, nous obtenons

$$\|u^{(k+1)} - u^*\|^2 \leq \|u^{(k)} - u^*\|^2 + M^2\rho_k^2\|u^{(k)} - u^*\|^2 - 2\rho_k\alpha\|u^{(k)} - u^*\|^2$$

c'est à dire

$$\|u^{(k+1)} - u^*\|^2 \leq (1 - 2\alpha\rho_k + M^2\rho_k^2)\|u^{(k)} - u^*\|^2. \tag{3.12}$$

Considérons maintenant la fonction $\varphi : \mathbb{R} \to \mathbb{R}$ donnée par

$$\varphi(\rho) = 1 - 2\alpha\rho + M^2\rho^2.$$

Le but est de montrer qu'il existe $\beta \in [0, 1[$ tel que

$$\varphi(\rho) \leq \beta^2 \quad \forall\, \rho \in [a_1, a_2]. \tag{3.13}$$

Comme $\varphi(0) = \varphi(\frac{2\alpha}{M^2}) = 1$ on voit facilement que

$$\varphi(\rho) < 1 \quad \forall\, \rho \in]0, \frac{2\alpha}{M^2}[. \tag{3.14}$$

Notons alors

$$\gamma_1 = \max_{\rho\in[a_1,a_2]} \varphi(\rho).$$

Le Théorème de Weierstrass nous dit que le maximum de φ est atteint sur $[a_1, a_2]$ et on déduit de (3.14) que

$$\gamma_1 < 1.$$

Notons maintenant

$$\gamma_2 = \max\{\gamma_1, 0\} \ \geq \ \gamma_1$$

et il est claire que $\gamma_2 \in [0, 1[$. On peut alors poser $\beta = \sqrt{\gamma_2}$ et on a

$$0 \leq \beta < 1.$$

Comme $\gamma_1 \leq \beta^2$ on déduit de la définition de γ_1 que (3.13) est satisfaite.
On obtient alors

$$\varphi(\rho_k) \leq \beta^2$$

ce qui avec (3.12) nous donne

$$\|u^{(k+1)} - u^*\| \leq \beta\|u^{(k)} - u^*\|, \quad \forall\, k \in \mathbb{N}.$$

Par reccurence, nous déduisons $\forall k \in \mathbb{N}^*$:

$$\|u^{(k)} - u^*\| \leq \beta\|u^{(k-1)} - u^*\| \leq \beta^2\|u^{(k-2)} - u^*\| \leq \cdots \leq \beta^k\|u^{(0)} - u^*\|$$

ce qui nous donne le résultat attendu. □

Définition :
Une méthode de gradient du type (3.7) est dite **à pas constant (ou fixe)** si ρ_k est constant (indépéndant du k). Dans le cas contraire, elle est dite **à pas variable**.
Remarque :
On déduit du Théorème 3.6 que si ρ est tel que

$$0 < \rho < \frac{2\alpha}{M^2}$$

alors la méthode de gradient à pas fixe

$$u^{(k+1)} = u^{(k)} - \rho \nabla J(u^{(k)})$$

est convergente (choisir $a_1 = a_2 = \rho$ dans le Théorème 3.6).
C'est la méthode de gradient la plus simple à utiliser.
Inconvenient : En général il est difficile de connaître α et M (en supposant qu'ils existent). Alors dans la pratique on prends un ρ assez petit pour être sur d'avoir la convergence. Mais dans ce cas la convergence peut être très lente !
Un exemple où on peut calculer α et M : Si J est quadratique avec une matrice A qui est SDP. Alors on peut prendre $\alpha = \lambda_{min} > 0$ (la plus petite valeur propre de A) et $M = \|A\|_2 = \rho(A) = \lambda_{max} > 0$ (la plus grande valeur propre de A).

Une alternative : une algorithme basé sur la recherche linéaire.
Cette méthode pour trouver ρ_k est une méthode intermédiaire entre l'algorithme de gradient à pas optimal et un algorithme de gradient à pas constant. Le principe en este le suivant :
On note $g : \mathbb{R} \to \mathbb{R}$ la fonction donnée par

$$g(\rho) = J(u^{(k)} - \rho \nabla J(u^{(k)})).$$

On cherche $r > 0$ et $s \in \mathbb{N},\ s \geq 2$, tels que

$$g(0) > g(r) > g(2r) > \cdots > g(sr)$$

et

$$g(sr) \leq g((s+1)r)$$

(pour tout $r > 0$ assez petit il existe au moins un nombre naturel s avec cette propriété, si on fait une hypothèse de coercivité sur J).
On pose alors

$$\rho_k = sr.$$

Idée intuitive : bien exploiter les possiblités de "descendre" dans la direction $-\nabla J(u^{(k)})$.
L'algorithme associé est le suivant :

faire $r = 1$
tant que $(g(2r) < g(r) < g(0)$ *est faux) faire* $r = r/2$
$\rho = 2r$
tant que $(g(\rho + r) < g(\rho))$ *faire* $\rho = \rho + r$
$\rho_k = \rho$.

3.3 La méthode des gradients conjugués

Rappels notations :

1. *Si* $v_1, v_2, \cdots v_l \in \mathbb{R}^n$ *on note par* $\mathcal{L}(v_1, v_2, \cdots v_l) = \{\sum_{i=1}^{l} \alpha_i v_i,\ \alpha_1, \cdots \alpha_l \in \mathbb{R}\}$ *l'espace vectoriel engendré par les vecteurs* $v_1, v_2, \cdots v_l$ *; c'est un sous-espace vectoriel de* $\mathbb{R}^n$.
2. *Si* $a \in \mathbb{R}^n$ *et* $M \subset \mathbb{R}^n$ *alors* $a + M$ *désigne l'ensemble* $\{a + x,\ x \in M\}$.

3.3.1 Le cas quadratique

Dans ce paragraphe on considère $J : \mathbb{R}^n \mapsto \mathbb{R}$ une fonction quadratique

$$J(u) = \frac{1}{2} < Au\,,\, u > - < b\,,\, u > + c, \quad \forall\, u \in \mathbb{R}^n$$

avec $A \in \mathcal{M}_n(\mathbb{R}),\ b \in \mathbb{R}^n,\ c \in \mathbb{R}$. On suppose que la matrice A est SDP.

Idée de la méthode : Rappellons que la méthode de gradient à pas optimal consiste à faire :

$$u^{(k+1)} = u^{(k)} - \rho_k \nabla J(u^{(k)}) = \min_{\rho \in \mathbb{R}} J\left(u^{(k)} - \rho \nabla J(u^{(k)})\right)$$

Ceci est équivalent avec :

$$u^{(k+1)} \in u^{(k)} + \mathcal{L}\left(\nabla J(u^{(k)})\right) \text{ est l'élément qui minimise } J \text{ sur } u^{(k)} + \mathcal{L}\left(\nabla J(u^{(k)})\right)$$

Dans la suite on va procéder de la manière suivante : On va noter pour tout $k \in \mathbb{N}$:

$$G_k = \mathcal{L}\left(\nabla J(u^{(0)}),\ \nabla J(u^{(1)}),\ \cdots \nabla J(u^{(k)})\right) \subset R^n$$

La méthode des **gradients conjugués** consiste à chercher $u^{(k+1)} \in u^{(k)} + G_k$ tel que

$$J(u^{(k+1)}) = \min_{v \in u^{(k)} + G_k} J(v) \tag{3.15}$$

(en supposant qu'un tel minimum existe).
On minimise donc sur un espace plus "grand" que dans la méthode de gradient à pas optimal. On s'attend alors à un "meilleur" minimum. Il reste à montrer que cette méthode est facile à implémenter et qu'elle donne le résultat attendu.

Comme J est elliptique et que $u^{(k)} + G_k$ est un ensemble fermé et convexe (car c'est un espace affine), alors il existe une solution unique du problème de minimisation (3.15).
En plus, le Théorème 2.17 nous donne

$$< \nabla J(u^{(k+1)})\,,\, v - u^{(k+1)} > \ \geq 0, \quad \forall\, v \in u^{(k)} + G_k. \tag{3.16}$$

Soit maintenant $w \in G_k$ arbitraire. Remarquons qu'on a :

$$u^{(k+1)} + w = u^{(k)} + \left[u^{(k+1)} - u^{(k)}\right] + w \in u^{(k)} + G_k \ \text{ car } \ u^{(k+1)} - u^{(k)} \in G_k.$$

On peut alors prendre $v = u^{(k+1)} + w$ en (3.16) et aussi $v = u^{(k+1)} - w$. On obtient

$$< \nabla J(u^{(k+1)}) \, , \, w > = 0, \quad \forall \, w \in G_k$$

(c'est à dire : $\nabla J(u^{(k+1)})$ est **orthogonal** sur tous les vecteurs de G_k). Ceci nous donne

$$< \nabla J(u^{(k+1)}) \, , \, \nabla J(u^{(l)}) > = 0, \quad \forall \, l = 0, 1, \cdots k. \tag{3.17}$$

(*rappellons que dans l'algorithme de gradient à pas optimal on avait seulement* $< \nabla J(u^{(k+1)}) \, , \, \nabla J(u^{(k)}) > = 0$.)

Conséquences :

1. Si les vecteurs $\nabla J(u^{(0)}), \nabla J(u^{(1)}), \cdots \nabla J(u^{(k)})$ sont tous $\neq 0$ alors ils sont indépendants (c'est une conséquence immédiate du résultat bien connu : si des vecteurs non-nuls sont orthogonaux par rapport à un produit scalaire, alors ils sont indépendants (facile à montrer)).
2. L'algorithme s'arrête en au plus n itérations, car il existe $k \in \{0, 1, \cdots n\}$ tel que

   $$\nabla J(u^{(k)}) = 0$$

   (sinon, on aurait $n+1$ vecteurs non-nuls indépendants en $\mathbb{R}^n$, ce qui est impossible). Alors $u^{(k)}$ est la solution rechérchée.

La question qui se pose maintenant est : **comment calculer $u^{(k+1)}$ à partir de $u^{(k)}$?** Supposons qu'on a

$$\nabla J(u^{(i)}) \neq 0, \quad \forall \, i = 0, 1, \cdots k$$

(sinon, l'algorithme s'arrêtait avant d'avoir à calculer $u^{(k+1)}$).
Nous avons l'expression suivante :

$$u^{(l+1)} = u^{(l)} + \Delta_l \quad \forall \, l = 0, 1, \cdots k. \tag{3.18}$$

avec $\Delta_l \in G_l$ que nous écrivons sous la forme

$$\Delta_l = \sum_{i=0}^{l} \alpha_i^l \, \nabla J(u^{(i)}) \tag{3.19}$$

avec $\alpha_i^l \in \mathbb{R}$ des coefficients à trouver. Rappellons que

$$\nabla J(x) = Ax - b \quad \forall \, x \in \mathbb{R}^n.$$

On utilisera souvent la formule suivante :

$$\nabla J\left(u^{(l+1)}\right) = \nabla J\left(u^{(l)}\right) + A\Delta_l. \tag{3.20}$$

(car $\nabla J(u^{(l+1)}) = Au^{(l+1)} - b = A(u^{(l)} + \Delta_l) - b = Au^{(l)} - b + A\Delta_l$).

Proposition 3.7. *On a*

1.

$$\Delta_l \neq 0, \quad \forall\, l = 0, 1, \cdots k.$$

2.

$$< A\Delta_l, \Delta_m >= 0 \quad si \;\; 0 \leq m < l \leq k.$$

Démonstration. En faisant le produit scalaire de l'égalité (3.20) par $\nabla J\left(u^{(l)}\right)$ on obtient

$$\|\nabla J(u^{(l)})\|^2 + < A\Delta_l\,,\, \nabla J(u^{(l)}) >= 0.$$

Comme $\|\nabla J(u^{(l)})\|^2 \neq 0$ on déduit $< A\Delta_l\,,\, \nabla J(u^{(l)}) > \neq 0$ donc $\Delta_l \neq 0$, ce qui finit la partie 1.
Pour montrer la partie 2. nous faisons le produit scalaire de (3.20) avec $\nabla J(u^{(i)}),\ i < l$. On obtient

$$< A\Delta_l\,,\, \nabla J(u^{(i)}) >= 0, \quad \forall\, i = 0, 1, \cdots l-1, \quad \text{donc} \;\; \forall\, i = 0, 1, \cdots m.$$

En multipliant par α_i^m et en sommant pour i de 0 à m on obtient l'égalité de la partie 2. □

Ceci nous amème à donner la définition suivante :

Définition 3.8. On dit que deux vecteurs $x, y \in I\!R^n$ sont **conjugués** par rapport à une matrice $B \in \mathcal{M}_n(I\!R)$ si

$$< Bx\,,\, y >= 0.$$

Remarque : Si B est une matrice SDP alors l'application

$$(x, y) \in I\!R^n \times I\!R^n \to < Bx\,,\, y > \in I\!R$$

est un produit scalaire en R^n (résultat admis !) qu'on appelle *produit scalaire associé à la matrice* B. (à noter que le produit scalaire habituel est un fait un produit scalaire associé à la matrice identité I_n). Alors la définition précédente nous dit que, dans le cas où B est SDP, deux vecteur sont conjugués par rapport à la matrice B s'ils sont orthogonaux par rapport au produit scalaire associé à B.
Comme les vecteurs $\Delta_0, \Delta_1, \cdots \Delta_k$ sont non-nuls et orthogonaux par rapport au produit scalaire associé à la matrice A (Proposition 3.7) on déduit immédiatement :

Proposition 3.9. *Les vecteurs* $\Delta_0, \Delta_1, \cdots \Delta_k$ *sont indépendants.*

Il est facile de voir qu'on a

$$\mathcal{L}(\Delta_0, \Delta_1, \cdots \Delta_l) = \mathcal{L}\left(\nabla J(u^{(0)}),\, \nabla J(u^{(1)}), \cdots \nabla J(u^{(l)})\right), \quad \forall\, l = 0, 1, \cdots k$$

(car l'inclusion "$\subset$" est évidente de (3.19) et en plus les deux espaces ont la même dimension : $l + 1$).
On déduit alors :

$$\alpha_l^l \neq 0, \quad \forall\, l = 0, 1, \cdots k \tag{3.21}$$

(car sinon, on aurait $\Delta_l \in \mathcal{L}\left(\nabla J(u^{(0)}), \nabla J(u^{(1)}), \cdots \nabla J(u^{(l-1)})\right) = \mathcal{L}(\Delta_0, \Delta_1, \cdots \Delta_{l-1})$, ce qui contredirait l'indépendance des $\Delta_0, \cdots, \Delta_l$).
On peut alors écrire

$$\Delta_k = \rho_k d^{(k)}$$

avec

$$d^{(k)} = \nabla J(u^{(k)}) + \sum_{i=0}^{k-1} \lambda_i^k \, \nabla J(u^{(i)}) \tag{3.22}$$

où on a posé

$$\begin{cases} \rho_k = \quad \alpha_k^k \\ \lambda_i^k = \quad \alpha_i^k / \alpha_k^k \end{cases} \tag{3.23}$$

(remarquons que dans (3.22) la somme n'existe pas si $k = 0$).
Nous avons donc :

$$u^{(k+1)} = u^{(k)} + \rho_k d^{(k)}$$

avec $d^{(k)}$ donné par (3.22) et (3.23).

Il nous reste à calculer les ρ_k et $d^{(k)}$.
Comme $< A\Delta_k \, , \, \Delta_l > = 0$ pour $l < k$, on déduit que $< Ad^{(k)} \, , \, \Delta_l > = 0$ ce qui nous donne, en utilisant aussi (3.20) :

$$0 = < d^{(k)} \, , \, A\Delta_l > = < d^{(k)} \, , \, \nabla J(u^{(l+1)}) - \nabla J(u^{(l)}) > \quad \text{pour} \;\; 0 \leq l \leq k-1.$$

A l'aide de (3.22) cela nous donne

$$< \nabla J(u^{(k)}) + \sum_{j=0}^{k-1} \lambda_j^k \, \nabla J(u^{(j)}) \, , \, \nabla J(u^{(l+1)}) - \nabla J(u^{(l)}) > = 0 \quad \text{pour} \;\; 0 \leq l \leq k-1. \tag{3.24}$$

En prennant $l = k - 1$ dans cette égalité on trouve

$$\|\nabla J(u^{(k)})\|^2 - \lambda_{k-1}^k \, \|\nabla J(u^{(k-1)})\|^2 = 0$$

ce qui nous donne

$$\lambda_{k-1}^k = \frac{\|\nabla J(u^{(k)})\|^2}{\|\nabla J(u^{(k-1)})\|^2}. \tag{3.25}$$

En prennant $l \leq k - 2$ en (3.24), on trouve la relation de reccurence

$$\lambda_{l+1}^k \|\nabla J(u^{(l+1)})\|^2 - \lambda_l^k \, \|\nabla J(u^{(l)})\|^2 = 0$$

ce qui donne

$$\lambda_l^k = \lambda_{l+1}^k \, \frac{\|\nabla J(u^{(l+1)})\|^2}{\|\nabla J(u^{(l)})\|^2}.$$

On en déduit facilement, en utilisant aussi (3.25)

$$\lambda_i^k = \frac{\|\nabla J(u^{(k)})\|^2}{\|\nabla J(u^{(i)})\|^2}, \quad i = 0, 1, \cdots k-1.$$

On obtient alors, à l'aide de (3.22) :

$$d^{(k)} = \nabla J(u^{(k)}) + \sum_{i=0}^{k-1} \frac{\|\nabla J(u^{(k)})\|^2}{\|\nabla J(u^{(i)})\|^2} \nabla J(u^{(i)}) =$$

$$\nabla J(u^{(k)}) + \frac{\|\nabla J(u^{(k)})\|^2}{\|\nabla J(u^{(k-1)})\|^2} \left[\nabla J(u^{(k-1)}) + \sum_{i=0}^{k-2} \frac{\|\nabla J(u^{(k-1)})\|^2}{\|\nabla J(u^{(i)})\|^2} \nabla J(u^{(i)}) \right]$$

Ceci nous donne la suite $\{d^{(k)}\}$ par reccurence sous la forme

$$\begin{cases} d^{(0)} = \nabla J(u^{(0)}) \\ d^{(k)} = \nabla J(u^{(k)}) + \frac{\|\nabla J(u^{(k)})\|^2}{\|\nabla J(u^{(k-1)})\|^2} \, d^{(k-1)}. \end{cases} \tag{3.26}$$

Il reste à détérminer les ρ_k.
Rappellons qu'on a la relation de reccurence :

$$u^{(k+1)} = u^{(k)} + \rho_k d^{(k)}$$

et que nous avons

$$J\left(u^{(k)} + \rho_k d^{(k)}\right) \leq J(u^{(k)} + y), \quad \forall\, y \in G_k$$

ce qui nous donne

$$J\left(u^{(k)} + \rho_k d^{(k)}\right) \leq J(u^{(k)} + \rho d^{(k)}), \quad \forall\, \rho \in I\!R$$

car $\rho d^{(k)} \in G_k$. On en déduit que ρ_k est un point de minimum sur $I\!R$ de l'application

$$\rho \in I\!R \to J(u^{(k)} + \rho d^{(k)}) \in I\!R.$$

donc ρ_k est un point où la dérivée de cette application s'annule, ce qui donne

$$< \nabla J(u^{(k)} + \rho_k d^{(k)}) \, , \, d^{(k)} > = 0$$

c'est à dire

$$< Au^{(k)} + \rho_k A d^{(k)} - b \, , \, d^{(k)} > = 0.$$

On obtient alors

$$\rho_k = -\frac{< \nabla J(u^{(k)}) \, , \, d^{(k)} >}{< Ad^{(k)} \, , \, d^{(k)} >} \tag{3.27}$$

Remarque : La Proposition 3.7 nous dit que $\Delta_k \neq 0$, ce qui nous donne $d^{(k)} \neq 0$. Ceci implique $< Ad^{(k)} \, , \, d^{(k)} > \, > 0$, car la matrice A est SDP.

L'algorithme des gradients conjugués (AGC) pour une fonction quadratique avec une matrice A qui est SDP est alors le suivant :

1. **pas 1.** *On pose* $k = 0$, *on choisit* $u^{(0)} \in I\!R^n$ *et on pose* $d^{(0)} = \nabla J(u^{(0)}) = Au^{(0)} - b$.
2. **pas 2.** *Si* $\nabla J(u^{(k)}) = 0$ *STOP "La solution* u^* *est* $u^{(k)}$. *Sinon, va au* **pas 3.**
3. **pas 3.** *On pose*

$$\rho_k = -\frac{< \nabla J(u^{(k)}) \, , \, d^{(k)} >}{< Ad^{(k)} \, , \, d^{(k)} >}$$

$$u^{(k+1)} = u^{(k)} + \rho_k d^{(k)}$$

$$\beta_k = \frac{\|\nabla J(u^{(k+1)})\|^2}{\|\nabla J(u^{(k)})\|^2}$$

$$d^{(k+1)} = \nabla J(u^{(k+1)}) + \beta_k d^{(k)}$$

faire $k = k + 1$
retour au **pas 2.**

3.3.2 Cas d'une fonction J quelconque

On suppose ici $J : I\!R^n \mapsto I\!R$ une fonction elliptique. Dans ce cas l'algorithme des gradients conjuguées est le suivant (c'est une "petite" modification de l'algorithme précédent) :

1. **pas 1.** *On pose* $k = 0$, *on choisit* $u^{(0)} \in I\!R^n$ *et on pose* $d^{(0)} = \nabla J(u^{(0)})$.
2. **pas 2.** *Si* $\nabla J(u^{(k)}) = 0$ *STOP "La solution* u^* *est* $u^{(k)}$. *Sinon, va au* **pas 3.**
3. **pas 3.** *On pose*

$$u^{(k+1)} = u^{(k)} + \rho_k d^{(k)}$$

où $\rho_k \in I\!R$ *est l'unique élément qui minimise la fonction*

$$\rho \in I\!R \to J(u^{(k)} + \rho d^{(k)}) \in I\!R$$

et ensuite

$$\beta_k = \frac{\|\nabla J(u^{(k+1)})\|^2}{\|\nabla J(u^{(k)})\|^2}$$

$$d^{(k+1)} = \nabla J(u^{(k+1)}) + \beta_k d^{(k)}$$

faire $k = k + 1$
retour au **pas 2.**

Ceci est la variante dite de **Fletcher-Reeves**.
Il y a aussi la variante dite de **Polack-Ribiere** avec comme seul changement

$$\beta_k = \frac{< \nabla J(u^{(k+1)}) - \nabla J(u^{(k)}) \, , \, \nabla J(u^{(k+1)}) >}{\|\nabla J(u^{(k)})\|^2}$$

Cette dernière variante donne des meilleurs résultats en pratique.
Remarque : Ces deux versions coincident dans le cas d'une fonction J quadratique, car dans ce cas on a

$$< \nabla J(u^{(k)}) \, , \, \nabla J(u^{(k+1)}) > = 0.$$

Chapitre 4

Optimisation avec contraintes

On rappelle qu'on se donne $U \subset \mathbb{R}^n$ où U est un ensemble fermé des contraintes, avec $U \neq \emptyset$ et $U \neq \mathbb{R}^n$. On se donne aussi la fonction

$$J : \mathbb{R}^n \mapsto \mathbb{R}.$$

On cherche à résoudre le problème de minimisation

$$\min_{u \in U} J(u) \tag{4.1}$$

(c'est à dire, on cherche $u^* \in U$ tel que

$$J(u^*) \leq J(u), \quad \forall u \in U \quad) \tag{4.2}$$

On va considérer deux cas pour U :

1. U est un **pavé**, c'est à dire, il est de la forme $U = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ (avec la convention qu'on peut avoir "$]-\infty$" à la place de "$[a_i$" ou "$+\infty[$" à la place de "$b_i]$", donc le pavé peut être de volume infinit).
2. U est de la forme

$$U = \{x \in \mathbb{R}^n,\ \varphi_i(x) \leq 0,\ \ i = 1, 2, \cdots m\} \tag{4.3}$$

   avec $m \in \mathbb{N}^*$ et $\varphi_1, \varphi_2, \cdots \varphi_m$ des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ données (on dit dans ce cas que U est donné par des **contraintes inégalités larges**).

Remarques :

1. Le premier cas est un cas particulier du deuxième
2. Dans le cas des contraintes inégalités larges, le problème de minimisation associé est aussi appelé **problème de programation nonlinéaire** si au moins une des fonctions $\varphi_1, \varphi_2, \cdots \varphi_m$ ou J est non affine. Si toutes ces fonctions sont affines, alors on a un **problème de programation linéaire** (vu en L3).

3. Dans la pratique on peut rencontrer des problèmes de minimisation avec contraintes égalités, c'est à dire, des contraintes de la forme

$$\tilde{U} = \{x \in I\!R^n,\ \varphi_i(x) = 0,\ \ i = 1, 2, \cdots m\}$$

ou des problèmes avec des contraintes mélangées (égalités et inégalités larges). Dans ce cas on peut toujours se ramener à des problèmes avec contraintes inégalités larges, ceci par deux méthodes :
· soit en éliminant des contraintes à l'aides des égalités
· soit en écrivant une égalité $\varphi_i(x) = 0$ comme une double inégalité : $\varphi_i(x) \leq 0$ et $-\varphi_i(x) \leq 0$.

Exemple : Considérons l'ensemble

$$U = \{x = (x_1, x_2, x_3), \quad x_1 + x_2 \leq x_3,\ \ 3x_1 - 2x_2 = x_3\}.$$

Alors on peut exprimer à l'aide de la dernière égalité de U une variable en fonction des 2 autres (par exemple : $x_3 = 3x_1 - 2x_2$). On remplace ensuite x_3 dans l'inégalité de U : $x_1 + x_2 \leq 3x_1 - 2x_2$ ce qui donne $-2x_1 + 3x_2 \leq 0$. On peut alors introduire l'ensemble à 2 variable seulement :

$$U_0 = \{x = (x_1, x_2), \quad -2x_1 + 3x_2 \leq 0\}.$$

Alors minimiser une fonction $J(x_1, x_2, x_3)$ sur U est équivalent au fait de minimiser sur U_0 une fonction J_0 de 2 variables définie par

$$J_0(x_1, x_2) = J(x_1, x_2, 3x_1 - 2x_2).$$

L'autre méthode (plus compliquée) est d'écrire U sous la forme équivalentes

$$U = \{x = (x_1, x_2, x_3), \quad x_1 + x_2 \leq x_3,\ \ 3x_1 - 2x_2 \leq x_3,\ \ x_3 \leq 3x_1 - 2x_2\}.$$

On peut montrer facilement le résultat suivant :

Proposition 4.1. *Soit U donné par (4.3). Si toutes les fonctions $\varphi_1, \varphi_2, \cdots, \varphi_m$ sont convexes alors U est un ensemble convexe.*

Démonstration. Laissée en exercice. □

4.1 Rappel sur les multiplicateurs de Lagrange

Soient $J, \theta_1, \theta_2, \cdots \theta_m : I\!R^n \mapsto I\!R$ des fonctions de classe C^1, avec $m \in I\!N^*$. On note

$$\tilde{O} = \{x \in I\!R^n,\ \theta_i(x) = 0,\ \ i = 1, 2, \cdots m\}.$$

On dit que $\tilde{O}$ est une **variété** de $I\!R^n$.

Définition 4.2. 1. Si $x \in \tilde{O}$ est tel que la famille des vecteurs $\{\nabla\theta_i(x)\}_{i=1,\cdots m}$ forme un **système libre** en $I\!R^n$ alors on dit que x est un **point régulier** de $\tilde{O}$.

2. La variété $\tilde{O}$ est dite **régulière** si tous les points de $\tilde{O}$ sont réguliers.

Remarque : Une condition équivalente de régularité de $\tilde{O}$ en x est : $rang(B) = m$ où B est la matrice Jacobienne donnée par

$$B_{ij} = \frac{\partial \theta_i}{\partial x_j}(x),\ i = 1, \cdots m,\ \ j = 1, \cdots n.$$

(on a alors nécéssairement : $m \leq n$).

On a le résultat suivant :

Théorème 4.3. *(Admis sans preuve !)*
Soit x^* un point régulier de $\tilde{O}$ tel que x^* soit un extremum local de J sur $\tilde{O}$ (minimum local ou maximum local). Alors il existe $\lambda_1^*, \lambda_2^*, \cdots \lambda_m^* \in I\!R$ (appellés **multiplicateurs de Lagrange***) tels que*

$$\nabla J(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla \theta_i(x^*) = 0 \tag{4.4}$$

Remarque : Le système (4.4) donne des conditions **nécessaires** d'optimalité (appellés aussi **conditions nécessaires d'optimalité de premier ordre**, car elles font intervenir des dérivées une fois uniquement). Ces conditions ne sont pas en général suffisantes. Le système (4.4) nous donne n équations avec $n + m$ inconnues. Mais le fait que $x^* \in \tilde{O}$ nous donne encore m équations : $\theta_i(x^*) = 0,\ i = 1, \cdots m$, ce qui nous fait au total $n + m$ équations avec $n + m$ inconnues.

4.2 Optimisation sous contraintes d'inégalités

On se donne de nouveau les fonctions $J, \theta_1, \theta_2, \cdots \theta_m : I\!R^n \mapsto I\!R$ de classe C^1, avec $m \in I\!N^*$.
Notons maintenant :

$$O = \{x \in I\!R^n,\ \theta_i(x) \leq 0,\ \ \forall\, i = 1, 2, \cdots m\}.$$

Notations :
On introduit la fonction à valeurs vectorielles $\theta : I\!R^n \mapsto I\!R^m$ donnée par

$$\theta(x) = (\theta_1(x),\ \theta_2(x),\ \cdots\ \theta_m(x))^T.$$

Pour un vecteur $b = (b_1, \cdots b_m)^T \in I\!R^m$ la notation $b \leq 0$ signifie $b_i \leq 0,\ \forall\, i = 1, \cdots m$. On peut donc écrire

$$O = \{x \in I\!R^n,\ \theta(x) \leq 0\}.$$

On va s'intéresser au problème de minimisation :

$$\min_{x \in O} J(x). \tag{4.5}$$

Hypothèse : On va supposer que les fonctions θ_i sont de telle manière que $O \neq \emptyset$ et aussi que O ne se réduit pas à un seul point ou à un nombre finit de points (dans quel cas le problème de minimisation (4.5) est très banal !)

4.2.1 Conditions d'optimalité de premier ordre : multiplicateurs de Karush-Kuhn-Tucker

On donnera des conditions d'optimalités similaires au système (4.4), mais pour des contraintes inégalités.

Définition 4.4. 1. On dit que la contrainte $\theta_i(u) \leq 0$ est **active** en $v \in O$ si on a

$$\theta_i(v) = 0.$$

On introduit alors l'ensemble

$$I(v) = \{i \in \{1, 2, \cdots m\},\ \theta_i(v) = 0\}$$

et la variété (définie uniquement si $I(v) \neq \emptyset$)

$$\tilde{O}(v) = \{u \in I\!R^n, \quad \theta_i(u) = 0,\ \forall\, i \in I(v)\}$$

(remarquons que $v \in \tilde{O}(v)$).

2. On dira que $v \in O$ est un **point régulier** de O si :
- soit $I(v) = \emptyset$
- soit v est un point régulier de la variété $\tilde{O}(v)$.

Le résultat principal de ce paragraph est le suivant :

Théorème 4.5. *Soit x^* un point régulier de O. Si x^* est un point de minimum local de J sur O alors il existe $p_1^*, p_2^*, \cdots p_m^* \in [0, +\infty[$ (appellés* **multiplicateurs de Karush-Kuhn-Tucker***) tels que :*

$$\nabla J(x^*) + \sum_{i=1}^{m} p_i^* \nabla \theta_i(x^*) = 0 \tag{4.6}$$

$$p_i^*\, \theta_i(x^*) = 0, \quad \forall\, i = 1, 2, \cdots m. \tag{4.7}$$

Remarque : Ce sont encore des conditions nécessaires d'optimalité de premier ordre appellés **conditions de Karush-Kuhn-Tucker**.

La preuve du Théorème 4.5 fait appel aux résultats suivants :

Lemme 4.6. *Soit $v \in O$ tel que $I(v) \neq \emptyset$ et soit $w \in I\!R^n$ tel que*

$$< \nabla \theta_j(v)\, ,\, w > < 0, \quad \forall\, j \in I(v).$$

Alors w est une direction admissible pour v en O, c'est à dire, il existe $t_0 > 0$ tel que

$$v + tw \in O, \quad \forall\, t \in [0, t_0]$$

(on a même plus : $v + tw \in \overset{\circ}{O} \quad \forall\, t \in]0, t_0[$.)

Démonstration. Il faut montrer : $\theta_j(v+tw) \leq 0, \ \ \forall\, j = 1, 2 \cdots, m$ si t est "proche" de 0.
Cas 1. Si $j \notin I(v)$.
Alors $\theta_j(v) < 0$ ce qui donne $\theta_j(v+tw) < 0$ si t est "proche" de 0 (on utilise la continuité de θ_j).
Cas 2. Si $j \in I(v)$.
Alors $\theta_j(v) = 0$ et en utilisant le developpement de Taylor à l'ordre 1 on obtient

$$\theta_j(v+tw) = \theta_j(v) + t \ < \nabla\theta_j(v)\,,\, w > \ + o(t). \tag{4.8}$$

Comme $\theta_j(v) = 0$ et en utilisant l'hypothèse du lemme, on obtient : $\theta_j(v+tw) < 0$ si $t > 0$ est "proche" de 0. Ceci nous donne le résultat. □

Remarque : On peut affaiblir l'hypothèse du lemme ci-dessus en prennant : $< \nabla\theta_j(v)\,,\, w > \leq 0$ si $j \in I(v)$ et si θ_j est **affine** (car dans ce cas dans (4.8) on a $o(t) = 0$).

Le corollaire suivant est une conséquence immédiate des lemmes **??** et 2.16 :

Corollaire 4.7. *Soit $v \in O$ tel que $I(v) \neq \emptyset$ et soit $w \in I\!R^n$ tel que*

$$< \nabla\theta_j(v)\,,\, w > < 0, \quad \forall\, j \in I(v).$$

Supposons que v est un minimum local de J sur O. Alors

$$< \nabla J(v)\,,\, w > \geq 0.$$

Lemme 4.8. *Soient $d_1, d_2, \cdots, d_l \in I\!R^n$ des vecteurs linéairement indépendents (donc forcément $l \leq n$). Alors la matrice $B \in \mathcal{M}(I\!R)$ telle que $B_{ij} = < d_i\,,\, d_j >, \ \ \forall\, i, j = 1, \cdots, l$ est inversible.*

Démonstration. Soit $y = (y_1, y_2, \cdots, y_l)^T \in I\!R^l$ telle que $B\,y = 0$ c'est à dire

$$\sum_{j=1}^{l} < d_i\,,\, d_j > \ y_j = 0, \quad \forall\, i = 1, 2, \cdots, l.$$

En multipliant chacune de ces égalités par y_i et en sommant sur i on obtient

$$\sum_{i=1}^{l}\sum_{j=1}^{l} < y_i d_i\,,\, y_j d_j > = 0 \quad \text{ c'est à dire } \quad \| \sum_{i=1}^{l} y_i d_i \|^2 = 0.$$

Ceci nous donne $\sum_{i=1}^{l} y_i d_i = 0$. Avec l'indépendence des vecteurs $d_1, d_2, \cdots, d_l$ on déduit $y_1 = y_2 = \cdots = y_l$, donc $y = 0$.
On a donc montré que $By = 0$ entraîne $y = 0$, donc B est inversible. □

Preuve du Théorème 4.5.
Soit $x^* \in O$ un point de minimum local de J sur O. Il y a deux situations :

Cas a) Si $I(x^*) = \emptyset$ (aucune contrainte n'est active en x^*). Alors x^* appartient à l'intérieur de O ce qui implique

$$\nabla J(x^*) = 0$$

et le Théoème 4.5 est vraie avec $p_1^* = p_2^* = \cdots = p_m^* = 0$.

Cas b) Si $I(x^*) \neq \emptyset$.
Pour simplifier supposons que $I(x^*) = \{1, 2, \cdots, l\}$ avec $1 \leq l \leq m$. Cela veut dire :

$$\begin{array}{lll} \theta_i(x^*) = 0 & \text{si} & 1 \leq i \leq l \\ \theta_i(x^*) < 0 & \text{si} & l+1 \leq i \leq m \end{array}$$

(avec la convention évidente que la dernière ligne est absente si $l = m$).
Introduisons les ensembles

$$\tilde{O}(x^*) = \{x \in I\!R^n,\ \theta_i(x) = 0 \quad \text{si} \quad 1 \leq i \leq l\}$$

et

$$\tilde{V}(x^*) = \{x \in I\!R^n,\ \theta_i(x) < 0 \quad \text{si} \quad l+1 \leq i \leq m\}$$

(prendre $\tilde{V}(x^*) = I\!R^n$ si $l = m$).
Il est clair que $\tilde{V}(x^*)$ est un ouvert et que $x^* \in \tilde{V}(x^*)$, ce qui nous dit que $\tilde{V}(x^*)$ est un voisinage de x^*.
D'autre part nous avons $\tilde{O}(x^*) \cap \tilde{V}(x^*) \subset O$. En plus il existe $W \subset I\!R^n$ avec W voisinage de x^* tel que x^* est un point de minimum de J sur $O \cap W$. Mais $\tilde{O}(x^*) \cap \tilde{V}(x^*) \subset O$ donc $\tilde{O}(x^*) \cap \tilde{V}(x^*) \cap W \subset O \cap W$ ce qui nous donne que x^* est un point de minimum de J sur $\tilde{O}(x^*) \cap \tilde{V}(x^*) \cap W$. Comme $\tilde{V}(x^*) \cap W$ est un voisinage de x^* alors x^* est un point de minimum local de J sur $\tilde{O}(x^*)$.
On utilise alors le Théorème des multiplicateurs de Lagrange, donc il existe $\tilde{p}_1, \cdots, \tilde{p}_l \in I\!R$ tels que

$$\nabla J(x^*) + \sum_{k=1}^{l} \tilde{p}_k \nabla \theta_k(x^*) = 0. \tag{4.9}$$

Nous posons alors $p_1^*, p_2^*, \cdots, p_m^* \in I\!R$ tels que

$$\begin{array}{lll} p_i^* = \tilde{p}_i & \text{si} & 1 \leq i \leq l \\ p_i^* = 0 & \text{si} & l+1 \leq i \leq m. \end{array}$$

Alors les égalités (4.6) et (4.7) du Théorème 4.5 sont satisfaites (observer qu'on a $p_i^* \theta_i(x^*) = 0,\ \forall\, i = 1, \cdots, m$).
Pour finir la preuve du Théorème, il reste encore à montrer :

$$p_i^* \geq 0, \quad \forall\, i = 1, \cdots, l. \tag{4.10}$$

Nous considérons deux sous-cas :
Cas b1) $l = 1$ (c'est à dire : $I(x^*) = \{1\}$).
Nous avons alors de (4.9) :

$$\nabla J(x^*) + p_1^* \nabla \theta_1(x^*) = 0.$$

En faisant le produit scalaire de cette égalité par $\nabla\theta_1(x^*)$ (remarquer que $\nabla\theta_1(x^*) \neq 0$) on obtient

$$p_1^* \|\nabla\theta_1(x^*)\|^2 = - < \nabla J(x^*)\,,\, \nabla\theta_1(x^*) > .$$

D'autre part, du Corollaire 4.7 avec $v = x^*$ et $w = -\nabla\theta_1(x^*)$ on déduit

$$< \nabla J(x^*)\,,\, -\nabla\theta_1(x^*) > \geq 0$$

On déduit alors immédiatement

$$p_1^* \geq 0.$$

Cas b2) $l \geq 2$.
On va montrer uniquement

$$p_1^* \geq 0, \tag{4.11}$$

les preuves des autre inégalités étant identiques.
L'idée de la preuve est de construire une suite $\{w^{(q)}\}_{q\in\mathbb{N}^*} \subset \mathbb{R}^n$ telle que

$$\begin{array}{l} < w^{(q)}\,,\, \nabla\theta_1(x^*) > = -1 \\ < w^{(q)}\,,\, \nabla\theta_k(x^*) > = -\frac{1}{q} \qquad \forall\ k = 2, \cdots, l \end{array} \tag{4.12}$$

(les directions $w^{(q)}$ seront alors admissibles).
Pour cela, on construit $w^{(q)}$ sous la forme

$$w^{(q)} = \sum_{j=1}^{l} \alpha_j \nabla\theta_j(x^*)$$

avec $\alpha_j \in \mathbb{R}$ à choisir de sorte que (4.12) soit vraie. Il est clair qu'il faut alors :

$$\sum_{j=1}^{l} \alpha_j < \nabla\theta_k(x^*)\,,\, \nabla\theta_j(x^*) > = \left\{ \begin{array}{lll} -1 & \text{si} & k = 1 \\ -1/q & \text{si} & k = 2, \cdots, l \end{array} \right.$$

Ceci est une égalité du type $B\alpha = b$ où B est la matrice carrée de taille l donnée par

$$B_{kj} = < \nabla\theta_k(x^*)\,,\, \nabla\theta_j(x^*) >$$

α est le vecteur inconnu $\alpha = (\alpha_1, \cdots, \alpha_l)^T$ et b est le vecteur dont les éléments b_k sont donnés par

$$b_k = \left\{ \begin{array}{lll} -1 & \text{si} & k = 1 \\ -1/q & \text{si} & k = 2, \cdots, l \end{array} \right.$$

Comme les vecteurs $\nabla\theta_1(x^*), \cdots, \nabla\theta_l(x^*)$ sont indépendents par hypothèse, la matrice B est inversible (voir Lemme 4.8) donc l'équation précédente admet une solution α unique. Donc la suite $w^{(q)}$ avec la propriété (4.12) existe.
D'autre part, l'égalité (4.9) nous donne

$$\nabla J(x^*) + \sum_{k=1}^{l} p_k^* \nabla\theta_k(x^*) = 0.$$

En multipliant cette égalité par $w^{(q)}$ on obtient

$$< \nabla J(x^*)\,,\, w^{(q)} > = -\sum_{k=1}^{l} p_k^* < \nabla\theta_k(x^*)\,,\, w^{(q)} >$$

ce qui donne à l'aide de (4.12)

$$< \nabla J(x^*)\,,\, w^{(q)} > = p_1^* + \frac{1}{q}\sum_{k=2}^{l} p_k^*. \tag{4.13}$$

Nous avons aussi

$$< \nabla J(x^*)\,,\, w^{(q)} > \geq 0$$

comme conséquence du Corollaire 4.7 avec $w = w^{(q)}$. En passant alors à la limite $q \to +\infty$ en (4.13), on obtient l'inégalité (4.11), ce qui finit la preuve du Théorème 4.5.

Remarque : Si on considère l'ensemble

$$O = \{x \in I\!R^n,\ \theta_i(x) \leq 0,\ i = 1, \cdots, m_1, \quad \varphi_j(x) = 0,\ j = 1, \cdots, m_2\}$$

avec $m_2 \geq 1$ alors on peut considérer O comme donné uniquement par des contraintes inégalités, en écrivant

$$O = \{x \in I\!R^n,\ \theta_i(x) \leq 0,\ i = 1, \cdots, m_1, \quad \varphi_j(x) \leq 0,\ j = 1, \cdots, m_2, \quad -\varphi_j(x) \leq 0,\ j = 1, \cdots, m_2\}$$

Il est facile de voir qu'aucun point x de O n'est régulier, car d'une part les deux contraintes $\varphi_j(x) \leq 0$ et $-\varphi_j(x) \leq 0$ sont actives et d'autre part, aucune famille de vecteurs qui contient à la fois $\nabla\varphi_j(x)$ et $-\nabla\varphi_j(x)$ ne peut être indépendente.
Conclusion : L'hypothèse de régularité n'est satisfaite pour aucun $x \in O$ si O comporte au moins une contrainte égalité transformée artificiellement en deux contraintes inégalités.
Exemple : L'ensemble

$$O = \{x \in I\!R^2,\ x_1^2 - x_2 \leq 0,\ x_2 - x_1 - 4 = 0\}$$

peut s'écrire sous la forme

$$O = \{x \in I\!R^2,\ x_1^2 - x_2 \leq 0,\ x_2 - x_1 - 4 \leq 0,\ -x_2 + x_1 + 4 \leq 0\}.$$

En introduisant les fonctions : $\theta_1, \theta_2, \theta_3 : I\!R^2 \to I\!R$ données par

$$\theta_1(x) = x_1^2 - x_2,\ \theta_2(x) = x_2 - x_1 - 4,\ \theta_3(x) = -x_2 + x_1 + 4$$

on peut écrire O sous la forme standard

$$O = \{x \in I\!R^2,\ \theta_i(x) \leq 0,\ i = 1, 2, 3\}.$$

Il est clair que pour tout $x \in O$ nous avons $\{2, 3\} \subset I(x)$. D'autre part, $\nabla\theta_2(x) = (-1, 1)^T$ et $\nabla\theta_3(x) = (1, -1)^T = -\nabla\theta_2(x)$. Donc aucune famille de vecteurs qui inclut $\nabla\theta_2(x)$ et $\nabla\theta_3(x)$ ne peut être indépendente, donc aucun $x \in O$ n'est régulier.

Pour rémedier à ce genre d'inconvenient, il y a deux solutions :

1. Eliminer les contraintes égalités en éliminant des variables (quand cela est possible)

2. Affaiblir l'hypothèse de régularité du Théorème 4.5.

Dans la suite on choisira cette dernière possibilité.

Définition 4.9. On dit que les contraintes de O sont **qualifiés** en $v \in O$ si
a) Soit $I(v) = \emptyset$
b) Soit $I(v) \neq \emptyset$ et il existe $w \in I\!R^n$ tel qu'on a $\forall\, i \in I(v)$:

$$< \nabla\theta_i(v)\,,\, w > \leq 0,$$

avec en plus

$$< \nabla\theta_i(v)\,,\, w > < 0 \quad \text{si } \theta_i \text{ est non-affine.}$$

Remarques :

1. Si θ_i est affine pour tout $i \in I(v)$ alors les contraintes de O sont qualifiés en v (prendre $w = 0$ dans la partie **b)** de la Définition 4.9). En particulier si les fonctions θ_i sont affines pour tout $i = 1, \cdots, m$ alors les contraintes sont qualifiées en tout point de O.
2. Comme conséquence de la Définition 4.9 le vecteur w "pointe" vers l'ensemble O (donc $\exists t_0 > 0$ tel que $v + tw \in O,\ \forall\, t \in [0, t_0]$) (voir Lemme 4.6 et la Remarque après).

Nous avons :

Proposition 4.10. *Si un point $v \in O$ est régulier alors les contraintes de O sont qualifiés en v. Autrement dit, la condition de qualification de la Définition 4.9 est plus* **faible** *que la condition de régularité de la Définition 4.4.*

Démonstration. Nous avons construit dans la preuve du Théorème 4.5 une suite $w^{(q)}$ telle que $< \nabla\theta_i(v)\,,\, w^{(q)} > < 0,\ i = 1, \cdots, l$. Ceci prouve le résultat. □

On a alors le résultat suivant, plus forte que le Théorème 4.5 ; ce sera un résultat admis !

Théorème 4.11. *Soit $x^* \in O$ tel que les contraintes de O sont qualifiés en x^*. Si x^* est un point de minimum local de J sur O alors il existe $p_1^*, p_2^*, \cdots p_m^* \in [0, +\infty[$ (appellés* **multiplicateurs de Karush-Kuhn-Tucker***) tels que :*

$$\nabla J(x^*) + \sum_{i=1}^{m} p_i^* \nabla \theta_i(x^*) = 0 \tag{4.14}$$

$$p_i^* \, \theta_i(x^*) = 0, \quad \forall\, i = 1, 2, \cdots m. \tag{4.15}$$

Remarque :
l'énoncé de ce dernier résultat s'obtient à partir de l'énoncé du Théorème 4.5 en remplaçant l'hypothèse "x^* *point régulier de* O" par l'hypothèse plus faible "*les contraintes de* O *sont qualifiés en* x^*".

Le résultat suivant nous donne une condition suffisante pour la qualification en tout point de O :

Proposition 4.12. *Supposons que les fonctions $\theta_1, \theta_2, \cdots, \theta_m$ sont* **convexes**. *Supposons aussi que*
- ou bien toutes les fonctions $\theta_1, \theta_2, \cdots, \theta_m$ sont affines
- ou bien il existe $y \in I\!R^n$ tel que pour tout $i = 1, \cdots, m$ on a

$$\theta_i(y) \leq 0$$

$$\theta_i(y) < 0 \qquad \textit{si } \theta_i \textit{ n'est pas affine.}$$

Alors pour tout $x \in O$ les contraintes de O sont qualifiés en x.

Démonstration. Soit $x \in O$. Supposons que $I(x) \neq \emptyset$ (sinon OK, x est qualifié). Supposons aussi que toutes les fonctions $\{\theta_i\}_{i=1,\cdots,m}$ ne sont pas affines (sinon OK, x est qualifié, voir la première remarque après la Définition 4.9). Alors il existe $y \in I\!R^n$ avec la propriété : $\theta_i(y) \leq 0, \quad \forall\, i = 1, \cdots, m$ et
pour tout $i = 1, \cdots, m$ tel que θ_i n'est pas affine, on a $\theta_i(y) < 0$.
Alors par convexité de θ_i on a

$$\theta_i(y) \geq \theta_i(x) + \, < \nabla \theta_i(x) \, , \, y - x > \quad \forall\, i = 1, \cdots m$$

En utilisant l'égalité $\theta_i(x) = 0$ pour tout $i \in I(x)$ on obtient

$$< \nabla \theta_i(x) \, , \, y - x > \leq \theta_i(y), \quad \forall\, i \in I(x).$$

En utilisant les propriétés de $\theta_i(y)$ on déduit

$$< \nabla \theta_i(x) \, , \, y - x > \leq 0, \quad \forall\, i \in I(x)$$

et

$$< \nabla \theta_i(x) \, , \, y - x > < 0, \quad \forall\, i \in I(x) \quad \text{si} \quad \theta_i \quad \text{est non-affine.}$$

Donc la partie **b)** de la Définition 4.9 est satisfaite avec $w = y - x$, ce qui finit la preuve. □

Exemple 1

$$O_1 = \{x \in I\!R^2,\ \ x_2 \geq x_1^2,\ \ x_2 - x_1 = 4\}.$$

Il est clair qu'on peut écrire O_1 sous la forme

$$O_1 = \{x \in I\!R^2,\ \ \theta_i(x) \leq 0,\ \ i = 1, 2, 3\}$$

avec $\theta_1, \theta_2, \theta_3 : I\!R^2 \to I\!R$ définies par

$$\theta_1(x) = x_1^2 - x_2, \quad \theta_2(x) = x_2 - x_1 - 4, \quad \theta_3(x) = -x_2 + x_1 + 4.$$

On observe que θ_1, θ_2 et θ_3 sont convexes, avec en plus θ_2 et θ_3 affines. En prennant $y = (0, 4)^T$ nous avons :

$$\theta_1(y) < 0, \quad \theta_2(y) = \theta_3(y) = 0.$$

Alors les hypothèses de la Proposition 4.12 sont satisfaites, donc tous les points de O_1 sont qualifiés.

Exemple 2

$$O_2 = \{x \in I\!R^2,\ \ x_2 = x_1^2,\ \ x_2 - x_1 \leq 4\}.$$

On écrit O_2 sous la forme

$$O_2 = \{x \in I\!R^2,\ \ \theta_i(x) \leq 0,\ \ i = 1, 2, 3\}$$

avec

$$\theta_1(x) = x_1^2 - x_2, \quad \theta_2(x) = -x_1^2 + x_2, \quad \theta_3(x) = x_2 - x_1 - 4.$$

Comme θ_2 n'est pas convexe, la Proposition 4.12 ne s'applique pas (ceci n'implique pas automatiquement la non-qualification !) Soit $x \in O_2$. Il est clair que $\{1, 2\} \subset I(x)$. Pour avoir la qualification, il faudrait qu'il existe $w \in I\!R^2$ tel que

$$< \nabla\theta_1(x)\,,\, w > < 0 \quad \text{et} \quad < \nabla\theta_2(x)\,,\, w > < 0$$

or ceci est impossible car $\nabla\theta_2(x) = -\nabla\theta_1(x)$. Donc aucun point de O_2 n'est qualifié.

4.2.2 Théorie générale du point selle

Soit U et P deux ensembles quelconques et l'application $\mathcal{L} : U \times P \to I\!R$

Définition 4.13. On dit que le couple $(u^*, p^*) \in U \times P$ est un **point selle** de $\mathcal{L}$ si on a

$$\mathcal{L}(u^*, p) \leq \mathcal{L}(u^*, p^*) \leq \mathcal{L}(u, p^*), \quad \forall\, u \in U,\ \ \forall\, p \in P$$

(Autrement dit : u^* est un point de minimum de la fonction $u \in U \to \mathcal{L}(u, p^*) \in I\!R$ et p^* est un point de maximum de la fonction $p \in P \to \mathcal{L}(u^*, p) \in I\!R$.)

Exemple : Prendre $U = P = I\!R$ et $\mathcal{L}(u,p) = u^2 - p^2$. Il est facile de voir que $(0,0)$ est un point selle de $\mathcal{L}$.

On introduit les fonctions suivantes :
$G : U \to \overline{I\!R}$ et $H : P \to \overline{I\!R}$ données par

$$G(u) = \sup_{q \in P} \mathcal{L}(u,q)$$

et respectivement par

$$H(p) = \inf_{v \in U} \mathcal{L}(v,p).$$

Remarque : (u^*, p^*) est un point selle de $\mathcal{L}$ si et seulement si

$$G(u^*) = \mathcal{L}(u^*, p^*) = H(p^*).$$

On a le résultats suivant :

Proposition 4.14.

$$\sup_{p \in P} \inf_{u \in U} \mathcal{L}(u,p) \leq \inf_{u \in U} \sup_{p \in P} \mathcal{L}(u,p)$$

(autrement dit : $sup_{p\in P} H(p) \leq \inf_{u\in U} G(u)$)

Démonstration. Soit $(u,p) \in U \times P$ fixé arbitraire. Il est clair qu'on a

$$\inf_{v \in U} \mathcal{L}(v,p) \leq \mathcal{L}(u,p) \leq \sup_{q \in P} \mathcal{L}(u,q)$$

ce qui nous donne

$$H(p) \leq G(u), \quad \forall\ (u,p) \in U \times P.$$

En passant au "suppremum" en p et au "infimum" en u on obtient le résultat. □

Nous avons :

Théorème 4.15. *Si (u^*, p^*) est un point selle de $\mathcal{L}$ sur $U \times P$ alors*

$$\sup_{p \in P} \inf_{u \in U} \mathcal{L}(u,p) = \mathcal{L}(u^*, p^*) = \inf_{u \in U} \sup_{p \in P} \mathcal{L}(u,p)$$

(autrement dit : $\sup_{p\in P} H(p) = \mathcal{L}(u^*,p^*) = \inf_{u\in U} G(u)$).

Démonstration. La définition du point selle nous dit

$$H(p^*) = \mathcal{L}(u^*, p^*) = G(u^*)$$

ce qui nous permet d'écrire

$$\sup_{p \in P} H(p) \geq \mathcal{L}(u^*, p^*) \geq \inf_{u \in U} G(u).$$

A l'aide de la Proposition 4.14 on obtient le résultat attendu. □

Exemple : Reprennons l'exemple précédent : $\mathcal{L} : I\!R \times I\!R \to I\!R, \quad \mathcal{L}(u,p) = u^2 - p^2$. Nous avons

$$G(u) = \sup_{p \in I\!R} (u^2 - p^2) = u^2$$

et

$$H(p) = \inf_{u \in I\!R} (u^2 - p^2) = -p^2$$

On aussi :

$$\inf_{u \in I\!R} G(u) = 0, \quad \sup_{p \in I\!R} H(p) = 0, \quad \mathcal{L}(0,0) = 0$$

(le point $(0,0)$ étant le point selle de $\mathcal{L}$), ce qui vérifie l'égalité du Théorème 4.15.

4.2.3 Applications de la théorie du point selle à l'optimisation

On utilise ici de nouveau les notations du debut de la Section 4.2, spécifiques aux problèmes d'optimisation.
Nous introduisons la fonction $L : I\!R^n \times I\!R^m_+ \to I\!R$ donnée par

$$L(x,p) = J(x) + \sum_{i=1}^{m} p_i \theta_i(x) \equiv J(x) + < p \, , \, \theta(x) >$$

(où on voit p comme un vecteur colonne de composantes $p_1, p_2, \cdots, p_m$). On a le résultat suivant :

Théorème 4.16. *Si (x^*, p^*) est un point selle de L sur $R^n \times I\!R^m_+$ alors x^* est un point de minimum de J sur O. En plus (x^*, p^*) satisfait les conditions de Karush-Kuhn-Tucker, c'est à dire*

$$\begin{cases} \nabla J(x^*) + \sum_{i=1}^m p_i^* \nabla \theta_i(x^*) = 0 \\ \\ p_i^* \theta_i(x^*) = 0, \qquad\qquad \forall \; i = 1, \cdots, m. \end{cases} \tag{4.16}$$

Démonstration. Nous avons

$$L(x^*, p^*) \geq L(x^*, p), \quad \forall \; p \in I\!R^m_+$$

ce qui donne

$$\sum_{i=1}^{m} p_i^* \theta_i(x^*) \geq \sum_{i=1}^{m} p_i \theta_i(x^*), \quad \forall \; p \in I\!R^m_+. \tag{4.17}$$

En prennant respectivement $p = 0$ et $p = 2p^*$ dans l'inégalité précédente on obtient

$$< p^* \, , \, \theta(x^*) > = 0. \tag{4.18}$$

Maintenant pour un $j \in \{1, 2, \cdots, m\}$ fixé prennons dans (4.17) $p = p^* + e_j$. On déduit alors $\theta_j(x^*) \leq 0$. Comme ceci est vrai pour tout j, on déduit

$$x^* \in O. \tag{4.19}$$

D'autre part, on se sert de l'inégalité

$$L(x^*, p^*) \leq L(x, p^*), \quad \forall\, x \in I\!R^n$$

c'est à dire

$$J(x^*) + < p^*\,,\, \theta(x^*) > \leq J(x) + < p^*\,,\, \theta(x) > \quad \forall\, x \in I\!R^n. \tag{4.20}$$

En utilisant l'égalité (4.18) et le fait que $< p^*\,,\, \theta(x) > \leq 0$ pour $x \in O$, on déduit

$$J(x^*) \leq J(x), \quad \forall\, x \in O.$$

Ceci nous dit que x^* est un point de minimum de J sur O.

D'autre part, de (4.20) on déduit que x^* est un point de minimum de la fonction $x \in I\!R^n \to J(x) + \sum_{i=1}^m p_i^* \theta_i(x)$. Ceci nous donne, en utilisant le Théorème 2.17, la première égalité de (4.16). Finalement, de (4.18) et (4.19) on a

$$\sum_{j=1}^m p_j^* \theta_j(x^*) = 0 \quad \text{et} \quad p_i^* \theta_i(x^*) \leq 0 \quad \text{pour tout } i \in \{1, \cdots, m\}$$

ce qui nous donne la deuxième partie de (4.16). Ceci finit la preuve du Théorème. □

Remarque : Ce résultat nous donne une condition suffisante pour avoir un minimum de J sur O. On verra aussi une reciproque, sous des hypothèses supplementaires. On cherchera alors à résoudre le problème appellé **le problème dual :**

$$\sup_{p \in I\!R^m_+} \inf_{u \in I\!R^n} L(u, p)$$

(justifié par le Théorème 4.15) plus simple à résoudre que le problème de minimisation de J sur O (qu'on appellera **problème primal**).

4.2.4 Le cas convexe

On verra ici des réciproques des Théorèmes 4.5 et 4.16 dans le cas où J et θ_j sont convexes.

Théorème 4.17. *(Réciproque du Théorème 4.5).*
Supposons que toutes les fonctions $J, \theta_1, \theta_2, \cdots, \theta_m$ sont **convexes**. *Supposons aussi qu'il existe $p_1^*, p_2^*, \cdots, p_m^* \in [0, +\infty[$ et $x^* \in O$ tels que les conditions de Karush-Kuhn-Tucker* (4.6) *et* (4.7) *du Théorème 4.5 soient satisfaites. Alors*
a) *x^* est un point de minimum de J sur O*
b) *(x^*, p^*) est un point selle de L*
(Rappel : $L : I\!R^n \times I\!R^m_+ \to I\!R, \quad L(x, p) = J(x) + < p\,,\, \theta(x) >$).

Démonstration. **a)** Nous avons

$$J(x^*) \leq J(x^*) - \sum_{i=1}^{m} p_i^* \theta_i(y), \quad \forall\, y \in O \ \ (\text{ car } \ \theta_i(y) \leq 0,\ p_i^* \geq 0).$$

Ensuite de (4.7) et de la convexité de θ_i on déduit

$$p_i^* \theta_i(y) = p_i^* [\theta_i(y) - \theta_i(x^*)] \geq p_i^* \ < \nabla \theta_i(x^*)\,,\, y - x^* >$$

On obtient alors des inégalités précédentes et de (4.6) :

$$J(x^*) \leq J(x^*) + < \nabla J(x^*),\ y - x^* >, \quad \forall\, y \in O.$$

En utilisant cette fois-ci la convexité de J nous obtenons

$$J(x^*) \leq J(y), \quad \forall\, y \in O$$

ce qui finit la preuve de **a)**.
b) Nous avons

$$L(x^*, p) = J(x^*) + < p\,,\, \theta(x^*) > \leq J(x^*) + < p^*\,,\, \theta(x^*) > = L(x^*, p^*), \quad \forall\, p \in I\!R^m_+$$

car $< p\,,\, \theta(x^*) > \leq 0$ et $< p^*, \theta(x^*) > = 0$.
D'autre part, comme les fonctions $J, \theta_1, \cdots, \theta_m$ sont convexes alors l'application $x \in I\!R^n \to L(x, p^*)$ est convexe. Mais l'égalité (4.6) nous dit que le gradient de cette application s'annule en x^*. On déduit alors (voir la remarque après Théorème 2.18) que x^* est un point de minimum de cette application et ceci finit la preuve. □

On finit cette section par les corollaires suivants :

Corollaire 4.18. *(Lien entre point selle et conditions de Karush-Kuhn-Tucker)*
Supposons que les fonctions $J, \theta_1, \cdots, \theta_m$ sont convexes. Alors (x^*, p^*) est point selle de L **si et seulement si** *$x^* \in O$ et les conditions* (4.6) *et* (4.7) *sont satisfaites.*

Démonstration. C'est une conséquence immédiate des Théorèmes 4.16 et 4.17 partie **b)**. □

Corollaire 4.19. *(Réciproque du Théorème 4.16)*
Supposons que les fonctions $J, \theta_1, \cdots, \theta_m$ sont convexes. Soit x^* un point de minimum de J sur O tel que les contraintes de O sont qualifiés en x^*. Alors il existe $p^* \in I\!R^m_+$ tel que (x^*, p^*) est un point selle de L.

Démonstration. C'est une conséquence immédiate des Théorèmes 4.11 et 4.17 partie **b)**. □

4.3 Algorithmes de minimisation avec contraintes

Le but de cette section est de donner des algorithmes numériques pour la résolution du problème de minimisation avec contraintes :

$$\min_{u\in U} J(u) \tag{4.21}$$

avec $U \subset I\!R^n$ un enesemble de contraintes.

On rappelle les deux cas de contraintes que nous considérons dans ce cours :

1. U est un pavé, c'est à dire, il est de la forme $U = [a_1,\, b_1] \times [a_2,\, b_2] \times \cdots \times [a_n,\, b_n]$ (avec la convention qu'on peut avoir "$]-\infty$" à la place de "$[a_i$" ou "$+\infty[$" à la place de "$b_i]$", donc le pavé peut être de volume infinit).
2. U est de la forme

   $$U = \{x \in I\!R^n,\ \theta_i(x) \leq 0,\ \ i = 1, 2, \cdots m\}$$

   avec $m \in I\!N^*$ et $\theta_1, \theta_2, \cdots \theta_m$ des fonctions de $I\!R^n$ dans $I\!R$ données (U est donné par des contraintes inégalités larges).

Dans le premier cas on utilisera des extension naturelles des méthodes de minimisation sur $I\!R^n$, vues en Chapitre 3. Dans le deuxième cas on utilisera des méthodes du type dualité.

4.3.1 Méthodes de relaxation

On suppose ici

$$U = [a_1,\, b_1] \times [a_2,\, b_2] \times \cdots \times [a_n,\, b_n] \equiv \Pi_{i=1}^n [a_i,\, b_i]. \tag{4.22}$$

L'algorithme dans ce cas ressemble à l'algorithme de relaxation pour minimisation sans contraintes.

Algorithme :

On se donne $u^{(k)} = \left(u_1^{(k)},\, u_2^{(k)}, \cdots, u_n^{(k)}\right)^T$ et on va calculer $u^{(k+1)} = \left(u_1^{(k+1)},\, u_2^{(k+1)}, \cdots, u_n^{(k+1)}\right)^T$ en n pas successifs. On a :

$$\begin{aligned}
J(u_1^{(k+1)},\, u_2^{(k)},\, \cdots u_n^{(k)}) &= \min_{y\in[a_1,\, b_1]} J(y,\, u_2^{(k)},\, \cdots u_n^{(k)}) \\
J(u_1^{(k+1)},\, u_2^{(k+1)}, u_3^{(k)}\, \cdots u_n^{(k)}) &= \min_{y\in[a_2,\, b_2]} J(u_1^{(k+1)},\, y, u_3^{(k)}\, \cdots u_n^{(k)}) \\
&\cdots \\
J(u_1^{(k+1)},\, \cdots u_n^{(k+1)}) &= \min_{y\in[a_n,\, b_n]} J(u_1^{(k+1)},\, \cdots u_{n-1}^{(k+1)}, y)
\end{aligned}$$

Comme dans le cas sans contraintes, on a le théorème de convergence suivant :

Théorème 4.20. *Si $J : I\!R^n \to I\!R$ est une fonction elliptique et l'ensemble U est donné par* (4.22) *alors la méthode de relaxation pour la minimisation de J sur U est bien définie et converge.*

Démonstration. La preuve du fait que la méthode est bien définie se fait exactement comme dans le Théorème 3.3. La convergence est admise ! □

4.3.2 Méthodes de projection

Rappellons d'abord le résultat fondamental suivant :

Théorème 4.21. *(Théorème de projection sur un convexe fermé)*
Soit $U \subset \mathbb{R}^n$ un ensemble convexe, fermé, non-vide et soit $v \in \mathbb{R}^n$. Alors il existe un unique element $Pv \in U$ (on peut le noter pour plus de précision par $P_U(v)$) tel que

$$\|v - P_U(v)\| = \inf_{u \in U} \|v - u\| = \min_{u \in U} \|v - u\|.$$

L'élément $P_U(v) \in U$ s'appelle **la projection de v sur U en $\mathbb{R}^n$**. *En plus on a*
a) *L'élément $P_U(v) \in U$ satisfait aussi*

$$< v - P_U(v)\,,\, u - P_U(v) > \leq 0, \quad \forall u \in U \tag{4.23}$$

b) *Si $w \in U$ est tel que*

$$< v - w\,,\, u - w > \leq 0, \quad \forall u \in U \tag{4.24}$$

alors $w = P_U(v)$ (donc $w = P_U(v)$ est l'unique élément de U satisfaisant (4.24)*.*
c) *On a*

$$\|P_U(v_1) - P_U(v_2)\| \leq \|v_1 - v_2\|, \quad \forall\, v_1, v_2 \in \mathbb{R}^n$$

(la fonction P_U n'augmente pas les distances). Ceci implique que P_U est une fonction lipschitzienne, donc **continue**.
d) *On a*

$$v = P_U(v) \quad \textit{si et seulement si} \quad v \in U$$

e) *Si U est le sous-espace affine fermé de $\mathbb{R}^n$ donné par $U = a + U_0$ avec $a \in \mathbb{R}^n$ et U_0 un sous-espace vectoriel fermé de $\mathbb{R}^n$ alors l'inégalité* (4.23) *devient l'égalité*

$$< v - P_U(v)\,,\, u > = 0, \quad \forall u \in U_0$$

(c'est à dire $v - P_U(v) \perp U_0$).

Idée de la preuve
Tout revient à minimiser sur U la fonction $g : \mathbb{R}^n \to \mathbb{R}$ définie par

$$g(u) = \|u - v\|^2.$$

Montrons d'abord que g est une fonction elliptique. Pour cela, on utilise les calculs et les résultats du Chapitre 2. On peut écrire

$$g(u) = < v - u\,,\, v - u > = < u, u > - 2 < u, v > + < v, v > .$$

ce qui donne

$$\nabla g(u) = 2u - 2v$$

$$\nabla^2 g(u) = 2I_n \quad (\text{ donc constante indépendente de } u).$$

Comme $2I_n$ est une matrice symmétrique et définie positive (matrice SDP) alors g est elliptique.
Comme U est fermé convexe alors il existe un unique élément $Pv = P_U(v) \in U$ tel que $g(P_U(v)) \leq g(u), \ \forall\, u \in U$ c'est à dire

$$\|v - P_U(v)\| \leq \|v - u\|, \quad \forall\, u \in U.$$

ce qui donne le résultat principal du théorème. Nous avons aussi (voir Théorème 2.17) :

$$< \nabla g(P_U(v)) \,,\, u - P_U(v) > \geq 0, \quad \forall\, u \in U$$

ce qui nous donne facilement

$$< v - P_U(v) \,,\, u - P_U(v) > \leq 0, \quad \forall\, u \in U.$$

ce qui prove **a)**. La partie **b)** est une conséquence de l'ellipticité et de l'unicité.
D'autre part, soient $v_1, v_2 \in I\!R^n$. On a alors

$$< v_1 - Pv_1 \,,\, u - Pv_1 > \leq 0, \quad \forall\, u \in U$$

$$< v_2 - Pv_2 \,,\, u - Pv_2 > \leq 0, \quad \forall\, u \in U$$

En prennant $u = Pv_2$ dans la première inégalité et $u = Pv_1$ dans la deuxième et en faisant la somme, on trouve

$$< Pv_2 - Pv_1 \,,\, v_1 - Pv_1 - v_2 + Pv_2 > \leq 0$$

ce qui donne

$$\|Pv_2 - Pv_1\|^2 \leq < Pv_2 - Pv_1 \,,\, v_2 - v_1 > \leq \|Pv_2 - Pv_1\| \, \|v_2 - v_1\|$$

(on a utilisé l'inégalité de Cauchy-Schwarz). Ceci nous donne **c)** par simplification.
La partie **d)** est évidente.
Pour la partie **e)**, soit U_0 un sous-espace vectoriel de $I\!R^n$, $a \in I\!R^n$, $U = a + U_0$ et $h \in U_0$ arbitraire et fixé. On va prendre en (4.23) : $u = Pv \pm h$ qui est un élément de U car $Pv \in U$ et $h \in U_0$. Ceci donne

$$\pm < v - Pv \,,\, h > \leq 0$$

d'où on déduit

$$< v - Pv \,,\, h > = 0$$

ce qui finit la preuve.

Exemple 1.
Si $n = 1$ et $U = [a,\, b]$ (avec $-\infty \leq a < b \leq +\infty$) alors il est facile de voir que pour tout $v \in I\!R$ on a

$$P_U(v) = \begin{cases} a & \text{si} \quad v < a & \text{(partie inexistente si } \ a = -\infty) \\ v & \text{si} \quad v \in U & \\ b & \text{si} \quad v > b & \text{(partie inexistente si } \ b = +\infty) \end{cases}$$

On a évidement le cas particulier : si $U = [0,\ +\infty[$ alors

$$P_U(v) = \begin{cases} v & \text{si} \quad v \geq 0 \\ 0 & \text{si} \quad v < 0 \end{cases}$$

On peut aussi noter dans ce cas

$$P_U(v) = v^+, \quad \forall\, v \in I\!R.$$

Exemple 2.
Supposons que U est un pavé, donc

$$U = \Pi_{i=1}^n [a_i,\, b_i] \in I\!R^n$$

avec $-\infty \leq a_i < b_i \leq +\infty, \quad i = 1, \cdots, n.$ Alors $\forall\, v = (v_1, v_2, \cdots, v_n)^T \in I\!R^n$ on a

$$P_U(v) = (P_1 v_1, P_2 v_2, \cdots, P_n v_n)^T$$

avec $P_k v_k =$ la projection de v_k en $I\!R$ sur l'intervalle $[a_k,\, b_k]$ (voir TD pour la preuve).
Cas particulier : Si $U = [0, , +\infty[^n$ alors $P_U(v) = (v_1^+, v_2^+, \cdots, v_n^+)^T$.
Remarque : Dans le cas où U est donné par des contraintes inégalité générales, alors il peut être difficile en pratique de calculer la projection en $I\!R^n$ sur U.

Revenons au problème de minimisation (4.21). On supposera dans la suite que l'ensemble U des contraintes est **convexe** et **fermeé**.
Supposons que $x^* \in U$ est tel que

$$J(x^*) = \min_{x \in U} J(x).$$

Alors

$$< \nabla J(x^*)\,,\, x - x^* > \geq 0, \quad \forall\, x \in U \tag{4.25}$$

qui est équivalente pour tout $\rho > 0$ à l'inégalité

$$\rho\ < -\nabla J(x^*)\,,\, x - x^* > \leq 0, \quad \forall\, x \in U.$$

Ceci donne

$$< [x^* - \rho \nabla J(x^*)] - x^*\,,\, x - x^* > \leq 0, \quad \forall\, x \in U.$$

En utilisant le Théorème de projection cette dernière inégalité est équivalente à l'égalitè

$$x^* = P_U(x^* - \rho \nabla J(x^*)).$$

Alors la solution du problème de minimisation (4.21) est un point fixe de l'application $g : U \mapsto O$ donnée par

$$g(x) = P_U(x - \rho \nabla J(x)).$$

Une idée naturelle pour approcher numériquement ce point fixe est d'utiliser une méthode iterative, qui consiste à construire une suite $x^{(k)} \in U$ par reccurence :

$$x^{(k+1)} = P_U(x^{(k)} - \rho \nabla J(x^{(k)})), \quad k \in I\!N$$

avec $\rho > 0$ donné. Il est aussi possible de faire varier ρ à chaque pas. Donc l'algorithme général sera dans ce cas :

$$\begin{cases} x^{(k+1)} = P_U(x^{(k)} - \rho_k \nabla J(x^{(k)})), \quad k \in I\!N \\ x^{(0)} \quad \text{donné } \in U \end{cases} \tag{4.26}$$

avec $\rho_k > 0$ données. C'est la **méthode de gradient avec projection à pas variable.** On l'appelle **méthode de gradient avec projection à pas fixe** si ρ_k est indépendent de k.
On a le résultat suivant :

Théorème 4.22. *Soit $U \subset I\!R^n$ un ensemble fermé, convexe et non vide et $J : I\!R^n \mapsto I\!R$ une fonction elliptique telle que ∇J soit lipschitzienne*
(Donc il existe $M > 0$ et $\alpha > 0$ tels que $\forall x, y \in I\!R^n$ on a

$$< \nabla J(x) - \nabla J(y)\,,\, x - y > \geq \alpha \|x - y\|^2$$

et

$$\|\nabla J(x) - \nabla J(y)\| \leq M\|x - y\|. \quad)$$

Supposons qu'il existe des nombres réels a_1 et a_2 satisfaisant

$$0 < a_1 \leq a_2 < \frac{2\alpha}{M^2}$$

tels que

$$a_1 \leq \rho_k \leq a_2 \quad \forall\, k \in I\!N.$$

Alors la méthode (4.26) *converge et la convergence est géométrique*
(c'est à dire, il existe $C \geq 0$ et $\beta \in [0, 1[$ tels que

$$\|x^{(k)} - x^*\| \leq C\beta^k, \quad \forall\, k \in I\!N. \quad)$$

Démonstration. Par définition nous avons

$$x^{(k+1)} = P_U\left(x^{(k)} - \rho_k \nabla J(x^{(k)})\right)$$

et aussi

$$x^* = P_U\left(x^* - \rho_k \nabla J(x^*)\right)$$

En faisant la différence et en utilisant la Partie **c)** du Théorème 4.21 (Théorème de projection), on déduit

$$\|x^{(k+1)} - x^*\| \leq \|x^{(k)} - x^* - \rho_k \left[\nabla J(x^{(k)}) - \nabla J(x^*)\right] \|$$

c'est à dire

$$\|x^{(k+1)}-x^*\|^2 \leq \|x^{(k)}-x^*\|^2+\rho_k^2\|\nabla J(x^{(k)})-\nabla J(x^*)\|^2-2\rho_k < x^{(k)}-x^* , \nabla J(x^{(k)})-\nabla J(x^*) > .$$

En utilisant les hypothèses sur J on déduit

$$\|x^{(k+1)} - x^*\|^2 \leq (1 - 2\alpha\rho_k + M^2\rho_k^2)\|x^{(k)} - x^*\|^2.$$

La suite de la preuve est exactement comme dans la preuve du Théorème 3.6 (à partir de l'inégalité (3.12)). □

Remarque : Cette méthode est applicable uniquement si on peut calculer facilement la projection P_U, par exemple si U est un pavé en $\mathbb{R}^n$. Pour un ensemble U donné par des contraintes inégalités larges, il n'est pas facile en général d'utiliser cette méthode.

Exemple : Supposons que J est une fonction quadratique associée à une matrice SDP, donc $J : \mathbb{R}^n \to \mathbb{R}$ est donnée par

$$J(x) = \frac{1}{2} < Ax , x > - < b, x > + c$$

avec $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée réelle et SDP, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.
Supposons aussi que $U = [0, +\infty[^n$.
Rappellons qu'on a dans ce cas :

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

$$P_U(x_1, x_2, \cdots, x_n) = (x_1^+, x_2^+, \cdots, x_n^+)$$

Alors (4.26) devient

$$x_i^{(k+1)} = \max\{x_i^{(k)} - \rho_k \left(Ax^{(k)} - b\right)_i , 0\},\ i = 1, \cdots, n \ \text{ et } \ k \in \mathbb{N}.$$

4.3.3 Méthodes de pénalisation exterieure

On considère toujours le problème de minimisation (4.21).
On introduira une fonction $\psi : \mathbb{R}^n \mapsto \mathbb{R}$ ayant les propriétés suivantes :

$$\begin{cases} \psi \quad \text{fonction continue et convexe} \\ \psi(x) \geq 0, \quad \forall\, x \in \mathbb{R}^n \\ \psi(x) = 0 \quad \text{si et seulement si} \quad x \in U. \end{cases} \tag{4.27}$$

Remarque : Une fonction ψ avec les propriétés de (4.27) est appellée **fonction de pénalisation extérieure pour** U. On introduit ensuite pour tout $\epsilon > 0$ "assez petit" une fonction $J_\epsilon : \mathbb{R}^n \mapsto \mathbb{R}$ définie par

$$J_\epsilon(x) = J(x) + \frac{1}{\epsilon}\psi(x). \tag{4.28}$$

La fonction J_ϵ s'appelle **fonction pénalisée de** J.
La **méthode de pénalisation exterieure** du problème (4.21) consiste à minimiser sur $\mathbb{R}^n$ la fonction J_ϵ, avec un $\epsilon > 0$ qui "tend vers 0". La partie $\frac{1}{\epsilon}\psi(x)$ pénalise l'éloignement de x par rapport à U.
On réduit donc le problème de minimisation avec contraintes (de J sur U) à un problème de minimisation sans contraintes (de J_ϵ sur $\mathbb{R}^n$). On applique pour la minimisation sans contraintes de J_ϵ tout algorithme vu en Chapitre 3.

Le résultat suivant de convergence justifie cette méthode :

Théorème 4.23. *Soit $J : \mathbb{R}^n \mapsto \mathbb{R}$ une fonction continue, coercive sur $\mathbb{R}^n$ et strictement convexe, $U \subset \mathbb{R}^n$ un ensemble convexe, fermé et non-vide et $\psi : \mathbb{R}^n \mapsto \mathbb{R}$ une fonction de pénalisation extérieure de U (donc satisfaisant les hypothèses de* (4.27)*).*
Alors pour tout $\epsilon > 0$ il existe un unique élément $x_\epsilon \in \mathbb{R}^n$ vérifiant

$$J_\epsilon(x_\epsilon) = \min_{x \in \mathbb{R}^n} J_\epsilon(x).$$

En plus on a

$$x_\epsilon \to x^* \quad \textit{pour} \quad \epsilon \mapsto 0$$

où x^* est l'unique point de U satisfaisant

$$J(x^*) \leq J(x), \quad \forall\, x \in U.$$

Démonstration. Il est facile de voir que pour tout $\epsilon > 0$ fixé, J_ϵ est une fonction continue, coercive et strictement convexe sur $\mathbb{R}^n$; donc il existe un unique point de minimum de J_ϵ sur $\mathbb{R}^n$ qu'on va noter par x_ϵ. Il est facile de voir que

$$J_\epsilon \geq J \quad \text{et} \quad J_\epsilon = J \ \text{ sur } \ U.$$

On déduit alors

$$J(x_\epsilon) \leq J_\epsilon(x_\epsilon) \leq J_\epsilon(x) = J(x), \quad \forall\, x \in U. \tag{4.29}$$

Ceci nous dit que la suite $J(x_\epsilon)$ est bornée, ce qui implique que la suite x_ϵ est bornée (car sinon, il y aurait une sous-suite de x_ϵ dont la norme tend vers $+\infty$, ce qui impliquerait, par coercivité de J, que $J(x_\epsilon) \mapsto +\infty$; ceci est impossible à cause de (4.29)).
Comme x_ϵ est borné en $\mathbb{R}^n$, le Théorème de Bolzano-Weierstrass nous dit qu'il existe une sous-suite $x_{\epsilon'}$ de x_ϵ tel que $x_{\epsilon'}$ converge vers un élément x^* de $\mathbb{R}^n$. En passant à la limite $\epsilon' \mapsto 0$ en (4.29) on déduit, grâce à la continuité de J que

$$J(x^*) \leq J(x), \quad \forall\, x \in U. \tag{4.30}$$

Pour montrer que x^* est un point de minimum de J sur U il reste à montrer que $x^* \in U$; grâce à la troisième propriété de (4.27) ceci revient à montrer que $\psi(x^*) = 0$. En effet nous avons :

$$0 \leq \psi(x_{\epsilon'}) = \epsilon' \left[J_{\epsilon'}(x_{\epsilon'}) - J(x_{\epsilon'})\right] \leq \epsilon' \left[J(x) - J(x_{\epsilon'})\right]$$

pour un $x \in U$ fixé. Par continuité de J nous avons que $J(x_{\epsilon'})$ converge, donc le term $J(x) - J(x_{\epsilon'})$ est borné. Ceci implique que $\epsilon' [J(x) - J(x_{\epsilon'})] \to 0$ pour $\epsilon' \to 0$. On obtient alors $\psi(x_{\epsilon'}) \mapsto 0$. Par continuité de ψ on déduit $\psi(x^*) = 0$, c'est à dire $x^* \in U$.
Comme J est strictement convexe, x^* est l'unique point de minimum de J sur U.

Il nous reste à démontrer que toute la séquence x_ϵ converge vers x^*. Montrons ce résultat par absurd : supposons que x_ϵ ne converge pas vers x^* ; ceci implique qu'il existe une sous-suite $x_{\epsilon''}$ de x_ϵ et un $\delta > 0$ tels que

$$\|x_{\epsilon''} - x^*\| \geq \delta. \tag{4.31}$$

En réproduisant pour $x_{\epsilon''}$ le même argument que pour x_ϵ, on peut extraire une sous-suite $x_{\epsilon'''}$ de $x_{\epsilon''}$ qui converge vers x^* (l'unicité de x^* est ici essentielle) ; ceci contredit (4.31) et finit la preuve du Théorème.

□

Remarque : Cette méthode est utile s'il est facile de construire une fonction ψ avec les propriétés (4.27).
Exemple : On appelle **distance** d'un point $x \in I\!R^n$ à l'ensemble U, notée $\text{dist}(x, U)$ la quantité donnée par

$$\text{dist}(x, U) = \|x - P_U(x)\|.$$

(rappellons que U est fermé et convexe). On définit alors la fonction $\psi : I\!R^n \mapsto I\!R$ par

$$\psi(x) = \text{dist}^2(x, U) = \|x - P_U(x)\|^2.$$

Il est très facile de montrer toutes les propriétés de (4.27) sauf la convexité de ψ. Pour la convexité de ψ dans un cas particulier de U voir TD.

4.3.4 Méthode d'Uzawa

La méthode d'Uzawa est basée sur la formulation duale et la recherche des points selle et elle est bien adaptée pour des ensembles de contraintes définies par des inégalités larges.

On va considérer ici des ensembles de contraintes comme dans la Section 4.2, c'est à dire

$$U = O = \{x \in I\!R^n, \;\; \theta_i(x) \leq 0, \quad \forall\, i = 1, 2, \cdots, m\}$$

avec $\theta_1, \theta_2, \cdots, \theta_m : I\!R^n \mapsto I\!R$, $m \in I\!N^*$. On peut encore écrire

$$U = O = \{x \in I\!R^n, \;\; \theta(x) \leq 0\}.$$

Rappellons que nous considèrons le problème de minimisation (4.21).

Nous faisons dans cette partie les hypothèses suivantes :

1. J est elliptique, donc J est de classe C^1 et il existe $\alpha > 0$ tel que

$$< J(x) - J(y)\,,\, x - y > \geq \alpha \|x - y\|^2, \quad \forall\, x, y \in I\!R^n$$

2. Pour tout $k = 1, \cdots, m$ on a que θ_k est une fonction convexe et de classe C^1. En plus θ est lipschitzienne, donc il existe $M > 0$ tel que

$$\|\theta(x) - \theta(y)\| \leq M\|x - y\|, \quad \forall\, x, y \in I\!R^n.$$

3. Les contraintes de O sont qualifiés en tout point $x \in O$.

Nous introduisons l'application $L : I\!R^n \times I\!R^m_+ \mapsto I\!R$ donnée par

$$L(x,p) = J(x) + < p, \theta(x) > = J(x) + \sum_{k=1}^{m} p_k \theta_k(x), \quad \forall\, x \in I\!R^n, \ \ \forall\, p \in I\!R^m_+.$$

Grâce à la convexité de toutes les fonctions θ_k l'ensemble U est convexe (Proposition 4.1). Comme l'ensemble U est fermé et convexe et J elliptique, il existe un unique point de minimum $x^* \in U$ de J sur U. Le Corollaire 4.19 nous dit alors qu'il existe $p^* = (p_1^*, p_2^*, \cdots, p_m^*)^T \in I\!R^m_+$ tel que (x^*, p^*) est un point selle de L. Du Théorème 4.15 on déduit alors que

$$\sup_{p \in I\!R^m_+} \inf_{x \in I\!R^n} L(x,p) = L(x^*, p^*) = \inf_{x \in I\!R^n} \sup_{p \in I\!R^m_+} L(x,p).$$

Introduisons la fonction $H : I\!R^m_+ \mapsto \overline{I\!R}$ par

$$H(p) = \inf_{x \in I\!R^n} L(x,p).$$

La Proposition 2.12 et les hypothèses faites sur J et les θ_k nous disent que pour tout $p \in I\!R^m_+$ fixé l'application

$$x \in I\!R^n \mapsto L(x,p) \in I\!R$$

est elliptique. Alors il existe un unique point de minimum de cette aplication sur $I\!R^n$, que nous allons noter $\bar{x}^p$. Donc $\bar{x}^p \in I\!R^n$ est tel que

$$L(\bar{x}^p, p) \leq L(x,p) \quad \forall\, x \in I\!R^n.$$

On déduit alors que H prend ses valeurs en $I\!R$ et que

$$H(p) = \min_{x \in R^n} L(x,p) = L(\bar{x}^p, p), \quad \forall\, p \in I\!R^m_+.$$

Le problème **dual** de (4.21) sera

$$\text{Trouver } \ \tilde{p}^* \in I\!R^m_+ \ \text{ tel que } \quad H(\tilde{p}^*) = \sup_{p \in I\!R^m_+} H(p) \tag{4.32}$$

(le problème (4.21) sera appellé problème **primal**).
Remarque : Le problème de maximisation (4.32) peut s'écrire comme un problème de minimisation :

$$-H(\tilde{p}^*) = \inf_{p \in I\!R^m_+} [-H(p)].$$

Supposons que $\tilde{p}^*$ est bien une solution du problème de maximisation (4.32). Alors on a

$$\sup_{p\in I\!R^m_+}\inf_{x\in I\!R^n} L(x,p) = \sup_{p\in I\!R^m_+} H(p) = H(\tilde{p}^*) = L(\tilde{x}^*,\tilde{p}^*)$$

où on a noté $\tilde{x}^* = \bar{x}^{\tilde{p}^*}$.
Alors le point selle (x^*,p^*) que nous cherchons, se trouve parmi les solutions $(\tilde{x}^*,\tilde{p}^*)$ provenant de (4.32).

Comme l'ensemble des contraintes du problème dual (4.32) est un pavé (c'est l'ensemble $I\!R^m_+$) on peut alors utiliser un algorithme de gradient avec projection à pas variable :

$$p^{(k+1)} = P_{I\!R^m_+}[p_i^{(k)} + \rho_k \nabla H(p^{(k)})]$$

ce qui donne, en tenent compte de l'expression de la projection sur $I\!R^m_+$:

$$p^{(k+1)} = \max\left\{p_i^{(k)} + \rho_k \frac{\partial H}{\partial p_i}(p^{(k)}),\ 0\right\} \quad i = 1,\cdots,m \tag{4.33}$$

avec $\rho_k > 0$.

Il nous reste à calculer $\frac{\partial H}{\partial p_i}$.
Nous avons le **calcul formel** suivant, en supposant que l'application $p \mapsto \bar{x}^p$ est de classe C^1 : comme $H(p) = J(\bar{x}^p) + \sum_{j=1}^m p_j\theta_j(\bar{x}^p)$ alors

$$\frac{\partial H}{\partial p_i} = <\nabla J(\bar{x}^p)\,,\,\frac{\partial \bar{x}^p}{\partial p_i}> \ +\theta_i(\bar{x}^p) + \sum_{j=1}^m p_j <\nabla\theta_j(\bar{x}^p)\,,\,\frac{\partial \bar{x}^p}{\partial p_i}>$$

Mais comme $\nabla J(\bar{x}^p) + \sum_{j=1}^m p_j\nabla\theta_j(\bar{x}^p) = 0$ (conséquence immédiate du fait que $\bar{x}^p$ est le point de minimum de la fonction $x \mapsto L(x,p)$) alors il reste

$$\frac{\partial H}{\partial p_i} = \theta_i(\bar{x}^p).$$

Le résultat précis avec démonstration rigoureuse est donné dans le lemme suivant :

Lemme 4.24. *La fonction H est de classe C^1 et on a $\forall p \in I\!R^m_+$:*

$$\frac{\partial H}{\partial p_i}(p) = \theta_i(\bar{x}^p), \quad i = 1,2,\cdots,m.$$

Démonstration. Soit $p \in I\!R^m_+$ et $t \in I\!R$ tel que $p + te_i \in I\!R^m_+$. Le but principal est de montrer que la limite pour $t \mapsto 0$ de $\frac{1}{t}[H(p+te_i) - H(p)]$ existe et de calculer cette limite.

Etape 1. On montre d'abord que l'application $p \mapsto \bar{x}^p$ est continue.
Soit $p \geq 0$ fixé et $p' \geq 0$ tel que p' converge vers p. Les vecteurs $\bar{x}^p$ et $\bar{x}^{p'}$ satisfont respectivement

$$\nabla J(\bar{x}^p) + \sum_{j=1}^m p_j\nabla\theta_j(\bar{x}^p) = 0$$

et

$$\nabla J(\bar{x}^{p'}) + \sum_{j=1}^{m} p'_j \nabla \theta_j(\bar{x}^{p'}) = 0$$

En faisant la différence entre ces deux égalités et en faisant le produit scalaire avec $\bar{x}^{p'} - \bar{x}^p$ on trouve

$$< \nabla J(\bar{x}^{p'}) - \nabla J(\bar{x}^p)\,,\, \bar{x}^{p'} - \bar{x}^p > + \sum_j p'_j \; < \nabla \theta_j(\bar{x}^{p'}) - \nabla \theta_j(\bar{x}^p)\,,\, \bar{x}^{p'} - \bar{x}^p >$$
$$+ \sum_j (p'_j - p_j) \; < \nabla \theta_j(\bar{x}^p)\,,\, \bar{x}^{p'} - \bar{x}^p > = 0$$

On utilise ensuite l'inégalitée d'ellipticité de J et le fait que les fonctions θ_j sont convexes et on déduit

$$\alpha \, \|\bar{x}^{p'} - \bar{x}^p\|^2 \leq - \sum_j (p'_j - p_j) \; < \nabla \theta_j(\bar{x}^p)\,,\, \bar{x}^{p'} - \bar{x}^p >$$

En utilisant l'inégalité de Cauchy-Schwarz, on trouve

$$\alpha \, \|\bar{x}^{p'} - \bar{x}^p\|^2 \leq \sum_j |p'_j - p_j| \, \|\nabla \theta_j(\bar{x}^p)\| \cdot \|\bar{x}^{p'} - \bar{x}^p\|$$

ce qui nous donne immédiatement

$$\bar{x}^{p'} \mapsto \bar{x}^p \quad \text{ si } \quad p' \mapsto p.$$

On a donc la continuité souhaitée.

Etape 2. Nous avons d'une part

$$H(p) = L(\bar{x}^p, p) \leq L(\bar{x}^{p+te_i}, p)$$

et d'autre part

$$H(p + te_i) = L(\bar{x}^{p+te_i}, p + te_i) \leq L(\bar{x}^p, p + te_i).$$

Ceci nous donne

$$H(p+te_i) - H(p) \leq L(\bar{x}^p, p+te_i) - L(\bar{x}^p, p) = J(\bar{x}^p) + \sum_j p_j \theta_j(\bar{x}^p) + t\theta_i(\bar{x}^p) - J(\bar{x}^p) - \sum_j p_j \theta_j(\bar{x}^p)$$

c'est à dire

$$H(p + te_i) - H(p) \leq t\theta_i(\bar{x}^p). \tag{4.34}$$

Nous avons aussi

$$H(p + te_i) - H(p) \geq L(\bar{x}^{p+te_i}, p + te_i) - L(\bar{x}^{p+te_i}, p) = J(\bar{x}^{p+te_i}) +$$
$$\sum_j p_j \theta_j(\bar{x}^{p+te_i}) + t\theta_i(\bar{x}^{p+te_i}) - J(\bar{x}^{p+te_i}) - \sum_j p_j \theta_j(\bar{x}^{p+te_i})$$

c'est à dire

$$H(p+te_i) - H(p) \geq t\theta_i(\bar{x}^{p+te_i}). \tag{4.35}$$

En divisant (4.34) et (4.35) par $t \neq 0$ on trouve

$$\theta_i(\bar{x}^{p+te_i}) \leq \frac{H(p+te_i) - H(p)}{t} \leq \theta_i(\bar{x}^p), \quad \text{si} \quad t > 0$$

et

$$\theta_i(\bar{x}^p) \leq \frac{H(p+te_i) - H(p)}{t} \leq \theta_i(\bar{x}^{p+te_i}), \quad \text{si} \quad t < 0.$$

En passant à la limite $t \mapsto 0$ et en utilisant le fait que $\bar{x}^{p+te_i} \mapsto \bar{x}^p$ (conséquence de l'Etape 1) on déduit

$$\frac{H(p+te_i) - H(p)}{t} \to \theta_i(\bar{x}^p) \quad \text{si} \quad t \mapsto 0.$$

En utilisant de nouveau le résultat de l'Etape 1 on obtient le résultat attendu. □

On peut alors remplacer l'expression de $\frac{\partial H}{\partial p_i}$ dans (4.33) et on trouve

$$p^{(k+1)} = \max\left\{p_i^{(k)} + \rho_k\theta_i(\bar{x}^{p^{(k)}}),\ 0\right\} \quad i = 1, \cdots, m$$

On va noter dans la suite $x^{(k)} = \bar{x}^{p^{(k)}}$. Alors **l'algorithme d'Uzawa** est le suivant :

pas 1. *On pose $k = 0$, on choisit $p^{(0)} \in \mathbb{R}^m_+$ (par exemple $p^{(0)} = 0$) et on choisit $k_{max} \in \mathbb{N}$ assez grand (par exemple $k_{max} = 1000$) et $\epsilon > 0$ assez petit (par exemple $\epsilon = 10^{-6}$).*

pas 2. *Calculer $x^{(k)} \in \mathbb{R}^n$ qui minimise sur $\mathbb{R}^n$ la fonction $x \in \mathbb{R}^n \to J(x) + \sum_{j=1}^m p_j^{(k)}\theta_j(x)$ (appliquer l'un des algorithme de minimisation sans contrainte vus en Chapitre 3).*
Si $(k \geq 1)$ alors
Si $\|x^{(k)} - x^{(k-1)}\| \leq \epsilon$ alors $x^{(k)}$ est le point de minimum recherché (l'algorithme a convergé)
Sinon, va au **pas 3.**

pas 3.
Si $(k = k_{max})$ alors l'algorithme diverge.
Sinon, faire :

$$p_i^{(k+1)} = \max\left\{p_i^{(k)} + \rho_k\theta_i(x^{(k)}),\ 0\right\} \quad i = 1, \cdots, m$$

ensuite $k = k + 1$, retour au **pas 2.**

Nous finissons par le résultat de convergence suivant :

Théorème 4.25. *(convergence de l'algorithme d'Uzawa)*
Soient $a_1, a_2 \in \mathbb{R}$ tels que

$$0 < a_1 \leq a_2 < \frac{2\alpha}{M^2}.$$

Supposons que les facteurs ρ_k sont tels que

$$a_1 \leq \rho_k \leq a_2 \quad \forall\, k \in I\!N.$$

Alors la suite $x^{(k)}$ générée par l'algorithme d'Uzawa converge vers x^* l'unique solution du problème de minimisation (4.21).

Démonstration. Il est clair que $x^{(k)}$ et x^* existent et sont uniques, grâce aux hypothèses du début du paragraph 4.3.4. On rappelle aussi qu'il existe $p^* \in I\!R^m_+$ tel que (x^*, p^*) soit un point selle de L.
Etape 1. Nous montrons les inégalités suivantes :

$$< \nabla J(x^{(k)})\,,\, x - x^{(k)} > \; + \sum_{j=1}^{m} p_j^{(k)} [\theta_i(x) - \theta_i(x^{(k)})] \geq 0, \quad \forall\, x \in I\!R^n \tag{4.36}$$

et

$$< \nabla J(x^*)\,,\, x - x^* > \; + \sum_{j=1}^{m} p_j^{*} [\theta_i(x) - \theta_i(x^{*})] \geq 0, \quad \forall\, x \in I\!R^n \tag{4.37}$$

Preuve de (4.36) *:* Dans l'inégalité

$$J(z) + \sum_{j=1}^{m} p_j^{(k)} \theta_j(z) \geq J(x^{(k)}) + \sum_{j=1}^{m} p_j^{(k)} \theta_j(x^{(k)}), \quad \forall\, z \in I\!R^n$$

prenons $z = x^{(k)} + t(x - x^{(k)})$ avec $x \in I\!R^n$ arbitraire et $t \in [0, 1]$. Grâce à la convexité des θ_j nous avons

$$\theta_j(x^{(k)} + t(x - x^{(k)})) - \theta_j(x^{(k)}) \leq t[\theta_j(x) - \theta_j(x^{(k)}]$$

ce qui donne avec l'inégalité précédente :

$$J(x^{(k)} + t(x - x^{(k)})) - J(x^{(k)}) + t \sum_{j=1}^{m} p_j^{(k)} [\theta_j(x) - \theta_j(x^{(k)}] \geq 0, \quad \forall t \in]0, 1].$$

En divisant par t et en passant à la limite $t \mapsto 0$ on obtient (4.36).
La preuve de (4.37) est analogue (remplacer $p^{(k)}$ et $x^{(k)}$ par p^* et x^* respectivement).
Etape 2. En posant $x = x^*$ en (4.36) et $x = x^{(k)}$ en (4.37) et en faisant la somme des deux inégalités, on obtient :

$$< \nabla J(x^*) - \nabla J(x^{(k)})\,,\, x^{(k)} - x^* > \; + \sum_{j=1}^{m} (p_j^* - p_j^{(k)}) [\theta_i(x^{(k)}) - \theta_i(x^*)] \geq 0$$

ce qui donne, en utilisant l'hypothèse d'ellipticité de J :

$$< p^{(k)} - p^*\,,\, \theta(x^{(k)}) - \theta(x^*) > \leq -\alpha \|x^{(k)} - x^*\|^2. \tag{4.38}$$

D'autre part, nous avons par hypothèse

$$p^{(k+1)} = P_{I\!R^m_+}\left(p^{(k)} + \rho_k \theta(x^{(k)}).\right) \tag{4.39}$$

Comme p^* réalise le minimum sur $I\!R^*_+$ de la fonction
$p \to -J(x^*) - < p^* \, , \, \theta(x^*) >$ nous avons

$$p^* = P_{I\!R^m_+}\left(p^* + \rho_k \theta(x^*)\right) \tag{4.40}$$

En faisant la différence entre (4.39) et (4.40) et en utilisant la partie **c)** du Théorème de projection, on déduit

$$\|p^{(k+1)} - p^*\| \leq \|p^{(k)} - p^* + \rho_k[\theta(x^{(k)}) - \theta(x^*)]\|$$

c'est à dire,

$$\|p^{(k+1)} - p^*\|^2 \leq \|p^{(k)} - p^*\|^2 + \rho_k^2 \|\theta(x^{(k)}) - \theta(x^*)\|^2 + 2\rho_k \, < p^{(k)} - p^* \, , \, \theta(x^{(k)}) - \theta(x^*) >$$

En utilisant le fait que θ est lipschitzienne, on déduit

$$\|p^{(k+1)} - p^*\|^2 \leq \|p^{(k)} - p^*\|^2 + \rho_k^2 M^2 \|x^{(k)} - x^*\|^2 + 2\rho_k \, < p^{(k)} - p^* \, , \, \theta(x^{(k)}) - \theta(x^*) > \tag{4.41}$$

On obtient ensuite facilement des inégalités (4.38) et (4.41) :

$$\|p^{(k+1)} - p^*\|^2 \leq \|p^{(k)} - p^*\|^2 + \rho_k^2 M^2 \|x^{(k)} - x^*\|^2 - 2\rho_k \alpha \, \|x^{(k)} - x^*\|^2$$

c'est à dire

$$\|p^{(k+1)} - p^*\|^2 \leq \|p^{(k)} - p^*\|^2 + (\rho_k^2 M^2 - 2\rho_k \alpha) \|x^{(k)} - x^*\|^2 \tag{4.42}$$

Introduisons la fonction $\psi : I\!R \to I\!R$ donnée par

$$\psi(\rho) = 2\alpha\rho - M^2\rho^2$$

On observe que $\psi(\rho) > 0, \;\; \forall \rho \in \left]0, \frac{2\alpha}{M^2}\right[$ et que

$$\psi(\rho) \geq \min\{\psi(a_1), \; \psi(a_2)\} \quad \forall \, \rho \in [a_1, a_2].$$

En choisissant alors $\beta = \min\{\psi(a_1), \; \psi(a_2)\} > 0$ on déduit

$$2\alpha\rho_k - M^2\rho_k^2 \geq \beta \quad \forall k \in I\!N.$$

Ceci nous permet de déduire de (4.42) :

$$\|p^{(k+1)} - p^*\|^2 \leq \|p^{(k)} - p^*\|^2 - \beta \|x^{(k)} - x^*\|^2 \tag{4.43}$$

On obtient alors

$$\|p^{(k+1)} - p^*\|^2 \leq \|p^{(k)} - p^*\|^2$$

ce qui nous dit que la suite réelle $\|p^{(k)} - p^*\|^2$ est décroissante. Comme elle est aussi positive, elle est convergente, et soit $l \geq 0$ sa limite. On obtient de (4.43) :

$$\|x^{(k)} - x^*\|^2 \leq \frac{1}{\beta}\left[\|p^{(k)} - p^*\|^2 - \|p^{(k+1)} - p^*\|^2\right]$$

Comme le membre de droite de cette inégalité converge vers 0, on déduit que $x^{(k)}$ converge vers x^*. Ceci finit la preuve. □

FIN

Bibliographie

[1] D. Azé, J.B. Hiriart-Urruty, *Analyse variationnelle et optimisation*, Cépaduès, 2010

[2] M. Bergounioux, *Optimisation et contrôle des systèmes linéaires*, Dunod, Paris, 2001.

[3] P. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, Paris, 1982.

[4] P. Ciarlet, B. Miara, J.M. Thomas, *Exercices l'analyse numérique matricielle et d'optimisation*, Masson, Paris, 1987.