Méthodes numériques d'optimisation

Marie Postel[1]
Laboratoire Jacques-Louis Lions
Sorbonne Université

20 janvier 2024

1. marie.postel@upmc.fr, http ://www.ljll.math.upmc.fr/postel

Table des matières

[1] Bonnans, J.F., Gilbert, J. C., Lemaréchal, C. et Sagastizábal. C. (2006) Numerical optimization, theoretical and practical aspects. Springer Verlag. *En français chez Springer : Optimisation numérique, Aspects théoriques et pratiques*

[2] Chemin,, J.-Y. (2012) Topologie et Calcul Différentiel. Polycopié UE 3M360 `https://www.licence.math.upmc.fr/UE/LM360/fichiers/10/1/LM360_2012defI-hyper.pdf`

[3] Delbos, F. et Gilbert, J. Ch. (2005) Global linear convergence of an augmented Lagrangian algorithm for solving convex quadratic optimization problems. Journal of Convex Analysis, 12, 45D69. 25, 26, 31

[4] Lelong-Ferrand, J. et Arnaudies, J.-M. (1977) Cours de mathématiques, Tome 2, Analyse, Dunod Université

[5] Nocedal, J. et Wright, S. J. (2006) Numerical Optimization, Second Edition, Springer

[6] Privat. Y. Cours d'Optimisation : aspects théoriques et algorithmes. `https://www.ljll.math.upmc.fr/privat/`

1 Introduction

1.1 Définition du problème d'optimisation

Dans le cadre de ce cours on étudie les problèmes d'optimisation continue qui consistent à trouver les extrema d'une fonction $f(x)$ définie sur $K \in V$ à valeur dans $\mathbb{R}$, où V est un espace vectoriel normé.
On s'intéresse à la détermination de

$$(P) \qquad \begin{cases} \inf & f(x) \\ s.c. & c^E x) = 0 \\ s.c. & c^I(x) \preceq 0 \\ & x \in \mathbb{R}^n \end{cases}$$

avec

$$\begin{array}{rcl} f & : & \mathbb{R}^n \longrightarrow \mathbb{R}, \\ c^E & : & \mathbb{R}^n \longrightarrow \mathbb{R}^m, \\ c^I & : & \mathbb{R}^n \longrightarrow \mathbb{R}^p, \\ f & c & \text{régulières.} \end{array}$$

Les contraintes d'égalité ou d'inégalité s'exprimeront parfois également sous la forme d'une contrainte d'appartenance à une partie convexe $K \subset \mathbb{R}^n$. Dans le cas où le minimum de f est atteint on utlisera la notation

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

La notation $c^I(x) \preceq 0$ n'est pas universelle. Dans ce document elle signifie $c^I_j(x) \leq 0$ pour tout $j = 1, \ldots, p$.
Les applications pratiques sont diverses

- *Etats d'équilibre de systèmes physiques :* la propagation d'ondes (lumineuses ou acoustiques) peut se modéliser par des parcours minimisant le temps de trajet ; la structure d'une protéine est celle qui minimise l'énergie potentielle intramoléculaire ; le problème de la chaînette décrit la courbe d'équilibre d'une ligne (chaîne ou câble) suspendue entre deux points, homogène, inextensible, sans rigidité en flexion, soumise à son seul poids.
- *Etats d'équilibre économiques :* du point de vue du consommateur (équilibre entre la consommation et le travail), du point de vue du producteur (optimisation d'un plan de production en fonction des prix de revient et de vente des biens produits)
- *Contrôle* : la fonction f mesure les performances ou le coût d'un système physique, le vecteur x mesure les paramètres du système qu'on peut faire varier. *Exemple* : f consommation d'une voiture, x forme du véhicule, puissance du moteur, vitesse.
- *Identification de paramètres* : la fonction f mesure la différence entre des observations dépendant de paramètres inconnus et les mêmes quantités calculées à partir d'un modèle. *Exemple :* échographie. f est la différence entre les signaux sonores mesurés et ceux renvoyés par un obstacle connu. x module de compressibilité acoustique du milieu observé.

Dans tous les cas, il est nécessaire de relier les paramètres à la fonction à minimiser par l'intermédiaire d'un modèle.

1.1.1 Définition du minimum

Appelons

$$Y = \{x \in K,\ c^E(x) = 0,\ c^I(x) \preceq 0\} \tag{1.1}$$

l'ensemble des vecteurs vérifiant toutes les contraintes. On dira que $x^\star \in Y$ **réalise**

- un *minimum local* s'il existe $\varepsilon > 0$ tel que

$$f(x^\star) \leq f(x) \quad \text{pour tout} \quad x \in Y \quad \text{t.q. } \|x - x^\star\| \leq \varepsilon.$$

— un *minimum local strict ou isolé* s'il existe $\varepsilon > 0$ tel que

$$f(x^\star) < f(x) \quad \text{pour tout} \quad x \in Y \quad \text{t.q. } x \neq x^\star \text{ et } \|x - x^\star\| \leq \varepsilon.$$

— un *minimum global* si

$$f(x^\star) \leq f(x) \quad \text{pour tout} \quad x \in Y.$$

— un *minimum global strict ou isolé* si

$$f(x^\star) < f(x) \quad \text{pour tout} \quad x \in Y \quad \text{t.q. } x \neq x^\star.$$

On dit parfois que $x^\star$ **est un minimum** de $f(x)$, mais c'est un abus de langage. Le terme exact, si $x^\star$ **réalise** un minimum de f, est qu'il est un **minimiseur** de f, qu'on note

$$x^\star = \underset{x \in Y}{\operatorname{argmin}} f(x)$$

Dans le cas où la fonction objectif présente un minimum de régularité, on peut énoncer le résultat suivant

Théorème 1.1. *Soit une fonction f continue sur un sous ensemble C fermé de $\mathbb{R}^n$. Si l'une des hypothèses suivantes est vérifiée*

— *C est borné,*

— *C est non borné et f coercive (*$\lim_{\|x\| \to \infty} f(x) = +\infty$*),*

alors f admet un minimum sur C

Preuve On commence par montrer qu'il existe x_0 tel que l'ensemble $C_{x_0} = \{x, f(x) \leq f(x_0)\}$ est borné.

— Si C est borné, $C_{x_0} \subset C$ est borné pour tout x_0.

— Si C n'est pas borné et f coercive, supposons que $\forall x_0 \in C$ C_{x_0} soit non borné, d'après la coercivité il existe $M(x_0) > 0$, tel que si $\|x\| \geq M(x_0)$ $f(x) \geq f(x_0)$. Choisissons $y \in C_{x_0}$ non borné, tel que $\|y\| \geq M(x_0)$ pour avoir une contradiction.

Donc f continue atteint ses bornes sur $x^\star$ sur l'ensemble $C_{x_0} = \{x, f(x) \leq f(x_0)\}$ fermé borné. Donc il existe $x^\star \in C_{x_0}$ tel que $f(x^\star) = \min_{x \in C_{x_0}} f(x) = \min_{x \in C} f(x)$. ■

1.1.2 Deux classes de méthodes

— Méthodes déterministes : utilisation des propriétés de régularité de la fonction objectif, méthodes de descente
  — avantages : rapidité
  — inconvénient : possibilité d'être piégé près d'un minimum local

— Méthodes stochastiques : exploration aléatoire du domaine de recherche, pas besoin de régularité
  — avantages : robustesse, minimum global
  — inconvénient : lenteur

Le chapitre 2 présente les méthodes déterministes pour l'optimisation sans contraintes. Les méthodes d'optimisation avec contraintes sont présentées au chapitre **??** pour les contraintes d'égalité et au chapitre pour les contraintes d'inégalité. Avant cela, quelques rappels sur la différentiabilité des fonctions de plusieurs variables et sur la convexité des ensembles et des fonctions.

1.2 Rappels de calcul différentiel

Les méthodes d'optimisation utilisant les propriétés de régularité des fonctions qu'on cherche à minimiser, nous commençons par des rappels de ces propriétés. On renvoie au polycopié de calcul différentiel de L3 [2] pour la plupart des démonstrations.

1.2.1 Différentiabilité au premier ordre

Définition 1.1. Soit f une application de $\mathbb{R}^n$ dans $\mathbb{R}^{n'}$ On dit que f est différentiable au sens de Fréchet en x s'il existe une application linéaire L continue de $\mathbb{R}^n$ dans $\mathbb{R}^{n'}$ telle que pour tout $h \in \mathbb{R}^n$

$$f(x+h) = f(x) + L(h) + o(\|h\|),$$

et on note $Df(x) = L$ la différentielle de f au point x.

Définition 1.2. On dit que f est différentiable au sens de Gâteaux en x si pour tout $h \in \mathbb{R}^n$, la fonction $g(t) = f(x+th)$ est dérivable. On note $Df(x)$ l'application différentielle de f en x qui s'applique à $h \in \mathbb{R}^n$

$$Df(x)h = \frac{df(x+th)}{dt}_{|t=0}.$$

Quelques définitions et propriétés s'appliquant aux fonctions différentiables

— Si une fonction est différentiable (au sens de Fréchet) alors sa différentielle au sens de Gâteaux existe (la réciproque n'étant pas toujours vraie).

— Si V est un espace de Hibert, si f est différentiable, le théorème de représentation de Riez conduit à la définition du gradient de $f : \nabla f(x) \in V$

$$\langle \nabla f(x), y \rangle = Df(x)y.$$

— Dérivée directionnelle de f dans la direction $d \in \mathbb{R}^n$

$$\lim_{\alpha \to 0} \frac{f(x+\alpha d) - f(x)}{\alpha} = \langle \nabla f(x), d \rangle.$$

— Direction de descente $d \in \mathbb{R}^n$

$$\langle \nabla f(x), d \rangle < 0.$$

— Si $V = \mathbb{R}^n$ le gradient est le vecteur des dérivées partielles $\left(\frac{\partial f(x)}{\partial x_j} \right)_{j=1,\dots n}$.

— On dit qu'une fonction $f : U \subset \mathbb{R}^n \to \mathbb{R}$ est de classe C^k si toutes ses dérivées partielles jusqu'à l'ordre k existent et sont continues sur U

Définition 1.3. Définition de la matrice jacobienne E et F evn, dimensions n et m, bases $\mathcal{B}$ et $\mathcal{B}'$.
Soit $U \subset E$ ouvert et $f : U \to F$, $f(x) = (f_1(x), \dots, f_m(x))$ différentiable en $a \in U$.
$Df(a) \in \mathcal{L}(E,F)$ donc il existe une unique **matrice jacobienne** $Jf(a)$, $m \times n$, qui représente $Df(a)$ dans les bases $\mathcal{B}$ et $\mathcal{B}'$.
Soit $h = (h_1, \dots, h_n) \in E$ on a $Df(a).h = Jf(a)h$.
$Df(a) = \begin{pmatrix} Df_1(a) \\ \vdots \\ Df_m(a) \end{pmatrix}$, soit $Df(a)h = \begin{pmatrix} Df_1(a)h \\ \vdots \\ Df_m(a)h \end{pmatrix} = J_f(a) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$.
Le $j$eme vecteur colonne de $J_f(a)$ est le vecteur $\begin{pmatrix} Df_1(a)e_j \\ \vdots \\ Df_m(a)e_j \end{pmatrix}$
(avec $(e_j)_m = \delta_{jm}$), et $Df_i(a)e_j = \frac{\partial f_i}{\partial x_j}(a)$,
ou $Jf(a) = (Jf(a)_{i,j})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} = (\frac{\partial f_i}{\partial x_j})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$

Et enfin rappelons les formules de Taylor

Théorème 1.2. *Formules de Taylor au premier ordre : Soit U un ouvert d'un espace vectoriel E et $f : U \longrightarrow F$ une application différentiable de U dans un espace vectoriel F. S'il existe $k \geq 0$ tel que $\|df(x)\| \leq k$ pour tout $x \in U$ alors quels que soient $x, y \in U$ tels que le segment $[x, y] \subset U$ on a*

$$\|f(y) - f(x)\| \leq k\|y - x\|. \tag{1.2}$$

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable sur S centrée en x. Pour tout $d \in \mathbb{R}^n$ t.q. $x + d \in S$ il existe $\alpha \in [0, 1]$ t. q.

$$f(x + d) = f(x) + \langle \nabla f(x + \alpha d), d\rangle. \tag{1.3}$$

Exercice 1.1. Montrer que les fonctions suivantes sont différentiables au premier ordre et calculer leur différentielle

- $f : \mathbb{R} \to \mathbb{R}$, dérivable sur $\mathbb{R}$
- $L : E \to F$, linéaire avec E et F evn
- $A : E \to F$, affine : $A(x) = L(x) + b$ avec L linéaire et $b \in F$
- $f(X) = \|X\|_2^2$, avec $X \in \mathbb{R}^n$

Corrigé :

- $f : \mathbb{R} \to \mathbb{R}$, dérivable sur $\mathbb{R}$
  $(\forall x \in \mathbb{R},\ f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ existe et est finie). $Df(x) = f'(x)$, $Df(x)(h) = f'(x)h$
- $L : E \to F$, linéaire avec E et F evn $f(X + h) - f(X) = f(h) = Df(X)h$ avec $Df(x) = f$. Si E dim n et F dim m, $f(X) = AX$ avec A matrice $m \times n$. Donc $df(x)h = Ah$
- $A : E \to F$, affine : $A(x) = L(x) + b$ avec L linéaire et $b \in F$
  Idem $df(x)h = Ah$
- $f(X) = \|X\|_2^2$, avec $X \in \mathbb{R}^n$ $f(X+h) - f(X) = \|X + h\|_2^2 - \|X\|_2^2 = \langle 2x, h\rangle + \|h\|^2 = \sum_{i=1}^n (2x_i + h_i)h_i = 2\sum_{i=1}^n x_i h_i + \sum_{i=1}^n h_i^2 = Df(X)h + \|h\|_2^2$
  avec $Df : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$, $(Df(X))_{ij} = \delta_{ij}x_i$ et $Df(X)h = \langle 2X, h\rangle$

■

Autres exemples

- $f : \mathbb{R}^2 \to \mathbb{R} : (x_1, x_2) \mapsto \begin{cases} 0 & \text{si } (x_1, x_2) = (0, 0) \\ \frac{x_1 x_2^3}{x_1^2 + x_2^6} & \text{sinon} \end{cases}$
- Différentielle d'une application bilinéaire : Soient E_1, E_2 et F trois evn. Alors toute application bilinéaire continue $B : E_1 \times E_2 \to F$ est différentiable en tout point $(a_1, a_2) \in E_1 \times E_2$ et sa différentielle est l'application linéaire $E_1 \times E_2 \to F$ définie par $(h, k) \mapsto B(a_1, k) + B(h, a_2)$.
- Différentielle d'une application multiliénaire : Toute application multilinéaire continue $L : E_1 \times \ldots \times E_k \to F$ est différentiable en tout point et sa différentielle au point $(a_1, \ldots, a_k) \in E_1 \times \ldots \times E_k$ est l'application linéaire de $E_1 \times \ldots \times E_k$ dans F :

$$\begin{aligned}(h_1, \ldots, h_k) \quad &\mapsto \quad L(h_1, a_2, \ldots, a_k) + L(a_1, h_2, a_3, \ldots, a_k) \\ &\quad\quad + \ldots + L(a_1, \ldots, a_{k-1}, h_k)\end{aligned}$$

Pour obtenir les différentielles des exemples ci-dessus on utilise le théorème suivant

Théorème 1.3. *Soit $L : E_1 \times \ldots \times E_k \to F$ $k - linaire$.*
Les conditions suivantes sont équivalentes :

1. *L est continue en tout point.*
2. *L est continue en $(0_{E_1}, \ldots, 0_{E_k})$.*

3. *il existe une constante* $C > 0$ *telle que pour tout* $(x_1, \ldots, x_k) \in E_1 \times \ldots \times E_k$, $\|L(x_1, \ldots, x_k)\|_F \leq C\|x_1\|_{E_1} \ldots \|x_k\|_{E_k}$.

On rappelle enfin la règle de calcul de la différentielle d'une composition :

Proposition 1.1. *Soient* E, F *et* G *trois espaces vectoriels normés. Soient* $U \subset E$ *et* $V \subset F$ *deux ouverts. Soient* $f : U \to V$ *et* $g : V \to G$ *deux applications telle que* $f(U) \subset V$.
Si f *est différentiable en* a *et* g *est différentiable en* $f(a)$, *alors* $g \circ f$ *est différentiable en* a *et*

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a).$$

A titre d'exemple, on montre comment on peut appliquer cette règle au calcul des dérivées partielles sur un evn de dimension finie : soit E un espace vectoriel de dimension n
$B = (e_1, \ldots, e_n)$ une base de E.
$f : E \to R^m$ une application différentiable en $a \in E$.
$\phi : R^n \to E$, $(x1, ..., xn) \mapsto \phi(x_1, ..., x_n) = \sum_{i=1}^n x_i e_i$.
$\phi^{-1}(e_j)$ est le j-ème vecteur de la base canonique de R^n.
On pose $g = f \circ \phi : R^n \to R^m$ et $b = \phi^{-1}(a)$. Alors

$$\begin{aligned} \frac{\partial g}{\partial x_j}(b) &= D_{\phi^{-1}(e_j)} g(b) = Df \circ \phi(b)(\phi^{-1}(e_j)) \\ &= Df(\phi(b)) \circ D\phi(b)(\phi^{-1}(e_j)) \\ &\quad = f(\phi(b) \circ \phi(\phi^{-1}(e_j)) = Df(a)(e_j) = D_{e_j} f(a) \end{aligned}$$

$D_{e_j} f(a)$ n'est autre que la j-ème dérivée partielle de f en a. Donc, les dérivées partielles de f en a sont égales aux dérivées partielles de g en b, par suite $Jf(a) = Jg(b)$. Ainsi le calcul des matrices jacobiennes, en dimension finie, se ramène au calcul d'une jacobienne dans R^n.

Exercice 1.2. 1. 0n considère $E = \mathbb{R}_3[X] = \{a_0 + a_1 X + a_2 X^2 + a_3 X^3, a_i \in \mathbb{R}\}$
dim $E = 4$, base $B = \{1, X, X^2, X^3\}$
On considère la fonction $f : E \to \mathbb{R}$, définie par

$$f(a_0 + a_1 X + a_2 X^2 + a_3 X^3) = e^{a_0} + \cos(a_1) + \cos(a_2) + a_2^2 \sin(a_3)$$

— Montrer que f différentiable
— Calculer les dérivées partielles de f au point $P_0 = 1 + 2X^2 + X^3$

Corrigé : On applique la méthode précédente : $g : \mathbb{R}_3[X] \to \mathbb{R}^4$ est l'isomorphisme canonique qui à un polynôme de degré 4 associe ses coordonnées dans la base canonique. Sa différentielle est Dg telle que $Dg(P)H = g(H)$.

Ici $P_0 = 1 + 2X^2 + X^3$ donc $g(P_0) = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$

$h : \mathbb{R}^4 \to \mathbb{R}$ est la forme linéaire qui à $a \in \mathbb{R}^4$ associe $e^{a_0} + \cos(a_1) + \cos(a_2) + a_2^2 \sin(a_3)$, sa différentielle est Dh telle que $Dh(a)\delta =< \nabla h(a), \delta >$ avec $\nabla h(a) = \begin{pmatrix} e^{a_0} \\ -sina_1 \\ -\sin(a_2) + 2a_2 \sin(a_3) \\ a_2^2 \cos(a_3) \end{pmatrix}$. Donc $\nabla h(g(P_0)) = \begin{pmatrix} e \\ 0 \\ -\sin(2) + 4\sin(1) \\ 4\cos(1) \end{pmatrix}$

Finalement comme $Df(P) = Dh(g(P))oDg(P)$ on obtient $Df(P_0)H = eH_0 + (4\sin(1) - \sin(2))H_2 + 4\cos(1)H_3$
■

Proposition 1.2. *Différentielle d'une combinaison linéaire : Soient E et F deux espaces vectoriels normés et soit $U \subset E$ un ouvert. Soient f et g deux applications définies dans U à valeurs dans F et λ et μ deux scalaires.*
Si f et g sont différentiables en a, il en est de même de $\lambda f + \mu g$ et

$$D(\lambda f + \mu g)(a) = \lambda Df(a) + \mu Dg(a).$$

Composition avec une application linéaire : Soient E, F et G trois espaces vectoriels normés et $a \in E$. Soit f une application d'un voisinage de a à valeurs dans F, différentiable en a.
1) Si $u \in \mathcal{L}(F, G)$, $u \circ f$ est différentiable en a et

$$D(u \circ f)(a) = u \circ Df(a)$$

2) Si $v \in \mathcal{L}(G, E)$ et si $v(b) = a$ alors $f \circ v$ est différentiable en b et

$$D(f \circ v)(b) = Df(a) \circ v$$

Exercice 1.3. Calculer la différentielle de $f(x) = \|Ax + b\|^2$ avec $A \in M_{m\times n}(\mathbb{R})$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$

Corrigé : On peut poser $v(x) = Ax + b$ affine, différentiable $Dv(x)h = Ah$, $g(x) = \|x\|^2$ différentiable $Dg(x)h = 2\langle x, h\rangle$. Donc $f = g \circ v$ différentiable et $Df(x) = Dg(v(x)) \circ Dv(x)$ donc $Df(x)h = 2\langle Ax + b, Ah\rangle$ ■

La composition avec une application bilinéaire suit la règle suivante :

Proposition 1.3. *Soient E, F et G trois espaces vectoriels normés. Soient U un ouvert de E, $f_1 : U \to F$ et $f_2 : U \to F$ deux applications différentiables en $a \in U$, à valeurs dans F.*
Alors pour toute application bilinéaire continue $\phi : F \times F \to G$, l'application de $E \to G$, $x \mapsto \phi(f_1(x), f_2(x))$ est différentiable en a et pour tout h in E

$$D\phi(f_1, f_2)(a)h = \phi(Df_1(a)h, f_2(a)) + \phi(f_1(a), Df_2(a)h)$$

Exercice 1.4. Calculer la différentielle de la fonction f qui à x associe $f(x) = \langle L(x), x\rangle$ avec $x \in E$ evn et $L \in \mathcal{L}(E)$ continue.

Corrigé : $\langle x, y\rangle$ est bilinéaire donc différentiable et $L(x)$ linéaire donc différentiable.

$$\begin{aligned} f(x+h) - f(x) &= \langle L(x+h), x+h\rangle - \langle L(x), x\rangle \\ &= \langle L(x) + L(h), x+h\rangle - \langle L(x), x\rangle \\ &= \langle L(x), h\rangle + \langle L(h), x\rangle + \langle L(h), h\rangle = g_x(h) + f(h) \end{aligned}$$

avec $g_x(h) = \langle L(x), h\rangle + \langle L(h), x\rangle$ linéaire en h. Or L continue donc $\langle L(h), h\rangle \leq \|L(h)\|\|h\| \leq C\|h\|^2$ donc $Df(x)h = \langle L(x), h\rangle + \langle L(h), x\rangle$ ■

Rappelons enfin quelques règles qu'on connait par coeur en dimension un mais peut-être moins en dimension quelconque

Proposition 1.4. *Différentielle d'un produit : Soient $f : U \to \mathbb{R}^m$ et $g : U \to \mathbb{R}$ deux applications différentiables en $a \in U$.*
Alors l'application $fg : U \to \mathbb{R}^m$ est différentiable et

$$D(fg)(a) = g(a).Df(a) + f(a).Dg(a).$$

Proposition 1.5. *Différentielle d'un quotient :*
Soit f une application définie dans U à valeurs dans $\mathbb{R}$, différentiable au point $a \in U$.
Si $f(a) \neq 0$ alors l'application $\dfrac{1}{f}$ est définie dans un voisinage de a et est différentiable en a :

$$D(\frac{1}{f})(a) = -\frac{1}{f(a)^2}Df(a).$$

Exercice 1.5. Calculer la différentielle de $f(x) = \frac{\langle Ax,x\rangle}{\|x\|_2^2}$ avec $x \in \mathbb{R}^n$ et $A \in S^n$, $x \neq 0_{\mathbb{R}^n}$.

Corrigé : On applique les règles précédentes : on pose $g(x) = \langle Ax, x\rangle$ qui est différentiable avec $Dg(x)h = 2\langle Ax, h\rangle$. On pose aussi $r(x) = \|x\|_2^2$ qui est différentiable avec $Dr(x)h = 2\langle x, h\rangle$. Avec ces notations $s(x) = \frac{1}{r(x)}$ pour $x \neq 0_{\mathbb{R}^n}$ est différentiable et $Ds(x)h = -\frac{2\langle x,h\rangle}{\|x\|_2^4}$. Finalement $f(x) = g(x)s(x)$ est différentiable et $Df(x) = s(x)Dg(x) + g(x)Ds(x)$ donc

$$\begin{aligned} Df(x)h &= \frac{2\langle Ax, h\rangle}{\|x\|^2} - \frac{2\langle Ax, x\rangle\langle x, h\rangle}{\|x\|_2^4} \\ &= 2\frac{\langle Ax, h\rangle\|x\|^2 - \langle Ax, x\rangle\langle x, h\rangle}{\|x\|_2^4} \end{aligned}$$

■

1.2.2 Différentiabilité au second ordre

Définition 1.4. Soit une fonction f différentiable sur V. f est deux fois différentiable en x_0 s'il existe une application linéaire $L(x_0) : V \to V'$ telle que

$$Df(x_0 + h) = Df(x_0) + L(x_0)h + o(\|h\|_V) \in V',$$

où V' désigne le dual topologique de V. La différentielle seconde de f, notée $D^2 f(x_0)$, est l'application $L(x_0) : V \to V'$

Définition 1.5. Soit une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$ deux fois différentiable sur $\mathbb{R}^n$. Le gradient de f est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^n$. On peut calculer ses dérivées partielles et former ainsi la *matrice hessienne* ou le *hessien*[a] de f, de $\mathbb{R}^n$ dans $\mathbb{R}^{n\times n}$.

$$Hf(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\dots n}.$$

a. En français, hessien ou hessienne ne prend pas de majuscule ni comme adjectif ni comme nom, contrairement à l'anglais

Théorème 1.4 (de Schwarz)**.** *Si la fonction f est deux fois différentiable, sa matrice hessienne est symétrique.*

Théorème 1.5 (Formule de Taylor au second ordre)**.** *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ deux fois différentiable sur S centrée en x.*

— *pour tout $d \in \mathbb{R}^n$ t.q. $x + d \in S$, il existe $\alpha \in [0, 1]$ t. q.*

$$f(x + d) = f(x) + \langle \nabla f(x), d\rangle + \frac{1}{2}\langle Hf(x + \alpha d)d, d\rangle. \tag{1.4}$$

On rappelle enfin le résultat général ([4])

Théorème 1.6 (Formule de Taylor à l'ordre p)**.** *Soit $f : E \longrightarrow F$, p fois différentiable et telle que $\|f^{(p)}(x)\| \leq k$*

sur $S \subset E$ centrée en x. Alors pour tout d tel que $x + d \in S$ on a

$$\|f(x+d) - f(x) - \sum_{q=1}^{p-1} f^{(q)}(x) \underbrace{(d, \ldots, d)}_{q \text{ fois}} \| \leq k \frac{1}{p!} \|d\|^p.$$

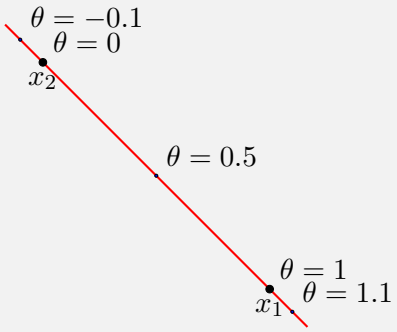
1.3 Rappels de convexité

1.3.1 Ensembles convexes

Dans ce paragraphe on rappelle quelques propriétés et définitions sur les ensembles convexes.

Définition 1.6. Ensemble affine Soit une *droite* passant par deux points x_1 et x_2 dans $\mathbb{R}^n$

$$x = \theta x_1 + (1-\theta) x_2, \quad \theta \in \mathbb{R}.$$



Un ensemble A est affine s'il contient toutes les droites passant par deux points quelconque de A

— Exemple : l'ensemble des solutions d'un système linéaire $A = \{x, Ax = b\}$

— *(Inversement tout ensemble affine peut s'exprimer comme l'ensemble des solutions d'un système linéaire*

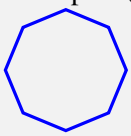
Définition 1.7. Ensemble convexe

Segment de droite reliant deux points x_1 et x_2

$$x = \theta x_1 + (1-\theta) x_2, \quad 0 \leq \theta \leq 1.$$

— Un ensemble C est convexe s'il contient tous les segments de droite reliant deux points quelconques de C

— Exemples (un *convexe*, deux **non convexes**)

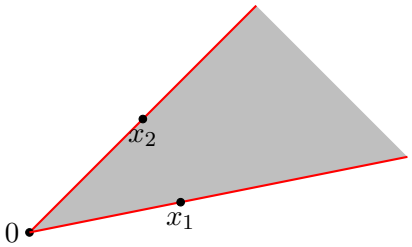


Définition 1.8. Combinaison convexe et enveloppe convexe Une *combinaison convexe* de k points de $\mathbb{R}^n$, $x_1, \ldots, x_k$ est un point $x \in \mathbb{R}^n$ tel que

$$x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k = \sum_{i=1}^{k} \theta_i x_i$$
$$\text{avec } \theta_1 + \cdots + \theta_k = 1 \text{ et } \theta_i \geq 0 \; \forall i = 1, \ldots, k$$

L'enveloppe convexe d'un ensemble E, noté *conv E*, est l'ensemble de toutes les combinaisons convexes de points de E

Définition : une *combinaison (conique) positive* de k points $(x_i)_{i=1,\dots,k}$ de $\mathbb{R}^n$ est un point de $\mathbb{R}^n$ de la forme
$x = \sum_{i=1}^{k} \theta_i x_i$
avec $\theta_i \geq 0$
$\forall i = 1, \dots, k$

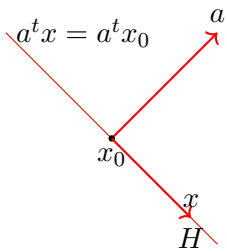


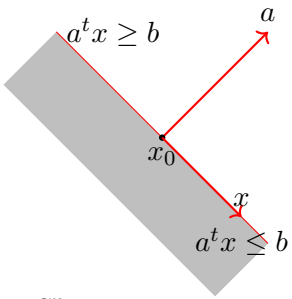
Définition : un sous ensemble $\hat{C}$ de $\mathbb{R}^n$ est un *cône convexe* si

$$\forall\, x_1, x_2 \in \hat{C},\ \forall\, \theta_1, \theta_2 \in \mathbb{R}^+,\ \theta_1 x_1 + \theta_2 x_2 \in \hat{C}$$

Définition : l'enveloppe conique d'un sous ensemble E de $\mathbb{R}^n$ est l'ensemble de toutes les combinaisons positives de points de E. On la note cône(E).
Propriété : l'enveloppe conique de E est le plus petit cône convexe contenant E.

Définition : un *hyperplan* est un ensemble de la forme
$H = \{x, a^t x = b\}$ avec $a \neq 0$
Définition : a est le *vecteur normal*
Propriété : un hyperplan est affine et convexe





Définition : un *demi-espace* est un ensemble de la forme
$E = \{x, a^t x \leq b\}$ avec $a \neq 0$
Propriété : un demi-espace est convexe

Notation : • S^n l'ensemble des matrices symétriques $n \times n$

$$A \in S^n \Leftrightarrow A = A^T$$

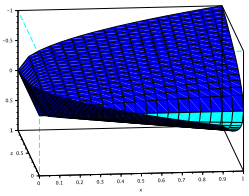
• S^n_+ l'ensemble des matrices symétriques semi définies positives $n \times n$

$$A \in S^n_+ \Leftrightarrow \begin{cases} A \in S^n \text{ et} \\ \forall\, x \in \mathbb{R}^n \langle Ax, x\rangle \geq 0 \end{cases}$$

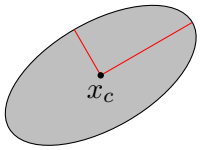
Propriété : S^n_+ est un cone convexe
• S^n_{++} l'ensemble des matrices symétriques définies positives $n \times n$

$$A \in S^n_{++} \Leftrightarrow \begin{cases} A \in S^n_+ \text{ et} \\ \langle Ax, x\rangle = 0 \Rightarrow x = 0 \end{cases}$$

Exemple : S^2_{++}



Définition : boule (euclidienne) de centre x_c et de rayon r
$B(x_c, r) = \{x, \|x - x_c\|_2 \leq r\} = \{x_c + ru, \|u\|_2 \leq 1\}$
Définition : un *ellipsoïde* est un ensemble de la forme
$E = \{x, (x - x_c)^t P^{-1}(x - x_c) \leq 1\}$ avec $P \in S^n_{++}$



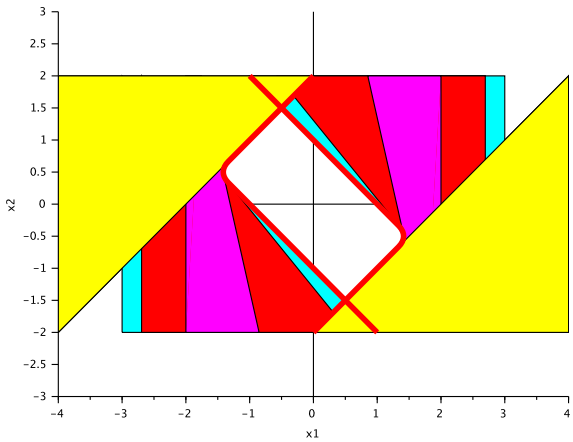
Propriété : $E = \{x_c + Au, \|u\|_2 \leq 1\}$ avec A carrée inversible
Définition : un *polyèdre* est un ensemble de la forme $P := \{x \in \mathbb{R}^n, Ax \preceq b\}$ avec $A \in \mathbb{R}^{m \times n}$ et $b \in \mathbb{R}^m$ fixés.
Propriété : un polyèdre est une intersection finie de demi espaces fermés de $\mathbb{R}^n$.

Propriété : Critère de convexité d'un ensemble C

- Vérifier la définition $\forall x_1, x_2 \in C, \quad \forall 0 \leq \theta \leq 1, \quad \theta x_1 + (1 - \theta)x_2 \in C$
- Montrer que C peut être obtenu à partir d'ensembles convexes simples (hyperplans, demi espaces, boules,...) par des opérations préservant la convexité
  - intersection
  - fonctions affines

Propriété : L'intersection d'un nombre quelconque de convexes est un convexe

Exemple 1.1. Par exemple considérons l'ensemble $S = \{x \in \mathbb{R}^m, |p(t)| \leq 1, \forall |t| \leq \pi/3\}$
avec $p(t) = x_1 \cos t + x_2 \cos\ 2t + \ldots + x_m \cos\ mt$. La figure ci-dessous en donne une représentation dans le cas $m = 2$. Il est trivial de prouver sa convexité en utilisant la propriété, mais c'est beaucoup plus difficile si on part de la définition.



Soit une fonction affine sur $\mathbb{R}^n$ $f(x) = Ax + b$ avec $A \in \mathbb{R}^{m \times n}$ et $b \in \mathbb{R}^m$
Propriété : L'image d'un convexe par f est un convexe

$$C \subset \mathbb{R}^n \text{ convexe } \Rightarrow f(C) = \{f(x), x \in C\} \text{ convexe}$$

Propriété : L'image réciproque d'un convexe par f est un convexe

$$C \subset \mathbb{R}^m \text{ convexe } \Rightarrow f^{-1}(C) = \{x \in \mathbb{R}^n,\ f(x) \in C\} \text{ convexe}$$

1.3.2 Fonctions convexes

Définition 1.9. Une fonction $f : C \longrightarrow \mathbb{R}$ est

— convexe ssi pour tout $x, y \in C$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall \alpha \in [0,1].$$

— strictement convexe ssi pour tout $x, y \in C$, $x \neq y$

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y), \quad \forall \alpha \in]0,1[.$$

Exercice 1.6. Montrer que les fonctions suivantes sont convexes

— affine $f(x) = a^t x + b$, avec $x, a, b \in \mathbb{R}^n$

— normes $\|x\|_p = (\sum_{i=1}^n x_i^p)^{1/p}$ for $p \geq 1$, $\|x\|_\infty = \max_k |x_k|$ avec $x \in \mathbb{R}^n$

— Fonctions affines sur $\mathbb{R}^{m\times n}$

$$f(X) = tr(A^T X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$

— Norme spectrale : valeur singulière maximale, pour $x \in \mathbb{R}^{m\times n}$

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{max}(X^T X))^{1/2}$$

Restriction d'une fonction convexe sur les droites $f : \mathbb{R}^n \to \mathbb{R}$
$g_{x,v}(t) = f(x + tv)$ avec $x, v \in \mathbb{R}^n$, $t \in \mathbb{R}$
f est convexe ssi $g_{x,v}(t)$ convexe pour tous $x, v \in \mathbb{R}^n$

Exercice 1.7. *Soit $f(X) = \log det\, X$, définie sur dom $f = S^n_{++}$*
Montrer que f est concave

Corrigé :

$$\begin{aligned}
g_{X,M}(t) &= f(X+Mt) = \log\det(X+Mt) \\
&= \log\det X^{1/2}(X^{-1/2} + X^{-1/2}Mt) \\
&= \log\det X^{1/2}(I + X^{-1/2}MX^{-1/2}t)X^{-1/2} \\
&= \log[\det X^{1/2}\det(I + X^{-1/2}MX^{-1/2}t)\det X^{-1/2}] \\
&= \log\det X + \log\det(I + X^{-1/2}MX^{-1/2}t) \\
&= \log\det X + \sum_{i=1}^{n}\log(1+t\lambda_i)
\end{aligned}$$

avec λ_i les vp de $X^{-1/2}MX^{-1/2} \in S^n_{++}$ ■

Convexité d'une fonction différentiable

Proposition 1.6. *Soit f une fonction définie sur un convexe $C \subset \mathbb{R}^n$ différentiable à valeurs réelles. La fonction f est*

1. *convexe si et seulement si $\forall (x,y) \in C^2, \langle \nabla f(x), y - x\rangle \leq f(y) - f(x)$.*
2. *convexe si et seulement si $\forall (x,y) \in C^2, \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0$.*
3. *strictement convexe si et seulement si $\forall (x,y) \in C^2, x \neq y, \langle \nabla f(x), y - x\rangle < f(y) - f(x)$.*

Preuve convexe $\to$ 1 : En effet si f convexe pour tout $t \in]0,1]$ on a

$$\begin{aligned}
f(x + t(y-x)) - f(x) &\leq t(f(y) - f(x) \\
\frac{f(x + t(y-x)) - f(x)}{t} &\leq f(y) - f(x)
\end{aligned}$$

On fait tendre t vers 0 et on obtient

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

$1 \to 2$: On a

$$\begin{aligned} \langle \nabla f(x), y-x\rangle &\leq f(y)-f(x) \\ \langle -\nabla f(y), y-x\rangle = \langle \nabla f(y), x-y\rangle &\leq f(x)-f(y) \end{aligned}$$

On somme et on obtient

$$\langle \nabla f(x)-\nabla f(y), y-x\rangle \quad \leq \quad 0$$

$2 \to$ convexe : On étudie $g(t) = f(ty+(1-t)x)$ on montre que $g'(t)$ est croissante donc que $g(t)$ est convexe et puis on écrit que $(t1+(1-t)0) \leq tg(1)+(1-t)g(0)$ ce qui est f convexe. ■

La propriété de convexité forte (ou ellipticité) permettra d'obtenir par la suite des résultats de convergence de certains algorithmes

Définition 1.10. Soit $K \subset V$ un convexe. Une fonction $f : K \to \mathbb{R}$ est dite **fortement convexe** ou **uniformément convexe** ou α**-convexe** ou α**-elliptique** s'il existe $\alpha > 0$ tel que

$$\forall (x,y) \in K^2, \forall \lambda \in [0,1], \quad f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)-\frac{\alpha}{2}\lambda(1-\lambda)\|x-y\|^2.$$

Suivant la régularité de la fonction f on a des caractérisations de sa convexité forte grace au théorème suivant

Théorème 1.7. *1. Si $f : V \to \mathbb{R}$ est continue, les propriétés suivantes sont équivalentes*

(a) f est α-elliptique

(b) Pour tout $(x,y) \in V^2$, $f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} - \frac{\alpha}{8}\|x-y\|^2$

2. Si $f : V \to \mathbb{R}$ est différentiable, les propriétés suivantes sont équivalentes

(a) f est α-elliptique

(b) Pour tout $(x,y) \in V^2$, $f(y) \geq f(x)+\langle \nabla f(x), y-x\rangle + \frac{\alpha}{2}\|x-y\|^2$

(c) Pour tout $(x,y) \in V^2$, $\langle \nabla f(y)-\nabla f(x), y-x\rangle \geq \alpha\|x-y\|^2$

3. Si $f : V \to \mathbb{R}$ est deux fois différentiable, les propriétés suivantes sont équivalentes

(a) f est α-elliptique

(b) Pour tout $(x,h) \in V^2$, $\langle H\, f(x)h, h\rangle \geq \alpha\|h\|^2$

Preuve Pour montrer 2c⇒ 2a) on étudie la fonction scalaire

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

on montre que la dérivée est croissante donc g est convexe. D'où on déduit f strictement convexe. ■

Exercice 1.8. Etudier la convexité des fonctions suivantes

1. Fonction quadratique

$$f(x) = \frac{1}{2}x^tPx + q^tx + r$$

avec $P \in S^n$

2. Moindres carrés $f(x) = \|Ax - b\|_2^2$

3. Quadratique sur linéaire $f(x, y) = x^2/y$ sur $\mathbb{R} \times \mathbb{R}^{+\star}$

4. Log-sum-exp $f(x) = \log \sum_{k=1}^{n} \exp x_k$

5. Moyenne géométrique

$$f(x) = (\prod_{k=1}^{n} x_k)^{1/n} \text{ pour } x \in \mathbb{R}_+^{n\star}$$

Corrigé : Pour la question 4)

$$\nabla^2 f(x) = \frac{1}{1^t z}\text{diag}(z) - \frac{1}{(1^t z)^2} z z^t, \quad \text{avec } z_k = \exp x_k$$

$\nabla^2 f(x) \geq 0 \Leftrightarrow v^t \nabla^2 f(x) v \geq 0$ pour tout v

$$v^t \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2)} \geq 0$$

En effet par cauchy schwarz on a

$$\sum_k z_k v_k = \sum_k \sqrt{z_k}(\sqrt{z_k} v_k) \leq (\sum_k z_k)^{1/2} (\sum z_k v_k^2)^{1/2}$$

■

Exercice 1.9. On définit les ensembles de niveau d'une fonction convexe $C_\alpha = \{x \in \text{dom } f, f(x) \leq \alpha\}$
Montrer que les C_α sont des convexes
On définit l'épigraphe de $f : \mathbb{R}^n \to \mathbb{R}$

$$epi\ f = \{(x, t) \in \mathbb{R}^{n+1}, x \in \text{dom } f,\ f(x) \leq t\}$$

Montrer que f convexe si et seulement si epi f est convexe

Corrigé : Il s'agit de montrer que pour tout $x, y \in C_\alpha$ et pour tout $\theta in [0, 1]$ on a $\theta x + (1 - \theta) y \in C_\alpha$.
Comme $x \in C_\alpha$ alors $f(x) \leq \alpha$, de même comme $y \in C_\alpha$ alors $f(y) \leq \alpha$. Comme f est convexe pour tout $\theta in [0, 1]$ on a

$$f(\theta x + (1 - \theta) y) \quad \leq \quad \theta f(x) + (1 - \theta) f(y) \leq \theta\alpha + (1 - \theta)\alpha = \alpha$$

donc $\theta x + (1 - \theta) y \in C_\alpha$.
D'après ce qui précède si f est convexe $epi\ f$ est convexe. En effet soit (x, t) et (y, s) dans $epi\ f$. Soit $\theta \in [0, 1]$ comme f est convexe

$$f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y) \leq \theta t + (1 - \theta) s$$

Réciproquement, pour tout x, y on a trivialement $(x, f(x)) \in epi\ f$ et $(y, f(y)) \in epi\ f$ donc si $epi\ f$ est convexe alors $f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y)$ ■

Exercice 1.10. Etudier la convexité de la fonction

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^t x), \quad \text{dom } f = \{x, a_i^t x < b_i,\ i = 1, \ldots, m\}$$

Corrigé : On peut montrer que chacun des termes de la somme est concave, en calculant son gradient $\frac{-a_i}{b_i - a_i^T x}$ puis en utilisant la propriété 1.6 ■

Exercice 1.11. Calculer le gradient et le hessien des fonctions suivantes en 0 et dire si la matrice hessienne est définie positive en ce point.

1. $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$, $f(x) = \sin(x_1)e^{x_2}(1 + x_3^2)$
2. $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$, $f(x) = e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right)$.

Corrigé :

1. Pour la première fonction on a pour le gradient

$$\nabla f(x) = \begin{pmatrix} \cos(x_1)e^{x_2}(1 + x_3^2) \\ \sin(x_1)e^{x_2}(1 + x_3^2) \\ 2x_3 \sin(x_1)e^{x_2} \end{pmatrix}, \quad \text{d'où } \nabla f(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

et pour le hessien

$$Hf(x) = \begin{pmatrix} -\sin(x_1)e^{x_2}(1 + x_3^2) & \cos(x_1)e^{x_2}(1 + x_3^2) & 2x_3 \cos(x_1)e^{x_2} \\ \cos(x_1)e^{x_2}(1 + x_3^2) & \sin(x_1)e^{x_2}(1 + x_3^2) & 2x_3 \sin(x_1)e^{x_2} \\ 2x_3 \cos(x_1)e^{x_2} & 2x_3 \sin(x_1)e^{x_2} & 2\sin(x_1)e^{x_2} \end{pmatrix}, \quad \text{d'où } Hf(0) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

donc la matrice hessienne est ni positive ($\langle Hf(0)[-1, 1, 0], [-1, 1, 0]\rangle = -2$), ni définie en 0 (valeur propre nulle).

2. Pour la deuxième fonction on remarque que

$$f(x) = e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right) = e^{x_1}(1 - x_2^2)\frac{\sin(x_3) + \cos(x_3)}{\cos(x_3) - \sin(x_3)}.$$

Par ailleurs on calcule le gradient

$$\nabla f(x) = \begin{pmatrix} e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right) \\ -2e^{x_1}x_2\text{tg}\left(x_3 + \frac{\pi}{4}\right) \\ e^{x_1}(1 - x_2^2)\left(1 + \text{tg}^2\left(x_3 + \frac{\pi}{4}\right)\right) \end{pmatrix}, \quad \text{d'où } \nabla f([0, 0, 0]) = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

et le hessien

$$Hf(x) = \begin{pmatrix} e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right) & -2e^{x_1}x_2\text{tg}\left(x_3 + \frac{\pi}{4}\right) & e^{x_1}(1 - x_2^2)\left(1 + \text{tg}^2\left(x_3 + \frac{\pi}{4}\right)\right) \\ -2e^{x_1}x_2\text{tg}\left(x_3 + \frac{\pi}{4}\right) & -2e^{x_1}\text{tg}\left(x_3 + \frac{\pi}{4}\right) & -2e^{x_1}x_2\left(1 + \text{tg}^2\left(x_3 + \frac{\pi}{4}\right)\right) \\ e^{x_1}(1 - x_2^2)\left(1 + \text{tg}^2\left(x_3 + \frac{\pi}{4}\right)\right) & -2e^{x_1}x_2\left(1 + \text{tg}^2\left(x_3 + \frac{\pi}{4}\right)\right) & 2e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right)\left(1 + \text{tg}^2\left(x_3 + \frac{\pi}{4}\right)\right) \end{pmatrix}$$

d'où

$$Hf([0, 0, 0]) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & -2 & 0 \\ 2 & 0 & 4 \end{pmatrix},$$

dont les valeurs propres sont $\{-2, 0, 5\}$ dont $Hf([0, 0, 0])$ est ni définie, ni positive

■

2 Minimisation sans contraintes

Dans ce chapitre on s'intéresse au problème d'optimisation sans contraintes, c'est à dire quand $Y = V$ dans (1.1). Dans une première partie on va énoncer des résultats permettant de caractériser $x^\star \in V$ réalisant un minimum local de $f(x)$. On présentera ensuite quelques algorithmes fondamentaux pour calculer $x^\star$ de manière approchée.

2.1 Conditions d'optimalité dans le cas sans contraintes

On commence par montrer un résultat important et très général

Théorème 2.1. *Soit $f : K \subset V \to \mathbb{R}$, où K est un convexe inclus dans V , un espace de Hilbert. On suppose que f est différentiable en $x^\star \in K$. Si $x^\star$ est un minimum local de f sur K , alors $x^\star$ vérifie l'inéquation d'Euler :*

$$Df(x^\star)(y - x^\star) \geq 0, \forall y \in K.$$

Preuve Si K est convexe, pour $y \in K$ et $\lambda \in]0, 1]$, $x^\star + \lambda(y - x^\star) \in K$ et donc

$$\frac{f(x^\star + \lambda(y - x^\star)) - f(x^\star)}{\lambda} \geq 0.$$

On en déduit l'inéquation d'Euler en faisant tendre λ vers 0. ■

Théorème 2.2. *Conditions nécessaires d'optimalité.*
Soit $x^\star$ réalisant un minimum local d'une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

1. *Condition nécessaire du premier ordre : si f est différentiable dans un voisinage ouvert V de $x^\star$, alors, $\nabla f(x^\star) = 0$*
2. *Condition nécessaire du second ordre : si, de plus, f est deux fois différentiable sur V , alors $Hf(x^\star)$ est semi définie positive et f est localement convexe en $x^\star$.*

Preuve (1) Si $x^\star$ est un minimiseur de f sur V, si $x^\star + h \in V$, $\forall \varepsilon$ on a

$$f(x^\star) \leq f(x^\star + \varepsilon h) \quad = \quad f(x^\star) + \langle \nabla f(x^\star), \varepsilon h\rangle + o(\|\varepsilon h\|) = f(x^\star) + \varepsilon\langle \nabla f(x^\star), h\rangle + o(\|\varepsilon h\|)$$

D'où on déduit, en divisant par ε, puis en faisant tendre ε vers 0, que $0 \leq \langle \nabla f(x^\star), h\rangle$. Ce résultat étant valable pour $-h$ également, on en déduit que $\nabla f(x^\star) = 0$.
(2) Si f est deux fois différentiable sur V , on fait cette fois-ci un développement de Taylor à l'ordre 2

$$f(x^\star) \leq f(x^\star + \varepsilon h) \quad = \quad f(x^\star) + \frac{1}{2}\langle Hf(x^\star)\varepsilon h, \varepsilon h\rangle + o(\|\varepsilon h\|^2) = f(x^\star) + \frac{\varepsilon^2}{2}\langle Hf(x^\star)h, h\rangle + o(\|\varepsilon h\|^2)$$

puisque $\nabla f(x^\star) = 0$. On divise par ε^2 et on fait tendre ε vers 0 pour obtenir

$$\langle Hf(x^\star)h, h\rangle \geq 0,$$

ce qui montre que le hessien est semi-défini positif.
Contre exemple pour défini positif : $f(x) = x^4$.
Contre exemple pour la réciproque : $f(x) = x^3$. ■

Théorème 2.3. *Conditions suffisantes d'optimalité*
Si f est deux fois différentiable en $x^\star$, et si $\nabla f(x^\star) = 0$ et si de plus

— *soit $Hf(x^\star)$ est définie positive*

— *soit f est deux fois différentiable dans un voisinage V de $x^\star$ et $Hf(x)$ est semi définie positive sur ce*

voisinage
alors $x^\star$ réalise un minimum local isolé de f.

Preuve

— Si $Hf(x^\star)$ est définie positive, il existe $\alpha > 0$ tel que $\langle Hf(x^\star)h, h\rangle \geq \alpha\|h\|^2$ pour tout $h \in \mathbb{R}^n$. La formule de Taylor à l'ordre deux en $x^\star$ donne

$$\begin{aligned} f(x^\star + h) &= f(x^\star) + \frac{1}{2}\langle Hf(x^\star)h, h\rangle + o(\|h\|^2) \\ &\geq f(x^\star) + \left(\frac{\alpha}{2} + \frac{o(\|h\|^2)}{\|h\|^2}\right) o(\|h\|^2) > 0 \end{aligned}$$

pour h suffisamment petit puisque $\frac{o(\|h\|^2)}{\|h\|^2} \to 0$

— On applique dans ce cas la formule de Taylor du Théorème 1.5 : pour tout $d \in \mathbb{R}^n$ t.q. $x^\star + d \in V$, il existe $\alpha \in [0, 1]$ t. q.

$$f(x^\star + d) = f(x^\star) + \frac{1}{2}\langle Hf(x^\star + \alpha d)d, d\rangle \geq f(x^\star).$$

■

Théorème 2.4. *Condition d'unicité dans le cas convexe*
(i) Si f est convexe sur une partie convexe $C \in \mathbb{R}^n$, tout minimum local de f sur C est global.
(ii) Si f est strictement convexe elle a au plus un minimum global.

Preuve (i) Soit $x^\star$, un minimum local de $f(x)$ sur C. Supposons qu'il existe $y \in C$ tel que $f(y) < f(x^\star)$. Soit $y_\lambda = \lambda y + (1-\lambda)x^\star$, avec $\lambda \in]0, 1[$. Pour λ suffisamment petit, y_λ est proche de $x^\star$ puisque $\|y_\lambda - x^\star\| = \lambda\|y - x^\star\|$ donc $f(y_\lambda) \geq f(x^\star)$. Comme f est convexe on en déduit

$$f(x^\star) \leq f(y_\lambda) \leq \lambda f(y) + (1-\lambda)f(x^\star) \Rightarrow f(x^\star) \leq f(y)$$

ce qui contredit l'hypothèse. Donc $x^\star$ minimise f sur C.
(ii) Si f est strictement convexe et $x_1 \neq x_2$ deux minimiseurs de f, alors

$$f(x_1) = f(x_2) \leq f\left(\frac{x_1 + x_2}{2}\right) < \frac{f(x_1) + f(x_2)}{2} = f(x_1)$$

ce qui est absurde. Cela implique donc $x_1 = x_2$. ■

Théorème 2.5. *Condition nécessaire et suffisante d'optimalité dans le cas convexe*
Si f est convexe sur $\mathbb{R}^n$ et de classe C^1, $x^\star \in \mathbb{R}^n$ réalise un minimum global de f si et seulement si $\nabla f(x^\star) = 0$.

Preuve La condition nécessaire $\nabla f(x^\star) = 0$ résulte du Théorème 2.2 pour avoir un minimum local, et du Théorème 2.4 pour l'équivalence entre minimum local et global dans le cas convexe. Pour la condition suffisante on utilise la Proposition 1.6

$$f(y) - f(x^\star) \geq \langle \nabla f(x^\star), y - x^\star\rangle = 0.$$

■

Théorème 2.6. *Conditions d'optimalité pour les problèmes quadratiques*
Considérons le problème

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle g, x\rangle + c$$

où Q est une matrice symétrique $n \times n$, $g \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

- *Si Q n'est pas semi définie positive, alors le problème ne possède pas de solution, c'est-à-dire qu'il n'existe aucun $x \in \mathbb{R}^n$ qui soit un minimum local.*
- *Si Q est définie positive, alors $x^\star = -Q^{-1}g$ est l'unique minimum. global*

Preuve f est deux fois différentiable et le gradient de f est $\nabla f(x) = \frac{1}{2}(Qx + (x^TQ)^T) + g = Qx + g$ puisque Q est symétrique, et le hessien est la matrice Q. On applique le Théorème 2.2 et on obtient le premier point par contraposée. Pour le deuxième point, si $x^\star$ est un minimiseur de f, $\nabla f(x^\star) = Qx^\star + g = 0$ et si Q est définie positive la seule solution est bien $x^\star = -Q^{-1}g$. ■

Exercice 2.1. Régression linéaire. Soient données N mesures x_i à des instants t_i, $i = 1, \ldots, N$. Pour vérifier l'hypothèse d'une dépendance affine de ces mesures en fontion du temps, on va minimiser la distance des mesures (x_i) à la droite d'équation $at + b$

$$f(a, b) = \sum_{i=1}^{N} |x_i - at_i - b|^2. \tag{2.1}$$

1. Donner les coefficients a et b de la droite minimisant l'écart "au sens des moindres carrés" (eq 2.1)
2. Reformuler le problème de manière vectorielle. Montrer que le vecteur $D = (a, b)^T$ est solution d'un système linéaire $AD = B$ où B et A s'expriment en fonction du vecteur $X = (x_i)_{i=1,\ldots,N}$ et de la matrice

$$T = \begin{pmatrix} t_1 & t_2 & \cdots & t_N \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Corrigé :

1. Les coefficients a et b minimisant $f(a, b)$ annulent les dérivées partielles. On résout le système

$$\begin{aligned} \frac{\partial f(a,)}{\partial a} &= \sum_{i=1}^{N} 2t_i(at_i + b - x_i) = 2a\sum_{i=1}^{N} t_i^2 + 2b\sum_{i=1}^{N} t_i - 2\sum_{i=1}^{N} t_i x_i = 0 \\ \frac{\partial f(a,)}{\partial b} &= \sum_{i=1}^{N} 2(at_i + b - x_i) = 2a\sum_{i=1}^{N} t_i + 2b\sum_{i=1}^{N} 1 - 2\sum_{i=1}^{N} x_i = 0 \end{aligned}$$

   et on obtient

$$\begin{aligned} a &= \frac{N\sum_{i=1}^N t_i x_i - \sum_{i=1}^N x_i \sum_{i=1}^N t_i}{N\sum_{i=1}^N t_i^2 - \left(\sum_{i=1}^N t_i\right)^2} \\ b &= \frac{\sum_{i=1}^N t_i^2 \sum_{i=1}^N x_i - \sum_{i=1}^N t_i \sum_{i=1}^N t_i x_i}{N\sum_{i=1}^N t_i^2 - \left(\sum_{i=1}^N t_i\right)^2} \end{aligned}$$

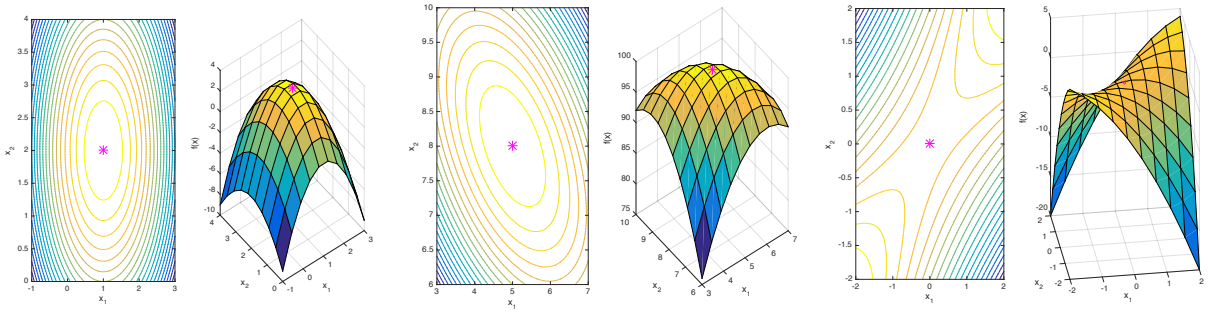
2. On peut reformuler le problème de manière vectorielle en écrivant que

$$f(a, b) = F(D) = \|X - T^T D\|^2.$$

   Donc $\nabla_D F(D) = 2T(T^T D - X)$. Le gradient est nul au minimum donc D est solution du système 2×2 $TT^T D = TX$.
   Remarque : cette méthode se généralise à l'identification de paramètres pour des modèles linéaires de dimension quelconque $Y = A^T X + b$.

Exercice 2.2. Etudier les optima des fonctions suivantes

$$\begin{aligned} f_1(x_1,x_2) &= -2x_1^2 - x_2^2 + 4x_1 + 4x_2 - 3, \\ f_2(x_1,x_2) &= -2x_1^2 - 2x_1x_2 - 2x_2^2 + 36x_1 + 42x_2 - 158, \\ f_3(x_1,x_2) &= -x_1^2 + 3x_1x_2 - x_2^2. \end{aligned}$$



f_1 f_2 f_3

FIGURE 2.1 – Lignes de niveau et visualisation en 3D des fonctions de l'exercice 2.2.

Corrigé :

1. Pour la première fonction on a

$$\begin{aligned} \partial_{x_1} f_1(x_1,x_2) &= -4x_1 + 4, \\ \partial_{x_2} f_1(x_1,x_2) &= -2x_2 + 4, \end{aligned}$$

qui s'annulent en $x_1 = 1$ et $x_2 = 2$. Le hessien

$$Hf_1(x_1,x_2) = \begin{pmatrix} -4 & 0 \\ 0 & -2 \end{pmatrix}$$

est une matrice définie négative (valeurs propres -4 et -2) donc f_1 est maximum en $(1,1)$.

2. Pour la deuxième fonction on a

$$\begin{aligned} \partial_{x_1} f_2(x_1,x_2) &= -4x_1 - 2x_2 + 36 = 2(-2x_1 - x_2 + 18), \\ \partial_{x_2} f_2(x_1,x_2) &= -4x_2 - 2x_1 + 42 = 2(-x_1 - 2x_2 + 21), \end{aligned}$$

qui s'annulent en $x_1 = 5$ et $x_2 = 8$. Le hessien

$$Hf_1(x_1,x_2) = 2A, \quad A = \begin{pmatrix} -2 & -1 \\ -1 & -2 \end{pmatrix}$$

A a pour valeurs propres les racines de $(-2-a)^2 - 1$ donc -1 et -3. C'est une matrice définie négative donc f_2 est maximum en $(5,8)$.

3. Pour la troisième fonction on a

$$\begin{aligned} \partial_{x_1} f_1(x_1,x_2) &= -2x_1 + 3x_2, \\ \partial_{x_2} f_1(x_1,x_2) &= 3x_1 - 2x_2 \end{aligned}$$

qui s'annulent en $x_1 = 0$ et $x_2 = 0$. Le hessien

$$Hf_1(x_1,x_2) = \begin{pmatrix} -2 & 3 \\ 3 & -2 \end{pmatrix}$$

a pour valeurs propres les racines de $(-2-a)^2 - 9$ donc 1 et -5. La matrice hessienne n'est ni positive ni négative donc $(0,0)$ est un col pour f_3.

2.2 Résolution de systèmes d'équations non linéaires

La condition d'optimalité $\nabla f(x) = 0$ nécessite la résolution d'un système d'équations non-linéaires de n équations à n inconnues (on verra par la suite qu'il y a d'autres méthodes pour chercher le minimum d'une fonction).
La méthode par excellence pour résoudre ce genre de problème est **l'algorithme de Newton** basé sur une linéarisation de l'équation $g(x) = 0$ utilisant la formule de Taylor.

2.2.1 Cas scalaire

Dans le cas scalaire, la recherche des zéros d'une fonction $g : \mathbb{R} \longrightarrow \mathbb{R}$ utilise la formule de Taylor

$$g(x^\star) \quad = \quad g(x) + g'(x)(x^\star - x) + o(\|x^\star - x\|).$$

On définit un **algorithme itératif** consistant à prendre comme approximation de la solution, l'intersection de la tangente à g au point courant avec l'axe des abscisses. On se ramène à l'algorithme du point fixe pour résoudre une équation non linéaire $\psi(x) = x$, avec $\psi(x) = x - g(x)/g'(x)$.

Algorithme 2.1 : Algorithme de Newton dans le cas scalaire

Données : La fonction $g(x)$ et sa dérivée $g'(x)$, tolérance ε, nombre maximum d'itérations $k_{\max}$
Résultat : $x^\star$ tel que $g(x^\star) = 0$
Initialisation : $k = 0$, x_0 approximation initiale de $g(x) = 0$.
tant que $|g(x_k)| > \varepsilon$ *et* $k \leq k_{\max}$ **faire**
  $x_{k+1} = x_k - \dfrac{g(x_k)}{g'(x_k)}$
  $k \leftarrow k + 1$
fin
$x^\star \leftarrow x_k$

Cet algorithme converge (quadratiquement) si la fonction g n'est pas "trop non-linéaire", si sa dérivée n'est pas trop proche de 0 dans un voisinage de la solution et si le point de départ x_0 n'est pas trop éloigné de la solution $x^\star$. On a en effet le théorème suivant

Théorème 2.7. *On suppose que g est de classe C^2 sur l'intervalle $I = [x^\star - r, x^\star + r]$ avec $g(x^\star) = 0$ et que $g' \neq 0$ sur I. Soit*

$$M = \max_{x \in I} \left| \frac{g''(x)}{g'(x)} \right|, \quad \textit{et } h = \min\left(r, \frac{1}{M}\right).$$

Alors pour tout point initial $x_0 \in]x^\star - h, x^\star + h[$ on a

$$|x_k - x^\star| \leq \frac{1}{M} \left(M|x_0 - x^\star|\right)^{2k},$$

d'où on déduit $\lim_{k \to +\infty} |x_k - x^\star| = 0$.

Preuve On introduit la fonction $u(x) = g(x)/g'(x)$. Comme la fonction g est monotone sur I et nulle en $x^\star$, elle a le même signe que $g'(x^\star)(x - x^\star)$ ce qui entraîne que $u(x)$ a le même signe que $x - x^\star$. De plus

$$u'(x) = 1 - \frac{g(x)g''(x)}{g'(x)^2} = 1 - \frac{g''(x)}{g'(x)} u(x),$$

donc $|u'(x)| \leq 1 + M|u(x)|$ sur I, d'où on déduit

$$|u(x)| \leq \frac{1}{M}\left(e^{M|x - x^\star|} - 1\right)$$

sur I. Par ailleurs on a $\psi(x) = x - u(x)$ donc $\psi'(x) = u(x)\frac{g''(x)}{g'(x)}$ donc en utilisant l'inégalité précédente

$$|\psi'(x)| \leq M|u(x)| \leq e^{M|x-x^\star|} - 1.$$

On laisse au lecteur le soin de vérifier que la fonction $m(t) = e^{|t|} - 1 - 2|t|$ est négative sur l'intervalle $[-1, 1]$. Donc pour $|x - x^\star| \leq \min(r, \frac{1}{M})$ on a $|\psi'(x)| \leq 2M|x - x^\star|$ et comme $\psi(x^\star) = x^\star$ on obtient par intégration entre $x^\star$ et $x \in [x^\star - h, x^\star + h]$

$$|\psi(x) - x^\star| \leq M|x - x^\star|^2.$$

En définissant la suite $x_{p+1} = \psi(x_p)$ on obtient $|M(x_{p+1} - x^\star)| \leq (M|x_p - x^\star|)^2$ soit par récurrence $|M(x_p - x^\star)| \leq (M|x_0 - x^\star|)^{2p}$. Pour $|x_0 - x^\star| < M$ on a bien $\lim_{k\to+\infty}(M|x_0 - x^\star|)^{2p} = 0$. ■

Avant d'aller plus loin, notons quelques caractéristiques de ce premier algorithme que l'on retrouvera dans n'importe quel algorithme itératif

- Une étape d'initialisation, qui, comme son nom l'indique, a lieu une seule fois au début.
- Un bloc répété (ou itéré) dans une boucle définie par
  - un test qui exprime la convergence attendue de l'algorithme. Par exemple ici le théorème nous assure que $\lim_{k\to+\infty} |x_k - x^\star| = 0$. Comme $x^\star$ est inconnu, nous utilisons la régularité de la fonction g pour en déduire $\lim_{k\to+\infty} g(x_k) = 0$, que nous approchons numériquement par $|g(x_k)| \leq \varepsilon$, avec ε choisi petit.
  - l'incrémentation d'un compteur d'itérations jusqu'à une valeur fixée d'avance. Par exemple ici la boucle s'arrête quand $k > k_{\max}$.
  - une combinaison des deux précédents. Pourquoi cela ? Pour prendre en compte le fait que les algorithmes peuvent avoir des conditions de convergence malaisées à vérifier en pratique, ou bien, une convergence si lente que la réalisation du test de convergence épuiserait notre patience ou nos ressources de calcul. Notons enfin que le choix de la tolérance doit être harmonisé avec les ressources que l'on alloue à l'algorithme. Si ε est trop petit par rapport $k_{\max}$, l'algorithme s'arrêtera systématiquement par "épuisement" et non parce qu'il aura convergé.

On introduit également, à l'occasion de cet algorithme, la notion de vitesse de convergence, qui mesure la rapidité avec laquelle la solution courante se rapproche de la cible.

Définition 2.1. Notons $e_k = x^k - x^\star$ l'erreur à l'itération k. On dit que

- l'algorithme converge si $\lim_{k\to\infty} \|e_k\| = 0$
- l'algorithme converge linéairement s'il existe $c \in]0, 1[$ tel que $\|e_k\| \leq c\|e_{k-1}\|$ pour $k > K(c)$
- la convergence est superlinéaire s'il existe $(c_k)_{k\in\mathbb{N}}$ avec $\lim_{k\to\infty} c_k = 0$ tel que $\|e_k\| \leq c_k\|e_{k-1}\|$
- la convergence est géométrique si la suite c_k est géométrique
- l'algorithme est d'ordre p s'il existe $c \in]0, 1[$ tel que $\|e_k\| \leq c\|e_{k-1}\|^p$ pour $k > K(c)$

On dit que la convergence est globale si elle ne dépend pas du point de départ x^0, locale si x^0 doit vérifier une condition de proximité avec la cible $x^\star$.

A titre d'exemple, la convergence de l'algorithme de Newton est quadratique et locale.

2.2.2 Méthode de Newton en dimension n

En dimension quelconque il s'agit de résoudre $G(x) = 0$ avec $G : \mathbb{R}^n \longrightarrow \mathbb{R}^n$. On note $J(x) \in \mathbb{R}^{n\times n}$ la matrice jacobienne de G au point x, $J_{i,j}(x) = \dfrac{\partial G_i(x)}{\partial x_j}$. Comme dans la méthode de Newton scalaire, l'idée est d'approcher G par sa partie linéaire au voisinage du point courant x_k

$$G(x_{k+1}) = G(x_k) + G'(x_k)(x_{k+1} - x_k) + O(||x_{k+1} - x_k)||.$$

Algorithme 2.2 : L'algorithme de Newton-Ralphson

Données : La fonction $G(x)$ et sa matrice jacobienne $J(x)$, tolérance ε, nombre maximum d'itérations $k_{\max}$
Résultat : $x^\star$ tel que $G(x^\star) = 0$
Initialisation : $k = 0$, x_0 approximation initiale de $G(x) = 0$.
tant que $\|G(x_k)\| > \varepsilon$ *et* $k < k_{\max}$ **faire**
  Résolution de $J(x_k)d_k = -G(x_k)$
  $x_{k+1} = x_k + d_k$
  $k \leftarrow k + 1$
fin
$x^\star \leftarrow x_k$

Là encore on a un résultat théorique nous assurant la convergence de l'algorithme sous certaines conditions :

Théorème 2.8. *On suppose que la fonction G est de classe C^2, que $G(x^\star) = 0$ et que l'application linéaire tangente $J(x^\star) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)$ est inversible. Alors $x^\star$ est un point fixe superattractif de $\Psi(x) = x - (G'(x))^{-1}G(x)$.*

Preuve On doit montrer que $\Psi'(x^\star) = 0$. On traduit le fait que G est deux fois différentiable en $x^\star$

$$\begin{aligned} G(x^\star + h) &= G(x^\star) + G'(x^\star)h + \frac{1}{2}hG''(x^\star)h + o(\|h\|^2) \\ &= G'(x^\star)\left[h + \frac{1}{2}G'(x^\star))^{-1}(hG''(x^\star)h + o(\|h\|^2))\right] \end{aligned}$$

et d'autre part le fait que G' est différentiable

$$\begin{aligned} G'(x^\star + h) &= G'(x^\star) + G''(x^\star)h + o(\|h\|) \\ &= G'(x^\star)\left[Id + G'(x^\star)^{-1}(G''(x^\star)h + o(\|h\|))\right] \end{aligned}$$

D'où

$$\begin{aligned} G'(x^\star + h)^{-1} &= \left[Id + G'(x^\star)^{-1}(G''(x^\star)h + o(\|h\|))\right]^{-1} G'(x^\star)^{-1} \\ &= \left[Id - G'(x^\star)^{-1}(G''(x^\star)h + o(\|h\|))\right] G'(x^\star)^{-1}. \end{aligned}$$

On a donc

$$\begin{aligned} G'(x^\star + h)^{-1}G(x^\star + h) &= \left[Id - G'(x^\star)^{-1}(G''(x^\star)h + o(\|h\|))\right] G'(x^\star)^{-1} \\ &\qquad G'(x^\star)\left[h + \frac{1}{2}(G'(x^\star))^{-1}(hG''(x^\star)h + o(\|h\|^2))\right] \\ &= h - \frac{1}{2}(G'(x^\star))^{-1}(hG''(x^\star)h + o(\|h\|^2)). \end{aligned}$$

On a donc

$$\Psi(x^\star + h) = x^\star + h - G'(x^\star + h)^{-1}G(x^\star + h) = x^\star + \frac{1}{2}G'(x^\star)^{-1}(hG''(x^\star)h + o(\|h\|^2))$$

Soit encore

$$\Psi(x^\star + h) - \Psi(x^\star) = \frac{1}{2}G'(x^\star)^{-1}(hG''(x^\star)h + o(\|h\|^2)$$

Donc par définition de la différentielle $\Psi'(x^\star) = 0$ et $\Psi''(x^\star) = G'(x^\star)^{-1}G''(x^\star)$. En particulier

$$\|\Psi(x^\star + h) - x^\star\| \leq \frac{1}{2}(M + \varepsilon(h))\|h\|^2)$$

où $M = |||\Psi''(x^\star)|||$. et $\varepsilon(h) \to 0$ quand $h \to 0$. ■

Exercice 2.3. Calculer

$$(x^\star, y^\star) = \underset{(x,y)\in\mathbb{R}^2}{\operatorname{argmin}} f(x,y)$$

avec

$$f(x,y) = (x-1)^4 + (y-2)^4$$

par la méthode de Newton. Démonter que la méthode converge pour tout $(x_0, y_0) \in \mathbb{R}^2$

Corrigé : The minimum of f is reached at $\tilde{X} = (1,2)^T$. On calcule le gradient et le hessien de $f(x,y)$:

$$\nabla f(x,y) = \begin{pmatrix} 4(x-1)^3 \\ 4(y-2)^3 \end{pmatrix}, \quad Hf(x,y) = \begin{pmatrix} 12(x-1)^2 & 0 \\ 0 & 12(y-2)^2 \end{pmatrix},$$

then step k in Newton algorithm is

$$\begin{aligned} \nabla f(x_k, y_k) &= -Hf(x_k, y_k) d_k \\ X^{k+1} &= X^k + d_k \end{aligned}$$

where $X^k = (x_k, y_k)^T$, which yields

$$\begin{aligned} x_{k+1} &= x_k - \frac{1}{3}(x_k - 1) \\ y_{k+1} &= y_k - \frac{1}{3}(y_k - 2) \end{aligned}$$

or, equivalently

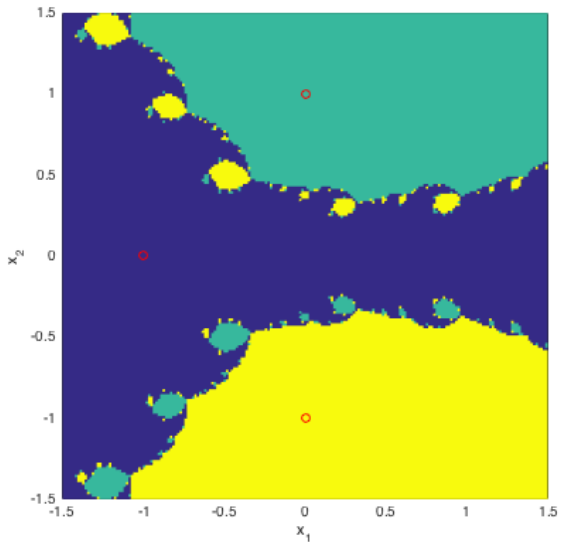
$$\begin{aligned} x_{k+1} - 1 &= \frac{2}{3}(x_k - 1) \\ y_{k+1} - 2 &= \frac{2}{3}(y_k - 2) \end{aligned}$$

Hence the sequence $X^k - \tilde{X}$ is contractant with constant $2/3$ and goes to zero as $k \to \infty$ whatever the initial value $X^0 = (x_0, y_0)^T$. ■

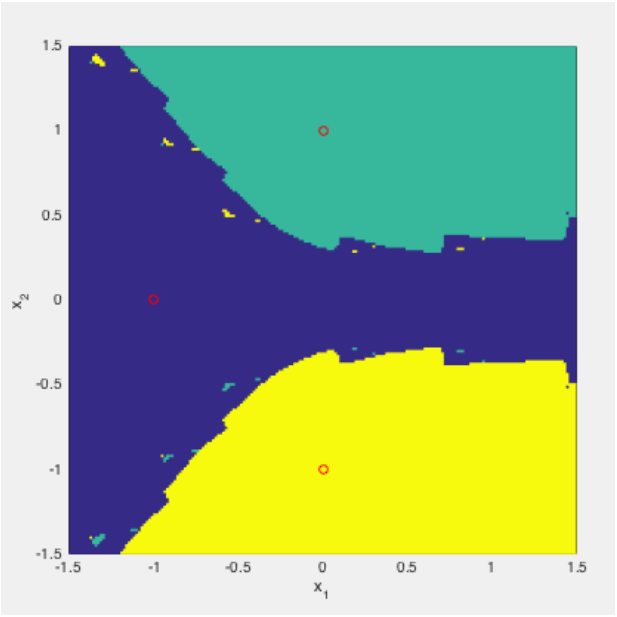
Exercice 2.4. Recherche de racines en dimension 2, bassins d'attraction fractals

1. Programmer l'algorithme de Newton pour résoudre $x^3 = 1$ dans $\mathbb{C}$
2. Tester pour différentes valeurs initiales
3. Comparer avec les résultats obtenus avec `fsolve`
4. Visualisation : on note $x^\star(x_0) \in \mathbb{C}$ la racine vers laquelle l'algorithme de Newton converge à partir de $x_0 \in \mathbb{C}$. On discrétize une zone englobant les trois racines et on affecte un code à chaque point de la zone, suivant la racine obtenue avec Newton en partant de ce point (voir l'exemple sur la Figure 2.2 pour l'équation $x^3 + x^2 + x + 1 = 0$).
   On programme l'algorithme suivant

Données : La fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, son jacobien $Jf : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$
Tableau $x^r \in \mathbb{R}^{2\times 3}$, $f(x^r_{:,k}) = 0$, pour $k = 1, 2, 3$
Résultat : Figure représentant les bassins d'attraction des racines de f
Initialisation : $N = 5$, $h = 1.5/N$ (on pourra augmenter N une fois sûr que le programme tourne)
pour $m = -N \nearrow N$,$n = -N \nearrow N$ **faire**
— Calculer $x_{m,n} = x^\star(mh + inh)$
— Calculer $C_{m,n} = \underset{k=1,2,3}{\operatorname{argmin}} \|x_{m,n} - x^r_{:,k}\|$
fin
Visualiser la matrice C avec la fonction matlab `pcolor` (vérifier la correspondance entre les lignes/colonnes et la représentation de $\mathbb{C}$, matérialiser la position des racines)



a) Newton-Ralphson algorithm



b) Matlab `fsolve` root finder

FIGURE 2.2 – Représentation des bassins d'attraction des racines complexes de l'équation $x^3 + x^2 + x + 1 = 0$

Corrigé :

1. $x^3 = 1$ dans $\mathbb{C}$ s'écrit

$$(x_1 + ix_2)^3 = x_1^3 + 3ix_1^2x_2 - 3x_1x_2^2 - ix_2^3 = 1$$

soit

$$\begin{aligned} x_1^3 - 3x_1x_2^2 - 1 &= 0, \\ 3x_1^2x_2 - x_2^3 &= 0. \end{aligned}$$

Il faut donc programmer une fonction matlab `x3_egal_1_dans_C(x)` qui renvoie la fonction et son jacobien

$$f(x) = \begin{pmatrix} x_1^3 - 3x_1x_2^2 - 1 \\ 3x_1^2x_2 - x_2^3 \end{pmatrix} \quad \text{and } Jf(x) = \begin{pmatrix} 3x_1^2 - 3x_2^2 & -6x_1x_2 \\ 6x_1x_2 & 3x_1^2 - 3x_2^2 \end{pmatrix}$$

2. Tester pour différentes valeurs initiales
3. Comparer avec les résultats obtenus avec `fsolve`

■

2.2.3 Méthode de quasi Newton

L'inconvénient majeur de la méthode de Newton est qu'elle nécessite le calcul de la dérivée, qui n'est pas toujours aisé. Une alternative consiste à approcher la dérivée par différences finies.
Dans le cas scalaire c'est la méthode de la *sécante* dont l'algorithme est le suivant :

Algorithme 2.3 : L'algorithme de la sécante

Données : La fonction $g(x)$, tolérance ε
Résultat : $x^\star$ tel que $g(x^\star) = 0$
Initialisation : $k = 0$, x_0 approximation initiale de $g(x) = 0$.
a_0 approximation initiale de $g'(x_0)$ (=1 par défaut)
tant que $|g(x_k)| > \varepsilon$ **faire**

$$x_{k+1} = x_k - \frac{g(x_k)}{a_k}$$
$$a_{k+1} = \frac{g(x_k) - g(x_{k+1})}{x_k - x_{k+1}}$$
$$k \leftarrow k + 1$$

fin
$x^\star \leftarrow x_k$

En dimension quelconque la difficulté est de définir une approximation de la matrice jacobienne à partir des itérés de la fonction elle-même. Plusieurs méthodes ont été proposées, elles proposent de partir de la matrice jacobienne exacte à l'étape initiale, puis de modifier son inverse à chaque itération, mais sans faire appel aux dérivées de la fonction. Notons W_k l'approximation de l'inverse de la matrice jacobienne à l'étape k et $y_k = g_{k+1} - g_k$ la différence entre deux gradients successifs. On aura l'algorithme suivant

Algorithme 2.4 : L'algorithme de quasi Newton

Données : La fonction $G : \mathbb{R}^n \longrightarrow \mathbb{R}^n$
La précision demandée $\varepsilon > 0$.
Résultat : $x^\star$ tel que $g(x^\star) = 0$
Initialisation : Une première approximation de la solution $x_0 \in \mathbb{R}^n$
Une première approximation de la matrice jacobienne $A_0 = J(x_0)$ ou directement de son inverse W_0
$x_1 = x_0 - W_0 G(x_0)$
$d_0 = x_1 - x_0$,
$y_0 = G(x_1) - G(x_0)$,
$k = 1$
tant que $\|G(x_k)\| > \varepsilon$ **faire**
Mise à jour : $W_k = W_{k-1} + B_{k-1}$
Calculer d_k solution de $d_k = -W_k G(x_k)$
$x_{k+1} = x_k + d_k$
$y_k = G(x_{k+1}) - G(x_k)$
$k \leftarrow k + 1$
fin
$x^\star \leftarrow x_k$

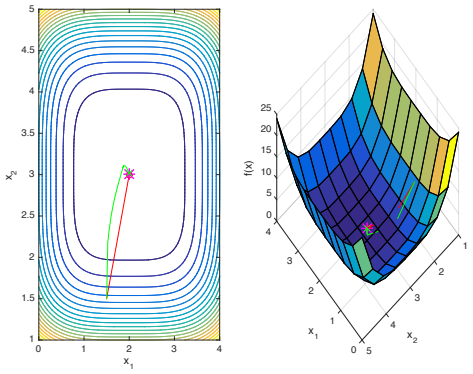


FIGURE 2.3 – Comparison des méthodes de Newton (rouge, 12 itérations) et quasi Newton (BFGS) (vert, 21 itérations) pour calculer le minimum de la fonction $f(x) = ((x_1 - 2)^4 + (x_2 - 3)^4)/2$

Pour assurer la convergence de l'algorithme, la mise à jour de la matrice W_k doit satisfaire certaines conditions

(i) W_k symétrique définie positive pour tout k (en effet la fonction $g(x)$ étant à l'origine un gradient d'une fonction $f(x)$ dont on cherche le minimum, le jacobien de $g(x)$ est le hessien de $f(x)$).

(ii) L'équation de quasi-Newton

$$W_k y_k = d_k \tag{2.2}$$

est satisfaite à chaque k

(iii) La différence entre deux approximations consécutives $W_{k+1} - W_k$ est minimum en un certain sens, par exemple au sens de la norme de Frobenius [2].

Ces conditions ne déterminent pas W_k de manière unique et plusieurs méthodes ont été proposées, en fonction de la norme choisie pour imposer la condition de stabilité (iii). On citera par exemple

— La méthode de Davidon-Fletcher-Powell

$$(DFP) \qquad W_{k+1} = W_k + \frac{d_k d_k^T}{\langle y_k, d_k\rangle} - \frac{W_k y_k y_k^T W_k}{\langle y_k, W_k y_k\rangle}.$$

— La méthode mise à jour par Broyden-Fletcher-Goldfarb-Shanno

$$(BFGS) \qquad W_{k+1} = W_k - \frac{d_k y_k^T W_k + W_k y_k d_k^T}{\langle y_k, d_k\rangle} + \left(1 + \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, d_k\rangle}\right)\frac{d_k d_k^T}{\langle y_k, d_k\rangle}.$$

A l'heure actuelle, c'est plutôt la méthode (BFGS) qui fait l'unanimité.

Exercice 2.5. Propriétés duales des formules de quasi Newton (BFGS) et (DFP)

1. Montrer que les matrices définies par (DFP) et (BFGS) satisfont l'équation de quasi-Newton (2.2).
2. On appelle formule duale de (DFP) ou (BFGS) la formule obtenue en remplaçant W_k par A_k, d_k par y_k et y_k par d_k. Montrer que la formule duale de (FDP) donne l'inverse de W_{k+1} calculé par (BFGS), et que la formule duale de (BFGS) donne l'inverse de W_{k+1} calculé par (FDP).
3. Montrer que la formule duale de (BFGS) s'écrit

$$(BFGS') \qquad A_{k+1} = \left(I - \frac{y_k d_k^T}{\langle y_k, d_k\rangle}\right) A_k \left(I - \frac{d_k y_k^T}{\langle y_k, d_k\rangle}\right) + \frac{y_k y_k^T}{\langle y_k, d_k\rangle}$$

4. Montrer que les deux formules (DFP) et (BFGS) conservent le caractère défini positif de la matrice W_k si $\langle y_k, d_k\rangle > 0$.

Corrigé :

1. Pour (DFP) on a

$$\begin{aligned} W_{k+1}y_k &= W_k y_k + \frac{d_k d_k^T y_k}{\langle y_k, d_k\rangle} - \frac{W_k y_k y_k^T W_k y_k}{\langle y_k, W_k y_k\rangle}, \\ &= W_k y_k + d_k \frac{\langle y_k, d_k\rangle}{\langle y_k, d_k\rangle} - W_k y_k \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, W_k y_k\rangle}, \\ &= W_k y_k + d_k - W_k y_k = d_k. \end{aligned}$$

Pour (BFGS) on a

$$\begin{aligned} W_{k+1}y_k &= W_k y_k - \frac{d_k y_k^T W_k y_k + W_k y_k d_k^T y_k}{\langle y_k, d_k\rangle} + \left(1 + \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, d_k\rangle}\right)\frac{d_k d_k^T y_k}{\langle y_k, d_k\rangle}, \\ &= W_k y_k - \frac{d_k\langle y_k, W_k y_k\rangle + W_k y_k\langle d_k, y_k\rangle}{\langle y_k, d_k\rangle} + \left(1 + \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, d_k\rangle}\right)\frac{d_k\langle d_k, y_k\rangle}{\langle y_k, d_k\rangle}, \\ &= W_k y_k - d_k \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, d_k\rangle} - W_k y_k \frac{\langle d_k, y_k\rangle}{\langle y_k, d_k\rangle} + d_k + d_k \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, d_k\rangle} = d_k. \end{aligned}$$

2. $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2} = \sqrt{trace(A * A)} = \sqrt{\sum_{i=1}^n \sigma_i^2}.$

2. Si on remplace W_k par A_k, y_k par d_k et d_k par y_k dans la formule (DFP) on obtient

$$A_{k+1} = A_k + \frac{y_k y_k^T}{\langle y_k, d_k\rangle} - \frac{A_k d_k d_k^T A_k}{\langle d_k, A_k d_k\rangle}$$

On fait le produit A_{k+1}^{DFP} avec W_{k+1} donné par (BFGS) on obtient (en laissant tomber les indices k)

$$\begin{aligned} A_{k+1}^{DFP} W_{k+1}^{BFGS} &= AW - \frac{Ady^TW}{y^Td} - \frac{AWyd^T}{y^Td} + \left(1 + \frac{y^TWy}{y^Td}\right)\frac{Add^T}{y^Td} \\ &\quad \frac{yy^TW}{y^Td} - \frac{yy^Tdy^TW}{(y^Td)^2} - \frac{yy^TWyd^T}{(y^Td)^2} + \left(1 + \frac{y^TWy}{y^Td}\right)\frac{yy^Tdd^T}{(y^Td)^2} \\ &\quad - \frac{Add^TAW}{d^TAd} + \frac{Add^TAdy^TW}{(d^TAd)(y^Td)} + \frac{Add^TAWyd^T}{(d^TAd)(y^Td)} \\ &\quad - \left(1 + \frac{y^TWy}{y^Td}\right)Add^TAdd^T(d^TAd)(y^Td) \\ &= Id. \end{aligned}$$

Si on remplace W_k par A_k, y_k par d_k et d_k par y_k dans la formule (BFGS) on obtient

$$A_{k+1}^{BFGS} = A_k - \frac{y_k d_k^T A_k + A_k d_k y_k^T}{\langle y_k, d_k\rangle} + \left(1 + \frac{\langle d_k, A_k d_k\rangle}{\langle y_k, d_k\rangle}\right)\frac{y_k y_k^T}{\langle y_k, d_k\rangle}$$

On fait le produit A_{k+1}^{BFGS} avec W_{k+1} donné par (DFP) et on obtient

$$A_{k+1}^{BFGS} W_{k+1}^{DFP} = Id.$$

3. On développe la formule proposée

$$\begin{aligned} A_{k+1} &= \left(A_k - \frac{y_k d_k^T A_k}{\langle y_k, d_k\rangle}\right)\left(I - \frac{d_k y_k^T}{\langle y_k, d_k\rangle}\right) + \frac{y_k y_k^T}{\langle y_k, d_k\rangle} \\ &= A_k - \frac{y_k d_k^T A_k}{\langle y_k, d_k\rangle} - \frac{A_k d_k y_k^T}{\langle y_k, d_k\rangle} + \frac{y_k d_k^T A_k d_k y_k^T}{\langle y_k, d_k\rangle^2} + \frac{y_k y_k^T}{\langle y_k, d_k\rangle} \\ &= A_k - \frac{y_k d_k^T A_k + A_k d_k y_k^T}{\langle y_k, d_k\rangle} + \left(1 + \frac{\langle d_k, Ad_k\rangle}{\langle y_k, d_k\rangle}\right)\frac{y_k y_k^T}{\langle y_k, d_k\rangle} \end{aligned}$$

qui est bien la formule duale de (BFGS).

4. On vérifie que les deux formules conservent le caractère symétrique. On suppose W_k symétrique, on a alors pour (DFP)

$$\begin{aligned} W_{k+1}^T &= W_k^T + \frac{(d_k d_k^T)^T}{\langle y_k, d_k\rangle} - \frac{(W_k y_k y_k^T W_k)^T}{\langle y_k, W_k y_k\rangle} \\ &= W_k + \frac{d_k d_k^T}{\langle y_k, d_k\rangle} - \frac{W_k y_k y_k^T W_k}{\langle y_k, W_k y_k\rangle} \\ &= W_{k+1} \end{aligned}$$

De même pour (BFGS)

$$\begin{aligned} W_{k+1}^T &= W_k^T - \frac{(d_k y_k^T W_k)^T + (W_k y_k d_k^T)^T}{\langle y_k, d_k\rangle} + \left(1 + \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, d_k\rangle}\right)\frac{(d_k d_k^T)^T}{\langle y_k, d_k\rangle} \\ &= W_k - \frac{W_k y_k d_k^T + d_k y_k^T W_k}{\langle y_k, d_k\rangle} + \left(1 + \frac{\langle y_k, W_k y_k\rangle}{\langle y_k, d_k\rangle}\right)\frac{d_k d_k^T}{\langle y_k, d_k\rangle} \\ &= W_{k+1}. \end{aligned}$$

Par ailleurs, comme une matrice est définie positive si son inverse l'est aussi, grâce à la dualité montrée à la question précédente, il suffit de montrer que l'une des deux formules (DFP) et (BFGS) conserve le caractère défini positif de la matrice W_k si $\langle y_k, d_k\rangle > 0$. On utilise la formle duale de (DFP) montrée à la question précédente et on étudie $\langle A_{k+1}u, u\rangle$ pour $u \neq 0$

$$\begin{aligned}
\langle A_{k+1}u, u\rangle &= \langle \left(I - \frac{y_k d_k^T}{\langle y_k, d_k\rangle}\right) A_k \left(I - \frac{d_k y_k^T}{\langle y_k, d_k\rangle}\right) u, u\rangle + \frac{\langle y_k y_k^T u, u\rangle}{\langle y_k, d_k\rangle} \\
&= \langle A_k \left(I - \frac{d_k y_k^T}{\langle y_k, d_k\rangle}\right) u, \left(I - \frac{d_k y_k^T}{\langle y_k, d_k\rangle}\right) u\rangle + \frac{\langle y_k y_k^T u, u\rangle}{\langle y_k, d_k\rangle} \\
&= \langle A_k v, v\rangle + \frac{\langle y_k^T u, y_k^T u\rangle}{\langle y_k, d_k\rangle} \geq 0,
\end{aligned}$$

où on a posé $v = \left(I - \frac{d_k y_k^T}{\langle y_k, d_k\rangle}\right) u$. La somme de deux termes positifs est nulle si chacun des termes est nul, ce qui implique, si A_k est définie

$$\begin{aligned}
&\left(I - \frac{d_k y_k^T}{\langle y_k, d_k\rangle}\right) u = 0 \text{ et } y_k^T u = 0, \\
&u = \frac{y_k^T u}{y_k^T d_k} d_k \text{ et } y_k^T u = 0, \\
&u = 0,
\end{aligned}$$

donc A_{k+1} est définie.

■

2.3 Méthodes de descente

2.3.1 Direction de descente

Les méthodes du paragraphe précédent consistent à chercher un minimum (ou un maximum) d'une fonction en annulant son gradient. Ce sont des méthodes relativement coûteuses car elles nécessitent de connaître le hessien de la fonction à minimiser pour pouvoir utiliser l'algorithme de Newton. On a vu par ailleurs que ce dernier ne converge que dans une zone limitée autour de la cible. On va voir dans ce paragraphe des méthodes plus robustes et moins coûteuses, qui vont être pilotées par la décroissance de la fonction qu'on cherche à minimiser.

Définition 2.2. On dit que $d \in \mathbb{R}^n$ est une direction de descente en x pour la fonction f s'il existe $\alpha_d > 0$ tel que

$$f(x + \alpha d) < f(x) \quad \forall \alpha < \alpha_d.$$

Grâce à la formule de Taylor (1.2) on a la caractérisation suivante

Proposition 2.1. *Si $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable, $d \in \mathbb{R}^n$ est une direction de descente en x si et seulement si*

$$\langle \nabla f(x), d\rangle < 0.$$

On regroupe sous le terme de méthodes de descente les méthodes itératives

$$x_{k+1} = x_k + \alpha_k d_k$$

qui sont caractérisées par

— Le calcul de la *direction* d_k (donne le nom à la méthode, gradient, gradient conjugué, etc.)

— La détermination du *pas* α_k (recherche linéaire).

L'algorithme général d'une méthode de descente est donc

Algorithme 2.5 : Algorithme général de descente

Données : La fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$
La précision demandée $\varepsilon > 0$.
Résultat : $x^\star$ tel que $f(x^\star) = \min_x f(x)$
Initialisation : $k = 0$,
Une première approximation de la solution $x_0 \in \mathbb{R}^n$
tant que $\|\nabla f(x_k)\| > \varepsilon$ **faire**
  Choix de la direction de descente d_k (telle que $\langle \nabla f(x_k), d_k \rangle < 0$)
  Choix du pas α_k dans la direction d_k, tel que $f(x_k + \alpha_k d_k) \leq f(x_k)$
  $x_{k+1} = x_k + \alpha_k d_k$
  $k \leftarrow k + 1$
fin
$x^\star \leftarrow x_k$

Les méthodes de descente sont simples mais en pratique elles convergent parfois lentement, voire pas du tout. Par ailleurs les conditions de convergence sont difficiles à vérifier a priori. En effet on a bien le résultat suivant, mais les constantes permettant d'assurer la convergence de l'algorithme ne sont pas aisées à calculer en pratique.

Théorème 2.9. *Soit f une fonction C^1 de $\mathbb{R}^n$ dans $\mathbb{R}$ et $x^\star$ un minimiseur de f. Si les conditions suivantes sont vérifiées*

1. *L'application f est α-elliptique, c'est-à-dire que*

$$\exists \alpha > 0, \forall (x, y) \in (\mathbb{R}^n)^2, \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2$$

2. *L'application ∇f est lipschitzienne, c'est-à-dire que*

$$\exists L > 0, \forall (x, y) \in (\mathbb{R}^n)^2, \quad \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$

Pour toute suite $(\alpha_k)_{k \in \mathbb{N}}$ telle qu'il existe a, b réels, tels que

$$0 < a < \alpha_k < b < \frac{2\alpha}{L^2}, \quad \forall k \in \mathbb{N},$$

la méthode du gradient définie par

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

converge géométriquement pour tout choix de condition initiale, c'est à dire que

$$\exists \beta \in]0, 1[, \|x_k - x^\star\| \leq \beta^k \|x_0 - x^\star\|$$

Exemples de choix possibles pour la direction de descente

— Algorithme du gradient ou de la plus grande pente (*steepest descent* en anglais)

$$d_k = -\nabla f(x_k).$$

— L'algorithme de Newton basé sur la direction

$$d_k = -Hf(x_k)^{-1} \nabla f(x_k).$$

Remarque : d_k n'est pas forcément une direction de descente si les hypothèses du Théorème 2.8 ne sont pas satisfaites.

— L'algorithme de quasi-Newton avec

$$d_k = -W_k \nabla f(x_k),$$

où W_k est une approximation de l'inverse du hessien, mis à jour à chaque itération.

— Algorithme du gradient conjugué (dans le cas quadratique)

$$d_k = \begin{cases} -\nabla f(x_1) & \text{pour } k = 1 \\ -\nabla f(x_k) + \beta_k d_{k-1} & \text{pour } k > 1. \end{cases}$$

Exercice 2.6. On considère la fonction $J : \mathbb{R}^2 \to \mathbb{R}^+$, définie pour $x = (x_1, x_2)^t$ par

$$J(x) = (x_1 - 1)^2 + (x_2^2 - x_1)^2. \tag{2.3}$$

1. Quel est le minimum de $J(x)$? où est-il atteint ?
2. Le vecteur $(0, -1)^t$ est-il une direction de descente au point $(0, 1)^t$?
3. Si oui, trouver le minimum de J dans cette direction.

Corrigé :

1. Le gradient de la fonction $J(x)$ est

$$\nabla J(x) = \begin{pmatrix} 4x_1 - 2 - 2x_2^2 \\ 4x_2(x_2^2 - x_1) \end{pmatrix}$$

Les extrema de J sont atteints en des points où le gradient s'annule soit : en $x_1 = x_2^2 = 1$ ou $x_2 = 0, x_1 = 1/2$. Or $J(1, \pm 1) = 0$ et $J(1/2, 0) = 1/2$. Donc le minimum de J est 0 et il est atteint aux deux points $(1, \pm 1)$ (inutile de calculer le hessien puisque $J(x) \geq 0$).

2. $\nabla J((0, 1)^t) = (-4, 4)^t$, donc $\langle \nabla J((0, 1)^t), (0, -1)^t \rangle = -4 < 0$ Le vecteur $(0, -1)^t$ est bien une direction de descente au point $(0, 1)^t$.
3. Le minimum de J dans cette direction est le minimum de

$$h(\alpha) = J((0, 1)^t + \alpha(0, -1)^t) = 1 + (1 - \alpha)^4.$$

Il est atteint en $\alpha = 1$ soit au point $(0, 0)^t$. En ce point J vaut 1.

■

2.3.2 Détermination du déplacement dans une direction donnée

Une étape cruciale dans les algorithmes de descente est la recherche du pas optimal dans la direction d_k. Le pas optimal est trouvé en minimisant la fonction h_k

$$h_k : \alpha \mapsto h_k(\alpha) = f_k(x_k + \alpha d_k)$$

La Figure 2.4 montre le pas utilisé en appliquant cette règle, dite Règle de Cauchy

$$\alpha_k = \underset{\alpha > 0}{\operatorname{argmin}}\, h_k(\alpha)$$

Une valeur de pas moins coûteuse à calculer mais non optimale est donnée par la Règle de Curry

$$\alpha_k = \inf \left\{ \alpha > 0; h_k'(\alpha) = 0, h_k(\alpha) < h_k(0) \right\}$$

Si on est capable de calculer α_k par la règle de Curry, on a le résultat de convergence suivant, pour l'algorithme du gradient

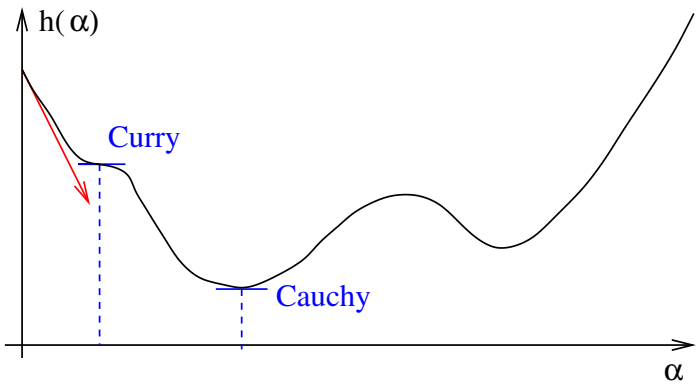


FIGURE 2.4 – Règles de Cauchy et de Curry

Théorème 2.10. *On suppose que $\nabla f(x)$ est lipschitzienne sur le domaine $\{x, f(x) \leq f(x^0)\}$. Alors l'algorithme 2.2 avec $d_k = \nabla f(x^k)$ et α_k déterminé par la règle de Curry satisfait*

- *soit $f(x^k)$ est non bornée inférieurement*
- *soit $\nabla f(x^k) \to 0$ quand $k \to \infty$.*

Preuve Voir [1] page 22 ■

La recherche du α optimal est un problème difficile et coûteux, sauf dans le cas quadratique, que nous allons détailler dans le prochain paragraphe. Nous verrons par la suite des algorithmes de calcul de pas non optimaux mais efficaces.

Cas quadratique Dans le cas où la fonction f est quadratique, c'est-à-dire qu'on peut l'écrire sous la forme $f(x) = \langle Ax, x\rangle - \langle b, x\rangle$ avec A matrice carrée $n \times n$ et b vecteur de $\mathbb{R}^n$, les méthodes de descente ci-dessus se précisent car on peut calculer explicitement le meilleur pas dans une direction donnée d. En effet si

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$$

avec de plus la matrice A symétrique, on peut calculer explicitement le minimum de la fonction

$$h(\alpha) = f(x + \alpha d) = f(x) + \frac{\alpha^2}{2}\langle Ad, d\rangle + \alpha\langle Ax - b, d\rangle$$

qui est réalisé pour

$$\alpha = -\frac{\langle g, d\rangle}{\langle Ad, d\rangle} \tag{2.4}$$

avec $g = \nabla f(x) = Ax - b$. La méthode du gradient devient la

Algorithme 2.6 : Méthode du gradient optimal

Données : La matrice A et le vecteur b, la tolérance ε
Résultat : $x^\star$ tel que $f(x^\star) = \min_x f(x)$
Initialisation : $k = 0$,
Une première approximation de la solution $x_0 \in \mathbb{R}^n$
$g_0 = Ax_0 - b$
tant que $\|g_k\| > \varepsilon$ **faire**
  $d_k = -g_k$
  $v_k = Ad_k$
  $\alpha_k = \dfrac{\langle d_k, d_k\rangle}{\langle v_k, d_k\rangle}$
  $x_{k+1} = x_k + \alpha_k d_k$
  $g_{k+1} = g_k + \alpha_k v_k$
  $k \leftarrow k + 1$
fin
$x^\star \leftarrow x_k$

Exercice 2.7. Reprendre les fonctions de l'exercice 1.11 et déterminer l'ensemble des directions de descente en $x = 0$.

Corrigé :

1. Pour $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$, $f(x) = \sin(x_1)e^{x_2}(1 + x_3^2)$ on a pour le gradient

$$\nabla f(x) = \begin{pmatrix} \cos(x_1)e^{x_2}(1 + x_3^2) \\ \sin(x_1)e^{x_2}(1 + x_3^2) \\ 2x_3 \sin(x_1)e^{x_2} \end{pmatrix}, \quad \text{d'où } \nabla f(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

   Les directions de descente en 0 sont les vecteurs $d \in \mathbb{R}^3$ t.q. $\langle \nabla f(0), d\rangle < 0$; il s'agit donc du demi-espace $\{x \in \mathbb{R}^3, x_1 < 0\}$

2. Pour la deuxième fonction $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$, $f(x) = e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right)$. on remarque que

$$f(x) = e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right) = e^{x_1}(1 - x_2^2)\frac{\sin(x_3) + \cos(x_3)}{\cos(x_3) - \sin(x_3)}.$$

   Par ailleurs on calcule le gradient

$$\nabla f(x) = \begin{pmatrix} e^{x_1}(1 - x_2^2)\text{tg}\left(x_3 + \frac{\pi}{4}\right) \\ -2e^{x_1}x_2\text{tg}\left(x_3 + \frac{\pi}{4}\right) \\ e^{x_1}(1 - x_2^2)\left(1 + \text{tg}^2\left(x_3 + \frac{\pi}{4}\right)\right) \end{pmatrix}, \quad \text{d'où } \nabla f([0,0,0]) = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

   Les directions de descente en 0 sont les vecteurs $d \in \mathbb{R}^3$ t.q. $\langle \nabla f(0), d\rangle < 0$; il s'agit donc du demi-espace $\{x \in \mathbb{R}^3, x_1 + 2x_3 < 0\}$

■

Exercice 2.8. 1. Programmer la visualisation des courbes de niveaux pour les fonctions test

$$J(X) = (x_1 - 2)^2 + (x_2 - 3)^2 \text{ et } J(X) = (x_1 - 2)^2 + (x_2 - 3)^4$$

puis sur la fonction de Rastrigin

$$J(X) = n + \|X\|^2 - \sum_{i=1}^{n} \cos(bx_i)$$

2. Tester la méthode du gradient avec la recherche du pas par la méthode Wolfe sur les trois fonctions tests. Etudier l'influence des différents paramètres dans la recherche du pas. Dans le cas de la fonction quadratique, comparer avec le pas optimal. L'algorithme de Wolfe est programmé dans le fichier `Wolfe.m`.

3. Comparer avec les résultats obtenus en utilisant la boite à outils `fminunc` de `Matlab`. En particulier on testera l'option 'GradObj'='on' et on affichera en sortie le nombre d'itérations et le nombre d'appels à la fonction objectif et à son gradient.
4. Programmer la méthode du gradient conjugué et la méthode de Polak-Ribière. Comparer les résultats dans le cas de la fonctionnelle quadratique.

Exercice 2.9. On considère le problème de minimisation

$$x^\star = \underset{\mathbb{R}^2}{\operatorname{argmin}}\, f(x), \quad \text{avec } f(x) = (x_1 - 1)^4 + (x_2 - 2)^4$$

1. Résoudre le problème de manière exacte.
2. Ecrire l'algorithme de Newton pour ce problème. Montrer qu'il converge pour n'importe quel choix de $x^0 \in \mathbb{R}^2$.
3. On considère maintenant l'algorithme du gradient pour ce problème. Pour choisir le pas α_k dans la direction du gradient

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

on considère la fonction $\varphi^k(\rho) = f(x^k - \rho \nabla f(x^k))$ et on propose de calculer $\rho_k = \operatorname{argmin}_{rho>0} p^k(\rho)$ avec $p^k(\rho)$ le polynome de degré deux qui interpole $\varphi(0)$, $\varphi'(0)$ et $\varphi''(0)$.

— Calculer ρ^k

— Montrer qu'il existe $\alpha > 0$ tel que si $\alpha_k = \alpha \rho_k$ alors

$$\|x^{k+1} - x^\star\| \leq \mu \|x^k - x^\star\| \text{ avec } 0 < \mu < 1$$

— En déduire que $\lim_{k\to\infty} x^k = x^\star$ et une estimation d'erreur par rapport à $\|x^0 - x^\star\|$.

Corrigé :

1. $X^\star = (1, 2)^T$.

$$\nabla f(x, y) = \begin{pmatrix} 4(x-1)^3 \\ 4(y-2)^3 \end{pmatrix}, \quad Hf(x, y) = \begin{pmatrix} 12(x-1)^2 & 0 \\ 0 & 12(y-2)^2 \end{pmatrix},$$

Etape k de l'algorithme de Newton

$$\begin{aligned} \nabla f(x_k, y_k) &= -Hf(x_k, y_k) d_k \\ X^{k+1} &= X^k + d_k \end{aligned}$$

d'où

$$\begin{aligned} x_{k+1} &= x_k - \frac{1}{3}(x_k - 1) \\ y_{k+1} &= y_k - \frac{1}{3}(y_k - 2) \end{aligned}$$

ou encore

$$\begin{aligned} x_{k+1} - 1 &= \frac{2}{3}(x_k - 1) \\ y_{k+1} - 2 &= \frac{2}{3}(y_k - 2) \end{aligned}$$

La suite $X^k - X^\star$ est contractante de constante $2/3$ donc $\to 0$ quand $k \to \infty$ pour tout $X^0 = (x_0, y_0)^T$.

2.

$$p^k(\rho) = \varphi(0) + \rho\varphi'(0) + \frac{1}{2}\rho^2\varphi"(0)$$

donc

$$\rho_k = \frac{-\varphi'(0)}{\varphi"(0)} = \frac{1}{12}\frac{(x_k-1)^6 + (y_k-2)^6}{(x_k-1)^8 + (y_k-2)^8}$$

$$\begin{aligned}
\|X^{k+1} - X^\star\|^2 &= \|X^k - \alpha\rho_k\nabla f(X^k) - X^\star\|^2 \\
&= \|X^k - X^\star\|^2 + \alpha^2\rho_k^2\|\nabla f(X^k)\|^2 - 2\alpha\rho_k\langle X^k - X^\star, \nabla f(X^k)\rangle \\
&= (x_k-1)^2 + (y_k-2)^2 + 16\alpha^2\rho_k^2((x_k-1)^6 + (y_k-2)^6) - 8\alpha\rho_k((x_k-1)^4 + (y_k-2)^4) \\
&= (x_k-1)^2(1 - 4\alpha\rho_k(x_k-1)^2)^2 + (y_k-2)^2(1 - 4\alpha\rho_k(y_k-2)^2)^2
\end{aligned}$$

On cherche donc une condition sur α pour que

$$|1 - 4\alpha\rho_k(x_k-1)^2| < 1 \text{ et } |1 - 4\alpha\rho_k(y_k-2)^2| < 1$$

ce qui équivaut à

$$2\alpha\rho_k(x_k-1)^2 < 1 \text{ et } 2\alpha\rho_k(y_k-2)^2 < 1.$$

Une condition suffisante est

$$\begin{aligned}
2\alpha\rho_k((x_k-1)^2 + (y_k-2)^2) &< 1 \\
\alpha\frac{1}{6}\frac{((x_k-1)^6 + (y_k-2)^6)((x_k-1)^2 + (y_k-2)^2)}{(x_k-1)^8 + (y_k-2)^8} &< 1
\end{aligned}$$

Posons $X = (x_k-1)^2$ et $Y = (y_k-2)^2$. On a

$$\alpha\frac{1}{6}\frac{(X^3+Y^3)(X+Y)}{X^4+Y^4} = \alpha\frac{1}{6}\left(1 + \frac{X^3Y + XY^3}{X^4+Y^4}\right) < 1$$

Or $X^3Y + XY^3 - X^4 - Y^4 = (X^3 - Y^3)(Y - X) \leq 0$. Donc si $\alpha < 3$

$$\alpha\frac{1}{6}\frac{(X^3+Y^3)(X+Y)}{X^4+Y^4} \leq \frac{1}{2}\left(1 + \frac{X^3Y + XY^3}{X^4+Y^4}\right) < 1.$$

■

La méthode du gradient, même dans un cas "idéal" comme une fonction quadratique, ne converge qu'asymptotiquement vers la solution. Regardons par exemple

$$f(x) = \frac{1}{2}(\alpha_1 x_1^2 + \alpha_2 x_2^2) \quad \text{avec } 0 < \alpha_1 < \alpha_2.$$

Le minimum de cette fonction est atteint en $x^\star = (0,0)$, et sauf en choisissant un point de départ sur un des deux axes, le gradient en x_k à l'itération k de l'algorithme du gradient n'est jamais colinéaire à x_k. Donc le minimum ne sera jamais exactement atteint. Pourtant, la recherche de ce minimum, dans le cas d'une fonction quadratique, revient à celle de la solution d'un système linéaire (celui annulant le gradient). Et on connait, si on a suivi un cours d'algèbre linéaire élémentaire, des méthodes pour résoudre un tel système en un nombre fini d'opérations... Il est clair que pour que la méthode d'optimisation soit compétitive, il faut utiliser une autre direction de descente que celle du gradient. Dans la méthode de Newton appliquée à une fonction quadratique, $Hf(x) = A$ et $\nabla f(x) = Ax - b$ d'où à chaque étape la détermination de la direction de descente par résolution de $Ad_k = -g_k$. En remplacant dans (2.4), on obtient $\alpha_k = 1$. Notons que la méthode ne présente pas d'intérêt puisqu'elle nécessite la résolution d'un système linéaire à chaque itération. Autant calculer directement le minimum de la fonction quadratique qui est la solution de $Ax = b$.

2.3.3 La méthode des gradients conjugués

La méthode des gradients conjugués est la base des méthodes rapides de minimisation des fonctions régulières et c'est aussi une méthode très utilisée pour résoudre les systèmes linéaires de grande dimension (issus par exemple d'une discrétisation par éléments finis d'une EDP). Le principe de cette méthode est d'accélérer la méthode du gradient en cherchant à l'étape k directement le minimum de la fonction $f(x)$ dans le plan (P) passant par le point x^k et engendré par les directions d^{k-1} et g^k, et non pas simplement g^k comme dans la méthode du gradient introduite plus haut. Notons V_k le sous espace vectoriel engendré par les gradients et les directions successives

$$V_k = \{g^i,\ i = 0, \ldots, k, \quad d^i,\ i = 0 \ldots k\} \tag{2.5}$$

et $U_k = \{x^k + \sum_{i=1}^{N_k} \alpha_i u_i, u_i \in B(V_k)\}$ l'espace affine engendré par V_k passant par x^k, où on a noté $B(V_k)$ une base de V_k et N_k la dimension de V_k

Proposition 2.2. *Les deux propriétés suivantes sont équivalentes*

- x^{k+1} *réalise le minimum de* f *sur* U_k,
- g^{k+1} *est orthogonal à* V_k.

L'espace vectoriel V_k *est engendré par* $\{g^i, i = 0, \ldots, k\}$.

Preuve Supposons $x^{k+1} = x^k + \sum \tilde{\alpha}_i u_i \in U_k$ et $f(x^{k+1}) \leq f(x), \forall x \in U_k$. C'est équivalent à dire que $\tilde{\alpha}$ minimise $h(\alpha) = f(x^k + \sum \alpha_i u_i)$ sur $\mathbb{R}^{N_k}$ donc le gradient de h doit être nul en $\tilde{\alpha}$

$$\frac{\partial h(\tilde{\alpha})}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{\partial f(x^{k+1})}{\partial x_i} \frac{\partial x_i^{k+1}}{\partial \alpha_j} = \langle \nabla f(x^{k+1}), u_j \rangle = 0.$$

Grâce à cette propriété, on montre par récurrence que V_k est en fait l'espace vectoriel engendré par les gradients successifs. A l'étape 0 c'est évident puisque $x^1 = x^0 + \alpha_0 g^0$. Supposons que $V_{k-1} = \{g^0, \ldots, g^{k-1}\}$ pour $k > 1$. On sait que $f(x^{k+1}) \leq f(x^k - t g^k)$ car x^{k+1} minimise f sur U_k contenant g^k et par ailleurs il existe $t > 0$ tel que $f(x^k - t g^k) \leq f(x^k)$ car g^k est une direction de descente (sinon si $g^k = 0$ on s'arrête). Comme $x^{k+1} \notin U_{k-1}$ (puisque $f(x^{k+1}) < f(x^k)$ et x^k minimum dans U_{k-1}), on en déduit que d^k n'est pas dans V_k et a forcément une composante non nulle sur g^k. Donc $d^k = \alpha_k g^k + u$ avec $\alpha_k \neq 0$ et $u \in V_{k-1}$, ou réciproquement $g^k = \frac{1}{\alpha_k}(d^k - u)$; on passe donc de V_{k-1} à V_k en rajoutant indifféremment d^k ou g^k à la famille génératrice de V_{k-1}. Par ailleurs, comme $V_k = \{g^0, \ldots, g^k\}$ avec les gradients orthogonaux deux à deux, la dimension de V^k est $k+1$ et l'algorithme s'arrête donc au plus tard pour $k = n - 1$. ■

On montre maintenant la propriété suivante

Théorème 2.11. *Soit* f *une fonction quadratique*

$$f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c,$$

avec A *symétrique définie positive. Supposons que* x^k *minimise* f *dans* U_{k-1}. *Alors les deux propriétés suivantes sont équivalentes :*

- *La direction* d^k *issue de* x^k *minimize* f *dans* U^k.
- $\langle d^k, Ad^i\rangle = 0$ *pour tout* $i = 0, \ldots k-1$.

Preuve D'après la propriété 2.2 si $f(x^k + t d^k)$ minimize f dans $V^k = \{d^0, \ldots, d^{k-1}, g^k\}$ alors $\nabla f(x^k + t d^k)$ est orthogonal à d^i, pour $i = 0, \ldots, k-1$, et à g^k. Or $\nabla f(y) = \nabla f(x) + A(y - x)$ donc $\nabla f(x^k + t d^k) = g^k + t A d^k$. Donc

$$\begin{aligned} \langle g^k + t A d^k, d^i\rangle &= 0, \quad i = 0, \ldots, k-1 \\ \langle g^k + t A d^k, g^k\rangle &= 0. \qquad\qquad (*) \end{aligned}$$

On a d'une part pour $i < k$ que $\langle g^k, d^i\rangle = 0$ toujours d'après la propriété 2.2, donc

$$\langle tAd^k, d^i\rangle = 0, \quad i = 0, \ldots, k-1.$$

Réciproquement, si $\langle d^k, Ad^i\rangle = 0$ pour tout $i = 0, \ldots k-1$, en utilisant l'équation (*) ci-dessus on obtient

$$t\langle Ad^k, g^k\rangle = -\langle g^k, g^k\rangle.$$

Comme $g^k \neq 0$ et $d^k \neq 0$ (sinon on a fini), $\langle g^k, g^k\rangle \neq 0$ et $\langle Ad^k, g^k\rangle \neq 0$ car A est définie positive, cette équation définit bien le minimum $x^{k+1} = x^k + td^k$. ■

Appliqué à une fonction quadratique cet algorithme converge donc en un nombre fini d'étapes (au plus N) et pour de grands systèmes il peut donner une très bonne précision en un nombre $p << N$ itérations.
Le calcul explicite des directions et des coefficients s'écrit le plus facilement en choisissant $(g^i)_{i=0,\ldots,k}$ comme base de V_k : on cherche $d^k = \sum_{i=0}^k b_i^k g^i$ tel que $f(x^k + \alpha_k d^k)$ minimise f sur U_k, en utilisant la propriété équivalente montrée dans le Théorème 2.11, c'est-à-dire $\langle d^k, Ad^i\rangle = 0$ pour $i = 0, \ldots, k-1$. Or pour $i > 0$, $d^i = x^{i+1} - x^i$ donc $Ad^i = g^{i+1} - g^i$, d'où, $\langle d^k, g^{i+1}\rangle = \langle d^k, g^i\rangle = \beta_k$ pour $i = 0, \ldots, k-1$. On a donc

$$\sum_{j=1}^k b_j \langle g^j, g^i\rangle = b_i^k \|g^i\|^2 \quad = \quad \beta_k \quad \forall i = 0, \ldots, k$$

La valeur de β_k n'est pas importante, puisqu'on cherche là la direction d^k (à laquelle il faudra ensuite associer un pas). Choisissons $\beta_k = -\|g^k\|^2$, on aura alors

Théorème 2.12. *La suite des directions est donnée par*

$$\begin{aligned} d^0 &= -g^0 \\ d^{k+1} &= -g^{k+1} + c_k d^k, \quad \textit{avec } c_k = \frac{\|g^{k+1}\|^2}{\|g^k\|^2}, \end{aligned}$$

et le pas optimal dans la direction d^k est alors

$$\alpha_k = -\frac{\langle g^k, d^k\rangle}{\langle Ad^k, d^k\rangle}$$

Preuve L'initialisation de la direction à $-g^0$ est immédiate. pour $k > 0$ on écrit

$$\begin{aligned} d^{k+1} &= \sum_{i=0}^{k+1} b_i^{k+1} g^i = \sum_{i=0}^{k+1} \frac{\beta_{k+1}}{\|g^i\|^2} g^i = -\sum_{i=0}^{k+1} \frac{\|g^{k+1}\|^2}{\|g^i\|^2} g^i \\ &= -g^{k+1} - \sum_{i=0}^{k} \frac{\|g^{k+1}\|^2}{\|g^i\|^2} g^i = -g^{k+1} - \frac{\|g^{k+1}\|^2}{\|g^k\|^2} \sum_{i=0}^{k} \frac{\|g^k\|^2}{\|g^i\|^2} g^i \\ &= -g^{k+1} + \frac{\|g^{k+1}\|^2}{\|g^k\|^2} \sum_{i=0}^{k} \frac{\beta_k}{\|g^i\|^2} g^i = -g^{k+1} + \frac{\|g^{k+1}\|^2}{\|g^k\|^2} \sum_{i=0}^{k} b_i^k g^i \\ &= -g^{k+1} + \frac{\|g^{k+1}\|^2}{\|g^k\|^2} d^k. \end{aligned}$$

On cherche maintenant le pas α_k qui minimise $h(\alpha) = f(x^k + \alpha d^k)$ soit

$$\begin{aligned} \langle \nabla f(x^k + \alpha_k d^k), d^k\rangle &= 0, \\ \langle A(x^k + \alpha_k d^k) + b, d^k\rangle &= 0 \\ \langle g^k + \alpha_k Ad^k, d^k\rangle &= 0, \\ \alpha_k &= -\frac{\langle g^k, d^k\rangle}{\langle Ad^k, d^k\rangle} \end{aligned}$$

On résume tous ces résultats dans l'algorithme suivant, où on remarquera l'introduction du vecteur $v^k = Ad^k$ qui évite de recalculer ce produit matrice vecteur pour la mise à jour du gradient.

Algorithme 2.7 : Méthode du gradient conjugué

Données : La matrice A et le vecteur b, la tolérance ε
Résultat : $x^\star$ tel que $f(x^\star) = \min_x f(x)$
Initialisation : $k = 0$,
Une première approximation de la solution $x^0 \in \mathbb{R}^n$
$g^0 = Ax^0 - b$
$d^0 = -g^0$
tant que $\|g^k\| > \varepsilon$ **faire**
— Calcul du minimum unidirectionnel :
$v^k = Ad^k$
$\alpha_k = -\dfrac{\langle g^k, d^k\rangle}{\langle v^k, d^k\rangle}$
$x^{k+1} = x^k + \alpha_k d^k$
— Mise à jour du gradient :
$g^{k+1} = g_k + \alpha_k v^k$
— Calcul de la nouvelle direction :
$c_k = \dfrac{\langle g^{k+1}, g^{k+1}\rangle}{\langle g^k, g^k\rangle}$
$d^{k+1} = -g^{k+1} + c_k d^k$
$k \leftarrow k + 1$
fin
$x^\star \leftarrow x^k$

On obtient pour les problèmes d'optimisation non quadratiques différentes variantes de la méthode du gradient conjugué en appliquant les formules du cas quadratique au cas d'une fonction quelconque (ce qui se justifie, dans le cas où une fonction est suffisamment régulière, par le fait qu'au voisinage du minimum elle est très proche d'une fonction quadratique). Ces méthodes sont proches des algorithmes de Quasi-Newton décrits au paragraphe 2.2.3, par exemple

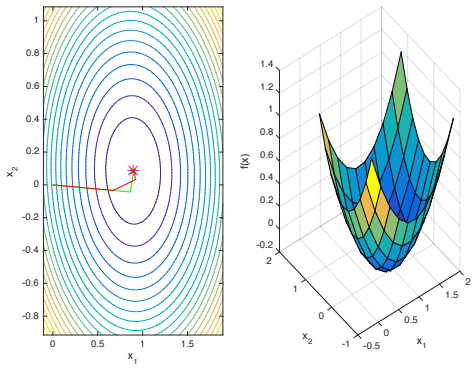


FIGURE 2.5 – Comparaison des méthodes du Gradient Conjugué (vert, 4 itérations) et de Polak-Ribière (rouge, 8 itérations) sur une fonction quadratique dans $\mathbb{R}^5$. Projection sur $(0, x_1, x_2)$.

Algorithme 2.8 : Méthode de Polak-Ribière

Données : La fonction f son gradient ∇f, la tolérance ε
Résultat : $x^\star$ tel que $f(x^\star) = \min_x f(x)$
Initialisation : $k = 0$,
Une première approximation de la solution $x^0 \in \mathbb{R}^n$
$g^0 = \nabla f(x^0)$
$d^0 = -g^0$
tant que $\|g^k\| > \varepsilon$ **faire**
- Calcul d'une approximation du minimum unidirectionnel vérifiant : $f(x^k + \alpha_k d^k) \leq f(x^k + \alpha d^k) < f(x^k)$ pour tout $0 < \alpha \leq \alpha_k$
- Calcul de la nouvelle position : $x^{k+1} = x^k + \alpha_k d^k$
- Calcul de la nouvelle direction :
$g^{k+1} = \nabla f(x^{k+1})$
$c_{k+1} = \dfrac{\langle g^{k+1} - g^k, g^{k+1}\rangle}{\langle g^k, g^k\rangle}$
$d^{k+1} = -g^{k+1} + c_{k+1} d^k$

$k \leftarrow k + 1$
fin
$x^\star \leftarrow x^k$

Malheureusement, si le point de départ est mal choisi, rien n'assure que la direction d^k reste une direction de descente. Des variantes de cet algorithme utilisent d'ailleurs des définitions différentes pour le coefficient c_k

$$\text{Fletcher-Reeves :} \quad c_{k+1} = \frac{\langle g^{k+1}, g^{k+1}\rangle}{\langle g^k, g^k\rangle}$$

$$\text{Hestenes-Stiefel :} \quad c_{k+1} = -\frac{\langle g^{k+1}, (g^{k+1} - g^k)\rangle}{\langle d^k, (g^{k+1} - g^k)\rangle}$$

2.3.4 Minimisation unidirectionnelle - Recherche linéaire

Comme on l'a annoncé au paragraphe 2.3.2 la détermination du pas α_k dans la direction d_k est un problème difficile dans le cas général (non quadratique). Nous donnons dans ce paragraphe deux méthodes abondamment utilisées en pratique car relativement robustes.

Règle d'Armijo Cette méthode consiste à linéariser la contrainte sur α_k (voir Figure 2.6)

$$f(x^k + \alpha_k d^k) < f(x^k) + \omega_1 \alpha_k \langle g^k, d^k\rangle \tag{2.6}$$

Algorithme 2.9 : Algorithme d'Armijo

Données : La fonction f, le point courant x, la direction de descente d, les coefficients $\tau \in]0, 1[$ et $\omega_1 \in]0, 1[$
Résultat : α tel que $f(x + \alpha d) < f(x)$
Initialisation : $k = 0$, la valeur initiale α_0
tant que $f(x + \alpha_k d) > f(x) + \omega \alpha_k \langle d, \nabla f(x)\rangle$ **faire**
Choisir $\alpha_{k+1} = \tau \alpha_k$
$k \leftarrow k + 1$
fin
$\alpha = \alpha_k$

En pratique, l'utilisation de la règle d'Armijo est rendue délicate par le choix de la valeur initiale et du coefficient ω_1. Il s'agit de faire les bons choix !...

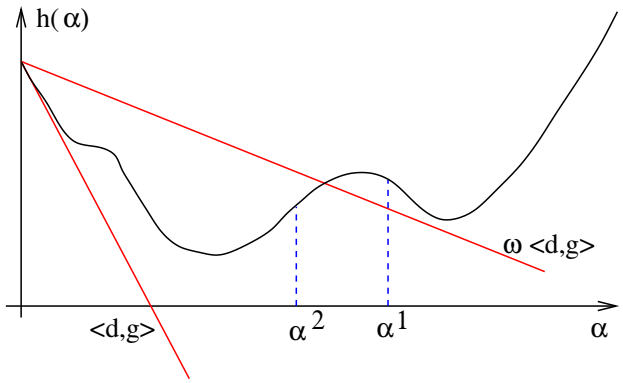


FIGURE 2.6 – Stratégie d'Armijo pour la recherche du pas. A la première itération $f(x^k + \alpha^1 d^k)$ est au dessus de la droite de coefficient $\omega_1 \langle g^k, d^k \rangle$ (en rouge), donc on diminue α_k. A la deuxième itération $f(x^k + \alpha^2 d^k)$ est en dessous de la droite et on choisit cette valeur pour α_k

La première valeur α_k^1 : On la fixe en faisant l'hypothèse d'un modèle quadratique pour $\varphi(\alpha) = f(x^k + \alpha d^k)$

$$\varphi(\alpha) = a_0 + a_1\alpha + a_2\alpha^2/2$$

avec

$$\begin{cases} a_0 = f(x^k) \\ a_1 = \langle d^k, \nabla f(x^k) \rangle \end{cases}$$

a_2 est déterminé en imposant la décroissance maximale de φ,

$$\Delta = \varphi(0) - \varphi_{\min} = a_1^2/(2a_2).$$

On pourra choisir par exemple un certain pourcentage de la valeur initiale. Attention cependant au choix de cette valeur si $f(x)$ n'est pas toujours positive.
Pas de Fletcher (voir Figure 2.7) α_k^1 est choisi pour donner la valeur minimale du modèle quadratique

$$\alpha_k^1 = \frac{-2\Delta}{\langle d^k, \nabla f(x^k) \rangle}$$

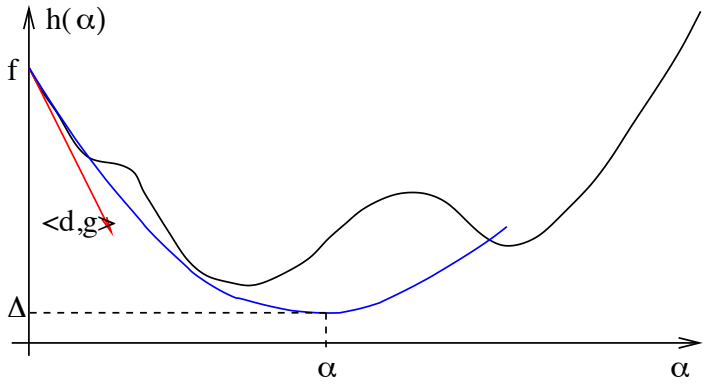


FIGURE 2.7 – Pas de Fletcher pour le démarrage de la stratégie d'Armijo

La stratégie d'Armijo est robuste, on a le résultat de convergence suivant

Théorème 2.13. *Si $f : \mathbb{R}^n \to \mathbb{R}$ est C^1 et que $\nabla f(x)$ est γ-lipschitzien alors la condition d'Armijo (2.6) est satisfaite pour tout*

$$\alpha \in [0, \omega], \quad \text{avec } \omega = \frac{(\omega_1 - 1)\langle \nabla f(x), d\rangle}{\gamma \|d\|^2}.$$

Inconvénient de la stratégie d'Armijo : on a toujours $\alpha_k^{i+1} < \alpha_k^i$, donc si α_k^1 est choisi trop petit, l'algorithme de descente converge très lentement, car à chaque itération le déplacement dans la direction de descente est trop petit.

Méthode de Wolfe Cette méthode permet de pallier l'inconvénient de la méthode d'Armijo en laissant aux valeurs successives du paramètre α_k la possibilité d'augmenter, si la valeur initiale est trop petite (voir Figures **??** et **??**) On rajoute pour cela à la condition d'Armijo (2.6), une borne inférieure sur la pente de la fonction en $x^k + \alpha_k d^k$

$$\langle \nabla f((x^k + \alpha_k d^k), d^k\rangle > \omega_2 \langle g^k, d^k\rangle, \text{ avec } 0 < \omega_1 < \omega_2 < 1. \tag{2.7}$$

Algorithme 2.10 : Algorithme de Wolfe

Données : La fonction f, son gradient ∇f, le point courant x^k, la direction de descente d^k, les coefficients $0 < \omega_1 < \omega_2 < 1$
Résultat : α_k tel que (2.6) et (2.7) sont satisfaites
Initialisation : Fixer $\alpha_D = -1$ et $\alpha_G = 0$.
$p = 0$, Fixer la valeur initiale de α_k^p (par exemple avec la méthode de l'approximation quadratique)
tant que $f(x^k + \alpha_k^p d^k) > f(x^k) + \omega_1 \alpha_k^p \langle d^k, \nabla f(x^k)\rangle$ *ou* $\langle \nabla f((x^k + \alpha_k^p d^k), d^k\rangle < \omega_2 \langle g^k, d^k\rangle$ **faire**
  si $f(x^k + \alpha_k^p d^k) > f(x^k) + \omega_1 \langle g^k, d^k\rangle \alpha_k^p$ **alors**
    $\alpha_D = \alpha_k^p$
  fin
  sinon
    $\alpha_G = \alpha_k^p$
  fin
  si $\alpha_D < 0$ *(pas encore fixé)* **alors**
    $\alpha_k^{p+1} = 2\alpha_G$
  fin
  sinon
    $\alpha_k^{p+1} = (\alpha_G + \alpha_D)/2$
  fin
  $p \leftarrow p + 1$
fin
$\alpha_k = \alpha_k^p$

On a également un résultat de convergence pour la méthode de Wolfe

Théorème 2.14. *"de Zoutendijk"*
Soit une fonction $f : \mathbb{R}^n \to \mathbb{R}$, minorée, continûment différentiable sur

$$\mathcal{N} = \{f(x) \leq f(x_0)\}$$

et dont le gradient $\nabla f(x)$ est L-lipschitzien. Alors, si les coefficients $(\alpha_k)_k$ vérifient les conditions (2.6) et (2.7)

$$\sum_k \cos\theta_k^2 \|\nabla f(x^k)\|^2 < \infty, \quad \text{avec } \cos\theta_k = \frac{-\langle d^k, \nabla f(x^k)\rangle}{\|d^k\|, \|\nabla f(x^k)\|}.$$

On déduit de ce théorème que l'algorithme converge vers un minimum de $f(x)$.
Preuve

Théorème 1. *de convergence (de Zoutendijk)*
On considère un algorithme de descente

$$\begin{aligned} x_0 &\in \mathbb{R}^n \\ x_{k+1} &= x_k + \alpha_k d_k, \end{aligned}$$

pour la minimisation d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$. On note

$$\cos\theta_k = \frac{-\langle \nabla f(x_k), d_k\rangle}{\|\nabla f(x_k)\|\|d_k\|}.$$

On suppose que

— *les coefficients α_k vérifient tous les conditions de Wolfe, avec $0 < \omega_1 < \omega_2 < 1$*

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \omega_1 \alpha_k \langle \nabla f(x_k), d_k\rangle \\ \langle \nabla f(x_k + \alpha_k d_k), d_k\rangle &\geq \omega_2 \langle \nabla f(x_k), d_k\rangle \end{aligned}$$

— *la fonction f est minorée, continument différentiable sur*

$$\mathcal{N} = \{f(x) \leq f(x_0)\}$$

— *le gradient de f est Lipschitzien sur $\mathcal{N}$*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \text{avec } L > 0$$

Alors

$$\sum_k \cos^2\theta_k \|\nabla f(x_k)\|^2 < \infty$$

Démonstration : α_k vérifie la deuxième condition de Wolfe, donc

$$\begin{aligned} \langle \nabla f(x_k + \alpha_k d_k), d_k\rangle &\geq \omega_2 \langle \nabla f(x_k), d_k\rangle \\ \langle \nabla f(x_{k+1}, d_k\rangle - \langle \nabla f(x_k), d_k\rangle &\geq (\omega_2 - 1)\langle \nabla f(x_k), d_k\rangle. \end{aligned}$$

Par ailleurs ∇f est Lipschitzien, d'où

$$\langle \nabla f(x_{k+1}) - \nabla f(x_k), d_k\rangle \leq \|\nabla f(x_{k+1}) - \nabla f(x_k)\|\|d_k\| \leq L\|x_{k+1} - x_k\|\|d_k\| = L\alpha_k\|d_k\|^2,$$

d'où on obtient

$$L\alpha_k\|d_k\|^2 \geq (\omega_2 - 1)\langle \nabla f(x_k), d_k\rangle.$$

Par ailleurs avec la première condition (Armijo) on a

$$f(x_{k+1}) - f(x_k) \leq \omega_1 \alpha_k \langle \nabla f(x_k), d_k\rangle,$$

donc, comme $\langle \nabla f(x_k), d_k\rangle < 0$,

$$f(x_{k+1}) - f(x_k) \leq \frac{\omega_1(\omega_2 - 1)}{L} \frac{\langle \nabla f(x_k), d_k\rangle^2}{\|d_k\|^2},$$

d'où, en sommant sur toutes les itérations,

$$f(x_k) - f(x_0) \leq -c \sum_k \frac{\langle \nabla f(x_k), d_k\rangle^2}{\|d_k\|^2},$$

avec

$$c = \frac{\omega_1(1-\omega_2)}{L} > 0.$$

En remarquant que

$$\begin{aligned} \frac{\langle \nabla f(x_k), d_k\rangle^2}{\|d_k\|^2} &= \left(\frac{-\langle \nabla f(x_k), d_k\rangle}{\|d_k\|\|\nabla f(x_k)\|}\right)^2 \times \|\nabla f(x_k)\|^2 \\ &= \cos\theta_k^2 \times \|\nabla f(x_k)\|^2, \end{aligned}$$

on a donc

$$f(x_k) - f(x_0) \leq -c\sum_k \cos\theta_k^2 \|\nabla f(x_k)\|^2,$$

d'où, puisque que f est minorée (par m),

$$c\sum_k \cos\theta_k^2 \|\nabla f(x_k)\|^2 \leq f(x_0) - m,$$

d'où le résultat.

Un exemple d'implémentation de la méthode de Wolfe dans Matlab est proposé dans l'annexe **??**.

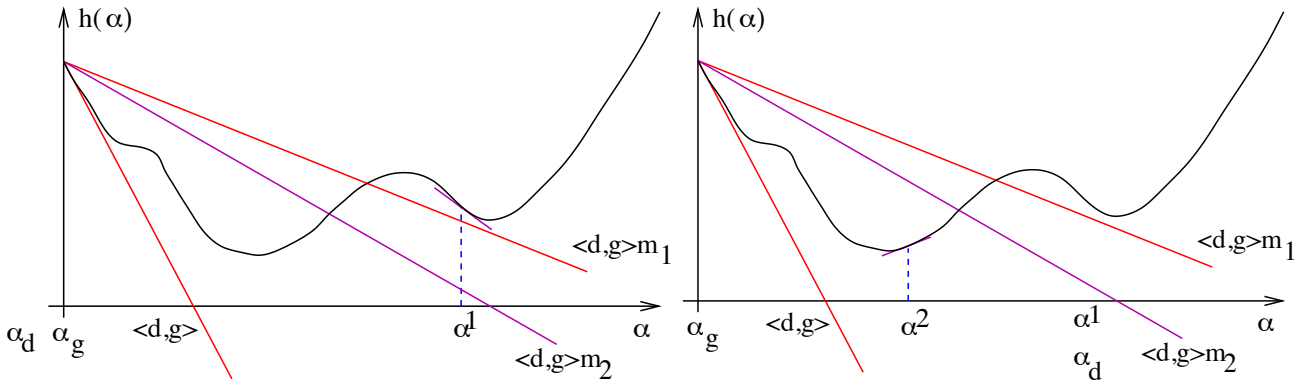


FIGURE 2.8 – Méthode de Wolfe, exemple 1, itérations 1 et 2. Dans ce cas Wolfe a fonctionné comme l'aurait fait Armijo. La première valeur de α conduisant à une valeur $f(x^k+\alpha^1 d^k)$ au dessus de la droite de coefficient $m_1\langle g^k, d^k\rangle$ (en rouge), on diminue α_k. A la deuxième itération $f(x^k + \alpha^1 d^k)$ est en dessous de la droite rouge et la pente en ce point est supérieur à la direction indiquée en violet, donc on choisit cette valeur α^2 pour α_k.

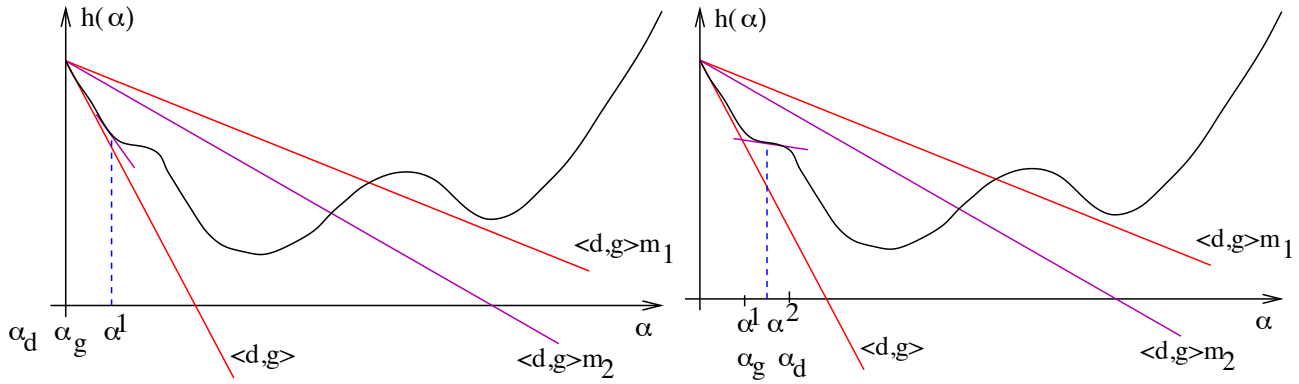


FIGURE 2.9 – Méthode de Wolfe, exemple 2, itérations 1 et 2. La première valeur de α conduisant à une valeur $f(x^k + \alpha^1 d^k)$ en dessous de la droite de coefficient $m_1\langle g^k, d^k\rangle$ (en rouge), mais avec une pente inférieure à la direction indiquée en violet. On augmente donc α_k. L'algorithme d'Armijo aurait arrêté dès la premiée itération, avec une valeur de α_k plus petite, donc moins optimale.

Exercice 2.10. Soit la fonction de $\mathbb{R}^2$ dans $\mathbb{R}$

$$f(x) = \tan(x_1)\sin(x_2 - \pi/2)$$

1. Calculer le gradient et le hessien de f.

2. Trouver une fonction quadratique $q(x)$ telle que

   $$\|f(x) - q(x)\| \leq C\|x\|^3$$

   dans un voisinage de $x = (\pi/4, \pi/2)$.

3. Trouver l'ensemble des directions de descente pour f au point $x = (\pi/4, \pi/2)$. Trouver parmi ces directions celle qui correspond à la descente la plus rapide. Justifier la réponse.

Corrigé : On peut commencer par ré-écrire $f(x)$ sous la forme

$$f(x) = -\tan(x_1)\cos(x_2)$$

1. Le gradient de f

$$\nabla f(x) = \begin{pmatrix} -\left(1+\tan(x_1)^2\right)\cos(x_2) \\ \tan(x_1)\sin(x_2) \end{pmatrix}$$

Le hessien de f

$$Hf(x) = \begin{pmatrix} -2\tan(x_1)\left(1+\tan(x_1)^2\right)\cos(x_2) & \left(1+\tan(x_1)^2\right)\sin(x_2) \\ \left(1+\tan(x_1)^2\right)\sin(x_2) & \tan(x_1)\cos(x_2) \end{pmatrix}$$

2. On remarque que

$$\nabla f(\bar{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad Hf(\bar{x}) = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$$

On applique la formule de Taylor

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x-\bar{x}\rangle + \frac{1}{2}\langle Hf(\bar{x})(x-\bar{x}), x-\bar{x}\rangle + O(\|x-\bar{x}\|^3)$$

donc

$$\|f(x) - q(x)\| \leq C\|x\|^3$$

avec

$$\begin{aligned} q(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x-\bar{x}\rangle + \frac{1}{2}\langle Hf(\bar{x})(x-\bar{x}), x-\bar{x}\rangle \\ &= (x_2 - \pi/2)(1-\pi+4x_1). \end{aligned}$$

3. Les directions de descente pour f au point $\bar{x} = (\pi/4, \pi/2)$ sont telles que

$$\langle d, \nabla f(\bar{x})\rangle < 0$$

soit $\{\mathbb{R}^{-*} \times \mathbb{R}\}$. Parmi ces directions celle qui correspond à la descente la plus rapide est la direction opposée au gradient donc $-e_1$. En effet, d'après la formule de Taylor au premier ordre on a

$$f(\bar{x} + \alpha d) = f(\bar{x}) + \alpha\langle d, \nabla f(\bar{x})\rangle + O(\|x-\bar{x}\|^2)$$

Comme

$$\langle d, \nabla f(\bar{x})\rangle = \cos(d, \widehat{\nabla f(\bar{x})})\|d\|\|\nabla f(\bar{x})\|,$$

pour le même incrément α, la diminution de f dans la direction d est donc maximum quand $d = -\nabla f(\bar{x})$.

■

Cas particulier de l'identification de paramètres On s'intéresse maintenant au cas où la fonction à minimiser mesure l'écart entre des données expérimentales et des valeurs calculées par un modèle dépendant d'un certain nombre de paramètres inconnus a priori. Le minimum de la fonction est réalisé pour le jeu de paramètres permettant d'expliquer au mieux les mesures expérimentales avec le modèle en question. Des techniques de minimisation ont été spécialement développées pour ce type de problème. Nous détaillons ici la méthode de Levenberg-Marquardt utilisée dans le cas où la fonctionnelle à minimiser correspond à un problème de moindres carrés

$$F(p) = \sum_{i=1}^{n} \|y_i - f(x_i, p)\|^2 \tag{2.8}$$

où

— le modèle $f(x,p)$ dépend de m paramètres p_j, $j = 1,\ldots,m$. La fonction f est à valeurs éventuellement vectorielles

$$\begin{array}{rcl} f \,:\, \mathbb{R}^d \times \mathbb{R}^m & \longrightarrow & \mathbb{R}^q \\ (x,p) & \mapsto & y \end{array}$$

— les données sont les n mesures y_i, $i = 1,\ldots,n$ à valeurs éventuellement vectorielles dans $\mathbb{R}^q$, correspondant aux n coordonnées x_i, $i = 1,\ldots,n$ éventuellement vectorielles dans $\mathbb{R}^d$

Si le modèle est linéaire, avec $m = l + 1$ paramètres

$$f(x,p) = Mp, \quad \text{avec } M = (x^T, 1)$$

à valeurs dans $\mathbb{R}$ ($q = 1$) on obtient pour la fonctionnelle

$$F(p) = \sum_{i=1}^{n} \|y_i - M_i p\|^2 = \langle Y - Mp, Y - Mp\rangle$$

C'est une forme quadratique en p. Le minimum est atteint pour p_{opt} solution de $Ap + b = 0$ avec $A = 2M^TM$ et $b = -2M^Ty$.
Dans le cas général le gradient et le hessien par rapport à p de la fonctionnelle sont respectivement

$$G = \nabla F(p) = \left(\frac{\partial F}{\partial p_k}\right)_{k=1,\ldots,m}, \quad \frac{\partial F}{\partial p_k} = 2\sum_{i=1}^{n}(f(x_i,p) - y_i)\frac{\partial f(x_i,p)}{\partial p_k} \tag{2.9}$$

$$HF(p) = \left(\frac{\partial^2 F}{\partial p_k \partial p_j}\right)_{k,j=1,\ldots,m}, \quad \frac{\partial^2 F}{\partial p_k \partial p_j} = 2\sum_{i=1}^{n}(f(x_i,p) - y_i)\frac{\partial^2 f(x_i,p)}{\partial p_k \partial p_j} + 2\sum_{i=1}^{n}\frac{\partial f(x_i,p)}{\partial p_k}\frac{\partial f(x_i,p)}{\partial p_j}$$

La méthode de Levenberg Marquardt consiste à combiner la méthode du gradient et la méthode de Newton, en privilégiant cette dernière dans les zones proches d'un minimum. Pour cela on commence par approcher la matrice hessienne par

$$H = (H_{k,j}), \quad H_{k,j} = 2\sum_{i=1}^{n}\frac{\partial f(x_i,p)}{\partial p_k}\frac{\partial f(x_i,p)}{\partial p_j} \tag{2.10}$$

c'est-à-dire en négligeant les termes en dérivées secondes. À l'étape k on définit la direction de descente d_k comme la solution du système linéaire

$$\tilde{H}_k d_k = -G_k, \text{ avec } \tilde{H}_{i,j} = H_{i,j}(1 + \lambda\delta_{ij}), \tag{2.11}$$

où δ_{ij} est le symbole de Kronecker et λ un paramètre qui va varier à chaque itération. Quand λ est très grand, la matrice $\tilde{H}$ est à diagonale dominante et le comportement de l'algorithme est proche de celui du gradient. Au contraire quand $\lambda \to 0$ c'est plutôt le comportement de l'algorithme de Newton qui est privilégié.

Algorithme 2.11 : Algorithme de Levenberg-Marquardt

Données : La fonction F, son gradient, son hessien, les coefficients λ, μ, ε, le nombre max d'itérations $k_{\max}$
Résultat : $p^\star$ tel que $F(p^\star) = \min_p F(p)$
Initialisation : $k = 0$, choix de p_0 initial,
Calcul de $F_0 = F(p_0)$, $G_0 = \nabla F(p_0)$
$d_0 = -G_0$
Choix initial pour $\lambda = 0.001$
tant que $\|d_k\| > \varepsilon$ *et* $k \leq k_{\max}$ **faire**
  Calcul de $\tilde{H}^k_{j,j} = HF_{i,j}(p_k)(1 + \lambda_k \delta_{ij})$
  Calcul de $G_k = \nabla F(p_0)$
  Calcul de d_k solution de $\tilde{H}^k d_k = -G_k$
  si $F(p_k + d_k) \geq F_k$ **alors**
    $\lambda_{k+1} = \mu \lambda_k$
  sinon
    $\lambda_{k+1} = \lambda_k / \mu$
    $p_{k+1} = p_k + d_k$
  $k \leftarrow k + 1$
fin
$p^\star = p_k$

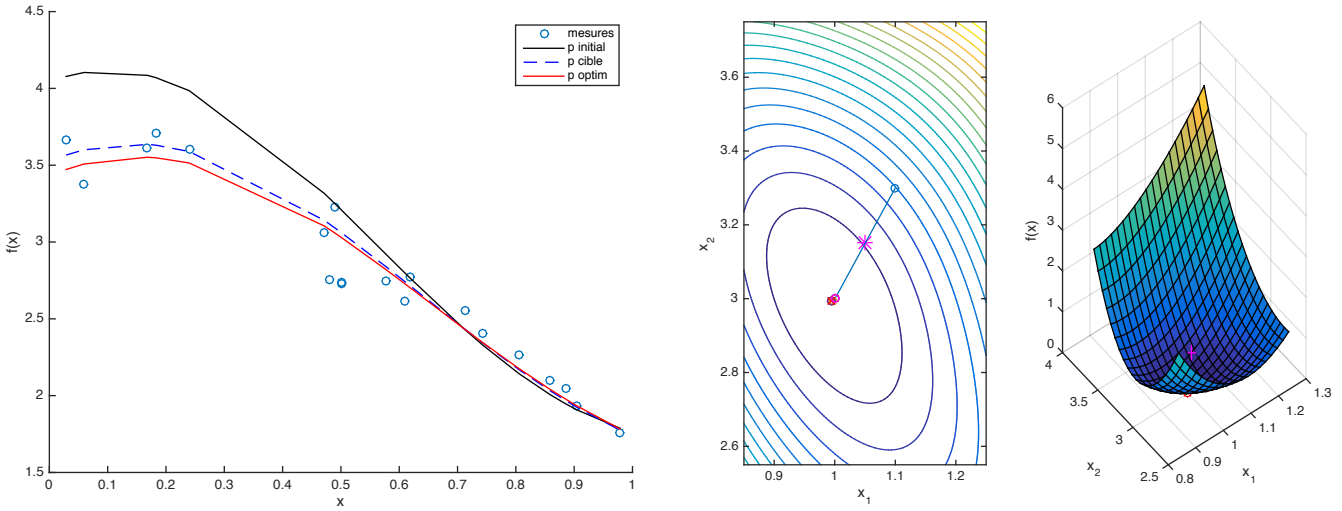


FIGURE 2.10 – Identification de paramètres pour le modèle $y = f(x,p) = p_1 \cos(p_2 x) + p_2 \sin(x + p_1)$ sur 20 mesures $y_i^{exp} = f(x_i^{exp}, p)$, $i = 1, \ldots, 20$, $p = (3,1)$, 5% bruit uniforme. Panel gauche : $y = f(x,p)$ avec p initial $(1.1, 3.3)$, exact $(1., 3.)$ et identifié $(0.995451, 2.99552)$. Panel droit : isovaleurs et valeurs du critère de vraisemblance et itérations successives de l'algorithme de Levenberg-Macquardt.

2.4 Algorithmes de minimisation sans dérivées

Dans beaucoup de cas, le calcul explicite de la différentielle de la fonction à minimiser est difficile, ou coûteux. Il existe toute une classe de méthodes de minimisation qui, même si leur convergence repose sur des hypothèses de régularité de la fonction, ne font pas intervenir explicitement le gradient. Nous donnons un exemple appelé la méthode de *Nelder-Mead* ou *downhill simplex* ou *amoeba*. Elle repose sur le concept du simplexe, qui est un cas particulier de polytopes à $n + 1$ vertex en dimension n (par exemple un segment sur la droite, un triangle dans un plan, etc.). Le principe de cet algorithme consiste à classer les sommets du polytope par ordre croissant de la valeur de la fonction, puis à le faire évoluer en explorant la demi-droite formée par le plus mauvais point (celui pour lequel la fonction est maximale) et le barycentre des autres points. La méthode de Nelder-Mead converge si la fonction est unimodale et régulière, sinon elle peut avoir un comportement oscillant. D'autres méthodes ne faisant pas appel aux dérivées ont

été mises au point depuis son invention, plus robustes et générales. Nous en donnons un aperçu au chapitre **??**.

Algorithme 2.12 : Algorithme de Nelder-Mead

Données : La fonction $f : \mathbb{R}^n \to \mathbb{R}$.
Résultat : $x^\star$ tel que $f(x^\star) = \min_x f(x)$
Initialisation : $k = 0$, calcul des valeurs de f en $n+1$ points non alignés $x_1^0, \dots, x_{n+1}^0$, ordonnés tels que $f(x_i^0) \leq f(x_{i+1}^0)$
tant que $k < K_{\max}$ *et* (x_i^k) *non alignés et* $f(x_i^k)$ *non identiques* **faire**
  calcul du barycentre : $x_c = \sum_{i=1}^n x_i^k$
  et du symétrique de x_{n+1}^0, $x_r = x_c + d$ avec $d = x_c - x_{n+1}^0$.
  si $f(x_r) < f(x_1^k)$ **alors**
    Expansion : $x_e = x_c + 2d$ et calcul de $f(x_e)$
    si $f(x_e) < f(x_r)$ **alors**
      $x_{n+1}^{k+1} = x_e$
    sinon
      $x_{n+1}^{k+1} = x_r$
  sinon si $f(x_1^k) < f(x_r) < f(x_n^k)$ **alors**
    $x_{n+1}^{k+1} = x_r$
  sinon si $f(x_n^k) < f(x_r) < f(x_{n+1}^k)$ **alors**
    Contraction externe : $x_{ce} = x_c + d/2$ et calcul de $f(x_{ce})$
    si $f(x_{ce}) < f(x_r)$ **alors**
      $x_{n+1}^{k+1} = x_{ce}$
    sinon
      $x_{n+1}^{k+1} = x_r$
  sinon si $f(x_{n+1}^k) < f(x_r)$ **alors**
    Contraction interne : $x_{ci} = x_c - d/2$ et calcul de $f(x_{ci})$
    si $f(x_{ci}) < f(x_{n+1}^k)$ **alors**
      $x_{n+1}^{k+1} = x_{ci}$
    sinon
      réduction vers x_1^k : $x_i^{k+1} = \frac{x_1^k + x_i^k}{2}$ pour $i = 2, n+1$
  Reclassement de $(x_i^{k+1})_{i=1,\dots,n+1}$ tels que $f(x_i^{k+1}) \leq f(x_{i+1}^{k+1})$
  $k \leftarrow k+1$
fin
$x^\star \leftarrow x_1^k$

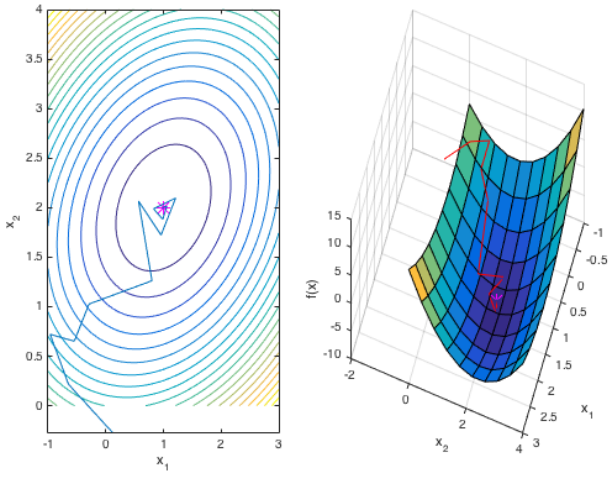


FIGURE 2.11 – Minimisation de $f(x) = \langle Ax, x\rangle/2 + \langle b, x\rangle$ avec l'algorithme de Nelder-Mead. ($A = [3, -1; -1, 5]$, $b = [-1; -9]$)

3 Optimisation avec contraintes

Dans un problème d'optimisation avec contraintes on cherche à minimiser une fonctionnelle sur une partie de son domaine de définition sous des contraintes d'égalité ou d'inégalité

$$\begin{cases} \min & f(x) \\ s.c. & c^E x) = 0 \\ s.c. & c^I(x) \preceq 0 \\ & x \in \mathbb{R}^n \end{cases} \tag{3.1}$$

avec

$$\begin{array}{rcl} f & : & \mathbb{R}^n \longrightarrow \mathbb{R}, \\ c^E & : & \mathbb{R}^n \longrightarrow \mathbb{R}^m, \\ c^I & : & \mathbb{R}^n \longrightarrow \mathbb{R}^p, \\ f & c & \text{régulières.} \end{array}$$

Les contraintes d'égalité ou d'inégalité s'exprimeront parfois également sous la forme d'une contrainte d'appartenance à une partie convexe $K \subset \mathbb{R}^n$.
On définit des notations pour le gradient et la hessienne de la i^{eme} contrainte

$$a_i^E(x) = \nabla c_i^E(x) \quad H_i^E(x) = Hess\, c_i^E(x), \quad a_i^I(x) = \nabla c_i^I(x) \quad H_i^I(x) = Hess\, c_i^I(x). \tag{3.2}$$

Les matrices jacobiennes des contraintes sont notées

$$A^E(x) = \nabla c^E(x) = \begin{pmatrix} a_1^E(x)^T \\ \vdots \\ a_m^E(x)^T \end{pmatrix}, \quad A^I(x) = \nabla c^I(x) = \begin{pmatrix} a_1^I(x)^T \\ \vdots \\ a_p^I(x)^T \end{pmatrix} \tag{3.3}$$

Enfin, soit y un vecteur de $\mathbb{R}^m$, z un vecteur de $\mathbb{R}^p$, qu'on appellera vecteurs des multiplicateurs de Lagrange, on définit le Lagrangien

$$\ell(x, y, z) = f(x) + \langle y, c^E(x)\rangle + \langle z, c^I(x)\rangle \tag{3.4}$$

et le gradient et la hessienne du Lagrangien par rapport à x sont

$$g(x, y, z) = \nabla_x \ell(x, y, z) = \nabla f(x) + \sum_{i=1}^{m} y_i a_i^E(x) + \sum_{i=1}^{p} z_i a_i^I(x) \tag{3.5}$$

$$H(x, y, z) = Hess_x \ell(x, y, z) = Hf(x) + \sum_{i=1}^{m} y_i H_i^E(x) + \sum_{i=1}^{p} z_i H_i^I(x) \tag{3.6}$$

Définition 3.1. Soit $x^\star$ un minimiseur de f.

— On dit que la i^{eme} contrainte d'inégalité est **active** si $c_i^I(x^\star) = 0$.

— On dit que les contraintes sont qualifiées en $x^\star$ si le rang de la matrice formée par la réunion de la matrice jacobienne des contraintes d'égalité et la matrice jacobienne des q contraintes d'inégalité actives en $x^\star$ est égal à $m + q$, alors appelé rang maximal.

Par définition les contraintes d'égalité sont toujours actives, alors que dans le cas des contraintes d'inégalité, la difficulté du problème consiste justement à déterminer si elle le sont ou pas. On va commencer par étudier le problème de minimisation avec des contraintes d'égalité, du point de vue théorique et algorithmique, puis on élargira les résultats au cas général avec des contraintes des deux types.

On introduit la **Fonction duale de Lagrange** $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$\begin{aligned} g(y,z) &= \inf_{x \in D_a} \ell(x,y,z) \\ &= \inf_{x \in D_a} \left(f(x) + \sum_{i=1}^{m} y_i c_i^E(x) + + \sum_{i=1}^{p} z_i c_i^I(x) \right) \end{aligned}$$

g est concave (peut-être infinie pour certains y,z)
Propriété : borne inférieure :
Si $z \geq 0$ alors $g(y,z) \leq p^\star = \inf_{x \in D_a} f(x)$
Cette propriété permet d'obtenir des informations intéressantes sur le minimum d'un problème, sans le résoudre entièrement. Voyons quelques exemples.

1. *Solution d'un système linéaire de norme minimale*
   Résoudre

$$p^\star = \inf_{Ax=b} x^T x$$

   — Lagrangien : $\ell(x,y) = x^T x + y^T(Ax - b)$
   — Pour minimiser $\ell(x,y)$ par rapport à x on annule le gradient

$$\nabla_x \ell(x,y) = 2x + A^T y = 0 \Leftarrow x = -A^T y/2$$

   — On injecte dans la définition de la fonction duale

$$g(y) = \ell(-A^T y/2, y) = -\frac{1}{4} y^T A A^T y - b^T y$$

   concave en y
   — Prop borne inférieure

$$p^\star \geq -\frac{1}{4} y^T A A^T y - b^T y \; \forall y$$

2. *Problème de programmation linéaire standard* Résoudre

$$p^\star = \begin{array}{c} \inf \quad c^T x \\ Ax = b \\ x \geq 0 \end{array}$$

   — Lagrangien : $\ell(x,y,z) = c^T x + y^T(Ax - b) - z^T x = -y^T b + (c + A^T y - z)x$
   — ℓ affine en x donc

$$g(y,z) = \inf_x \ell(x,y,z) = \begin{cases} -b^T y \text{ si } c + A^T y - z = 0 \\ -\infty \text{ sinon} \end{cases}$$

   — g est lineaire sur le domaine affine $\{(y,z), c + A^T y - z = 0\}$ donc concave
   — Prop borne inférieure $p^\star \geq -b^T y$ pour tout y t.q. $c + A^T y \geq 0$
3. *Partitionnement spectral* Résoudre

$$p^\star = \inf_{x_i^2 = 1, i=1,\ldots,n} x^T W x$$

   — Problème non convexe, combinatoire (card $D_a = 2^n$)
   — interprétation : la matrice W mesure la similarité
   — Lagrangien : $\ell(x,y) = x^T W x + \sum_{i=1}^n y_i(x_i^2 - 1) = x^T(W + diag(y))x - 1^T y$

—

$$g(y) = \inf_x \ell(x,y) = \begin{cases} -1^T y \text{ si } W + diag(y) \in \S_+^n \\ -\infty \text{ sinon} \end{cases}$$

— Prop borne inférieure $p^\star \geq -1^T y$ pour tout y t.q. $W + diag(y) \in S_+^n$.
Par exemple $y = -\lambda_{min}(W)1$ permet d'obtenir $p^\star \geq n\lambda_{min}(W)$

Sur ces exemples, la résolution du problème dual serait plus aisée que celle du problème de départ. Résoudre

$$d^\star = \sup_{y \in \mathbb{R}^m, z \in \mathbb{R}^p, z \geq 0} g(y,z)$$

Cependant, avant de s'y attaquer, il faut s'assurer qu'il est équivalent, ce qui n'est pas toujours le cas. On sait, avec le résultat de la "borne inférieure" que $p^\star \geq d^\star$, et par ailleurs, comme le problème dual est concave, on est assuré de l'existence d'une valeur optimale $d^\star$.
On dira qu'on est dans un cas de **Dualité faible** si $d^\star \leq p^\star$ et dans un cas de **Dualité forte** si $d^\star = p^\star$

Un critère de dualité forte est justement que les contraintes soient qualifiées. Notons que dans le cas d'un problème convexe

$$p^\star = \begin{array}{c} \inf \quad f(x) \\ Ax = b \\ c^I(x) \leq 0 \end{array}$$

s'il $\exists x$ t.q. $c^I(x) < 0$ et $Ax = b$ alors $d^\star = p^\star$.

3.1 Conditions d'optimalité pour l'optimisation avec contraintes d'égalité

Pour alléger les notations dans ce paragraphe on abandonne l'indice ^{E} indiquant qu'il s'agit de contraintes d'égalité. Le Lagrangien est une fonction de deux variables

$$\ell(x,y) = f(x) + \langle y, c(x) \rangle,$$

dont le gradient et la hessienne ont les expressions suivantes

$$g(x,y) = \nabla_x \ell(x,y) = \nabla f(x) + \sum_{i=1}^m y_i a_i(x), \tag{3.7}$$

$$H(x,y) = Hess_x \ell(x,y) = Hf(x) + \sum_{i=1}^m y_i H_i(x). \tag{3.8}$$

Avant d'énoncer les conditions d'optimalité on regarde le cas particulier de contraintes linéaires c'est à dire où

$$c(x) = (\langle c_1, x \rangle, \langle c_2, x \rangle, \ldots, \langle c_m, x \rangle)^T$$

avec les c_i une famille libre de vecteurs de $\mathbb{R}^n$. Le problème de minimisation $\min_{c(x)=0} f(x)$ se ramène alors à minimiser $f(x)$ sur l'espace vectoriel $K = \{x, \langle c_i, x \rangle = 0, i = 1, \ldots, m\}$ qui s'exprime sous la forme sans contrainte $\min_{\alpha \in \mathbb{R}^p} g(\alpha)$ avec $g(\alpha) = f(\sum_{j=1}^p \alpha_j k_j)$, où $(k_j)_{j=1}^p$ est une base de K. La condition d'optimalité dans le cas sans contrainte implique que si $x^\star$ est un minimiseur de f sur K, $\nabla_\alpha g(\alpha^\star) = 0$ où $x^\star = \sum_{j=1}^p \alpha_j^\star k_j$ et en utilisant la règle de dérivation directionnelle on en déduit $\langle \nabla_x f(x^\star), k_j \rangle = 0$ pour $j = 1, \ldots, p$. Autrement dit, $\nabla f(x^\star) \in K^\perp$ et une base de $K^\perp$ nous étant gracieusement offerte par la définition de K, on en déduit l'existence de $(y_i)_{i=1}^m$ telle que $\nabla f(x^\star) + \sum_{i=1}^m y_i c_i = 0$.
On peut maintenant généraliser ce résultat au cas de contraintes quelconques avec les conditions d'optimalité du 1er ordre exprimées par le théorème suivant

Théorème 3.1. *des extrema liés*
Soient f et c dans C^1, et $x^\star$ un minimiseur local de f vérifiant les contraintes d'égalité $c(x) = 0$. Si les contraintes sont qualifiées, il existe un vecteur de multiplicateurs de Lagrange $y^\star \in \mathbb{R}^m$ tel que

$$\begin{aligned} c(x^\star) &= 0 \quad \textit{faisabilité primale} \\ g(x^\star) + A^T(x^\star)y^\star &= 0 \quad \textit{faisabilité duale} \end{aligned}$$

Preuve On commence par donner une preuve constructive, dans le cas particulier où $n = 2$ et $m = 1$. L'idée consiste à se ramener à la minimisation sans contrainte d'une fonction d'une seule variable. La condition de qualification des contraintes s'écrit dans le cas où $m = 1$ $\nabla_x c_1(x^\star) \neq 0$, et on peut donc supposer (quitte à inverser x_1 et x_2) que $\partial_{x_2} c_1(x^\star) \neq 0$. Dans ce cas le théorème des fonctions implicites assure qu'il existe un voisinage $V_1 \times V_2$ de $x^\star$ et une fonction φ unique et différentiable en $x^\star$ tels que pour tout $x_1 \in V_1$ $c_1([x_1, \varphi(x_1)]) = 0$ et $x_2^\star = \varphi(x_1^\star)$ avec

$$\varphi'(x_1^\star) = \frac{-1}{\partial_{x_2} c_1(x^\star)} \partial_{x_1} c_1(x^\star).$$

On a donc, pour la fonction $\tilde{f}(x_1) = f([x_1, \varphi(x_1)])$, existence d'un minimum en $x_1^\star$. Les conditions d'optimalité du premier ordre pour $\tilde{f}$ conduisent à

$$\tilde{f}'(x_1^\star) = 0 \Leftrightarrow \frac{\partial f}{\partial x_1}([x_1^\star, \varphi(x_1^\star)]) + \varphi'(x_1^\star) \frac{\partial f}{\partial x_2}([x_1^\star, \varphi(x_1^\star)]) = 0.$$

On obtient l'expression recherchée en posant $y = -\frac{\partial_{x_2} f(x^\star)}{\partial_{x_2} c_1(x^\star)}$. ■

Dans le cas général, on utilise pour démontrer le théorème le résultat intermédiaire suivant

Lemme 3.1. *Soit g dans $C^1(\mathbb{R}^n, \mathbb{R}^m)$, le gradient de g est orthogonal à $\mathcal{C} = \{g(x) = cte\}$.*

Preuve Pour démontrer ce lemme on considère une paramétrisation $(x_i = A_i(\vec{y}))_{i=1,\dots,n}$ de $\mathcal{C} = \{g(x) = cte\}$, avec $\vec{y} = (y_1, \dots, y_k)$, $k = n - m$, dont l'existence est assurée par l'hypothèse de qualification des contraintes, qui permet d'appliquer le théorème des fonctions implicites. On dérive

$$g(A(\vec{y})) = c$$

on a donc pour tout $j = 1, \dots, k$

$$\begin{aligned} \frac{\partial g(A(\vec{y}))}{\partial y_j} &= \sum_{i=1}^{n} \frac{\partial g(A(\vec{y}))}{\partial x_i} \frac{\partial A_i(\vec{y})}{\partial y_j} \\ &= \langle \nabla g(x), \frac{\partial A}{\partial y_j} \rangle = 0 \end{aligned}$$

le vecteur gradient est bien orthogonal à tout vecteur tangent à C. ■

Preuve Revenons à la démonstration du Théorème **??**. Prenons maintenant comme fonction g la contrainte C et sa courbe d'isovaleur nulle correspondant à la contrainte $C_\ell(x) = 0$. On prend une paramétrisation $(x_i = A_i(\vec{y}))_{i=1,\dots,n}$ de cette courbe dans un voisinage de $x^\star$, donc pour $\vec{y} \in B(0, r)$ avec $A(B(0, r)) \in \mathcal{V}(x^\star)$. Si $x^\star = A(\vec{y}^\star)$ alors $\vec{y}^\star$ est solution de

$$y^\star = \underset{\vec{y} \in B(0,r)}{\operatorname{argmin}} \varphi(\vec{y}), \quad \text{avec } \varphi(\vec{y}) = f(A(\vec{y}))$$

Mis sous cette forme il s'agit d'un problème de minimisation sans contrainte par rapport à $\vec{y}$ pour lequel on peut appliquer le critère d'optimalité du 1er ordre puisque par composition d'applications C^1 la fonction $\varphi(\vec{y})$ est C^1. On a donc

$$\begin{aligned} \nabla \varphi(\vec{y}^\star) &= 0 \\ \frac{\partial \varphi(\vec{y}^\star)}{\partial y_j} &= \sum_{i=1}^{n} \frac{\partial f(^\star)}{\partial x_i} \frac{\partial A_i(\vec{y})}{\partial y_j}, \quad \forall j = 1, \dots, k \\ &= \langle (\nabla f(x^\star), \frac{\partial A(\vec{y})}{\partial y_j} \rangle = 0, \quad \forall j = 1, \dots, k. \end{aligned}$$

Les vecteurs $(\frac{\partial A(\vec{y})}{\partial y_j})_{j=1,\dots,k}$ engendrent le sous-espace tangent à $\{C_\ell(x) = 0\}$, donc cette égalité traduit que le gradient de f en $x^\star$ est orthogonal au sous-espace tangent à $\{C_\ell(x) = 0\}$. En utilisant le Lemme précédent on en déduit qu'il est colinéaire au gradient $\nabla_x C_\ell(x^\star)$, d'où l'existence de la combinaison linéaire dont les multiplicateurs de Lagrange sont les coefficients. ■

Avant d'aller plus loin on donne quelques exemples d'application de ce théorème.

Exemple 3.1. Considérons le problème

$$\min_{x_1^2+x_2^2=1} x_1^4 + x_2^4. \tag{3.9}$$

qu'on va résoudre de deux manières : par changement de variables en coordonnées polaires, et en utilisant le théorème précédent.

1. Posons $x_1 = \cos(\theta)$ $x_2 = \sin(\theta)$, le problème (**??**) devient $\min_{\theta\in[0,2\pi]}(\cos\theta^4+\sin\theta^4)$ dont la solution s'obtient en annulant la dérivée :

$$4\cos\theta\sin\theta(-\cos\theta^2+\sin\theta^2) = -2\sin(2\theta)\cos(2\theta) = 0,$$

dont les solutions sur $[0, 2\pi]$ sont $\pi/4 + k\pi/2$ et $k\pi/2$ pour $k = 0, \dots, 3$. En $k\pi/2$ la fonction $x_1^4 + x_2^4$ vaut 1 et en $\pi/4 + k\pi/2$ elle vaut $1/2$. La fonction a donc 4 minima locaux $(\pm\sqrt{2}/2, \pm\sqrt{2}/2)$, où elle vaut $1/2$, et 4 maxima locaux $\{(1, 0), (0, 1), (-1, 0), (0, -1))\}$, où elle vaut 1.

2. En appliquant le Théorème **??** on est conduit à rechercher $x^\star \in \mathbb{R}^2$ et $y^\star \in \mathbb{R}$ tels que

$$\begin{aligned}(x_1^\star)^2 + (x_2^\star)^2 &= 1\\ 4(x_1^\star)^3 + y^\star 2x_1^\star &= 0\\ 4(x_2^\star)^3 + y^\star 2x_2^\star &= 0\end{aligned}$$

ce qui conduit aux possibilités présentées dans le tableau suivant

	$x_1^\star = 0$	$y^\star = -2(x_1^\star)^2$
$x_2^\star = 0$	$(x_1^\star)^2 + (x_2^\star)^2 \neq 1$	$(x_1^\star)^2 = 1$ et $y^\star = -2$, $f(x^\star) = 1$
$y^\star = -2(x_2^\star)^2$	$(x_2^\star)^2 = 1$ et $y^\star = -2$ $f(x^\star) = 1$	$(x_1^\star)^2 = (x_1^\star)^2 = 1/2$ et $y^\star = -1$, $f(x^\star) = 1/2$

Exemple 3.2. Considérons maintenant le problème

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

où A est une matrice symétrique dans $\mathbb{R}^{n\times n}$. Posons $f(x) = \langle Ax, x\rangle$ et $c(x) = \|x\|^2 - 1$ on peut résoudre le problème en appliquant le Théorème **??**, l'existence d'un minimum étant assurée puisque f est continue et $\{c(x) = 0\}$ fermé borné. On a donc l'existence de $y^\star \in \mathbb{R}$ tel que $2Ax^\star + 2y^\star x^\star = 0$ donc l'existence d'au moins un couple valeur propre - vecteur propre, qui plus est minimisant $f(x)$. Pour retrouver le résultat bien connu de l'existence d'une base orthonormée de vecteurs propres de A avec leurs n valeurs propres associées on peut procéder par récurrence sur la dimension n : Pour $n = 1$ le résultat est trivial. Supposons l'hypothèse vraie au rang n, et considérons pour $A \in \mathbb{R}^{n+1\times n+1}$ le sous-espace $H = \{\text{vect}(x^\star)\}^\perp$. Cet espace est de dimension n et stable par A en effet si $\langle x^\star, x\rangle = 0$ alors $\langle x^\star, Ax\rangle = \langle Ax^\star, x\rangle = \langle -y^\star x^\star, x\rangle = 0$. On a donc l'existence d'une base de vecteurs propres orthonormée sur H et il suffit de diviser $x^\star$ par $\|x^\star\|$ pour compléter cette base de $\mathbb{R}^n$.

Les conditions d'optimalité du 2e ordre sont exprimées par le théorème suivant

Théorème 3.2. *Soient f et c dans C^2, et $x^\star$ un minimiseur local de f vérifiant les contraintes d'égalité $c(x) = 0$. Si les contraintes sont qualifiées, il existe un vecteur de multiplicateurs de Lagrange $y^\star \in \mathbb{R}^m$ tel que*

$$\langle s, H(x^\star, y^\star)s\rangle \geq 0 \quad \textit{pour tout } s \in \mathcal{N}$$

où

$$\mathcal{N} = \{s \in \mathbb{R}^n, A(x^\star)s = 0\}.$$

Remarques

— Le premier théorème, encore connu sous le nom de théorème de Lagrange, exprime qu'une solution (x^*, y^*) du problème minimise la fonction lagrangien par rapport à x. Attention, en général ce n'est pas vrai pour les variations par rapport à y.

— Les multiplicateurs de Lagrange (y) mesurent la sensibilité de minimum $x*$ par rapport à une contrainte donnée. Plus précisément, si on note $x^\star(\varepsilon)$ la solution du problème sous contraintes (**??**) où on a perturbé la i^{eme} contrainte $c_i(x) + \varepsilon = 0$ alors le i^{eme} multiplicateur de Lagrange y_i vérifie

$$y_i = \frac{d}{d\varepsilon} f(x^\star(\varepsilon))_{|\varepsilon=0}$$

Preuve Pour montrer ce résultat on calcule par composition des dérivées

$$\frac{d}{d\varepsilon} f(x^\star(\varepsilon)) = \nabla_x f(x^\star) \frac{dx^\star(\varepsilon)}{d\varepsilon} \tag{3.10}$$

et par ailleurs comme $C_j(x^\star(\varepsilon)) = -\varepsilon\delta_{i,j}$ en dérivant on a

$$\nabla_x C_j(x^\star(\varepsilon)) \frac{dx^\star(\varepsilon))}{d\varepsilon} = -\delta_{i,j} \tag{3.11}$$

On utilise l'identité remarquable

$$u \wedge (v \wedge w) = \langle u, w\rangle v - \langle u, v\rangle w$$

avec

$$u = \frac{dx}{d\varepsilon}, \quad v = \nabla f \quad \text{et } w = \sum_{j=1}^{m} y_j \nabla C_j,$$

on a donc en $x^\star$, $v \wedge w = 0$, d'où

$$\langle \frac{dx}{d\varepsilon}, \sum_{j=1}^{m} y_j \nabla C_j\rangle \nabla f - \langle \frac{dx}{d\varepsilon}, \nabla f\rangle \sum_{j=1}^{m} y_j \nabla C_j \quad = \quad 0, \quad j = 1, \ldots, m$$

d'où, en utilisant (**??**) et (**??**), en passant à la limite en ε

$$y_i \nabla f(x^\star) + \frac{d}{d\varepsilon} f(x^\star(\varepsilon))_{|\varepsilon=0} \sum_{j=1}^{m} y_j \nabla C_j = 0.$$

En identifiant avec le Théorème **??** on obtient l'identité recherchée, si $y_i \neq 0$. ■

Exemple 3.3. Application : minimisation d'une fonction quadratique sous contraintes linéaires d'égalité On se place dans le cas où

$$\begin{aligned} f(x) &= \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle \\ c(x) &= Bx - C \end{aligned}$$

avec A une matrice symétrique définie positive de $\mathbb{R}^{n\times n}$, b un vecteur de $\mathbb{R}^n$, B une matrice de $\mathbb{R}^{m\times n}$ et C un vecteur de de $\mathbb{R}^m$. Pour que les contraintes soient qualifiées il faut que la matrice B soit de rang m.
Le Lagrangien est

$$\ell(x, y) \quad = \quad \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \langle y, Bx - C\rangle$$

On écrit que son gradient par rapport à x est nul et que les contraintes sont vérifiées soit

$$\begin{aligned} Ax + b + B^t y &= 0 \\ Bx &= C \end{aligned}$$

Comme A est symétrique définie positive on peut inverser le premier système $x = -A^{-1}(b + B^t y)$. On reporte dans le deuxième système. Comme B est de rang m la matrice $BA^{-1}B^t$ est inversible on peut donc donc inverser le système

$$BA^{-1}B^t y = -(BA^{-1}b + C)$$

ce qui permet d'obtenir y puis x.

3.2 Conditions d'optimalité pour l'optimisation avec contraintes d'inégalité

Dans le cas de contraintes d'inégalité, seules celles qui sont actives interviennent effectivement dans la définition du minimiseur. Cette observation est traduite par les conditions d'optimalité suivantes dites de Kuhn-Tucker

Théorème 3.3. *$x^\star$ est un minimiseur local de f vérifiant les contraintes d'inégalité $c_I(x) \leq 0$ et les contraintes d'égalité $c_E(x^\star) = 0$. Si les contraintes sont qualifiées, il existe un vecteur $y^\star \in \mathbb{R}^m$ et un vecteur $z^\star \in \mathbb{R}^{+p}$ de multiplicateurs de Lagrange tels que*

$$\begin{aligned} & c^E(x^\star) = 0, c^I(x^\star) && \leq && 0 && \textit{faisabilité primale} \\ \forall x \in \mathbb{R}^n \quad & \ell(x^\star, y^\star, z^\star) \leq \ell(x, y^\star, z^\star) \textit{ et } z^\star && \geq && 0 && \textit{faisabilité duale} \\ & c_i^I(x^\star) z_i^\star && = && 0 && \textit{relaxation complémentaire} \end{aligned}$$

Essayons d'interpréter la dernière condition (de relaxation complémentaire) en utilisant les notions de dualité introduites plus haut : puis qu'on est dans le cas de la dualité forte (les contraintes sont qualifiées) alors $p^\star = d^\star$. Or

$$\begin{aligned} p^\star &= \inf_x f(x) + \langle y^\star, C^E(x) \rangle + \langle z^\star, C^I(x) \rangle \\ &\leq f(x^\star) + \langle y^\star, C^E(x^\star) \rangle + \langle z^\star, C^I(x^\star) \rangle \leq f(x^\star) \end{aligned}$$

donc

$$\langle z^\star, C^I(x^\star) \rangle = 0 \Rightarrow z_j^\star C^I(x^\star)_j = 0 \; \forall \, j = 1, \ldots, p.$$

Dans le cas où la fonctionnelle f et les contraintes c^E et c^I sont différentiables le théorème devient

Théorème 3.4. *Conditions de Karush-Kuhn-Tucker (KKT)*
Soient f, c^I et c^E dans C^1, et $x^\star$ un minimiseur local de f vérifiant les contraintes d'inégalité $c^I(x) \leq 0$ et les contraintes d'égalité $c^E(x^\star) = 0$. Si les contraintes sont qualifiées, il existe un vecteur $y^\star \in \mathbb{R}^m$ et un vecteur $z^\star \in \mathbb{R}^{+p}$ de multiplicateurs de Lagrange tel que

$$\begin{aligned} & c^E(x^\star) = 0, c^I(x^\star) && \leq && 0 && \textit{faisabilité primale} \\ g(x^\star) + {A^E}^T(x^\star) y^\star + {A^I}^T(x^\star) z^\star = 0 \textit{ et } z^\star && \geq && 0 && \textit{faisabilité duale} \\ \forall i = 1, \ldots, m \quad & c_i^I(x^\star) z_i^\star && = && 0 \forall i && \textit{relaxation complémentaire} \end{aligned}$$

On remarque par rapport au cas des contraintes d'égalité que les multiplicateurs de Lagrange doivent maintenant être positifs ou nuls. Comme prévu, les contraintes inactives correspondent à des multiplicateurs de Lagrange nuls.
Preuve On peut démontrer ce théorème à partir du th. des extrema liés. On remplace le problème d'optimisation sous contraintes par $\min x \in \mathbb{R}^n, t \in \mathbb{R}^p F(x,t)$ avec
$F(x,t) = f(x)$ avec et les contraintes d'égalités $c_i^E(x) = 0$ pour $i = 1, \ldots, m$ et $c_j^I(x) + t_j^2 = 0$ pour $j = 1, \ldots, p$.

On n'a plus de contraintes d'inégalités, mais on a augmenté d'autant le nombre d'inconnues primales (les t_i pour $i = 1, \ldots, p$. Le lagrangien du problème modifié est

$$L(x,t,y,z) = F(x,t) + \sum_{i=1}^{m} y_i c_i^E(x) + \sum_{j=1}^{p} z_j (c_j^I(x) + t_j^2)$$

Le théorème des extrema liés donne

$$\nabla_{x,t} F(x,t) + \sum_{i=1}^{m} y_i \nabla_{x,t} c_i^E(x) + \sum_{j=1}^{p} z_j \nabla_{x,t} (c_j^I(x) + t_j^2) = 0$$

$$\begin{aligned} \nabla_x f(x) + \sum_{i=1}^{m} y_i \nabla_x c_i^E(x) + \sum_{j=1}^{p} z_j \nabla_x (c_j^I(x) &= 0 \\ 2 z_j t_j &= 0, \quad j = 1, \ldots, p \end{aligned}$$

Pour retrouver la condition $z_j \geq 0$ on applique la condition d'optimalité du 2nd ordre sur le lagrangien de $F(x,t)$:

$$H_{x,t} L(x,t,y,z) = \begin{pmatrix} H_x \ell(x,y,z) & & 0 & & \\ & & \begin{pmatrix} 2z_1 & 0 & \ddots & 0 \\ 0 & 2z_2 & \ddots & 0 \\ \ddots & \ddots & \ddots & \ddots \\ 0 & \ddots & 0 & 2z_p \end{pmatrix} \\ 0 & & \end{pmatrix}$$

■

Nous allons illustrer les conditions KKT sur un exemple simple

Exemple 3.4. On regarde le problème de minimisation quadratique

$$\min_{x_1 + x_2 - 1 \leq 0} x_1^2 + x_2^2.$$

On voit facilement que $(0,0)$ vérifiant la contrainte d'inégalité, on est dans un cas où la contrainte est inactive en $x^\star$, et la solution du problème est la solution du problème non contraint c'est à dire $(0,0)$. Imaginons qu'on n'ait pas remarqué le caractère trivial de l'exercice, et sortons l'artillerie lourde, c'est à dire qu'on recherche $(x^\star, z^\star)$ avec $z^\star \geq 0$ tels que

$$\begin{aligned} \begin{pmatrix} 2x_1^\star \\ 2x_2^\star \end{pmatrix} + z^\star \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= 0 \\ z^\star (x_1^\star + x_2^\star - 1) &= 0. \end{aligned}$$

Des deux premières égalités on tire $x_1^\star = x_2^\star = -z^\star/2$, en remplaçant dans la troisième on obtient
soit $z^\star = 0$ qui conduit $x_1^\star = x_2^\star = 0$
soit $x_1^\star = x_2^\star = 1/2$ qui conduit à $z^\star = -1 < 0$ donc inacceptable.
Changeons maintenant la contrainte en $x_1 + x_2 + 1 \leq 0$. Cette fois-ci $(0,0)$ ne satisfait pas la contrainte, donc celle-ci sera active en $x^\star$. La troisième condition KKT est maintenant $z^\star(x_1^\star + x_2^\star + 1) = 0$, ce qui donne comme alternative :
soit $z^\star = 0$ qui conduit à $x_1^\star = x_2^\star = 0$ qui ne satisfait pas la contrainte,
soit $x_1^\star = x_2^\star = -1/2$ qui conduit à $z^\star = 1$, donc la bonne solution.
Pour s'en convaincre, on peut faire le changement de variable $x_2 = -1 - x_1$ dans la fonction à minimiser : $\min_{x_1} x_1^2 + (1 + x_1)^2$ est bien atteint en $x_1 = -1/2$.

On va caractériser les points vérifiant les conditions de Kuhn-Tucker à l'aide de la notion de *point selle*.

Définition 3.2. On appelle point selle ou point col de $l(x,y)$ (respectivement $l(x,y,z)$) un point $(x^\star,y^\star) \in \mathbb{R}^n \times \mathbb{R}^m$ (resp. $(x^\star,y^\star,z^\star) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{+p}$) tel que

$$\forall x \in \mathbb{R}^n,\ \ \forall y \in \mathbb{R}^m,\ \ l(x^\star,y) \leq l(x^\star,y^\star) \leq l(x,y^\star) \tag{3.12}$$

$$\text{(resp.)}\quad \forall x \in \mathbb{R}^n,\ \ \forall y,z \in \mathbb{R}^m \times \mathbb{R}^{+p},\ \ l(x^\star,y,z) \leq l(x^\star,y^\star,z^\star) \leq l(x,y^\star,z^\star). \tag{3.13}$$

On a une première condition suffisante

Théorème 3.5. *Soient $f(x)$, $c_E(x)$, $c_I(x)$ de classe C^1 alors si le triplet $(x^\star,y^\star,z^\star)$ est un point selle du Lagrangien alors il vérifie les conditions de Karush-Kuhn-Tucker du Théorème* **??**

et une condition nécessaire dans le cas convexe

Théorème 3.6. *Soient $f(x)$, $c_E(x)$, $c_I(x)$ convexes et de classe C^1 alors le triplet $(x^\star,y^\star,z^\star)$ est un point selle du Lagrangien si et seulement si il vérifie les conditions de Karush-Kuhn-Tucker du Théorème* **??**

Preuve On démontre ce résultat dans le cas où les contraintes sont linéaires. ■

Théorème 3.7. *Si $f(x)$ est convexe et si les contraintes sont linéaires alors le couple (x^*,y^*) solution de (P_E) est un point selle du Lagrangien, c'est-à-dire que*

$$(P_{PS})\quad \forall x \in \mathbb{R}^n,\ \ \forall y \in \mathbb{R}^m,\ \ \ell(x^*,y) \leq \ell(x^*,y^*) \leq \ell(x,y^*).$$

Preuve On démontre que si $(x^\star,y^\star)$ est solution de (P_{PS}) alors $x^\star$ est solution de (P_E) : Si

$$\ell(x^\star,y) \leq \ell(x^\star,y^\star) \quad \forall y$$

alors

$$\begin{aligned} f(x^\star) + \langle C(x^\star),y\rangle &\leq f(x^\star) + \langle C(x^\star),y^s tar\rangle \quad \forall y \\ \langle C(x^\star),y-y^\star\rangle &\leq 0 \quad \forall y \\ \langle C(x^\star),y-y^\star\rangle &= 0 \quad \forall y \\ C(x^\star) &= 0 \end{aligned}$$

Or d'après (P_{PS})

$$\forall x \quad f(x^\star) + \langle C(x^\star),y^\star\rangle \leq f(x) + \langle C(x),y^\star\rangle$$

Donc

$$\forall x \quad f(x^\star) \leq f(x) + \langle C(x),y^\star\rangle$$

ou encore

$$\forall x \text{ tel que } C(x) = 0 \quad f(x^\star) \leq f(x)$$

Donc $x^\star$ est solution de (P_E). ■

On montre maintenant l'équivalence entre les problèmes primal et dual, propriété qui sera utilisée pour définir l'algorithme d'Uzawa dans le paragraphe suivant. On définit la fonction **duale**

$$\phi(y) = \min_x \ell(x,y)$$

où on appelle **problème primal** le problème consistant à minimiser $\ell(x,y)$ par rapport à x, à y fixé. On a le théorème suivant :

Théorème 3.8. *La fonction $\phi(y)$ est concave, et on a*

$$\nabla_y \phi(y) = Bx(y) - C,$$

où $x(y)$ est la solution du problème primal

$$\ell(x, y) \leq l(z, y) \quad \forall z.$$

Et si la fonction duale $\phi(y)$ atteint son maximum en $y^\star$ alors $x(y^\star)$ est solution du problème (P_E).

Enfin on énonce des conditions d'optimalité faisant intervenir les hessiens de f, c^E et c^I dans le cas où elles sont de classe C^2.

Théorème 3.9. *Soient f, c^Eet c^I dans C^2, et $x^\star$ un minimiseur local de f vérifiant les contraintes d'inégalité $c^I(x) \leq 0$ et les contraintes d'égalité $c^E(x^\star) = 0$. Si les contraintes sont qualifiées, il existe un vecteur $y^\star \in \mathbb{R}^m$ et un vecteur $z^\star \in \mathbb{R}^{+p}$ de multiplicateurs de Lagrange tel que les critères de faisabilité primale et duale ainsi que la relaxation complémentaire soient vérifiés ainsi que*

$$\langle s, H(x^\star, y^\star, z^\star)s\rangle \geq 0 \quad \textit{pour tout } s \in \mathcal{N}_+$$

où

$$\mathcal{N}_+ = \left\{ s \in \mathbb{R}^n,\ \langle s, a_i^E(x^\star)\rangle = 0, i = 1, \ldots, m,\quad \begin{array}{l} \langle s, a_i^I(x^\star)\rangle = 0 \textit{ si } c_i^I(x^\star) = 0 \ \&\ z_i^\star > 0 \textit{ et} \\ \langle s, a_i^I(x^\star)\rangle \geq 0 \textit{ si } c_i^I(x^\star) \leq 0 \ \&\ z_i^\star = 0 \end{array} \right\}.$$

On peut énoncer une condition suffisante d'optimalité

Théorème 3.10. *Soient f et c^I dans C^2, $x^\star \in \mathbb{R}^n$ et des vecteurs de multiplicateurs de Lagrange $y^\star \in \mathbb{R}^m$, $z^\star \in \mathbb{R}^{+m}$ satisfaisant les conditions du théorème* **??** *ainsi que*

$$\langle s, H(x^\star, y^\star, z^\star)s\rangle > 0 \quad \textit{pour tout } s \in \mathcal{N}_+$$

où $\mathcal{N}_+$ est défini dans le théorème **??**. *Alors $x^\star$ est un minimiseur local isolé de $f(x)$ sous les contraintes $c^I(x) \leq 0$.*

Terminons ce paragraphe par quelques exercices

Exercice 3.1. **Problème de Tartaglia.** Décomposer le nombre 8 en somme de deux nombres positifs p_1 et p_2 tels que le produit de leur produit par leur différence soit maximum.

Corrigé : Poser le problème :

$$\begin{array}{c} \min \\ p_1 + p_2 = 8 \\ p_1 \geq 0 \\ p_2 \geq 0 \end{array} \quad f(p_1, p_2) = -p_1 p_2 (p_1 - p_2)$$

Conditions d'application du théorème KKT ? Contraintes actives ? Ecriture du lagrangien

$$\ell(p_1, p_2, y, z_1, z_2) = -p_1 p_2 (p_1 - p_2) + y(p_1 + p_2 - 8) - z_1 p_1 - z_2 p_2$$

Contraintes inactives $\Rightarrow z_1 = z_2 = 0$

$$\ell(p_1, p_2, y, z_1, z_2) = -p_1^2 p_2 + p_1 p_2^2 + y(p_1 + p_2 - 8)$$

Calcul du gradient

$$\nabla_p \ell(p_1, p_2, y, z_1, z_2) = \begin{pmatrix} -2p_2 p_1 + p_2^2 + y \\ 2p_1 p_2 - p_1^2 + y \end{pmatrix} \Rightarrow \left\{ \begin{array}{l} -2p_2 p_1 + p_2^2 + y = 0 \\ 2p_1 p_2 - p_1^2 + y = 0 \end{array} \right.$$

On soustrait et on obtient

$$p_2^2 + p_1^2 - 4p_1p_2 = (p_1 + p_2)^2 - 6p_1p_2 = 0$$

d'où comme $p_1 + p_2 = 8$, $p_1p_2 = 64/6 = 32/3$. Donc p_1, p_2 sont les racines du polynôme

$$x^2 - 8x + \frac{32}{3}$$

d'où

$$p_1 = 4 - \frac{4}{\sqrt{3}} \quad p_2 = 4 + \frac{4}{\sqrt{3}}$$

■

Exercice 3.2. **Exemple Problème de Kepler** Trouver le parallépipède de volume maximal inscrit dans l'ellipsoide

$$\mathcal{E} = \{(x, y, z) \in \mathbb{R}^3, x^2/a^2 + y^2/b^2 + z^2/c^2 = 1\}$$

Corrigé : Poser le problème

$$\begin{array}{c} \min \\ x^2/a^2 + y^2/b^2 + z^2/c^2 = 1 \\ x \geq 0 \\ y \geq 0 \\ z \geq 0 \end{array} \quad -xyz$$

Contraintes actives ? Lagrangien ? $x = 0$ ou $y = 0$ ou $z = 0 \Rightarrow f(x, y, z) = 0$! contraintes d'inégalité inactives

$$\ell(x, y, z) = -xyz + \lambda(x^2/a^2 + y^2/b^2 + z^2/c^2 - 1)$$

Calcul du gradient

$$\begin{aligned} -yz + 2\lambda x/a^2 &= 0 \\ -xz + 2\lambda y/b^2 &= 0 \\ -xy + 2\lambda z/c^2 &= 0 \end{aligned}$$

On multiplie chaque équation par x, y et z

$$\begin{aligned} -xyz + 2\lambda x^2/a^2 &= 0 \\ -xyz + 2\lambda y^2/b^2 &= 0 \\ -xyz + 2\lambda z^2/c^2 &= 0 \end{aligned}$$

On somme $-3xyz + 2\lambda = 0$ d'où $yz = 2\lambda/(3x)$ puis en remplaçant dans la 1ere équation

$$-2\lambda/(3x) + 2\lambda x/a^2 = 2\lambda(x/a^2 - 1/(3x))$$

$\lambda = 0$ est impossible (sinon $xyz = 0$) donc $3x^2 = a^2$. Similairement on obtient

$$x = \frac{a}{\sqrt{3}}, \quad y = \frac{b}{\sqrt{3}}, \quad z = \frac{c}{\sqrt{3}}$$

D'où finalement

$$Vol = 8xyz = 8\frac{abc}{3\sqrt{3}}, \quad \lambda = \frac{3xyz}{2} = \frac{abc}{2\sqrt{3}}$$

■

Exercice 3.3. On considère la fonction

$$f(x,y) = 2x^2 + 3y^2 + 2xy.$$

et le domaine $K = \{x^2 + 4y^2 \leq 1 \text{ et } x + y \geq 1\}$
On cherche $\min_{(x,y)\in K} f(x,y)$

1. Montrer que f a un minimum global sur $\mathbb{R}^2$ et le calculer
2. Montrer que K est convexe non vide.
3. Faire un dessin de K (en traits pleins) et des isovaleurs de $f(x,y)$ (approximativement et en pointillés).
4. Ecrire le lagrangien pour le problème (P).
5. Ecrire les conditions d'optimalité du 1er ordre (ou conditions KKT).
6. Utiliser le dessin et le résultat de la question 1 pour décider laquelle des deux contraintes sera active au minimum.
7. Calculer $(x^\star, y^\star)$ et les multiplicateurs de Lagrange associés.

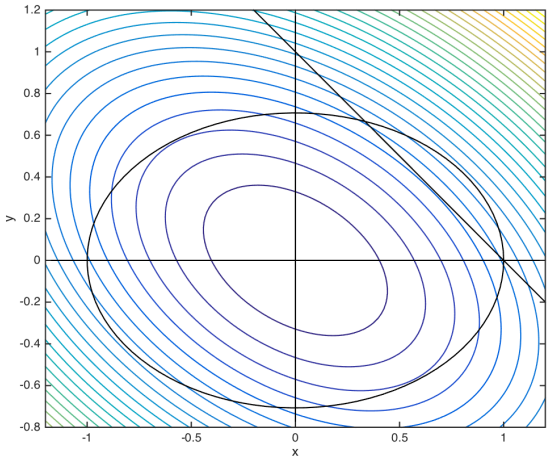
Corrigé :

1. Calculons $\nabla f(x,y) = (4x + 2y, 2x + 6y)^T$ et $Hf(x,y) = \begin{pmatrix} 4 & 2 \\ 2 & 6 \end{pmatrix}$. Les valeurs propres du hessien sont $(5 \pm \sqrt{5})/2$. Donc

$$f(x,y) = f(0,0) + \langle \nabla f(0,0), (x,y)^T \rangle + \frac{1}{2} \langle Hf(0,0)(x,y)^T, (x,y)^T \rangle$$

est une forme quadratique avec une matrice symétrique définie positive, donc le minimum 0 est atteint en $(0,0)$ et unique.

2. On considère $K = \{x^2 + 4y^2 \leq 1 \text{ et } x + y \geq 1\}$. Le point $(1,0) \in K \neq \emptyset$ L'ensemble $\{x^2 + 4y^2 \leq 1\}$ est l'ellipse de demi grand axe $x \in [-1,1]$, $y = 0$ et de demi petit axe $x = 0$, $y \in [-1/2, 1/2]$. Il est donc convexe. Le demi plan $\{x + y \geq 1\}$ est lui aussi convexe. L'intersection non vide de 2 convexes est convexe.



3. On applique la méthode des multiplicateurs de Lagrange :

$$\ell(x,y) = f(x,y) + a(x^2 + 4y^2 - 1) + b(1 - x - y)$$

4. si $(x^\star, y^\star)$ minimise f sur K il existe $a \geq 0$ et $b \geq 0$ tels que

$$\begin{aligned} \nabla_{x,y} \ell(x^\star, y^\star, a, b) &= \begin{pmatrix} 4x^\star + 2y^\star + 2ax^\star - b \\ 2x^\star + 6y^\star + 8ay^\star - b \end{pmatrix} = 0_{\mathbb{R}^2}, \\ a(x^{\star 2} + 4y^{\star 2} - 1) &= 0, \\ b(1 - x^\star - y^\star) &= 0 \end{aligned}$$

Comme le minimum global de $f(x,y)$ n'appartient pas à K, l'une des deux contraintes au moins est forcément active. Vu la convexité de f on prévoit que le minimum sera atteint sur le segment $\{x+y=1\}\cap K$ et le maximum sur l'arc $\{x^2+4y^2=1\}\cap K$.
Pour le minimum on cherche donc $x^\star$, $y^\star$, $b>0$ tels que

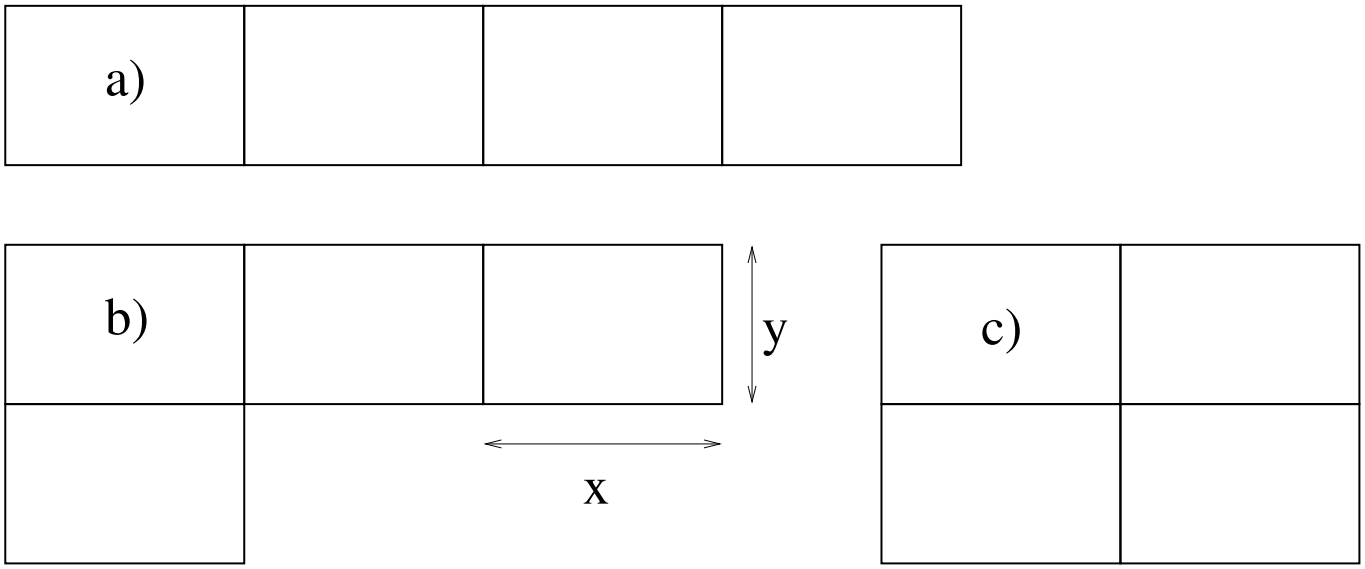
$$\begin{aligned} 4x^\star + 2y^\star &= b \\ 2x^\star + 6y^\star &= b \\ x^{\star 2} + 4y^{\star 2} &< 1 \\ x^\star + y^\star &= 1 \end{aligned}$$

ce qui conduit à $x^\star = 2/3$, $y^\star = 1/3$, b=10/3, $a=0$ et $f(x^\star,y^\star) = 5/3$.

■

Exercice 3.4. Un aviculteur dispose d'une longueur L de clôture grillagée pour délimiter quatre enclos rectangulaires identiques pour ses poules, canards, dindes et poulets. Les enclos peuvent avoir des clôtures mitoyennes. Trouver la disposition et les dimensions permettant de maximiser la surface des enclos.

Corrigé : On représente sur la figure ci-dessous les 3 configurations possibles



Notons S la surface totale des enclos et L la longueur de clôture nécessaire.

- Configuration a) $S=4xy$ et $L=8x+5y$. On élimine $y=(L-8x)/5$ on remplace dans $S=4x(L-8x)/5$. La dérivée $S'(x)=4L/5-64x/5$ s'annule en $x=L/16$ et la dérivée seconde est négative en ce point, donc la surface est maximale pour $(x,y)=(L/16,L/10)$ où elle vaut $L^2/40$.
- Configuration b) $S=4xy$ et $L=7x+6y$. On élimine $y=(L-7x)/6$ on remplace dans $S=4x(L-6x)/3$. La dérivée $S'(x)=4L/3-16x$ s'annule en $x=L/14$ et la dérivée seconde est négative en ce point, donc la surface est maximale pour $(x,y)=(L/14,L/12)$ où elle vaut $L^2/42$.
- Configuration c) $S=4xy$ et $L=6x+6y$. On élimine $y=(L-6x)/6$ on remplace dans $S=4x(L-6x)/6$. La dérivée $S'(x)=4L/6-8x$ s'annule en $x=L/12$ et la dérivée seconde est négative en ce point, donc la surface est maximale pour $(x,y)=(L/12,L/12)$ où elle vaut $L^2/36$.

Donc la troisième configuration est celle qui maximise la surface, avec des enclos carrés de côté $L/12$. ■

Exercice 3.5. On considère la fonction de $\mathbb{R}^3$ dans $\mathbb{R}$

$$f(x) = e^{x_2} + x_1x_2 + x_1^2 - 2x_1x_3 + x_3^2$$

et le problème de minimisation sous contraintes

$$(P_{IE}) \quad x^* = \begin{array}{c} \text{argmin} \\ C_E(x) = 0 \\ C_I(x) \leq 0 \end{array} f(x),$$

avec

$$\left\{ \begin{array}{rcl} C_I(x) & = & x_1^2 + x_2^2 + x_3^2 - 10, \\ C_E(x) & = & \langle a, x \rangle - 2, \end{array} \right.$$

avec $a \in \mathbb{R}^3$.
Déterminer a pour que $x^* = (1, 0, 2)^T$ soit solution de (P_{IE}).

Corrigé : Si $x^* = (1, 0, -1)^T$ est solution de (P_{IE}) il existe $y^* \in \mathbb{R}$ et $z^* \in \mathbb{R}^+$ tels que

$$\begin{array}{rcl} \nabla f(x^*) + y^* \nabla C_E(x^*) + z^* \nabla C_I(x^*) & = & 0 \\ z^* C_I(x^*) & = & 0 \end{array}$$

On calcule

$$\nabla f(x) = \begin{pmatrix} x_2 + 2x_1 - 2x_3 \\ e^{x_2} + x_1 \\ -2x_1 + 2x_3 \end{pmatrix} \text{ et } \nabla f(x^*) = \begin{pmatrix} -2 \\ 2 \\ 2 \end{pmatrix}$$

$$\nabla C_I(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{pmatrix} \text{ et } \nabla C_I(x^*) = \begin{pmatrix} 2 \\ 0 \\ 4 \end{pmatrix}$$

$$\nabla C_E(x) = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \forall x \in \mathbb{R}^3$$

La condition $z^* C_I(x^*) = 0$ implique que $z^* = 0$ car $C_I(x^*) = -5$.
On doit donc avoir $y^* \in \mathbb{R}$ tel que

$$\begin{pmatrix} -2 \\ 2 \\ 2 \end{pmatrix} + y^* \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 0$$

On a clairement $a_i \neq 0$ pour $i = 1, 2, 3$ et $y^* \neq 0$ d'où on peut exprimer les a_i en fonction de y^* et injecter dans la contrainte

$$\begin{array}{rcl} C_E(x^*) & = & 0 \\ a_1 + 2a_3 & = & 2 \\ \dfrac{2}{y^*} - \dfrac{4}{y^*} & = & 2 \\ y^* & = & -1 \end{array}$$

De là on obtient $a = (-2, 2, 2)^t$ ■

Exercice 3.6. **Exemple d'un problème de minimisation avec contraintes d'égalité.**
La discrétisation du problème de la chaînette (résolu en 1691 par Leibniz, Jean Bernoulli et Huygens). On considère un assemblage de $N+1$ barres articulées, fixé à ses deux extrémités dont on cherche la position d'équilibre. Les barres sont supposées identiques, toutes de longueur L, on note (x_{i-1}, y_{i-1}), (x_i, y_i) les positions des deux extrémités de la i^{eme} barre, $i = 1, \ldots, N$. On se donne les valeurs (x_0, y_0) et (x_{N+1}, y_{N+1}), les inconnues du problème sont les positions des extrémités intermédiaires $X = (x_i, y_i)_{i=1,\ldots,N}$. La position d'équilibre est celle qui minimise l'énergie, donc le centre de gravité du système

$$f(X) = \frac{1}{2(N+1)} \sum_{i=1}^{N+1} (y_{i-1} + y_i).$$

Par ailleurs on traduit les contraintes physiques imposant que les barres sont indéformables de longueur L

$$L^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2, \quad i = 1, \ldots, N+1.$$

1. Ecrire le problème sous forme matricielle.
2. Ecrire le Lagrangien ℓ pour ce problème et calculer ses gradients par rapport aux variables primales et duales.
3. Ecrire le Lagrangien pénalisé et calculer ses gradients par rapport aux variables primales et duales.

Corrigé :

1.

$$\begin{aligned} f(X) &= \frac{1}{2(N+1)} \sum_{i=1}^{N+1} (y_{i-1} + y_i) \\ &= \frac{1}{2(N+1)} y_0 + \frac{1}{(N+1)} \left(\frac{1}{2} y_1 + y_2 + \ldots + y_{N-1} + \frac{1}{2} y_N \right) + \frac{1}{2(N+1)} y_{N+1} \end{aligned}$$

y_0 et y_{N+1} sont des constantes donc c'est équivalent de rechercher le minimum de $f(X) =< P, X >$ avec le vecteur P défini par

$$P = \frac{1}{(N+1)} (0, 1/2, 0, 1, \ldots, 0, 1, 0, 1/2)^t.$$

Les contraintes $L^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2$ pour $i = 1, \ldots, N+1$ peuvent s'écrire $\langle B_i X, X \rangle = L^2$ avec B_i nulle partout sauf dans la sous-matrice B_i^s correspondant aux lignes et colonnes $2i-3$ à $2i$

$$B_i^s = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 \end{pmatrix}$$

2. le Lagrangien s'écrit pour $Z = (z_i)_{i=1,\ldots,N+1}$

$$\ell(X, Z) = \langle P, X \rangle + \sum_{i=1}^{N+1} z_i (\langle B_i X, X \rangle - L^2)$$

Les gradient du Lagrangien, par rapport aux variables primales

$$\nabla_X \ell(X, Z) = P + 2 \sum_{i=1}^{N+1} z_i B_i X,$$

et par rapport aux variables duales

$$(\nabla_Z \ell(X, Z))_i = \langle B_i X, X \rangle - L^2.$$

3. Méthode de pénalisation. On choisit un ε petit et on minimise

$$f_\varepsilon(X) = \frac{1}{2(N+1)} \sum_{i=1}^{N+1} (y_{i-1}+y_i) + \frac{1}{\varepsilon} \sum_{i=1}^{N+1} \left((x_i-x_{i-1})^2 + (y_i-y_{i-1})^2 - L^2\right)^2$$

Le gradient de la fonctionnelle pénalisée est $\nabla f_\varepsilon(X) = G_i(X)$ avec $i = 1, \ldots, 2N$. On doit calculer $G_{2i-1} = \partial f(X)/\partial x_i$ et $G_{2i} = \partial f(X)/\partial y_i$ pour $i = 1, \ldots, N$

$$\begin{aligned} G_{2i-1} &= \partial f(X)/\partial x_i = \frac{4}{\varepsilon}\left((x_i-x_{i-1})^2 + (y_i-y_{i-1})^2 - L^2\right)(x_i-x_{i-1}) \\ &\quad -\frac{4}{\varepsilon}\left((x_{i+1}-x_i)^2 + (y_{i+1}-y_i)^2 - L^2\right)(x_{i+1}-x_i) \\ G_{2i} &= \partial f(X)/\partial y_i = \frac{4}{\varepsilon}\left((x_i-x_{i-1})^2 + (y_i-y_{i-1})^2 - L^2\right)(y_i-y_{i-1}) \\ &\quad -\frac{4}{\varepsilon}\left((x_{i+1}-x_i)^2 + (y_{i+1}-y_i)^2 - L^2\right)(y_{i+1}-y_i) + \begin{cases} \frac{1}{2(N+1)} \text{ si } i = 1 \text{ ou } N \\ \frac{1}{(N+1)} \text{ sinon} \end{cases} \end{aligned}$$

■

Exercice 3.7. Une entreprise fabrique deux types de vélos. Le modèle X est vendu à 500 euros l'unité et le modèle Y est vendu à 1000 euros. Les coûts de production mensuels s'élèvent à

$$c(x,y) = 5x^2 + 5y^2 - 2.5xy + 1000$$

où x (respectivement y) désigne le nombre de vélos du modèle X (resp. Y) fabriqués. On suppose que tous les vélos fabriqués sont vendus aussitôt.

1. Donner la fonction de profit mensuel $p(x,y)$
2. On suppose que la capacité de production est de 150 vélos par mois, trouver la répartition entre les deux modèles permettant de maximiser le profit.
3. Le profit est-il maximum quand la capacité de production maximum est atteinte ?

Corrigé :

1. La fonction de profit mensuel $p(x,y)$ s'obtient en retranchant les coûts de production au produit de la vente

$$p(x,y) = 500x + 1000y - 5x^2 - 5y^2 + 2.5xy - 1000$$

2. La capacité de production maximum peut être exprimée comme une contrainte d'inégalité

$$(C) \qquad r(x,y) = x + y - 150 \leq 0.$$

Il s'agit donc de minimiser $-p(x,y)$ tout en vérifiant (C). On écrit le lagrangien

$$L(x,y,\lambda,\mu) = -500x - 1000y + 5x^2 + 5y^2 - 2.5xy + 1000 + \lambda(x + y - 150 + \mu^2)$$

et on cherche à annuler ses dérivées partielles par rapport à x, y, λ et μ.

$$\begin{aligned} \partial_x L(x,y,\lambda,\mu) &= -500 + 10x - 2.5y + \lambda = 0 \\ \partial_y L(x,y,\lambda,\mu) &= -1000 + 10y - 2.5x + \lambda = 0 \\ \partial_\lambda L(x,y,\lambda,\mu) &= x + y - 150 + \mu^2 = 0 \\ \partial_\mu L(x,y,\lambda,\mu) &= 2\lambda\mu = 0 \end{aligned}$$

En combinant les 3 premières équations on obtient $-375 + 2\lambda = 7.5\mu^2$ et la dernière équation implique alors $\mu = 0$. Donc $\lambda = 187.5$ d'où on obtient $x = 55$ et $y = 95$

3. La contrainte est alors réalisée $x + y = 150$ et donc le profit maximum est atteint quand la production est maximum

■

Exercice 3.8. Un récipient cylindrique doit contenir $20\pi m^3$. Le prix du matériau constituant le fond et le couvercle est de 10 euros $/m^2$, celui du matériau constituant les côtés est de 8 euros $/m^2$. Trouver les dimensions (rayon r et hauteur h) du récipient le plus économique, par la méthode des multiplicateurs de Lagrange.

Corrigé : Le prix à minimiser est

$$P(r,h) = 2\pi r^2 \times 10 + 2\pi rh \times 8 = 4\pi(5r^2 + 4rh)$$

La contrainte sur le volume est

$$c(r,h) = \pi r^2 h - 20\pi = 0$$

Le Lagrangien s'écrit

$$L(r,h,y) = p(r,h) + yc(r,h)$$

Les conditions d'optimalité sont $\nabla_{r,h} L = 0$ et $\nabla_y L = 0$ soit, après simplifications

$$\begin{aligned} 4(5r + 2h) + rhy &= 0 \\ 16r + r^2 y &= 0 \\ r^2 h - 20 &= 0 \end{aligned}$$

En résolvant ce système de trois équations à trois inconnues on trouve

$$\begin{aligned} r &= 2, \\ h &= 5, \\ y &= -8. \end{aligned}$$

Le prix du récipient le plus économique est alors 240π euros. ■

3.3 Algorithmes pour l'optimisation avec contraintes

On s'intéresse maintenant au problème général présenté au début de ce paragraphe, faisant intervenir m contraintes d'égalité notées $(c_i^E(x))_{i=1,\dots,m}$ et p contraintes d'inégalité notées $(c_i^I(x))_{i=1,\dots,p}$, avec les multiplicatuers de Lagrange associés, $y \in \mathbb{R}^m$ et $z \in \mathbb{R}^p$. *Plusieurs méthodes* peuvent être envisagées pour résoudre ce problème en pratique

- Changement d'inconnues
- Projection
- Pénalisation
- Méthode des multiplicateurs de Lagrange
- Méthode du Lagrangien augmenté

Exercice 3.9. **Cas pathologique pour** `Matlab` On s'intéresse au problème de minimisation sous contrainte d'inégalité dans $\mathbb{R}^2$ suivant

$$f(x) = \|x\|^2, \quad c(x) = 1 - x_1 - x_2^2$$

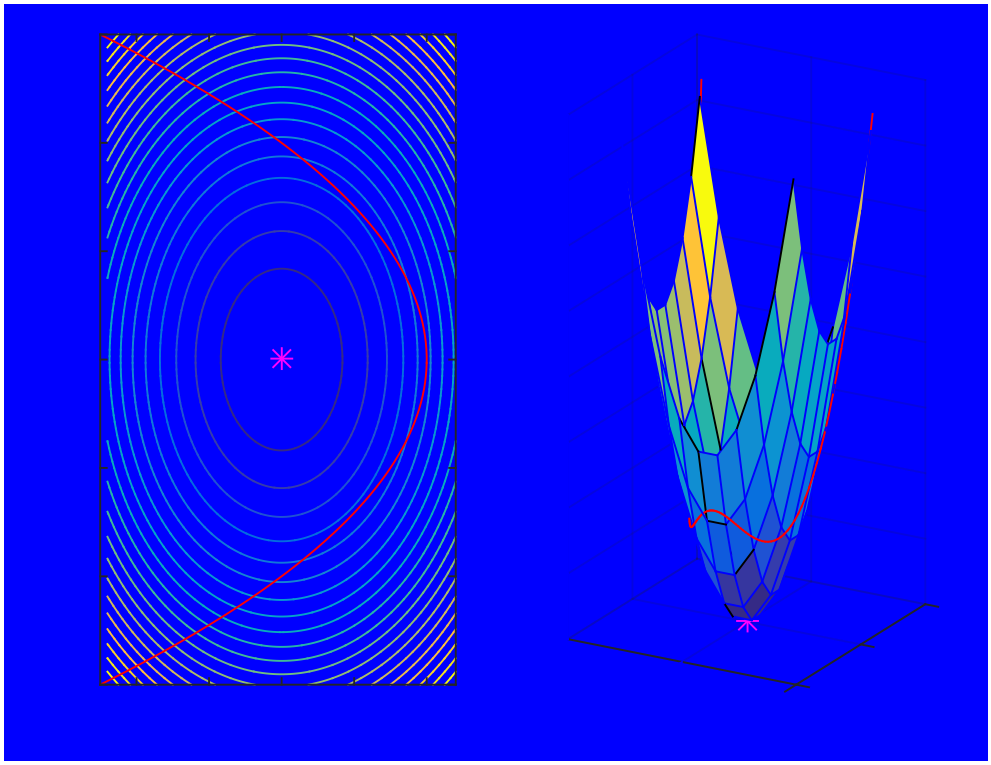
1. Dessiner le cercle de rayon unité centré à l'origine et la courbe délimitant la contrainte. En déduire graphiquement la solution du problème de minimisation

$$(P_I) \qquad \left\{ \begin{array}{ll} \min & f(x) \\ s.c. & c(x) \leq 0 \\ & x \in \mathbb{R}^2 \end{array} \right.$$

2. Calculer par la méthode duale la solution du problème de minimisation.
3. Commenter la solution calculée par `Matlab` , avec la fonction `fmincon` quand on fait varier la solution initiale passée en argument, en la plaçant par exemple sur l'axe $x_2 = 0$.

Corrigé :

1. Le minimum global de $\|x\|^2$ est atteint en 0 qui ne vérifie pas la contrainte, donc la contrainte sera activée. On voit que la parabole est tangente au cercle de centre 0 en 3 points, $(1,0)$, $(0.5, \sqrt{2}/2)$ et $(0.5, -\sqrt{2}/2)$, de norme respectives 1, et 0.75. Donc le minimum global de $\|x\|^2$ sur $\{c(x) \leq 0\}$ est atteint en $(0.5, \sqrt{2}/2)$ et $(0.5, -\sqrt{2}/2)$.



2. Calculons ce résultat par la méthode duale : si $x^\star \in \mathbb{R}^2$ est un minimum il existe $y^\star \geq 0$ tel que

$$\begin{array}{rcl} 2x^\star + y^\star(-1, -2x_2^\star)^T & = & 0, \Leftrightarrow \left\{ \begin{array}{l} 2x_1^\star - y^\star = 0 \\ 2x_2^\star - 2y^\star x_2^\star = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x_1^\star = y^\star/2 \\ x_2^\star(1 - y^\star) = 0 \end{array} \right. \\ y^\star(1 - x_1^\star - x_2^{\star 2}) & = & 0, \\ 1 - x_1^\star - x_2^{\star 2} & \leq & 0. \end{array}$$

   Si $y^\star = 1$, $x_1^\star = 0.5$, $x_2^\star = \pm\sqrt{2}/2$ verifie la contrainte $1 - x_1^\star - x_2^{\star 2} = 0$.
   Si $x_2^\star = 0$, la condition $y^\star(1 - x_1^\star - x_2^{\star 2}) = 0$ devient $y^\star(1 - y^\star/2) = 0$.
   La première solution $y^\star = x_1^\star = 0$ ne vérifie pas la contrainte $1 - x_1^\star - x_2^{\star 2} \leq 0$.
   La deuxième solution $y^\star = 2$, $x_1^\star = 1$ vérifie la contrainte $1 - x_1^\star - x_2^{\star 2} = 0$.
   $\|(1,0)\| = 1$ et $\|(0.5, \pm\sqrt{2}/2)\| = 0.75$, comme il n'y a pas d'autres extrema, la fonction est minimum en $\|(0.5, \pm\sqrt{2}/2)\| = 0.75$ et maximum en $(1,0)$.
3. Exemple d'utilisation de `fmincon`

```
function [y,yeq,g,geq]=contrainte(x)
%% Definition d'une contrainte d'egalite x1+x2^2=1
% yeq=1-x(1)-x(2)^2;
% y=[];
% if nargout>2
%     geq=[-1;-2*x(2)];
%     g=[];
% end
% Definition d'une contrainte d'inegalite x1+x2^2>=1
y=1-x(1)-x(2)^2;
yeq=[];
if nargout>2
    g=[-1;-2*x(2)];
    geq=[];
end
%
% Definition d'une contrainte d'inegalite x1+x2^2<=1
% y=-1+x(1)+x(2)^2;
% yeq=[];
% if nargout>2
%     g=[+1;+2*x(2)];
%     geq=[];
% end
%
%
```

quadcircle.m

```
%-----------------------------------------------%
% Methodes d'optimisation avec Matlab           %
% Master Mathematiques pour l'Entreprise - UPMC %
% Marie.Postel@upmc.fr                          %
%-----------------------------------------------%
% Test des performances de fmincon sur un cas pathologiue
clear
close all
% changer la valeur de la condition initiale et observer le changement de
% la solution fournie par matlab
x0=[2;0.0000];
nx=100;
PlotContour(@quadcircle,nx,nx,[0;0],1.2,1.5)
y=linspace(-1.5,1.5,nx);
x=1.-y.^2;
for i=1:nx
    z(i)=quadcircle([x(i);y(i)]);
end
hold on
subplot(1,2,1)
plot(x,y,'r')
subplot(1,2,2)
plot3(x,y,z,'r')
[x,fval,exitflag,output,lambda,grad,hessian] = fmincon(@quadcircle,x0
    ,[],[],[],[],[],[],@contrainte);
```

```
fprintf('appel 1 a fmincon x=[%f %f] f=%f exitflag=%d\n iterations=%d,
    funCount=%d algorithm=%s\n\n',...
     x,fval,exitflag,output.iterations,output.funcCount,output.algorithm);

options=optimset('GradObj','on');
[x,fval,exitflag,output,lambda,grad,hessian] = fmincon(@quadcircle,x0
    ,[],[],[],[],[],[],@contrainte,options);
fprintf('appel 2 a fmincon x=[%f %f] f=%f exitflag=%d\n iterations=%d,
    funCount=%d algorithm=%s\n\n',...
     x,fval,exitflag,output.iterations,output.funcCount,output.algorithm);

options=optimset('GradObj','on','GradConstr','on');
[x,fval,exitflag,output,lambda,grad,hessian] = fmincon(@quadcircle,x0
    ,[],[],[],[],[],[],@contrainte,options);
fprintf('appel 3 a fmincon x=[%f %f] f=%f exitflag=%d\n iterations=%d,
    funCount=%d algorithm=%s\n\n',...
     x,fval,exitflag,output.iterations,output.funcCount,output.algorithm);
```

■

3.3.1 Changement d'inconnues

On peut parfois faire un changement de variables permettant d'éliminer autant d'inconnues qu'il y a de contraintes, réduisant ainsi la dimension du problème (voir par exemple l'exemple **??**). Dans le cas général, une première méthode pour résoudre le problème (P_I) consiste à faire un changement d'inconnues qui impose automatiquement les contraintes d'inégalité $c^I(x) \leq 0$ ou d'appartenance à un convexe $x \in K$. On évite ainsi d'introduire des multiplicateurs de Lagrange donc d'augmenter la dimension du problème. En revanche, suivant la forme des contraintes, il n'est pas toujours facile ou même possible d'exprimer ce changement de variables.
Quelques *exemples :*

- $K = (\mathbb{R}+)^n \longrightarrow$ poser $x = y^2$ et optimiser sans contraintes par rapport à y.
- $K = \prod_{i=1,\ldots,n}[a_i, b_i] \longrightarrow$ poser $x_i = \frac{a_i+b_i}{2} + \frac{b_i-a_i}{2} cos\theta_i$ et optimiser sans contraintes par rapport aux θ_i

3.3.2 Méthode du gradient projeté

Cette méthode s'utilise dans le cas où les contraintes d'inégalité peuvent s'exprimer sous la forme d'une contrainte d'appartenance à un convexe. Soit K un convexe fermé non vide de $\mathbb{R}^n$ et $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction convexe continue et différentiable. On peut résoudre le problème d'optimisation avec contraintes

$$\min_{x \in K} f(x)$$

à l'aide de l'algorithme suivant

Algorithme 3.1 : Algorithme du gradient projeté

Données : Fonction f, convexe K, pas $(\alpha_k)_{k\geq 0}$, tolérance τ, nombre maximum d'itérations $k_{\max}$
Résultat : $\min_{x\in K} f(x)$
Initialisation : choix de $x_0 \in \mathbb{R}^n$
tant que $\|x^{k+1} - x^k\| \geq \tau$ *ou* $k < k_{\max}$ **faire**
  Résoudre $x^{k+1} = P_K(x^k - \alpha_k \nabla f(x^k))$, $k \leftarrow k+1$
fin
$x^\star = x_k$

où P_K est la projection sur K, dont on rappelle ci-dessous la définition.
Projection sur un convexe.

Définition 3.3. Soit K un convexe fermé non vide de $\mathbb{R}^n$. La projection d'un point $x \in \mathbb{R}^n$ sur K, notée $P_K(x)$, est définie comme l'unique solution de

$$\inf_{y \in K} \|x - y\|_2^2.$$

Propriété : De plus $P_K(x)$ est l'unique élément de K vérifiant

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

Preuve

- Existence de $P_K(x)$ comme minimum sur un fermé d'une fonction $f(y) = \|x - y\|^2$ coercive (Théorème 1.1)
- Unicité : $f(y)$ est convexe (Hessien =$2I_{\mathbb{R}^n}$) donc minimum unique
- Equivalence entre les deux propriétés :
  $\Rightarrow$ pour $y \in K$ et $\theta \in [0,1]$, $\theta y + (1-\theta) P_K(x) \in K$
  - développer $\|x - P_K(x)\|^2 \leq \|\theta y + (1-\theta)P_K(x) - x\|^2$
  - simplifier par θ
  - faire $\theta = 0$

  $\Leftarrow$ pour $y \in K$ développer

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

■

On a la propriété suivante :

Théorème 3.11. *Soit f différentiable sur $\mathbb{R}^n$ et $K \in \mathbb{R}^n$ un convexe fermé non vide. Notons x_k l'itéré courant de l'algorithme du gradient projeté et*

$$d(\alpha) = P_K(x_k - \alpha \nabla f(x_k)) - x_k$$

Si $d(\alpha) \neq 0$ alors $d(\alpha)$ est une direction de descente $\forall \alpha > 0$.

Preuve Soit $\alpha > 0$ fixé. Supposons : $d(\alpha) = p_K(x_k - \alpha \nabla f(x_k)) - x_k \neq 0$.
$d(\alpha)$ direction de descente de f en $x_k \Leftrightarrow \langle nabla f(x_k), d(\alpha) \rangle < 0$
Or $\forall y \in K, \Leftrightarrow \langle P_k(x_k - \alpha \nabla f(x_k)) - (x_k - \alpha \nabla f(x_k)), y - P_K(x_k - \alpha \nabla f(x_k)) \rangle \geq 0$
Donc pour tout $y \in K$ $\langle d(\alpha) + \alpha \nabla f(x_k)), y - x_k - d(\alpha) \rangle \geq 0$
Comme $x_k \in K$ on choisit $y = x_k$ on a
$\langle d(\alpha) + \alpha \nabla f(x_k), d(\alpha) \rangle \leq 0$ ou encore $\alpha \langle \nabla f(x_k), d(\alpha) \rangle \leq -\|d(\alpha)\|^2 < 0$ ■
n a le résultat de convergence suivant

Théorème 3.12. *Si f est différentiable, α-elliptique et de gradient C-Lipschitzien, l'Algorithme du gradient projeté* **??** *converge vers $x^\star$ quand $k \to \infty$ pour $(\alpha_k)_{k \geq 0}$ suffisamment petit : $\alpha_k \leq \dfrac{2\alpha}{C^2}$ où α et C sont deux constantes telles que*

$$\begin{aligned} (\nabla f(x) - \nabla f(y), x - y) &\geq \alpha \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, \\ \|\nabla f(x) - \nabla f(y)\| &\leq C \|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \end{aligned}$$

Preuve Par composition de deux applications contractantes

— $x \to x - \alpha \nabla f(x)$ avec les conditions sur α_k (voir Théorème gradient SC)

— $x \to P_K(x)$. En effet

$$\begin{aligned}
\|x-y\|^2 &= \|x-y-(P_K(x)-P_K(y))+(P_K(x)-P_K(y))\|^2 \\
&= \|x-y-(P_K(x)-P_K(y))\|^2 + \\
&\quad \langle x-y-(P_K(x)-P_K(y)), P_K(x)-P_K(y)\rangle + \|P_K(x)-P_K(y)\|^2 \\
&= \|x-y-(P_K(x)-P_K(y))\|^2 \\
&\quad +\langle \mathbf{x}-\mathbf{P_K(x)}, \mathbf{P_K(x)}-\mathbf{P_K(y)}\rangle \\
&\quad +\langle -\mathbf{y}+\mathbf{P_K(y)}), \mathbf{P_K(x)}-\mathbf{P_K(y)}\rangle + \|P_K(x)-P_K(y)\|^2
\end{aligned}$$

Or $\langle x - P_K(x), P_K(x) - P_K(y)\rangle \geq 0$ car $P_K(y) \in K$ Donc

$$\begin{aligned}
\|x-y\|^2 &\geq \|x-y-(P_K(x)-P_K(y))\|^2 + \|P_K(x)-P_K(y)\|^2 \\
&\geq \|P_K(x)-P_K(y)\|^2
\end{aligned}$$

■

On rappelle la définition/proposition de la projection sur un convexe :

Définition 3.4. Soit K un convexe fermé non vide de $\mathbb{R}^n$. La projection d'un point $x \in \mathbb{R}^n$ sur K, notée $P_K(x)$, est définie comme l'unique solution de

$$\inf_{y \in K} \|x-y\|_2^2.$$

Propriété : De plus $P_K(x)$ est l'unique élément de K vérifiant

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

Preuve

— Existence de $P_K(x)$ comme minimum sur un fermé d'une fonction $f(y) = \|x-y\|^2$ coercive (Théorème 1.1)

— Unicité : $f(y)$ est convexe (Hessien =$2I_{\mathbb{R}^n}$) donc minimum unique

— Equivalence entre les deux propriétés :

$\Rightarrow$ pour $y \in K$ et $\theta \in [0,1]$, $\theta y + (1-\theta)P_K(x) \in K$

— développer $\|x - P_K(x)\|^2 \leq \|\theta y + (1-\theta)P_K(x) - x\|^2$

— simplifier par θ

— faire $\theta = 0$

$\Leftarrow$ pour $y \in K$ développer

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

■

Dans le cas général la solution de ce problème peut-être aussi difficile à trouver que celle du problème d'origine. On n'utilisera donc la méthode du gradient projeté que dans les cas où la projection est une étape facile à résoudre.

Exemple 3.5. Projection sur une intersection de demi espaces.
Supposons que $K = \{x \in \mathbb{R}^n, x_i \geq a_i, i \in I, x_j \leq b_j, j \in J\}$ avec $I, J \subset \{1, \ldots, n\}$. On a alors

$$P_K(x)_i = \begin{cases} \max(a_i, x_i), \text{ pour } i \in I \backslash J \\ \min(b_i, x_i), \text{ pour } i \in J \backslash I \\ \min(b_i, \max(a_i, x_i)), \text{ pour } i \in I \cap J \end{cases}$$

Exemple 3.6. Projection sur une droite dans $\mathbb{R}^2$.
La projection de x sur $K = \{x \in \mathbb{R}^2, x_1 + 2x_2 = 1\}$ revient à minimiser $d(y) = \|x - y\|^2$, avec $y = (y_1, y_2)^T$ et $y_1 + 2y_2 = 1$. On remplace y_1 par $1 - 2y_2$ dans $d(y) = (x_1 - y_1)^2 + (x_2 - y_2)^2$ et oncherche le minimum d'une fonction d'une variable (y_2).
On en déduit l'algorithme du gradient projeté à pas fixe pour le problème

$$\begin{cases} f(x) = (x_1 - 2)^2 + (x_2 - 3)^2 \\ K = \{x_1 + 2x_2 = 1\} \end{cases}$$

Solution exacte : remplacer x_1 par $1 - 2x_2$ et minimiser $\tilde{f}(x_2) = (1 + 2x_2)^2 + (x_2 - 3)^2$ par rapport à x_2. On obtient le minimum pour $x_2 = 1/5$ et $x_1 = 1 - 2x_2 = 3/5$.
Gradient projeté : Le gradient de $f(x^k)$ à l'itération k est $2(x_1^k - 2, x_2^k - 3)^T$. Puis la projection orthogonale de $x^k - \alpha\nabla f(x^k)$ sur $\{x_1 + 2x_2 = 1\}$ est le point x^{k+1} tel que

$$\begin{aligned} x_1^{k+1} + 2x_2^{k+1} &= 1 \\ -2(x_1^k - 2\alpha(x_1^k - 2) - x_1^{k+1}) + (x_2^k - 2\alpha(x_2^k - 3) - x_2^{k+1}) &= 0 \end{aligned}$$

d'où on tire

$$\begin{aligned} x_1^{k+1} &= \frac{1}{5}(1 + 4x_1^k - 2x_2^k + 4\alpha(x_2^k - 2x_1^k + 1)) \\ x_2^{k+1} &= \frac{1}{5}(2 - 2x_1^k + x_2^k - 2\alpha(x_2^k - 2x_1^k + 1)) \end{aligned}$$

Pour vérifier que l'algorithme converge, calculons $x^{k+1} - x^\star$

$$\begin{aligned} x_1^{k+1} - \frac{3}{5} &= \frac{1}{5}(-2 + 4x_1^k - 2x_2^k + 4\alpha(x_2^k - 2x_1^k + 1)) \\ x_2^{k+1} - \frac{1}{5} &= \frac{1}{5}(1 - 2x_1^k + x_2^k - 2\alpha(x_2^k - 2x_1^k + 1)) \end{aligned}$$

qui peut se mettre sous la forme

$$\begin{aligned} x_1^{k+1} - \frac{3}{5} &= \frac{1 - 2\alpha}{5}\left(4(x_1^k - \frac{3}{5}) - 2(x_2^k - \frac{1}{5})\right) \\ x_2^{k+1} - \frac{1}{5} &= \frac{1 - 2\alpha}{5}\left(-2(x_1^k - \frac{3}{5}) + (x_2^k - \frac{1}{5})\right) \end{aligned}$$

soit sous forme matricielle $x^{k+1} - x^\star = M(x^k - x^\star)$ avec

$$M = \frac{1 - 2\alpha}{5}\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}$$

D'où on tire une condition suffisante de convergence $\|M\| < 1$ qui se traduit sur les valeurs de α par $\frac{1}{12} < \alpha < \frac{11}{12}$.

3.3.3 Méthode de pénalisation

Principe : il s'agit de remplacer le problème sous contraintes par un problème sans contraintes où la fonctionnelle à minimiser prend des valeurs très grandes quand la contrainte n'est pas réalisée. On se place au départ dans le cas où les contraintes d'inégalité reviennent à des contraintes d'appartenance à un convexe K. Il s'agit de trouver une fonction de pénalisation $p(x)$ telle que les problèmes

$$(P_K) \qquad \min_{x \in K} f(x),$$

avec $K \subset \mathbb{R}^n$ et

$$(P_{K_\varepsilon}) \qquad \min_{x \in \mathbb{R}^n} \Theta_\varepsilon(x), \quad \text{avec} \quad \Theta_\varepsilon(x) = f(x) + \frac{1}{\varepsilon}p(x)$$

soient équivalents. Dans ce cas on parle de pénalisation exacte :

Définition 3.5. Une fonction de pénalisation $p(x)$ associée au problème (P_K) est exacte si toute solution de (P_K) minimise Θ_ε.

On pourra par exemple choisir comme fonction de pénalisation

$$p(x) = \begin{cases} 0 \text{ if } x \in K \\ +\infty \text{ if } x \notin K \end{cases}$$

qui remplit bien la condition d'exactitude mais n'a pas de propriétés de régularité permettant d'utiliser les algorithmes de minimisation qu'on a vu au paragraphe précédent.
Si on relaxe cette condition en admettant que les solutions du problème pénalisé (P_{K_ε}) soient différentes de celles du problème initial (P_K), on parle de pénalisation inexacte, qui peut faire intervenir des fonctions $p(x)$ régulières. Parmi les fonctions de pénalisation rentrant dans cette catégorie, on distinguera

— les fonctions de pénalisation intérieures, dont les solutions $x_\varepsilon^\star \in K$.
Dans certains problèmes où f n'est pas définie à l'extérieur de K c'est la seule possibilité. L'idée est d'utiliser un terme de pénalisation dit "barrière" $p(x)$ qui tend vers l'infini lorsque x s'approche de la frontière ∂K de K.
On pourra par exemple associer la fonction de pénalisation intérieure suivante, dite logarithmique

$$\Theta_\varepsilon(x) = f(x) - \varepsilon \log(x).$$

construite avec $p(x) = -\log x$. Dans ce cas c'est quand le coefficient multipliant $p(x)$ tend vers 0 que la solution du problème pénalisé tend vers la solution du problème contraint. Dans le cas général avec une contrainte $C^I(x) \leq 0$ à valeurs dans $\mathbb{R}^p$, on utilise la fonction de pénalisation

$$\Theta_\varepsilon(x) = f(x) - \varepsilon \sum_{i=1}^{p} \log(-c_i^I(x))$$

— les fonctions de pénalisation extérieures dont $x_\varepsilon^\star \to x^\star$ avec $x_\varepsilon^\star \notin K$ pour $\varepsilon \neq 0$.
Pour ces dernières, on pourra par exemple choisir
  — si $K = \mathbb{R}^{+^n}$ $p(x) = \|x^-\|^2$, où x^- est le vecteur des parties négatives des composantes du vecteur x
  — si $K = \{x, c(x) = 0\}$, $p(x) = \|c(x)\|^2$
  — si $K = \{x, c(x) \leq 0\}$, $p(x) = \|c(x)^+\|^2$

On illustre la pénalisation intérieure et extérieure dans l'exemple suivant.

Exemple 3.7. Exemple de pénalisation

$$(P_K) \qquad \min_{x \in \mathbb{R}^+} f(x) = x + x^3,$$
$$p_{ext}(x) = (x^-)^2, \text{ avec } x^- = \max(-x, 0),$$
$$p_{int}(x) = -\log x.$$

Pour les fonctions de pénalisation inexactes ; on a le résultat de convergence suivant

Théorème 3.13. *Soit f continue et coercive sur un ensemble $K \in \mathbb{R}^n$ fermé. On suppose que $p(x)$ vérifie les conditions*

1. *$p(x)$ continue sur $\mathbb{R}^n$,*
2. *$\forall x \in \mathbb{R}^n$, $p(x) \geq 0$,*
3. *$p(x) = 0 \Leftrightarrow x \in K$.*

Alors

1. *$\forall \varepsilon > 0$ le problème (P_{K_ε}) a au moins une solution $x_\varepsilon^\star$,*
2. *La famille $(x_\varepsilon^\star)_{\varepsilon > 0}$ est bornée,*

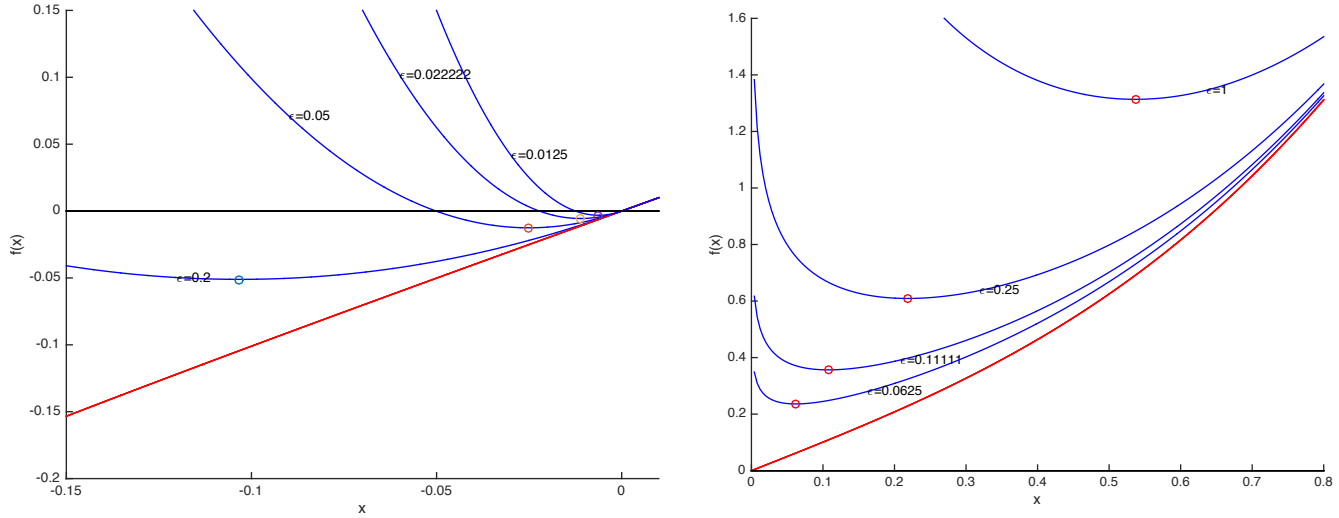


FIGURE 3.1 – Minimisation de $f(x) = x + x^3$ sur $\mathbb{R}^+$, chaque courbe bleue correspond à une valeur de ε différente, dont le minimum est indiqué par un cercle. La courbe rouge est le graphe de $x + x^3$. Panel gauche pénalisation extérieure de $f(x) + p_{ext}(x)/\varepsilon$ avec $p_{ext} = (x^-)^2$, Panel droit : pénalisation intérieure de $f(x) + \varepsilon p_{int}(x)$ avec $p_{int} = -\log x$.

3. *Toute sous-suite convergente extraite de* $(x^\star_\varepsilon)_{\varepsilon>0}$ *converge vers une solution de* (P_K) *quand* $\varepsilon \searrow 0$.

Ce théorème permet d'utiliser l'agorithme suivant

Algorithme 3.2 : Algorithme de pénalisation

Données : Fonction Θ_ε, tolérance τ, nombre maximum d'itérations $k_{\max}$
Résultat : $\min_{x\in K} f(x)$
Initialisation : choix de $x_0 \in \mathbb{R}^n$, $\varepsilon_0 > 0$
tant que $\|x_{k+1} - x_k\| \geq \tau$ *et* $k < k_{\max}$ **faire**
  Résoudre $x_{k+1} = \min_{x\in\mathbb{R}^n} \Theta_\varepsilon(x)$ avec x_k comme point de départ
  Choisir $\varepsilon_{k+1} < \varepsilon_k$
  $k \leftarrow k + 1$
fin
$x^\star = x_k$

3.3.4 Algorithmes basés sur la dualité

Du théorème **??** d'équivalence entre les problèmes primal et dual découle un algorithme pour trouver la solution (x^*, y^*), dit d'Uzawa.

Algorithme 3.3 : Algorithme d'Uzawa

Données : Le Lagrangien $\ell(x, y)$ et les gradients par rapport à x et y, le pas dual $\rho > 0$, la tolérance ε
Résultat : Solution $x^\star$ du problème P_E
Initialisation : Choisir une estimation y^0 et $\rho > 0$
tant que $\|g^k\| > \varepsilon\|g^0\|$ **faire**
  Chercher x^k solution du problème primal

$$\forall x \in \mathbb{R}^n, \quad \ell(x^k, y^k) \leq \ell(x, y^k)$$

  si $F^k = \ell(x^k, y^k) \leq F^{k-1}$ **alors**
    diminuer la valeur de ρ (à partir de la deuxième itération)
  Calculer $g^k = \nabla_y \ell(x^k, y^k)$
  Modifier $y^{k+1} = y^k + \rho g^k$
  $k \leftarrow k + 1$
fin
$x^\star = x_k$

Pour lequel on a le résultat de convergence suivant

Théorème 3.14. *On suppose que f est de classe C^1 et elliptique, c^E affine, c^I convexe et de classe C^1 et que f et c^I sont lipschitziennes. On suppose de plus que le Lagrangien $\ell(x, y, z)$ possède un point selle $(x^\star, y^\star, z^\star)$ dans $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$. Alors il existe ρ_1, ρ_2 avec $0 < \rho_1 < \rho_2$ tels que pour tout $\rho \in [\rho_1, \rho_2]$ la suite $(x^k)_{k\geq 0}$ générée par l'algorithme d'Uzawa converge vers $x^\star$.*

Même dans les conditions d'application du Théorème **??**, l'algorithme d'Uzawa converge très lentement. Il peut être accéléré grâce aux idées de la méthode de pénalisation, qui serviront également à convexifier localement la fonction. Comme on l'a vu dans le paragraphe **??**, la pénalisation exacte consiste à modifier la fonctionnelle à minimiser en lui rajoutant un terme qui devient très grand quand la contrainte n'est pas vérifiée, et qui disparaî,t quand la contrainte est vérifiée.
L'algorithme du Lagrangien augmenté découle de cette idée, qu'on présente ici dans le cas quadratique avec des contraintes linéaires d'égalités.

$$\left\{ \begin{array}{rcl} f_\varepsilon(x) & = & \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \frac{1}{2\varepsilon}\|Bx - C\|^2, \\ \nabla f_\varepsilon(x) & = & Ax + b + \frac{1}{\varepsilon}B^t(Bx - C). \end{array} \right. \tag{3.14}$$

On remplace le Lagrangien initial par sa régularisée de Yosida-Moreau :

$$\ell(x, y) \quad = \quad \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \langle y, Bx - C\rangle + \frac{\rho}{2}\|Bx - C\|^2. \tag{3.15}$$

Remarque ; le paramètre ρ dans (**??**) est l'inverse du paramètre de pénalisation dans (**??**).

Algorithme 3.4 : Lagrangien augmenté dans le cas quadratique avec contraintes linéaires

Données : Le Lagrangien $\ell(x, y)$ et les gradients par rapport à x et y, la tolérance ε
Résultat : Solution $x^\star$ du problème P_E
Initialisation : Choisir une estimation y^0 et $\rho^0 > 0$
tant que $\|g^k\| > \varepsilon\|g^0\|$ **faire**
  Chercher x^k solution du problème primal

$$\forall x \in \mathbb{R}^n, \quad \ell(x^k, y^k) \leq \ell(x, y^k)$$

  Calculer $g^k = Bx^k - C$
  Modifier $y^{k+1} = y^k + \rho^k g^k$
  Mettre à jour le paramètre ρ^k
  si $F^k = \ell(x^k, y^k) \leq F^{k-1}$ **alors**
    $\rho^{k+1} = \alpha\rho^k$ avec $0 < \alpha < 1$
  sinon
    $\rho^{k+1} = \beta\rho^k$, avec $\beta > 1$
  $k \leftarrow k + 1$
fin
$x^\star = x_k$