

الكلية متعددة التخصصات - وزازات
+•YELU•+ +•X+EMKE+ - U•OЖ•Ж•+
FACULTÉ POLYDISCIPLINAIRE DE OUARZAZATE



Université Ibn Zohr

Faculté Polydisciplinaire Ouarzazate

Module Modélisation

Polycopie pédagogique

Filière Sciences Mathématiques et Informatiques (Semestre 6).

Par

Abdelkrim EL MOUATASIM Professeur Habilité

Années Universitaires : 2015–2021

Chapitre 1

Outils d'optimisation

1.1 Calcul différentiel

Les notations

Espaces, points, fonctions :

- V, W , espaces vectoriels normés. Ex. : $\mathbb{R}^n, \mathcal{M}_n(\mathbb{R}), C([0, 1]), C^1([0, 1])...$
- x, y, u des éléments de V ou W . Ex. : $x = (x_1, \dots, x_n)^t, u : u(t) = \sin t$.
- f : application de V dans W . Ex. : $V = W = \mathcal{M}_n(\mathbb{R}), f(M) = M^{-1}$.
- F : applications de V dans \mathbb{R} . Ex. : $V = C^1([0, 1]), F(u) = \int_0^1 \sqrt{1 + (u'(t))^2} dt$.

Définition 1 (Différentielle au sens de Fréchet) Une application $f(x)$ de V dans W est différentiable en $x \in V$ si il existe une application linéaire continue de V dans W , notée $Df(x)$ telle que

$$\forall h \in V \quad f(x + h) = f(x) + Df(x).h + \epsilon(h) \|h\|$$

avec $\lim_{h \rightarrow 0} \|\epsilon(h)\| = 0$.

Exemple 1 Soit $V = \mathcal{M}_n(\mathbb{R})$ l'ensemble des matrices réelles de dimension n et $f(M) = M^2$.

Il suffit de calculer

$$f(M + h) = (M + h)(M + h) = M^2 + (Mh + hM) + h^2$$

et de ne retenir que les termes linéaires en h :

$$Df(M).h = Mh + hM$$

Exemple 2 Soit $u \in V = C([0, 1])$ et

$$F(u) = \int_0^1 \frac{u^4}{4} dx$$

Il vient, si $h \in V$

$$F(u + h) = \int_0^1 \frac{(u + h)^4}{4} dx = \int_0^1 \frac{u^4}{4} + u^3 h + \dots dx$$

où les termes suivants sont d'ordre supérieur ; d'où la différentielle

$$DF(u).h = \int_0^1 u^3 h dx$$

Définition 2 (Différentielle au sens de Gâteaux) Une fonction $F(x)$ de V dans \mathbb{R} est une fonction différentiable au sens de Gâteaux en $x \in V$ si il existe une forme linéaire sur V , notée $DF(x)$, telle que

$$\forall h \in V \quad \frac{d}{d\lambda} F(x + \lambda h)|_{\lambda=0} = DF(x).h \quad (1.1)$$

Theorem 1 (Lien avec Fréchet) Si F est différentiable au sens de Fréchet, F est différentiable au sens de Gâteaux.

Exemple 3 Soit $V = \mathbb{R}^n$

$$f(x) = \frac{1}{2} \langle Ax, x \rangle + \sum_j \frac{x_j^4}{4}$$

$$Df(x).h = \frac{d}{dt} f(x + th)|_{t=0} = \langle Ax, h \rangle + \sum_j x_j^3 h_j$$

$$Df(x).h = \langle Ax, h \rangle + \langle x^3, h \rangle = \langle Ax + x^3, h \rangle$$

Exemple 4 Soit $V = \mathbb{R}^n$ et $f(x) = Ax + x^3$, où A est une matrice. Soit $\Delta(x)$ la matrice diagonale telle que $\Delta_{i,i} = 3x_i^2$.

$$Df(x).h = \frac{d}{dt} f(x + th)|_{t=0} = Ah + \Delta(x)h$$

Définition 3 (Gradient) Soit V un espace de dimension finie muni d'un produit scalaire $\langle \cdot, \cdot \rangle$. Il existe un vecteur, le gradient, noté $\nabla F(x)$, tel que

$$DF(x).h = \langle \nabla F(x), h \rangle$$

:

Si $V = \mathbb{R}^n$ et $\langle x, y \rangle = \sum_i x_i y_i$ on retrouve la définition usuelle du gradient

$$\nabla F(x) = \left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_i}, \dots, \frac{\partial F}{\partial x_n} \right)^t$$

le gradient définit la direction de plus grande pente \perp à la surface équipotentielle.

Exemple 5

$$Df(x).h = \langle Ax + x^3, h \rangle$$

$$\nabla F(x) = Ax + x^3$$

Définition 4 (Matrice jacobienne) Si $V = W = \mathbb{R}^n$ la différentielle $Df(x)$ est une application linéaire représentée donc par une matrice $Jf(x)$, appelée matrice jacobienne, de coefficients

$$Jf(x)_{i,j} = \frac{\partial f_i(x)}{\partial x_j}$$

Exemple 6 Soit $V = \mathbb{R}^n$ et $f(x) = Ax + x^3$, où A est une matrice.

$$Df(x).h = Ah + \Delta(x)h = (A + \Delta(x))h$$

$$Jf = A + \Delta(x)$$

Soit $V = \mathbb{R}^n$ et $F(x)$ une fonction différentiable en tout point. Si $f(x) = F'(x) = \nabla F(x)$ est une application différentiable de \mathbb{R}^n dans \mathbb{R}^n , la matrice jacobienne $\mathbf{J}f(x)$ est symétrique.

Définition 5 (Hessien) Le Hessien de la fonction $F(x)$, supposée deux fois différentiable, est la matrice $\mathbf{H}F(x)$ telle que

$$\frac{d^2}{d\lambda^2} F(x + \lambda h)_{\lambda=0} = \langle \mathbf{H}F(x)h, h \rangle$$

Ce qui implique

$$\mathbf{H}F(x)_{i,j} = \frac{\partial^2 F(x)}{\partial x_i \partial x_j}$$

Exemple 7 Soit $V = \mathbb{R}^n$

$$F(x) = \frac{1}{2} \langle \mathbf{A}x, x \rangle + \sum_j \frac{x_j^4}{4}$$

$$f(x) = \nabla F(x) = \mathbf{A}x + x^3$$

$$\mathbf{J}f = \mathbf{A} + \Delta(x)$$

$$\mathbf{H}F = \mathbf{A} + \Delta(x)$$

1.1.1 Applications

Définition 6 Une matrice réelle D de dimension $n \times n$ est semi-définie positive si $x^T D x \geq 0$ pour tout $x \in \mathbb{R}^n$.

Définition 7 Une matrice réelle D de dimension $n \times n$ est définie positive si $x^T D x > 0$ pour tout $x \in \mathbb{R}^n$, $x \neq 0$.

Caractérisation des minimums

On note $B_r(x)$ la boule de centre x et de rayon r .

Theorem 2 Si $x \in V$ est un minimum local de $F(x)$ et (si $V = \mathbb{R}^n$) le Hessien $\mathbf{H}F(x)$ est semi-défini positif.

Réciproquement :

Si $\nabla F(x) = 0$ et si il existe une boule $B_r(x)$ telle que le Hessien est semi-défini positif alors x est un minimum local. Si

$$\forall h \in V \quad F(x + h) = F(x) + \langle \nabla F(x), h \rangle + \frac{1}{2} \langle \mathbf{H}F(x)h, h \rangle + \epsilon(h) \|h\|^2$$

où $\lim_{h \rightarrow 0} \|\epsilon(h)\| = 0$.

1.1.2 Approximation

Dérivation numérique

- Méthode "naïve" :

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

- en théorie, la formule est vraie pour $h \approx 0$
- en pratique, attention au choix de h !
 - h trop grand : calcul trop approximatif
 - h trop petit : problèmes d'arrondis

Méthode des différences centrales

- Taylor :
 - $f(x + y) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f'''(x) + \dots$

- On connaît f sur un ensemble de points $\{x_i, y_i\}$
 - $h = x_{i+1} - x_i$
 - $f(x+h) \rightarrow y_{i+1} = y_i + hf'(x_i) + \frac{h^2}{2!}f''(x_i) + \frac{h^3}{3!}f'''(x_i) + \dots$
 - $f(x-h) \rightarrow y_{i-1} = y_i - hf'(x_i) + \frac{h^2}{2!}f''(x_i) - \frac{h^3}{3!}f'''(x_i) + \dots$

$f(x+h)-f(x-h) \Rightarrow y_{i+1}-y_{i-1} = 2hf'(x_i) + \frac{2h^3}{3!}f'''(x_i) + \dots$ en négligeant les termes en $h^3 : f'(x_i) \approx \frac{y_{i+1}-y_{i-1}}{2h}$
meilleure approximation que la méthode "naïve" (h^3/h^2)

Calcule des dérivées d'ordre supérieur :

$f''(x_i)$?

$$\begin{aligned} y_{i+1} &= y_i + hf'(x_i) + \frac{h^2}{2!}f''(x_i) + \frac{h^3}{3!}f'''(x_i) + \dots \\ y_{i-1} &= y_i - hf'(x_i) + \frac{h^2}{2!}f''(x_i) - \frac{h^3}{3!}f'''(x_i) + \dots \\ y_{i+1} + y_{i-1} &= 2y_i + \frac{2h^2}{2!}f''(x_i) + \dots \end{aligned}$$

- en négligeant les termes en h^4 :

$$f''(x_i) \approx \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}$$

- et pour les autres dérivées ?

1.2 Fonctions convexes et concaves

Soient a et b deux points dans \mathbb{R}^n .

Le segment de droite joignant ces deux points est l'ensemble des points

$$\{x \in \mathbb{R}^n \mid \exists \lambda \in [0, 1] \text{ tel que } x = a + \lambda(b - a) = \lambda b + (1 - \lambda)a\}.$$

Une fonction est *convexe* si tout point sur le segment de droite joignant $f(a)$ et $f(b)$ se trouve au-dessus du graphe de f :

$$f(\lambda b + (1 - \lambda)a) \leq \lambda f(b) + (1 - \lambda)f(a), \quad (1.2)$$

pour toute paire de points (a, b) .

Si cette inégalité est stricte, la fonction f est dite *strictement convexe*.

La Figure 1.1 illustre une fonction strictement convexe.

En remplaçant le sens de l'inégalité dans (1.2), la fonction f est dite *concave* :

$$f(\lambda b + (1 - \lambda)a) \geq \lambda f(b) + (1 - \lambda)f(a),$$

pour toute paire (a, b) .

Si cette dernière inégalité est stricte, la fonction f est *strictement concave*.

La Figure 1.2 illustre une fonction strictement concave.

Propriété 1 Soit $X \subset \mathbb{R}^n$ un ensemble convexe. Soit $f : X \rightarrow \mathbb{R}^1$ une fonction convexe. Si \bar{x} est un minimum local de f sur X , alors \bar{x} est un minimum global de f sur X .

Démonstration : La preuve se fait par contradiction en supposant qu'il existe un point $\hat{x} \in X$ tel que $f(\hat{x}) < f(\bar{x})$. Puisque f est convexe, $f(\theta\hat{x} + (1 - \theta)\bar{x}) \leq \theta f(\hat{x}) + (1 - \theta)f(\bar{x}) < \theta f(\bar{x}) + (1 - \theta)f(\bar{x}) = f(\bar{x})$ pour tout $\theta \in (0, 1]$. Or pour $\theta > 0$ suffisamment petit,

$$x(\theta) = \theta\hat{x} + (1 - \theta)\bar{x} \in B_\epsilon(\bar{x}) \cap X.$$

Ainsi $f(x(\theta)) < f(\bar{x})$ où $x(\theta) \in B_\epsilon(\bar{x}) \cap X$, contredisant que \bar{x} est un minimum local de f sur X .

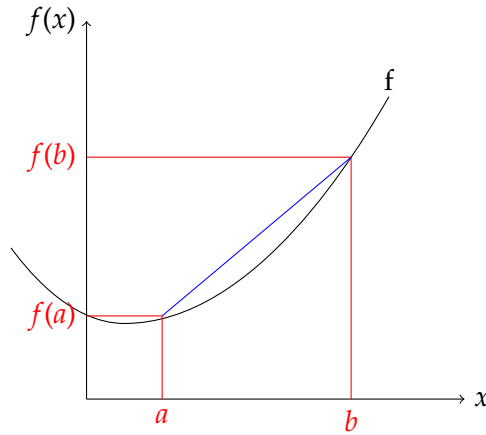


FIGURE 1.1 – Fonction convexe

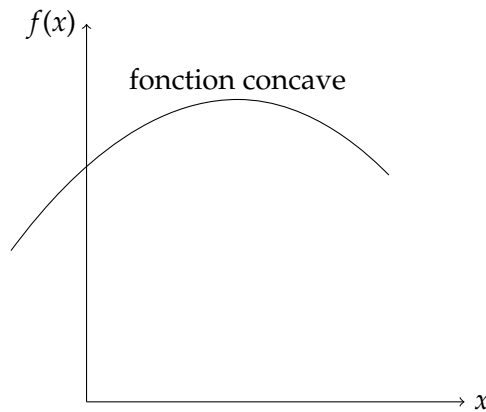


FIGURE 1.2 – Fonction concave

Propriété 2 (Inégalité du gradient) Soit $X \subset \mathbb{R}^n$ un ensemble convexe. Soit $f : X \rightarrow \mathbb{R}^1$ une fonction de classe C^1/X . Alors f est convexe sur X si et seulement si

$$f(x) \geq f(y) + \nabla f(y)^T(x - y)$$

pour tout paire de points $x, y \in X$.

Corollaire :

Soit $X \subset \mathbb{R}^n$ un ensemble convexe. Soit $f : X \rightarrow \mathbb{R}^1$ une fonction convexe de classe C^1/X . Alors si $\nabla f(x^*)^T(x - x^*) \geq 0$ pour tout $x \in X$, alors x^* est un minimum global de f sur X .

Démonstration : Le résultat découle directement de l'inégalité de gradient. En effet, puisque f est convexe sur X , alors

$$f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*)$$

pour tout $x \in X$. Puisque par hypothèse $\nabla f(x^*)^T(x - x^*) \geq 0$, alors

$$f(x) - f(x^*) \geq \nabla f(x^*)^T(x - x^*) \geq 0$$

d'où $f(x) \geq f(x^*)$ pour tout $x \in X$. Donc x^* est un minimum global de f sur X . **Conséquence :**

Si $\nabla f(x^*) = 0$, alors x^* est un minimum global de f sur X .

1.2.1 Tests de convexité et de concavité

Supposons que f est une fonction deux fois dérivable d'une seule variable. Alors,

1. f est convexe si et seulement si

$$\forall x, \frac{d^2 f(x)}{dx^2} \geq 0;$$

2. f est strictement convexe si

$$\forall x, \frac{d^2 f(x)}{dx^2} > 0;$$

3. f est concave si et seulement si

$$\forall x, \frac{d^2 f(x)}{dx^2} \leq 0;$$

4. f est strictement concave si

$$\forall x, \frac{d^2 f(x)}{dx^2} < 0.$$

Supposons que f est une fonction deux fois dérivable de deux variables. Dans ce cas, nous utilisons la Table 1.1 comme conditions nécessaires.

	Convexe	Strict. convexe	Concave	Strict. concave
$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2}$	≥ 0	> 0	≤ 0	< 0
$\frac{\partial^2 f(x_1, x_2)}{\partial x_2^2}$	≥ 0	> 0	≤ 0	< 0
$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} \cdot \frac{\partial^2 f(x_2, x_2)}{\partial x_2^2} - \left(\frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} \right)^2$	≥ 0	> 0	≥ 0	> 0

TABLE 1.1 – Test de convexité et de concavité pour une fonction à deux variables deux fois dérivable

Ces tests se généralisent à des fonctions de plus de deux variables en utilisant le Hessien (la matrice des dérivées secondes). Deux résultats supplémentaires sont utiles pour tester la convexité :

- une fonction qui s'exprime comme la somme de fonctions convexes (concaves) est convexe (concave);
- l'opposé d'une fonction concave est convexe et vice-versa.

Pour les fonctions f de classe C^2 , il existe un critère pour vérifier la convexité qui est basé sur le Hessien $\nabla^2 f$.

Propriété 3 Soit $X \subset \mathbb{R}^n$ un ensemble convexe dont l'intérieur est non vide. Soit $f : X \rightarrow \mathbb{R}^1$ une fonction de classe C^2/X . Alors f est convexe sur X si et seulement si son Hessien $\nabla^2 f$ est une matrice semi définie positive pour tout $x \in X$.

1.2.2 Valeurs propres

Theorem 3 Une matrice symétrique définie positive possède des valeurs propres strictement positives.

Démonstration :

1. Supposons A symétrique définie positive. Soit λ_i une valeur propre associée au vecteur propre x_i . On a donc $Ax_i = \lambda_i x_i$, et $x_i^T Ax_i = \lambda_i x_i^T x_i = \lambda_i \|x_i\|^2 > 0$.
2. Réciproquement, si toutes les valeurs propres sont positives, on a pour tout vecteur propre x_i $x_i^T Ax_i > 0$. Ces derniers formant une base orthonormée, on écrit pour tout $x \neq 0$

$$x^T Ax = \left(\sum_{i=1}^n \alpha_i x_i \right)^T A \left(\sum_{i=1}^n \alpha_i x_i \right) = \sum_{i=1}^n \alpha_i^2 \lambda_i > 0.$$

1.3 Ensembles convexes

Un ensemble est convexe si, pour toute paire de points de l'ensemble, le segment de droite joignant ces points est contenu dans l'ensemble. Il est possible de démontrer qu'un ensemble est convexe grâce aux propriétés suivantes :

- si f est convexe, alors $\{x \mid f(x) \leq b\}$ est convexe ;
- si f est concave, alors $\{x \mid f(x) \geq b\}$ est convexe ;
- l'intersection d'ensembles convexes est convexe.

Exemple 8 (Wyndor Glass) Soit le problème d'optimisation suivant :

$$\max_x z = 3x_1 + 5x_2$$

sous les contraintes

$$\begin{aligned} x_1 &\leq 4 && \text{(usine 1)} \\ 2x_2 &\leq 12 && \text{(usine 3)} \\ 3x_1 + 2x_2 &\leq 18 && \text{(usine 3)} \end{aligned}$$

$$x_1 \geq 0, x_2 \geq 0 \text{ (non-négativité)}$$

Exemple 9 (Wyndor Glass) Modifions l'exemple Wyndor Glass en remplaçant certaines contraintes par une contrainte non linéaire.

$$\begin{aligned} \max z &= 3x_1 + 5x_2 \\ \text{s.c. } x_1 &\leq 4 \\ 9x_1^2 + 5x_2^2 &\leq 216 \\ x_1 &\geq 0, x_2 \geq 0. \end{aligned}$$

Le problème ainsi modifié peut se représenter comme sur la Figure 1.3.

Remarquons que l'objectif est concave (car linéaire) et que le domaine réalisable est convexe : c'est un modèle de *programmation convexe*.

Dans le cas présent, la solution optimale est sur la frontière du domaine réalisable, mais ne correspond pas à un coin (l'intersection de deux contraintes).

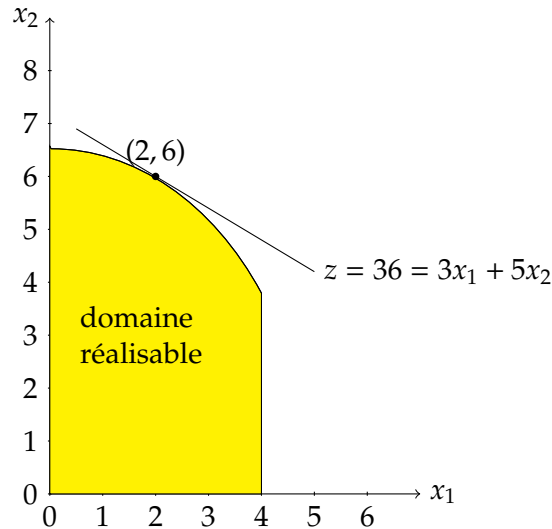


FIGURE 1.3 – exemple Wyndor Glass

Modifions encore l'exemple Wyndor Glass, mais cette fois en remplaçant l'objectif lineaire par une fonction non lineaire, comme illustré sur la Figure 1.4 :

$$\begin{aligned} \max z &= 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2 \\ x_1 &\leq 4 \\ 2x_2 &\leq 12 \\ 3x_1 + 2x_2 &\leq 18 \end{aligned}$$

Nous pouvons à nouveau remarquer que l'objectif est concave et que le domaine réalisable est convexe : c'est aussi un modèle de programmation convexe. Ici, la solution optimale $(8/3, 5)$ est sur la frontière du domaine réalisable, mais ne correspond pas à un point extrême du domaine réalisable.

Considérons le même exemple, mais avec une fonction objectif différente, tel que représenté sur la Figure 1.5 :

$$\begin{aligned} \max z &= 54x_1 - 9x_1^2 + 78x_2 - 13x_2^2 \\ x_1 &\leq 4 \\ 2x_2 &\leq 12 \\ 3x_1 + 2x_2 &\leq 18 \end{aligned}$$

Il s'agit à nouveau d'un modèle de programmation convexe, avec comme solution optimale $(3, 3)$, laquelle se trouve à l'intérieur du domaine réalisable. La fonction objectif est la somme de deux fonctions d'une seule variable. Si nous annulons les dérivées de chacune de ces fonctions, nous obtenons la solution unique $(3, 3)$, qui se trouve à l'intérieur du domaine ; c'est donc nécessairement la solution optimale.

Revenons sur la fonction objectif initiale, mais introduisons une contrainte non-linéaire, qui définit

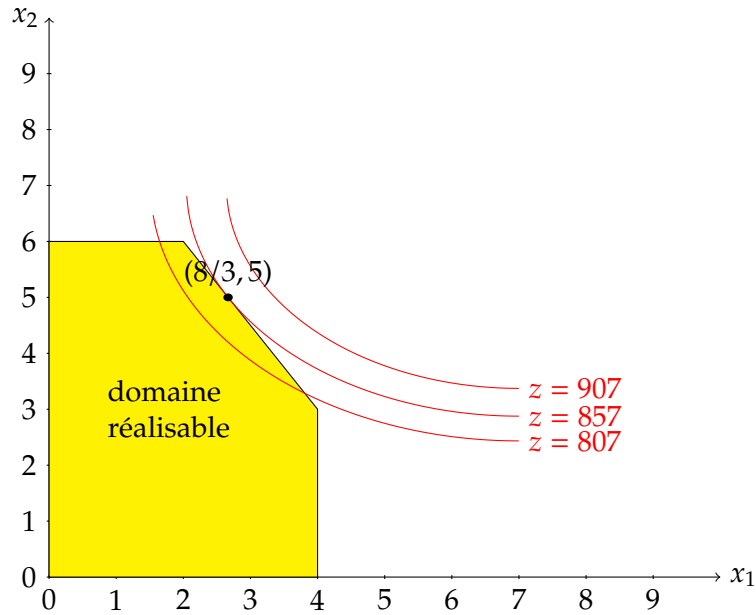


FIGURE 1.4 – exemple Wyndor Glass - cas 2

un domaine réalisable non-convexe, comme illustré sur la Figure 1.6 :

$$\begin{aligned}
 \max z &= 3x_1 + 5x_2 \\
 \text{s.c. } x_1 &\leq 4 \\
 x_2 &\leq 7 \\
 8x_1 - x_1^2 + 14x_2 - x_2^2 &\leq 49 \\
 x_1 &\geq 0, x_2 \geq 0.
 \end{aligned}$$

Le problème ainsi modifié peut se représenter comme sur la Figure 1.6.

Dans ce modèle de *programmation non convexe*, nous remarquons la présence de deux maxima locaux :

- x est un *maximum local* si $f(x) \geq f(a)$ pour tout a réalisable suffisamment près de a ;
- $(4, 3)$ et $(0, 7)$ sont des maxima locaux, mais $x^* = (0, 7)$ est la *maximum global* : $f(x) \geq f(a)$ pour tout a réalisable.

Les méthodes classiques de programmation non-linéaire permettent d'identifier un maximum local, mais pas nécessairement le maximum global, alors qu'en programmation convexe, tout maximum local est global.

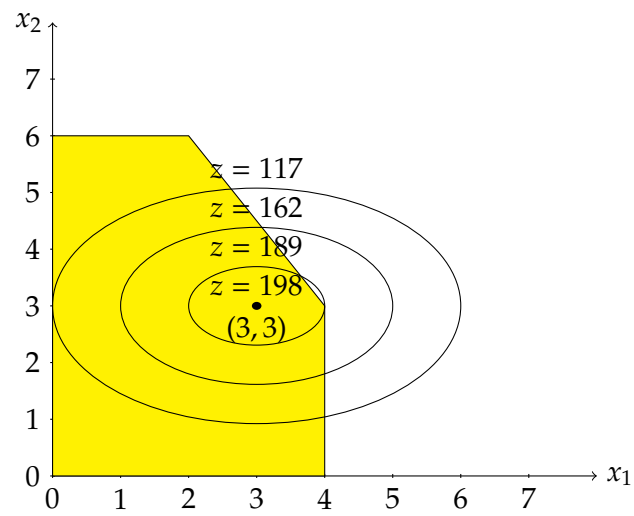


FIGURE 1.5 – exemple Wyndor Glass - cas 3

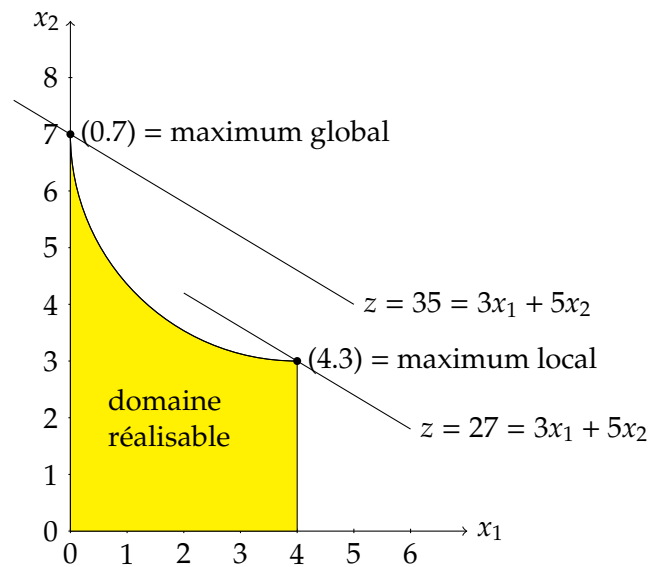


FIGURE 1.6 – exemple Wyndor Glass - cas 4

Chapitre 2

Optimisation sans contrainte

2.1 Introduction

2.1.1 Formulation générale des problèmes d'optimisation

La forme générale d'optimisation est la suivante :

$$\left\{ \begin{array}{ll} \min & f(x) \\ \text{s.c.} & g(x) \leq 0 \\ & h(x) = 0 \\ & x \in \mathbb{R}^n. \end{array} \right. \quad (2.1)$$

Il va de soi que la plupart des problèmes réels ou industriels ne sont pas initialement sous la forme proposée. C'est pourquoi un des premiers travaux consiste en général à mettre le problème initial sous une forme standard. Par exemple, un problème donné sous forme

$$\max_{x \in \mathbb{R}^n} g(x),$$

se mettra sous la forme standard d'optimisation sans contraintes de (2.1) en posant $f(x) = -g(x)$.

Considérons tout d'abord le cas d'un modèle de programmation non-linéaire dans lequel il n'y a aucune contrainte :

$$\min f(x_1, x_2, \dots, x_n). \quad (2.2)$$

Il est alors possible de montrer que si $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ est un minimum local, alors

$$\frac{\partial f(x^*)}{\partial x_j} = 0, \quad j = 1, 2, \dots, n.$$

En d'autres mots, lorsque la dérivée s'annule en un point donné, ce point *peut être* un maximum local, mais l'inverse n'est pas nécessairement vérifié : ce point peut aussi être un minimum local ou un point de selle. Par contre, si f est convexe, un point où la dérivée s'annule est nécessairement un minimum global. De plus, si f est strictement convexe, un tel minimum global est unique.

Dans le cas d'une fonction non convexe, afin de trouver un minimum global, il faut

- identifier tous les minima locaux ;
- identifier celui de plus petite valeur ;
- vérifier que la fonction est bornée inférieurement (sinon, il n'y a pas de minimum global).

Le problème est qu'identifier tous les minima locaux peut être extrêmement difficile.

2.1.2 Résultats d'existence

Considérons notre problème d'optimisation (2.1), que l'on écrira un peu différemment, en mettant les contraintes sous la forme $x \in K \subset \mathbb{R}^n$:

$$\min_{x \in K} f(x). \quad (2.3)$$

Définition 8 Une ensemble $K \subset \mathbb{R}^n$ est dit compact si, de toute suite $\{x_k\}$, où $x_k \in K$, $\forall k$, on peut extraire une sous-suite convergente.

Theorem 4 Un ensemble $K \subset \mathbb{R}^n$ est compact si et seulement si il est fermé et borné.

Exemple 10 Dans \mathbb{R} , les intervalles fermés du type $[a, b]$ sont compacts. La notion de fermeture signifie qu'une suite $\{x_k\}$, où $x_k \in K$, $\forall k$, doit converger vers une limite $x \in K$. Pour illustrer sur un exemple qu'un intervalle ouvert dans \mathbb{R} ne peut pas être compact : soit $K =]0, 1[$ et la suite $x_k = \frac{1}{k}$, on a bien $x_k \in K$ mais $\lim_{k \rightarrow 0} x_k = 0 \notin K$.

Théorèmes généraux d'existence

Theorem 5 Si $f : K \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est continue et si de plus K est un ensemble compact, alors le problème (2.3) admet une solution optimale $x^* \in K$, qui vérifie donc

$$f(x^*) \leq f(x), \quad \forall x \in K.$$

Theorem 6 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue sur \mathbb{R}^n . Si

$$\lim_{\|x\| \rightarrow \infty} f(x) = \infty,$$

alors (2.1) admet une solution optimale x^* .

Démonstration : Soit $x_0 \in \mathbb{R}^n$. Puisque $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$ il existe $M > 0$ tel que $\|x\| > M \Rightarrow f(x) > f(x_0)$, donc

$$\exists M > 0, \quad f(x) \leq f(x_0) \Rightarrow \|x\| \leq M.$$

Puisque x^* est caractérisé par $f(x^*) \leq f(x)$, $\forall x \in \mathbb{R}^n$, on a donc forcément $\|x^*\| \leq M$. Donc x^* est solution du problème

$$\min_{\|x\| \leq M} f(x),$$

et le théorème précédents s'applique, la boule $\{x \in \mathbb{R}^n, \|x\| \leq M\}$ étant compacte.

2.1.3 Unicité

Theorem 7 Si $f : K \subset \mathbb{R}^n \rightarrow \mathbb{R}$ strictement convexe sur K convexe. Le minimum de f sur K , s'il existe, est unique

Démonstration : Soit donc $x^* \in K$ tel que $f(x^*) \leq f(x)$, $\forall x \in K$. Supposons qu'il existe $y^* \neq x^*$ tel que $f(y^*) \leq f(x)$, $\forall x \in K$. Formons pour $\lambda \in]0, 1[$ le vecteur

$$u = \lambda y^* + (1 - \lambda)x^*.$$

D'après la stricte convexité de f et puisque nécessairement $f(y^*) = f(x^*)$ on a

$$f(u) < \lambda f(y^*) + (1 - \lambda)f(x^*) = f(x^*),$$

ce qui contredit le fait que x^* soit un minimum. On a donc $x^* = y^*$.

2.1.4 Conditions d'optimalité

Conditions nécessaires

Theorem 8 Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable et x^* vérifiant

$$f(x^*) \leq f(x), \quad \forall x \in \mathbb{R}^n,$$

alors on a nécessairement

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

Démonstration : Pour tout $t \in \mathbb{R}^*$ et pour $h \in \mathbb{R}^n$ on a $f(x^*) \leq f(x^* + th)$. On a donc

$$\lim_{t \rightarrow 0^+} \frac{f(x^*) - f(x^* + th)}{t} = \nabla f(x^*)^T h \leq 0,$$

et

$$\lim_{t \rightarrow 0^-} \frac{f(x^*) - f(x^* + th)}{t} = \nabla f(x^*)^T h \geq 0,$$

donc $\nabla f(x^*)^T h = 0, \quad \forall h \in \mathbb{R}^n$, donc $\nabla f(x^*)^T = 0$ (prendre par exemple $h = \nabla f(x^*)$).

Conditions nécessaires et suffisantes

Theorem 9 Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ convexe et différentiable. Si x^* vérifiant

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

alors on

$$f(x^*) \leq f(x), \quad \forall x \in \mathbb{R}^n.$$

Démonstration : Soient $x \in \mathbb{R}^n$ et $\lambda \in [0, 1]$. Puisque f est convexe on a

$$f(\lambda x + (1 - \lambda)x^*) \leq \lambda f(x) + (1 - \lambda)f(x^*).$$

On retranche $f(x^*)$ de chaque côté de l'inégalité, on note que $\lambda x + (1 - \lambda)x^* = x^* + \lambda(x - x^*)$, puis on divise par λ , ce qui donne l'inégalité

$$\frac{f(x^* + \lambda(x - x^*)) - f(x^*)}{\lambda} \leq f(x) - f(x^*).$$

Et si on fait tendre λ vers 0 on obtient $\nabla f(x^*)^T (x - x^*) \leq f(x) - f(x^*)$, donc $0 \leq f(x) - f(x^*)$.

Theorem 10 Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ deux fois différentiable. Si

$$\begin{cases} \nabla f(x^*) = 0, \\ \nabla^2 f(x^*) \text{ est définie positive,} \end{cases}$$

alors x^* est un minimum local de f .

Démonstration : On a

$$\begin{aligned} f(x^* + th) &= f(x^*) + t \nabla f(x^*)^T h + \frac{t^2}{2} h^T \nabla^2 f(x^*) h + t^2 \|h\|^2 \epsilon(th), \\ &= f(x^*) + \frac{t^2}{2} h^T \nabla^2 f(x^*) h + t^2 \|h\|^2 \epsilon(th). \end{aligned}$$

On a donc pour $t > 0$

$$\frac{f(x^* + th) - f(x^*)}{t^2} = \frac{1}{2} h^T \nabla^2 f(x^*) h + \|h\|^2 \epsilon(th).$$

Donc si t est suffisamment petit on aura bien $f(x^* + th) - f(x^*) > 0$ puisque $h^T \nabla^2 f(x^*) h > 0$.

Cas de fonctions à deux variables

Tout comme

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

on a :

$$\Delta_1 = \frac{\partial^2 f(x, y)}{\partial x^2},$$

$$\Delta_2 = \frac{\partial^2 f(x, y)}{\partial x^2} \frac{\partial^2 f(x, y)}{\partial y^2} - \left(\frac{\partial^2 f(x, y)}{\partial x \partial y} \right)^2,$$

:

Ainsi

- $\Delta_1 > 0$ et $\Delta_2 > 0 \rightarrow$ minimum.
- $\Delta_1 < 0$ et $\Delta_2 > 0 \rightarrow$ maximum.
- Δ_1 quelconque et $\Delta_2 < 0 \rightarrow$ point-selle.
- Δ_1 quelconque et $\Delta_2 = 0 \rightarrow$ on ne peut pas conclure.

2.1.5 Algorithme

Résolution d'équations

- L'identification des points critiques revient à résoudre

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

- Il s'agit d'un système de n équations à n inconnues.
- On verra que les conditions d'optimalité pour les problèmes avec contraintes se ramènent également à un système d'équations.
- Analysons d'abord les algorithmes permettant de résoudre ces systèmes d'équations.
- Ils seront ensuite adaptés pour les problèmes d'optimisation.

2.1.6 Histoire

Histoire d'algorithme



- Abu Ja'far Muhammad ibn Musa Al-Khwarizmi (780-840)
- Traité al Kitab almukhtasar fi hisab **al-jabr** w'al muqabala, qui est à l'origine de l'algèbre
- traduction latine de cet ouvrage, intitulée Algoritmi de numero indorum

2.2 Fonctions à une seule variable

2.2.1 Méthode de la bisection

Considérons en premier lieu le cas le plus simple : la fonction objectif f est concave et comporte une seule variable. Il est alors possible de montrer que, si x^* est une solution optimale et il existe x^i et x^u tels que $x^i \leq x^* \leq x^u$ et $f(x^i) \neq f(x^*)$, $f(x^u) \neq f(x^*)$, alors il existe a et b tels que $a \leq x^* \leq b$ et

$$\frac{df(x)}{dx} > 0, \text{ si } x < a,$$

$$\frac{df(x)}{dx} = 0, \text{ si } x = x^*,$$

$$\frac{df(x)}{dx} < 0, \text{ si } x > b.$$

Si f est strictement concave en x^* , alors $a = x^* = b$, comme illustré sur la Figure 2.1.

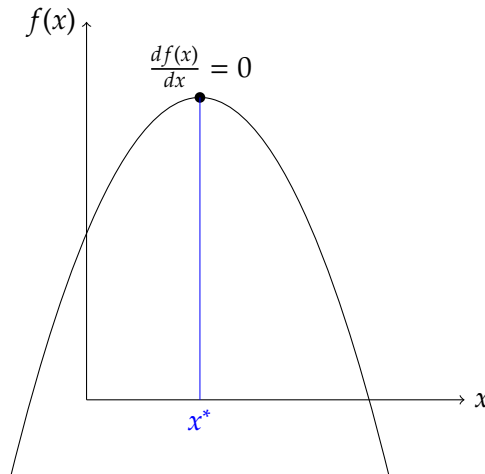


FIGURE 2.1 – Maximum d'une fonction strictement concave

Algorithme 1 (Méthode de la bisection)

- 1 Nous fixons d'abord une borne inférieure x^i pour laquelle la dérivée en ce point est strictement positive.
- 2 Nous déterminons également une borne supérieure x^u pour laquelle la dérivée en ce point est strictement négative.
- 3 Si les deux bornes ne sont pas suffisamment près l'une de l'autre, i.e. la distance $x^u - x^i$ est trop grande, nous prenons le point milieu entre les deux bornes comme point candidat x^c :

$$x^c = \frac{x^i + x^u}{2},$$

et nous passons à l'étape suivante. Sinon, c'est-à-dire si

$$|x^u - x^i| \leq 2\epsilon,$$

avec $\epsilon > 0$ suffisamment petit, nous nous arrêtons : x^c est une approximation suffisamment précise de la solution optimale.

- 4 Si la dérivée en x^c est nulle, arrêt : x^c est la solution optimale.
- 5 Si la dérivée en x^c est positive, nous posons $x^i = x^c$.
- 6 Si la dérivée en x^c est négative, nous posons $x^u = x^c$.
- 7 Retour à l'étape 3.

critères d'arrêts

1. L'erreur absolue : $|r - x_m| \simeq \frac{|x_1 - x_2|}{2} < \epsilon_{abs}$
2. L'erreur relative : $\frac{|r - x_m|}{|r|} \simeq \frac{|x_1 - x_2|}{|x_m|} < \epsilon_{rel}$
3. On peut arrêter l'algorithme si $|f'(x_m)| < \epsilon_f$

Méthode bi-section : Erreurs

Soit $[x_1 - x_2] = [a, b]$ l'intervalle de départ de longueur $L = b - a$.
Après une itération on a $x_m = \frac{x_1 + x_2}{2}$ et le nouvel intervalle $[x_1 - x_2]$ est de longueur $\frac{L}{2}$. A l'étape n , la longueur est $\frac{L}{2^n}$. On sait que $r \in [x_1, x_2]$ et

$$|r - x_m| \leq \frac{L}{2^n}.$$

Étant donnée une erreur absolue Δr , c'est quoi la valeur de n (nombre d'itérations) pour avoir

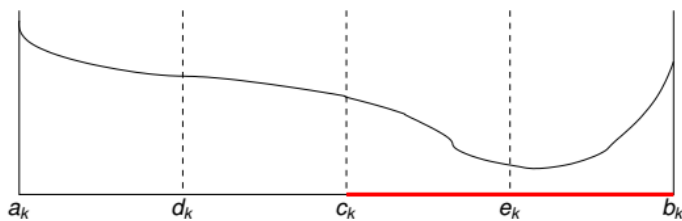
$$|r - x_m| \leq \frac{L}{2^n} < \Delta r$$

La réponse est : $n > \frac{\ln(\frac{L}{\Delta r})}{\ln 2}$.

2.2.2 Méthode de la règle d'Or

L'algorithme par dichotomie :

- à l'itération k : intervalle $[a_k, b_k]$
- $d_k = \frac{3a_k + b_k}{4}$ $c_k = \frac{a_k + b_k}{2}$ $e_k = \frac{a_k + 3b_k}{4}$
- $f(c_k) > f(e_k) \implies a_{k+1} = c_k$ et $b_{k+1} = b_k$
- $f(d_k) > f(c_k) \implies a_{k+1} = a_k$ et $b_{k+1} = c_k$
- sinon $a_{k+1} = d_k$ et $b_{k+1} = e_k$
- arrêt : quand $b_k - a_k \leq \epsilon$



2.2.3 Méthode de Newton

Soit une équation à résoudre de la forme :

$$f'(x) = 0$$

À partir d'une valeur initiale x_0 de la solution, on cherche une correction δx telle que :

$$0 = f'(x_0 + \delta x) \approx f'(x_0) + f''(x_0)\delta x$$

on peut alors isoler la correction recherchée :

$$\delta x = -\frac{f'(x_0)}{f''(x_0)}$$

la correction δx est en principe la quantité que l'on doit ajouter à x_0 pour annuler la fonction $f'(x)$. puisque nous avons négligé les termes d'ordre supérieur ou égal à 2 dans le développement de Taylor, cette correction n'est pas parfaite et l'on pose :

$$x_1 = x_0 + \delta x$$

Interprétation géométrique de la méthode de Newton

Menons par le point $(x_n, f'(x_n))$ la tangente à la courbe $y = f'(x)$ fournie par

$$D(x) = f'(x_n) + f''(x_n)(x - x_n)$$

si on cherche le point d'intersection de la tangente, $D(x) = 0$, avec l'axe des x , on retrouve le point x_{n+1} tel que défini par l'algorithme.

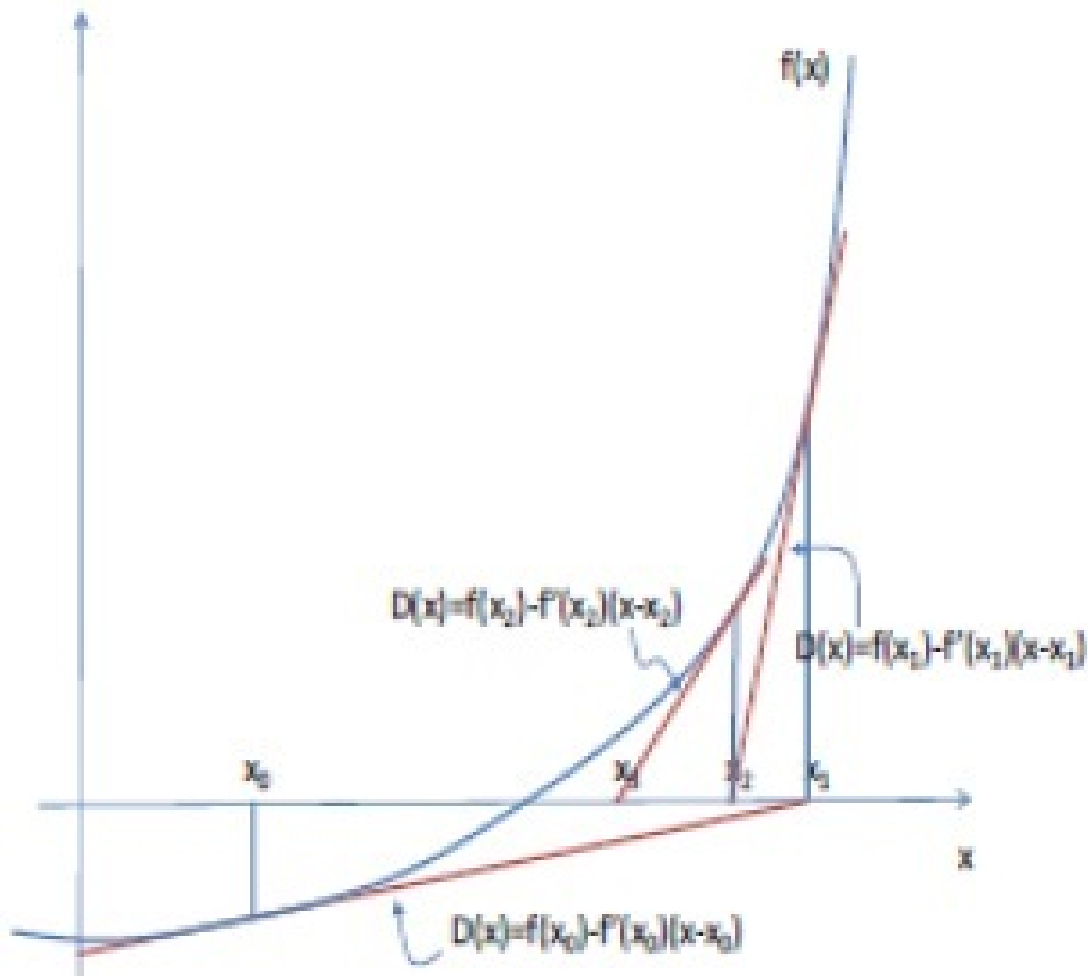
Algorithme de la méthode de Newton

1. Étant donné ϵ_a , un critère d'arrêt
2. Étant donné N , le nombre maximal d'itérations
3. Étant donné x_0 , une valeur initiale de la solution
4. Effectuer : $x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$
5. si $\frac{|x_{n+1} - x_n|}{|x_{n+1}|} < \epsilon_a$:
 - convergence atteint
 - écrire la solution x_{n+1} : arrêt
6. si le nombre maximal d'itérations N est atteint :
 - convergence non atteinte en N itérations : arrêt
7. retour à l'étape 4.

Étude de convergence

On peut associer la méthode de Newton à l'application de la méthode de point fixe sur une fonction g particulière

$$g(x) = x - \frac{f'(x)}{f''(x)}$$



On retrouve les résultats de convergence obtenue pour le point fixe. On peut cependant revoir les résultats en fonction de f' puisque la relation en f' et g est maintenant fixée.

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

Pour une racine r de f' on aura donc

$$g'(r) = 0$$

et on a ainsi la convergence quadratique que l'on recherche.

Principe

Déterminer un intervalle $[a, b]$ sur lequel il y a convergence de Newton n'est pas toujours facile. Il existe cependant un théorème qui donne une condition suffisante.

Theorem 11 (convergence globale de la méthode de Newton) Si la fonction $f'(x)$ définie sur $[a, b]$ vérifie :

$$1. f'(a)f'(b) < 0$$

2. $\forall x \in [a, b] \quad f''(x) \neq 0$ (stricte monotonie de f')
3. $\forall x \in [a, b] \quad f'''(x) \neq 0$ (concavité de f' dans le même sens.)

Alors en choisissant x_0 de $[a, b]$ tel que $f'(x_0)f'''(x_0) > 0$, les itérations de Newton convergent vers l'unique solution x^* de $f'(x) = 0$ dans $[a, b]$. De plus, la convergence est quadratique.

Remarque 1 • La fonction f' doit être dérivable.

- x_{k+1} peut ne pas être calculable si $f''(x_k) = 0$ ou si x_k n'est pas dans le domaine de définition de f' .
- chaque itération nécessite une évaluation de f' et une évaluation de f'' .
- cette méthode est souvent appelée aussi méthode de Newton-Raphson.
- la méthode de Newton est une méthode de point fixe puisque x_{k+1} peut s'écrire sous la forme $x_{k+1} = g(x_k)$ avec

$$g(x) = x - \frac{f'(x)}{f''(x)}.$$

2.3 Fonctions à plusieurs variables

2.3.1 Méthode du gradient

Principe des méthodes de descente

Définition 9 Soit $f : \mathbb{R}^n \Rightarrow \mathbb{R}$. On dira qu'un vecteur d est une direction de descente en x s'il existe $\bar{t} > 0$ tel que

$$f(x + td) < f(x), \quad t \in]0, \bar{t}].$$

Le principe d'une méthode de descente consiste à faire les itérations suivantes

$$x_{k+1} = x_k + t_k d_k, \quad t_k > 0, \tag{2.4}$$

tout en assurant la propriété

$$f(x_{k+1}) < f(x_k).$$

Le vecteur d_k est la direction de descente en x_k . Le scalaire t_k est appelé le **pas** de la méthode à l'itération k . On peut caractériser les directions en x_k à l'aide du gradient :

Theorem 12 Soit $d \in \mathbb{R}^n$ vérifiant

$$\nabla f(x)^T d < 0,$$

alors d est une direction de descente en x .

Démonstration : On a pour $t > 0$

$$f(x + td) = f(x) + t \nabla f(x)^T d + t\epsilon(t),$$

donc si on écrit

$$\frac{f(x + td) - f(x)}{t} = \nabla f(x)^T d + \epsilon(t),$$

on voit bien que pour t suffisamment petit on aura $f(x + td) - f(x) < 0$. \square

Dans la méthode (2.4) le choix de t_k est lié à la fonction

$$\varphi(t) = f(x_k + t d_k),$$

en particulier, une façon de choisir t_k peut être de résoudre le problème d'optimisation (à une seule variable)

$$\min_{t \geq 0} \varphi(t).$$

Le pas \hat{t}_k obtenu ainsi s'appelle le pas optimal. La fonction $\varphi(t) = f(x_k + t d_k)$ étant différentiable, on a alors nécessairement

$$\varphi'(\hat{t}_k) = \nabla f(x_k + \hat{t}_k d_k)^T d_k = 0.$$

Considérons une fonction concave f de *plusieurs variables*. La méthode de la bisection ne peut plus s'appliquer, aussi nous devons nous tourner vers une meilleure alternative. Le gradient de f au point x' est défini comme

$$\nabla_x f(x') = \begin{pmatrix} \frac{\partial f(x')}{\partial x_1} \\ \frac{\partial f(x')}{\partial x_2} \\ \vdots \\ \frac{\partial f(x')}{\partial x_n} \end{pmatrix}.$$

Il est possible de montrer que le gradient correspond à une direction d'augmentation de la valeur de f . Ainsi, dans la méthode du gradient, nous nous déplaçons dans la direction du gradient en tentant d'augmenter au maximum la valeur de l'objectif. À partir d'un point initial x' , nous effectuons un déplacement dans la direction du gradient vers un nouveau point x :

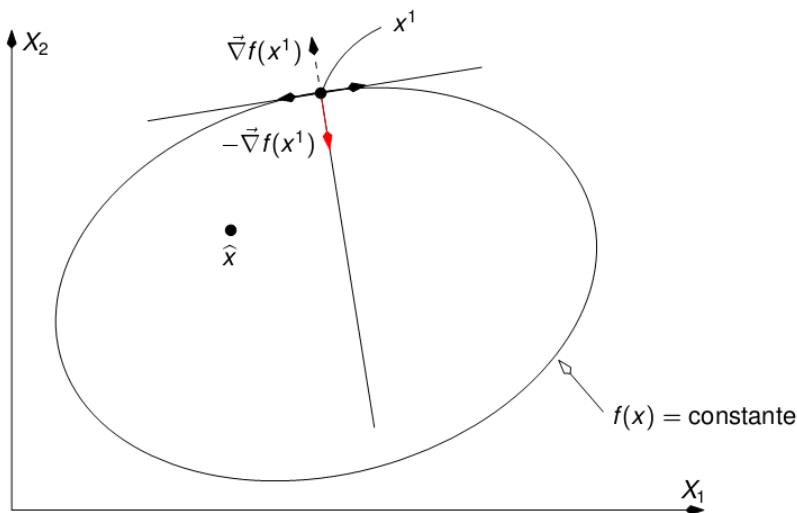
$$x = x' + s,$$

où $s = t^* \nabla_x f(x)$. Dans cette formule, t^* est la solution du problème de maximisation suivant :

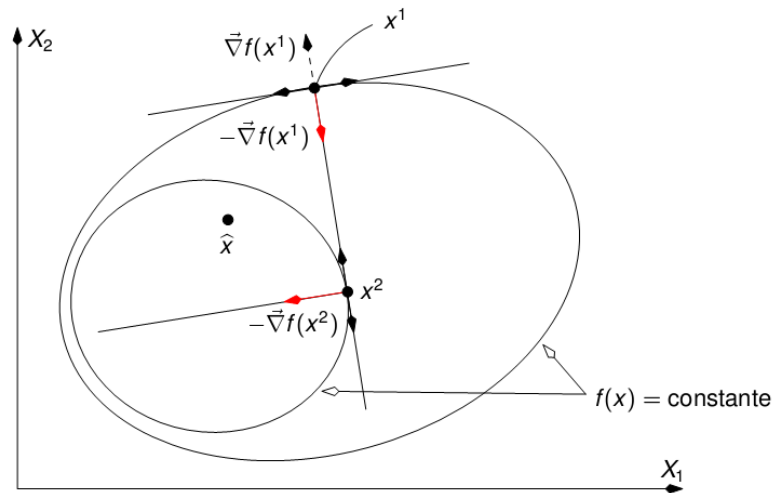
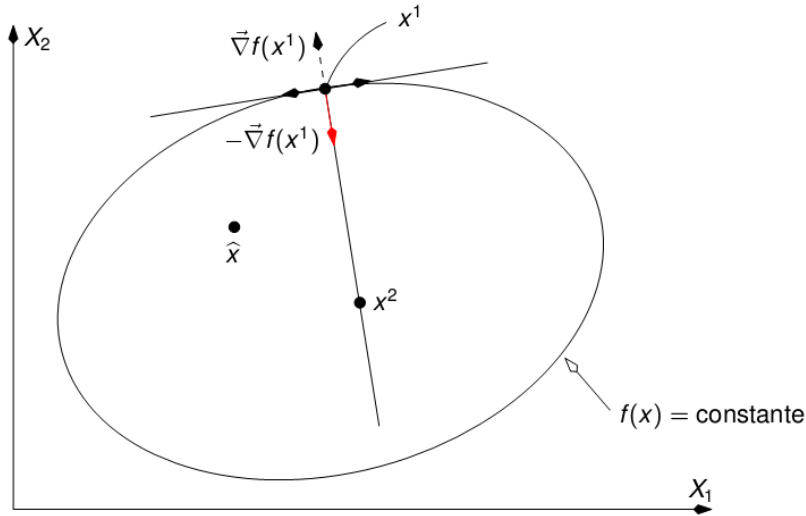
$$\max_{t \geq 0} f(x' + t \nabla_x f(x')).$$

C'est un problème de maximisation d'une fonction concave d'une seule variable : nous pouvons par conséquent le résoudre, par exemple en employant la méthode de la bisection

Nous posons ensuite $x' = x$, et nous itérons ainsi jusqu'à ce que le gradient s'annule (ou presque).



Remarque 2 La valeur de f diminue à mesure que les courbes de niveau se rapprochent du centre des ellipses !



Nous avons par conséquent l'algorithme suivant.

Algorithme 2 (Méthode du gradient) 1. Déterminer une solution initiale x_0 . Posons $k = 0$. Nous fixons également une constante $\epsilon > 0$ suffisamment petite.

2. Résoudre le problème suivant :

$$\max_{t \geq 0} f(x_k + t \nabla_x f(x_k)). \quad (2.5)$$

3. Soit t^* la solution optimale de (2.5). Posons

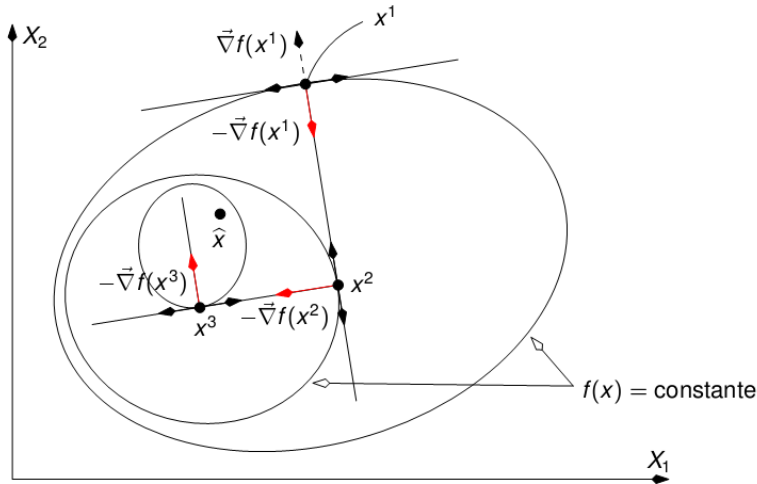
$$x_{k+1} = x_k + t^* \nabla_x f(x_k),$$

et incrémentons k .

4. Si

$$\left| \frac{\partial f(x_k)}{\partial x_j} \right| \leq \epsilon, \quad j = 1, 2, \dots, n,$$

arrêt. Sinon, nous retournons en 2.



Convergence

Theorem 13 Soit $f \in C^1$. Tout point limite x^* d'une sous suite convergent de la suite générée par la méthode du gradient est tel que $\nabla f(x^*) = 0$.

Démonstration. Considérons la sous suite $\{x^{k_j}\}$ de $\{x^k\}$ telle que

$$\{x^{k_j}\} \longrightarrow x^*.$$

Par conséquent, se référant aux suites

$$\{x^k\} = \{\dots, x^{k_j}, x^{k_j+1}, \dots, x^{k_j+\tau}, \dots\}$$

$$\{x^{k_j}\} = \{\dots, x^{k_j}, x^{k_{j+1}}, \dots\} \text{ où } x^{k_{j+1}} = x^{k_j+\tau}.$$

Ainsi, se référant à la suite $\{x^k\}$

$$f(x^{k_{j+1}}) = f(x^{k_j+\tau}) \leq \dots \leq f(x^{k_j+1}) = \min_{t \geq 0} \{f(x^{k_j} - t \nabla f(x^{k_j}))\}.$$

Donc

$$f(x^{k_{j+1}}) \leq f(x^{k_j} - t \nabla f(x^{k_j})) \quad \forall t \geq 0.$$

Puisque $f \in C^1$, f est continue sur \mathbb{R}^n et

$$\lim_{j \rightarrow \infty} f(x^{k_{j+1}}) = f(x^*)$$

$$\lim_{j \rightarrow \infty} [f(x^{k_j} - t \nabla f(x^{k_j}))] = f(x^* - t \nabla f(x^*)) \quad \forall t \geq 0.$$

Donc

$$f(x^*) \leq f(x^* - t \nabla f(x^*)) \quad \forall t \geq 0. \tag{2.6}$$

Par conséquent $\nabla f(x^*) = 0$ car autrement si $\nabla f(x^*) \neq 0$, alors $-\nabla f(x^*)$ est une direction de descente, donc il existe $t > 0$ tq $f(x^* - t \nabla f(x^*)) < f(x^*)$ ce qu'est contradiction avec la relation (2.6).

Déplacement en zig-zag

Theorem 14 Si $x^{k+1} = x^k - t_k \nabla f(x^k)$ où t_k est tel que $f(x^k - t_k \nabla f(x^k)) = \min_{t \geq 0} \{f(x^k - t \nabla f(x^k))\}$, alors $\nabla f(x^{k+1})^T \nabla f(x^k) = 0$.

Démonstration : Si $\nabla f(x^k) = 0$, alors la preuve est complétée.
Si $\nabla f(x^k) \neq 0$, considérons la fonction $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}$

$$\varphi(t) = f(x^k - t \nabla f(x^k)).$$

Par dérivation en chaîne,

$$\varphi'(t) = -\nabla f(x^k)^T \nabla f(x^k - t \nabla f(x^k)).$$

Puisque t_k est un minimum local de $\varphi(t)$, alors $\varphi'(t_k) = 0$, et $\nabla f(x^k)^T \nabla f(x^{k+1}) = 0$. \square

Notes :

1. L'ordre de convergence de la méthode du gradient est linéaire.
2. Au début de son application, la méthode permet de faire diminuer la valeur de la fonction économique relativement rapidement, mais son efficacité à ce chapitre diminue au cours des itérations.
3. La méthode du gradient est la méthode la plus rapide marginalement ("steepest descent methode") :

$$\nabla f(x)^T d = \|\nabla f(x)\| \|d\| \cos \theta$$

où θ est l'angle entre d et $\nabla f(x)$. Ainsi, parmi les directions d de norme 1, celle minimisant $\nabla f(x)^T d$ a un angle de 180 car $\cos(180) = -1$. Ainsi, cette direction d correspond à

$$-\frac{\nabla f(x)}{\|\nabla f(x)\|}.$$

Point faible de la méthode du gradient

\Rightarrow Pour éviter les zigzags et accélérer la convergence, on peut avoir recours à l'un des procédés suivants :

— **diminuer le pas ne pas aller jusqu'en \tilde{x} :**

Polyak (1966) : effectuer des pas prédéterminés en imposant une suite (t^k) telle que $t^k \downarrow 0$ et $\sum_k t^k = +\infty \Rightarrow (x^k)$ tend vers \hat{x}

— **utiliser d'autres directions :**

Forsythe (1986), Luenberger (1973) : toutes les m itérations, au lieu de partir dans la direction de l'anti-gradient en x^k , "couper" en partant dans la direction $\delta = x^k - x^{k-m}$

— utiliser des directions "conjuguées"

2.3.2 La méthode de Newton

La méthode de Newton permet de construire un algorithme permettant de résoudre le système d'équations non-linéaires

$$g(x) = 0,$$

où $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est différentiable : on se donne $x_0 \in \mathbb{R}^n$ et on fait les itérations

$$x_{k+1} = x_k - Jg(x_k)^{-1}g(x_k), \quad (2.7)$$

où $Jg(x)$ est la dérivée (au jacobienne) de g au point x . L'application de cette méthode au problème d'optimisation

$$\min_{x \in \mathbb{R}^n} f(x), \quad (2.8)$$

consiste à l'utiliser pour résoudre le système d'optimalité du problème (2.8), c'est à dire que l'on pose $g(x) = \nabla f(x)$ dans (2.7) : on obtient les itérations

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k). \quad (2.9)$$

Algorithme de Newton

L'algorithme de Newton combine la stratégie de Newton de choix de la direction de descente avec, à chaque étape, une recherche du pas optimal :

Algorithme 3 *Newton($f, x_0, tolerance$)*

$$x \leftarrow x_0, \quad g \leftarrow \nabla f(x_0), \quad H \leftarrow \nabla^2 f(x_0)$$

Tant que : $g^T H^{-1} g \geq tolerance$

Calculer le pas optimal t dans la direction : $u = -H^{-1}g$

$$x \leftarrow x + tu, \quad g \leftarrow \nabla f(x), \quad H \leftarrow \nabla^2 f(x)$$

Retourner x

Convergence

Proposition :

Si le point initial x^0 est choisi dans un voisinage d'un minimum local où le hessien est défini positif, alors $\{x^k, k \geq 0\}$ converge vers ce minimum local.

1. Convergence rapide : quadratique (plus rapide que le gradient à pas optimal).
2. "dans un voisinage" : ... initialisation difficile.
3. Combiner les méthodes : gradient (pour s'approcher du voisinage) et Newton (pour accélérer la convergence une fois dans le voisinage).

La méthode de Newton est intéressante car sa convergence est quadratique au voisinage de la solution, c'est à dire que l'on a

$$\|x_{k+1} - \hat{x}\| \leq \gamma \|x_k - \hat{x}\|^2, \quad \gamma > 0,$$

mais la convergence n'est assurée que si x_0 est suffisamment proche de \hat{x} , ce qui en limite l'intérêt. Pour résoudre le problème de convergence locale de la méthode de Newton, on peut penser à lui ajouter une phase de recherche linéaire, dans la direction

$$d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k).$$

Cela est possible uniquement si d_k est une direction de descente en x_k , soit

$$\nabla f(x_k)^T d_k = -\nabla f(x_k)^T \nabla^2 f(x_k)^{-1} \nabla f(x_k) < 0,$$

ce qui sera le cas si $\nabla^2 f(x_k)$ est une matrice définie positive, ce qui n'est pas garanti (on sait tout au plus que $\nabla^2 f(\hat{x}) > 0$).

Chapitre 3

Modélisation Statistique

3.1 Corrélation et régression simple

3.1.1 Corrélation entre deux variables aléatoires.

Si deux variables X et Y sont indépendantes, alors

$$E(XY) = E(X)E(Y)$$

Dans le cas général, cette relation n'est pas vraie, on définit la covariance entre X et Y par

$$C(X, Y) = E(XY) - E(X)E(Y)$$

Cette valeur n'est pas "normalisée" (si X et Y ont des unités physiques par exemple ce nombre a une dimension), on définit donc la corrélation par

$$c(X, Y) = \frac{C(X, Y)}{\sigma(X)\sigma(Y)}$$

ce nombre est sans unité. On montre qu'il appartient à l'intervalle $[-1, 1]$. Si $Y = X$ alors $c = 1$, si $Y = -X$ alors $c = -1$. Si les variables sont indépendantes, alors $c = 0$. Mais la réciproque est fausse. On utilise toutefois ce calcul pour estimer si deux variables sont corrélées ou indépendantes. Plus précisément, si deux variables paraissent corrélées, on peut essayer d'exprimer l'une en fonction de l'autre, c'est ce qu'on va faire dans la section suivante.

3.1.2 Régression linéaire avec un modèle linéaire

Soit $(X_i)_{1 \leq i \leq n}$ et Y_i deux séries statistiques. On cherche à savoir s'il est raisonnable de prévoir Y_i en fonction de X_i , avec par exemple une relation linéaire $Y_i = a + bX_i$. Pour cela on minimise $f(a, b) = \sum (a + bX_i - Y_i)^2$ par rapport à a et b , ce qui donne les équations :

$$\begin{cases} \frac{\partial f}{\partial a} = 0 &= 2 \sum_{i=1}^n (a + bX_i - Y_i) \\ \frac{\partial f}{\partial b} = 0 &= 2 \sum_{i=1}^n X_i (a + bX_i - Y_i) \end{cases}$$

équivalentes à :

$$\begin{cases} a \sum_{i=1}^n + b \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i \\ a \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i Y_i \end{cases} \quad (3.1)$$

Si on note U la matrice de première colonne remplie de 1, et de deuxième colonne remplie par X_i (en ligne i),

$$U = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \dots & \dots \\ 1 & X_n \end{pmatrix}$$

et si on note U^* la transposée de U ,

$$U^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{pmatrix}$$

on observe que

$$U^*U = \begin{pmatrix} \sum_{i=1}^n 1 & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}$$

le système (3.1) devient

$$U^*U \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} = U^*Y, \quad \text{où } Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}$$

Donc la solution est donnée par

$$A = \begin{pmatrix} a \\ b \end{pmatrix} = \text{inv}(U^*U)(U^*Y)$$

On peut montrer que les coefficients de la droite sont donnés par :

$$b = \frac{E(XY) - E(X)E(Y)}{\sigma(X)^2}, \quad a = \bar{Y} - b\bar{X}$$

et

$$f(a, b) = n(1 - r^2)\sigma(Y)^2, \quad r = \frac{E(XY) - E(X)E(Y)}{\sigma(X)\sigma(Y)}$$

On montre que cette méthode se généralise (dans sa formulation matricielle), si on veut prévoir par exemple Y_i à partir de deux séries statistiques x_i et X_i avec une relation linéaire de la forme $Y_i = a + bx_i + cX_i$, alors (a, b, c) est obtenu en calculant $\text{inv}(U^*U)U^*Y$, où cette fois la matrice U possède 3 colonnes, la colonne de 1, la colonne des x_i et la colonne des X_i . Les calculs pratiques se font de manière identique.

Autres régressions linéaires

On peut aussi supposer que Y dépend de manière non linéaire de X , par exemple logarithmique, exponentielle, polynomiale, etc. mais toujours de manière linéaire des paramètres à déterminer et calculer ces paramètres en utilisant la même méthode (minimiser l'erreur quadratique). Ceci peut servir par exemple à modéliser la consommation d'une denrée non renouvelable (par exemple les combustibles fossiles) par une fonction logistique ou par une gaussienne.

Corrélations entre gaz à effet de serre et température

Si on fait une corrélation température-CO2 sur les 400 000 dernières années, on trouve un rapport $f(a,b)/n$ sur variance de 0.35, soit une qualité de corrélation de 0.65 (65% de la variance de la température est expliquée par le taux de CO2) et un coefficient d'augmentation de température de 9 degrés pour 100 ppm de CO2. On peut aussi faire une corrélation température en fonction du CO2 et du CH4, on trouve alors 6 degrés pour 100 ppmv de CO2 et 1.5 degré pour 100 ppbv de CH4 avec une qualité de corrélation de 0.68.

Bien entendu, une corrélation statistique ne suffit pas à conclure qu'une des variables dépend (au moins partiellement) de l'autre, mais ici la physique nous montre que l'augmentation du taux de CO2 augmente les radiations solaires capturées au niveau du sol, la corrélation permet de quantifier cet effet (calcul que l'on pourra comparer à celui obtenu par d'autres méthodes).

3.2 Régression linéaire multiple

Généralités

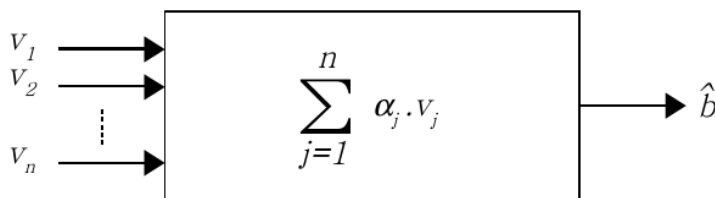
L'étude d'un phénomène peut, le plus souvent, être schématisé de la manière suivante : on s'intéresse à une grandeur b , que nous appellerons par la suite **réponse** ou **variable expliquée**, qui dépend d'un certain nombre de variables v_1, v_2, \dots, v_n que nous appellerons **facteurs** ou variables explicatives.

Notion de modèle

On cherche à mettre en évidence la liaison (relation fonctionnelle) pouvant exister entre la variable expliquée b et les variables explicatives v_1, v_2, \dots, v_n . On s'intéresse aux modèles dits linéaires, i.e. aux modèles du type :

$$b = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = \sum_{j=1}^n \alpha_j v_j$$

les α_i sont des réels appelés **coefficients du modèle**.



Pourquoi \hat{b} et pas b ?

On cherche à expliquer la variable b par n autres variables v_1, v_2, \dots, v_n mais on n'est pas certain que b ne dépend que de ces variables. Dans l'idéal $\hat{b} = b$ mais le plus souvent $\hat{b} \approx b$ avec $\hat{b} \neq b$.

3.2.1 Critère des moindres carrés- formulation

Critère

On cherche donc un modèle qui nous permet d'obtenir un \hat{b} le plus « proche » possible de b . Pour cela, on effectue m mesures ($m > n$) des variables v_1, v_2, \dots, v_n et de b . On cherche alors $\alpha_1, \alpha_2, \dots, \alpha_n$ tel que, pour $i = 1 \dots m$:

$$\hat{b}_i = \alpha_1 v_{i,1} + \alpha_2 v_{i,2} + \dots + \alpha_n v_{i,n} = \sum_{j=1}^n \alpha_j v_{i,j}$$

soit le plus « proche » possible de b_i .

En utilisant les notations matricielles, le système :

$$\begin{cases} \hat{b}_1 &= \alpha_1 v_{1,1} + \alpha_2 v_{1,2} + \cdots + \alpha_n v_{1,n} \\ \hat{b}_2 &= \alpha_1 v_{2,1} + \alpha_2 v_{2,2} + \cdots + \alpha_n v_{2,n} \\ \vdots & \\ \hat{b}_m &= \alpha_1 v_{m,1} + \alpha_2 v_{m,2} + \cdots + \alpha_n v_{m,n} \end{cases}$$

s'écrit :

$$\hat{b} = \underbrace{\begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n} \\ \vdots & \vdots & & \vdots \\ v_{m,1} & v_{m,2} & \cdots & v_{m,n} \end{pmatrix}}_A \times \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}}_x$$

Ainsi, on cherche $x = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ tel que Ax soit le plus « proche » possible de b .

On comprend alors que la notion de distance apparaît. On rappelle que la distance euclidienne usuelle est définie comme suite :

$$\forall x, y \in \mathbb{R}^m, d(x, y) = \sqrt{\|x - y\|^2}$$

où $\|\cdot\|$ est la norme euclidienne :

$$\forall x \in \mathbb{R}^m, \|x\|^2 = x^T x = \sum_{j=1}^m m x_j^2$$

On souhaite que $d(\hat{b} = Ax, b)$ soit minimale, ce qui s'écrit :

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

Chapitre 4

Optimisation avec contraintes linéaire

4.1 Introduction

Considérons le problème d'optimisation :

$$\begin{cases} \min f(\mathbf{x}) \\ \text{s.c. } A\mathbf{x} = b \\ \mathbf{x} \geq 0 \end{cases} \quad (4.1)$$

où

- A est une matrice de taille $m \times n$ (m lignes, n colonnes) de $\text{rang}(A) = m < n$.
- f est une fonction de \mathbb{R}^n à valeur dans \mathbb{R} est de classe C^1 .
- b un vecteur de dimension m .

:

Rappelons que, quelque soit le polyèdre initial, il est toujours possible de se ramener à ce type de contraintes.

4.2 Méthode du Frank-Wolfe

Considérons le problème de programmation linéaire suivant :

$$\begin{cases} \min \nabla f(\mathbf{x}_k)\mathbf{y} \\ \text{s.c. } A\mathbf{y} = b \\ \mathbf{y} \geq 0 \end{cases} \quad (4.2)$$

Supposons que l'on résolve ce problème en un point quelconque, qui n'est pas un minimum local du programme (4.1).

Le programme linéaire (4.2) possédera une solution y_k telle que

$$\nabla f(\mathbf{x}_k)(y_k - x_k) < 0,$$

c'est-à-dire, la direction $(y_k - x_k)$ constitue une direction de descente réalisable pour le programme (4.1). Nommant d cette direction, nous allons choisir

$$x_{k+1} = x_k + \theta d$$

de sorte que θ minimise f sur la ligne joignant x_k et y_k , tout en demeurant réalisable. Alors, $f(x_{k+1}) < f(x_k)$, et nous pouvons recommencer le processus.

Nous en déduisons la suite de quantités suivants :

$$\begin{aligned} y_k &= \arg \min_y \nabla f(x_k)y : Ay = b, y \geq 0; \\ d_k &= y_k - x_k; \\ \theta_k &= \arg \min_{0 \leq \theta \leq 1} f(x_k + \theta d_k); \\ x_{k+1} &= x_k + \theta_k d_k. \end{aligned}$$

Les observations précédentes montrent que la suite $f(x_k)$ est décroissante. Voir :

Jean-Pierre Dussault. Optimisation mathématique. Université de Sherbrooke, Canada. 2015.

L'algorithme de Frank-Wolfe utilise dans son calcul de direction la minimisation d'une fonction linéaire et donc résolue par un algorithme de programmation linéaire, i.e.

$$y_k = \arg \min_{y \in E} \nabla f(x_k)y$$

et $d = y_k - x_k$. Si on résout le programme linéaire avec l'algorithme du simplexe, la solution calculée sera toujours un point extrême de domaine réalisable E et donc on peut toujours choisir $\theta_{\max} = 1$. La condition d'optimalité est considérée satisfaite lorsque $\|\nabla f(x_k)(y_k - x_k)\|$ est assez petit.

4.3 Méthode du Gradient Réduit (RG)

4.3.1 Principe de la méthode RG

Méthode du Gradient Réduit

On définit une partition des variables en $B \subseteq \{1, \dots, n\}$ et $N = \{1, \dots, n\} - B$ telle que $|B| = m$ et A_B , la matrice formée des colonnes de A indexées dans B , soit inversible.

Quitte à permuter les colonnes, A se décompose en

$$(A_B \ A_N)$$

où A_N est la matrice formée des colonnes de A indexées dans N .

Quitte à permuter les lignes, tout x se décompose en

$$\begin{pmatrix} x_B \\ x_N \end{pmatrix}$$

où x_B est formé des composantes de x indexées dans B et x_N est formé des composantes de x indexées dans N .

Ainsi

$$Ax = b \Leftrightarrow (A_B \ A_N) \begin{pmatrix} x_B \\ x_N \end{pmatrix} = A_B x_B + A_N x_N = b$$

et si A_B est inversible alors on peut exprimer x_B en fonction de x_N par la relation

$$x_B = A_B^{-1}b - A_B^{-1}A_N x_N.$$

Exemple 11 Soit le système $\begin{cases} x_1 + 2x_2 + 3x_3 - x_4 = 6 \\ x_1 + 3x_2 + 2x_3 + 5x_4 = 4 \end{cases}$

Ici $A = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 1 & 3 & 2 & 5 \end{pmatrix}$

Choisissons $B = \{1, 3\}$, $N = \{2, 4\}$. Alors $A_B = \begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix}$, $A_N = \begin{pmatrix} 2 & -1 \\ 3 & 5 \end{pmatrix}$, $x_B \begin{pmatrix} x_1 \\ x_3 \end{pmatrix}$, $x_N \begin{pmatrix} x_2 \\ x_4 \end{pmatrix}$

$$\begin{aligned} (A_B \ A_N) \begin{pmatrix} x_B \\ x_N \end{pmatrix} &= \begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} + \begin{pmatrix} 2 & -1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} \\ &= \begin{pmatrix} x_1 + 3x_3 + 2x_2 - x_4 \\ x_1 + 2x_3 + 3x_2 + 5x_4 \end{pmatrix} \end{aligned}$$

En respectant cette partition, le problème (4.1) peut s'écrire comme suit :

$$\begin{cases} \min f(\mathbf{x}_B, \mathbf{x}_N) \\ \text{s.c. (i) } B\mathbf{x}_B + N\mathbf{x}_N = b, \\ \quad \quad \quad (ii) \ \mathbf{x}_B > 0, \\ \quad \quad \quad (iii) \ \mathbf{x}_N \geq 0. \end{cases} \quad (4.3)$$

On peut considérer \mathbf{x}_N comme un vecteur de variables indépendantes et \mathbf{x}_B comme un vecteur de variables dépendantes.

En effet, lorsque \mathbf{x}_N est fixé, \mathbf{x}_B est unique et est donné par (4.3.i). **Hypothèse :**

- On fait l'hypothèse que le problème (4.1) admet au moins une solution.
- On fera l'hypothèse de non-dégénérescence à savoir $\forall B$ t.q. A_B inversible, $A_B^{-1}b > 0$.

Comme pour la programmation linéaire, il est plus commode d'utiliser la transposée du gradient de f (vecteur ligne) que le gradient de f (vecteur colonne).

On note

$$\frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) = \nabla f^t$$

(On aura reconnu la matrice jacobienne de f).

De manière similaire aux notations précédentes on note

$$\frac{\partial f}{\partial x_N} = \left(\frac{\partial f}{\partial x_{j_1}}, \dots, \frac{\partial f}{\partial x_{j_{n-m}}} \right) \quad j_k \in N$$

et

$$\frac{\partial f}{\partial x_B} = \left(\frac{\partial f}{\partial x_{i_1}}, \dots, \frac{\partial f}{\partial x_{i_m}} \right) \quad i_k \in B.$$

Le principe de la méthode du gradient réduit est de considérer le problème (4.3) uniquement en terme de variables \mathbf{x}_N .

On résout localement le problème :

$$\min_{\mathbf{x}_N \in \mathbb{R}^m} g(\mathbf{x}_N); \quad \mathbf{x}_N \geq 0, \quad (4.4)$$

avec $g(\mathbf{x}_N) = f(\mathbf{x}_B(\mathbf{x}_N), \mathbf{x}_N)$.

Le gradient de g est appelé gradient réduit de f et sa transposée s'exprime par

$$\frac{\partial g}{\partial x_N} = \frac{\partial f}{\partial x_N} - \frac{\partial f}{\partial x_B} A_B^{-1} A_N. \quad (4.5)$$

Le gradient réduit $r(x)$ correspond au vecteur des coûts réduits en programmation linéaire :

$$r(x) = \frac{\partial f}{\partial x}(x) - \frac{\partial f}{\partial x_B}(x)A_B^{-1}A.$$

On a : $r_N(x) = \frac{\partial g}{\partial x_N}(x)$ et $r_B(x) = 0$.

Exemple 12 (suite) Soit le système

$$\begin{cases} x_1 + 2x_2 + 3x_3 - x_4 = 6 \\ x_1 + 3x_2 + 2x_3 + 5x_4 = 4 \end{cases}$$

et

$$f(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2.$$

Choisissons $B = \{1, 3\}$, $N = \{2, 4\}$.

On exprime x_1 en fonction de x_2, x_4 , on obtient

$$g(x_2, x_4) = (-5x_2 - 17x_4)^2 + x_2$$

et

$$\left(\frac{\partial g}{\partial x_2} \quad \frac{\partial g}{\partial x_4} \right) = (52x_2 + 170x_4 \quad 170x_2 + 578x_4).$$

:

Comme dans l'algorithme du simplexe, toute variable de base qui s'annule quitte la base pour être remplacée par une variable hors base. Le choix de la variable hors base n'est pas toujours unique ; en effet, même si la solution de base courante n'est pas dégénérée, il se peut qu'il y ait plusieurs variables hors base positives (ces variables sont appelées *superbasiques*). Il se peut aussi que le minimum de f dans la direction d soit atteint sans qu'aucune variable de base ne s'annule. On répète alors le processus sans changer de base.

Définition 10 On appelle algorithme du **simplexe convexe** la variante de l'algorithme du gradient réduit qui consiste à ne modifier qu'une seule variable hors base à chaque itération.

Principe de la méthode

Plutôt que de considérer f , fonction de n variables, on considère g fonction de $n - m$ variables. Outre le fait que le nombre de variables est ainsi moins important, l'intérêt est que ces $n - m$ variables sont soumises seulement aux contraintes relativement simple de non négativité.

On part d'un point x réalisable et on choisit une partition des variables (B, N) vérifiant les critères énoncés plus haut.

On va effectuer un déplacement dans la direction opposée au gradient réduit mais en prenant soin de ne pas rendre négatives les variables indexées dans N et nulles.

Définition 11 On définit la direction d , se décomposant en d_B et d_N , de la manière suivant :

$$— d_j = \begin{cases} 0 & \text{si } r_j(x) > 0 \text{ et } x_j = 0 \\ -r_j(x) & \text{sinon} \end{cases} \quad (j \in N)$$

— d_N étant défini, d_B est déterminé via les contraintes d'égalité par $d_B = -A_B^{-1}A_N d_N$.

4.3.2 Convergence de la méthode GR

Principe de la méthode

Si $d_N \neq 0$ le déplacement se fait jusqu'à atteindre le minimum de f dans la direction d mais en respectant les contraintes de non négativité des variables. On obtient un nouveau point x' . Si une de ses composantes dans B , disons s , s'annule on construit une nouvelle partition des variables

$$B' = B - \{s\} + \{r\} \text{ et } N' = N + \{s\} - \{r\}$$

où r est choisi de sorte que $A_{B'}$ soit inversible et que x'_r soit non nul puis on réitère. Cette opération, appelée changement de base, est nécessaire car sinon on risque de faire du surplace, la nullité de x'_s bloquant le déplacement suivant. Le fait de faire passer s dans N permet de contrôler la variation de la coordonnée s qui être que positive ou nulle par construction de la direction de déplacement.

Critère d'arrêt :

Si $d_N = 0$ les conditions de Kuhn-Tucker sont satisfaites et on arrête.

Theorem 15 *Si la méthode de gradient réduit construit une direction de déplacement nulle, i.e. $d_N = 0$, le point courant x vérifie les conditions de Kuhn-Tucker.*

Démonstration

Les conditions de Kuhn-Tucker s'écrivent :

$$\begin{cases} \frac{\partial f}{\partial x}(x) + \mu A - \lambda = 0 \\ \lambda \geq 0 \\ \lambda_i x_i = 0 \quad i = 1, \dots, n \end{cases}$$

Soient $\mu = -\frac{\partial f}{\partial x_B}(x)A_B^{-1}$, $\lambda_B = 0$, $\lambda_N = r_N(x)$. x, μ, λ vérifient-ils ces conditions? Pour la première égalité un simple calcul le montre. La deuxième condition ($\lambda \geq 0$) est vérifiée car $d_N = 0 \Rightarrow r_N(x) \geq 0$ d'après la définition de d_N . La troisième condition (complémentarité) est vérifiée car $d_j = 0 \Rightarrow (r_j > 0 \text{ et } x_j = 0) \text{ ou } (-r_j = 0) \quad \forall j \in N$ toujours par définition de d_N .

Algorithme du gradient réduit

On se donne une base initiale et un vecteur initial $\mathbf{x}^0 = (\mathbf{x}_B^0, \mathbf{x}_N^0)$ tel que $\mathbf{x}_B^0 > 0$, k_{\max} , on pose $k = 0$ et $d_N^0 \neq 0$.

Tant que ($d_N^k \neq 0$ et $k < k_{\max}$) **faire**

 Calcul du gradient réduit (4.5). **Pour** $i \in N$ **faire**

Si ($x_i^k = 0$ et $r_i(\mathbf{x}^k) \geq 0$) **Alors**

$d_i^k = 0$

Sinon

$d_i^k = -r_i(\mathbf{x}^k)$

Fin Si

Fin Pour

$$d_B^k = B^{-1} N d_N^k.$$

$$\mu_k = \arg \min_{0 \leq \mu \leq \mu_{\max}} f(\mathbf{x}^k + \mu d^k) \text{ avec } \mu_{\max} = \min_{j \in B, d_j^k < 0} \{-\frac{x_j^k}{d_j^k}\}.$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mu_k d^k.$$

Si ($\exists s \in B \ x_s = 0$) **Alors**

 soit $r \in N$ t.q. A'_B inversible et x_r maximum

$$B = B' \text{ et } N = N'$$

Fin Si

$$k = k + 1$$

Fait

Exemple 13

$$\begin{cases} \min f(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - 2x_1 - 3x_4 \\ \text{s.c.} & 2x_1 + x_2 + x_3 + 4x_4 = 7 \\ & x_1 + x_2 + 2x_3 + x_4 = 6 \\ & x \geq 0 \end{cases}$$

Soient $\mathbf{x}^0 = (2, 2, 1, 0)^t$, x_1 et x_2 les variables de base et $B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ la base.

$$\text{On a : } \nabla f(\mathbf{x}^0) = (2, 4, 2, -3)^T$$

$$r(\mathbf{x}^0) = (2, 4, 2, -3) - (2, 4) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 & 4 \\ 1 & 1 & 2 & 1 \end{pmatrix} = (0, 0, -8, -1),$$

par conséquent, $d_3 = 8$ et $d_4 = 1$.

On a :

$$d_B = B^{-1} N d_N = (5, -22)^T$$

$$\text{et } \mu \leq \mu_{\max} = \min_{j \in B, d_j^k < 0} \{-\frac{x_j^k}{d_j^k}\} = \frac{1}{11}.$$

On vérifie que le minimum de la fonction $f(\mathbf{x}^0 + \mu d)$ pour $\mu \in [0, \frac{1}{11}]$ est obtenu pour $\mu = \frac{1}{11}$.

La variable x_2 quitte alors la base pour y être remplacée par x_3 ou x_4 , au choix.

Remarque pratique :

au cours de l'algorithme, il faut effectuer le changement de base $B' = B - \{s\} + \{r\}$ $N' = N + \{s\} - \{r\}$

et il faut alors calculer $A_{B'}^{-1}$. Ce calcul peut être fait à moindre coût.

On a l'identité $A_{B'}^{-1} = (A_B^{-1}A_{B'})^{-1}A_B^{-1}$ étant obtenue à partir de A_B en remplaçant la colonne $A_{\{s\}}$

par la colonne $A_{\{r\}}$, on a $A_B^{-1}A_{B'} = \begin{pmatrix} 1 & 0 & y_1 & 0 \\ 0 & \ddots & \vdots & \\ \vdots & & p & \\ & & \vdots & \ddots \\ 0 & & y_m & 1 \end{pmatrix}$ où $A_B^{-1}A_{\{r\}} = \begin{pmatrix} y_1 \\ \vdots \\ p \\ \vdots \\ y_m \end{pmatrix}$

si $p \neq 0$ alors $(A_B^{-1}A_{B'})^{-1} = \begin{pmatrix} 1 & 0 & \frac{-y_1}{p} & 0 \\ 0 & \ddots & \vdots & \\ \vdots & & \frac{1}{p} & \\ & & \vdots & \ddots \\ 0 & & \frac{-y_m}{p} & 1 \end{pmatrix}$

On peut donc calculer $A_{B'}^{-1}$ en prémultipliant A_B^{-1} par une matrice simple à obtenir. p est appelé pivot de la colonne r .

Exemple 14 Reconsidérons le système $\begin{cases} x_1 + 2x_2 + 3x_3 - x_4 = 6 \\ x_1 + 3x_2 + 2x_3 + 5x_4 = 4 \end{cases}$

$$A = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 1 & 3 & 2 & 5 \end{pmatrix}$$

Choisissons $B = \{1, 3\}$, $N = \{2, 4\}$. Alors $A_B = \begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix}$, $A_B^{-1} = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}$

Effectuons le changement de base $B' = \{2, 3\}$, $N' = \{1, 4\}$ $A_B^{-1}A_{\{2\}} = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \end{pmatrix} p = 5 \neq 0$

et donc $A_{B'} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$ est inversible $(A_B^{-1}A_{B'})^{-1} = \begin{pmatrix} \frac{1}{5} & 0 \\ \frac{1}{5} & 1 \end{pmatrix}$ et $A_{B'}^{-1} = \begin{pmatrix} \frac{1}{5} & 0 \\ \frac{1}{5} & 1 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} \frac{-2}{5} & \frac{3}{5} \\ \frac{1}{5} & \frac{2}{5} \end{pmatrix}$

Sous l'hypothèse de non-dégénérescence, l'opération changement de base est toujours possible.

En effet $x_B + A_B^{-1}A_N x_N = A_B^{-1}b > 0$ et si on note p_j le pivot de la colonne j (relatif à x_s) on a :

$$s \in B \text{ t.q. } x_s = 0 \Rightarrow \sum_{j \in N} p_j x_j > 0 \Rightarrow \exists r \in N \text{ t.q. } p_r x_r \neq 0.$$

Exemple 15 (suite) Pour $B = \{1, 3\}$, on a $\begin{pmatrix} x_1 \\ x_3 \end{pmatrix} + A_B^{-1}A_N \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} = A_B^{-1}b$ soit $\begin{pmatrix} x_1 \\ x_3 \end{pmatrix} +$

$$\begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \end{pmatrix} \text{ soit encore } \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} + \begin{pmatrix} 5 & 17 \\ -1 & -6 \end{pmatrix} \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$x_3 = 0 \Rightarrow -x_2 - 6x_4 = 2$$

Les pivots des colonnes A_2 et A_4 (relatifs à x_3) sont -1 et -6 . S'ils étaient nuls on aurait $0 = 2$. On voit que la condition de non-dégénérescence n'est pas nécessaire puisqu'ici $A_B^{-1}b$ a une composante (celle associée à x_1) nulle et pourtant on peut effectuer les changements de base $B - \{1\} + \{2\}$ et $B - \{1\} + \{4\}$, les pivots (5 et 17) étant non nuls.

Exemple 16 (Récapitulatif) On considère le problème d'optimisation sous contraintes suivante :

$$\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.c.} & x_1 + x_2 \geq 1 \\ & x_1 \geq 0, x_2 \geq 0 \end{cases} \quad (P)$$

(P) revient à le pb d'optimisation avec contraintes d'égalité

$$\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.c.} & x_1 + x_2 + x_3 = 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{cases}$$

Choisissons le point initial $P_0 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$ et $B = \{3\}$, $N = \{1, 2\}$.

Itération 1.

On a $x_3 = -1 + x_1 + x_2$. On en déduit $g(x_1, x_2) = f(x_1, x_2, -1 + x_1 + x_2)$, $\frac{\partial g}{\partial x_N} = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}) = (2x_1, 2x_2)$ et

$r(P_0) = (4 \ 0)$. La direction de déplacement est $\begin{pmatrix} -4 \\ 0 \\ -4 \end{pmatrix}$.

$$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} -4 \\ 0 \\ -4 \end{pmatrix} \geq 0 \Rightarrow \begin{cases} \alpha \leq \frac{1}{2} \\ \alpha \leq \frac{1}{4} \end{cases} \Rightarrow \alpha_{\max} = \frac{1}{4}$$

Le minimum de $f(2 - 4\alpha, 0, 1 - 4\alpha)$ sous la contrainte $0 \leq \alpha \leq \frac{1}{4}$ est atteint en $\alpha_1 = \frac{1}{4}$.

Le nouveau point est $P_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Sa coordonnée 3 est nulle. Il faut effectuer un changement de base. L'indice 1 est l'indice dans N associé à la coordonnée de P_1 la plus grande : on forme la nouvelle décomposition $B = \{1\}$ et $N = \{2, 3\}$.

Itération 2.

On a $x_1 = 1 - x_2 + x_3$ d'où $g(x_2, x_3) = f(1 - x_2 + x_3, x_2, x_3)$, $\frac{\partial g}{\partial x_N} = (-\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_3}) = (-2x_1 + 2x_2 \ 2x_1)$

et $r(P_1) = (-2 \ 2)$. $d_N = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ et $d_B = -2$ donc la direction de déplacement est $\begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix}$.

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix} \geq 0 \Rightarrow \begin{cases} \alpha \leq \frac{1}{2} \\ \alpha \leq 0 \end{cases} \Rightarrow \alpha_{\max} = \frac{1}{2}$$

Le minimum de $f(1 - 2\alpha, 2\alpha, 0)$ sous la contrainte $0 \leq \alpha \leq \frac{1}{2}$ est atteint en $\alpha_2 = \frac{1}{4}$.

Le nouveau point est $P_2 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}$.

Itération 3.

Le gradient réduit en P_2 est $r(P_2) = (0 \ 1)$ donc $d_N = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. On arrête.

Vérifions que les conditions de Kuhn-Tucker sont satisfaites : ici $A = (1 \ 1 \ -1)$

soient $\mu = -\frac{\partial f}{\partial x_1}(\frac{1}{2}, \frac{1}{2}, 0)A_1^{-1} = -2 \times \frac{1}{2} = -1$ (car $A_1 = 1$), $\lambda_1 = 0$, $(\lambda_2 \ \lambda_3) = r(P_2) = (0 \ 1)$

$\frac{\partial f}{\partial x}(\frac{1}{2}, \frac{1}{2}, 0) + \mu A - \lambda = 0 \Leftrightarrow (2 \times \frac{1}{2} \ 2 \times \frac{1}{2} \ 0) - (1 \ 1 \ -1) - (0 \ 0 \ 0) = 0$, donc vérifié.

On aura remarqué que l'on a une écriture "transposée" des équations de Kuhn-Tucker.

$\lambda \geq 0$ et les conditions de complémentarité $\lambda_1 \times \frac{1}{2} = 0$, $\lambda_2 \times \frac{1}{2} = 0$, $\lambda_3 \times 0 = 0$ sont vérifiées.

Remarque 3 en partant du même point on obtient des cheminements différents selon le choix du B initial.

Si l'on choisit $B = \{1\}$ et $N = \{2, 3\}$, on passe par les 3 points $P_1 = \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}$, $P_2 = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$, $P_3 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ sans

changer de base B .

4.4 Autre méthodes

4.4.1 Méthode du gradient projeté

Une idée assez naturelle, pour adapter les méthodes d'optimisation sans contraintes aux problèmes avec contraintes, est de projeter à chaque étape le déplacement sur la frontière du domaine, afin de s'assurer que le nouveau point obtenu appartienne à l'ensemble des solutions réalisables. L'ensemble des contraintes linéaires forme alors un domaine de points réalisables qui s'écrit sous la forme

$$S = \{x \in \mathbb{R}^n : \sum_{j=1}^n a_{ij}x_j - b_i \leq 0, \quad i = 1, 2, \dots, m\},$$

où les a_{ij} ont été normalisés afin que

$$\|a_i\|^2 = \sum_{j=1}^n (a_{ij})^2 = 1 \quad i = 1, 2, \dots, m.$$

Soit A_q la sous matrice de A constituée par les lignes des contraintes saturées en un point x de S ,

$$A_q \equiv [a_1 \ a_2 \ \dots a_q]^t.$$

Puisque les vecteur a_i sont supposés linéairement indépendants, la matrice symétrique $A_q A_q^t$ est non singulière. Par conséquent, son inverse $(A_q A_q^t)^{-1}$ existe.

On définit une matrice symétrique $n \times n$ par

$$P_q = I - A_q^t (A_q A_q^t)^{-1} A_q.$$

En modifiant ainsi le moins possible la méthode d'origine, on peut espérer conserver son efficacité. De nombreux algorithmes basés sur ce schéma général ont été proposés. Un des premiers algorithmes bâtis suivant ce principe est la méthode du gradient projeté de Rosen qui consiste à projeter le gradient sur l'intersection des contraintes saturées.

La méthode du gradient projeté consiste à prendre comme direction de déplacement la projection du vecteur gradient sur la frontière du domaine et dont l'algorithme s'écrit sous la forme générale suivante :

- (i) On part de l'initialisation : $x^0 \in S$.
- (ii) On génère la suite de vecteurs $\{x^k\}_{k \geq 1} \in S$ définis par :

$$\forall k \geq 0 \quad x^{k+1} = x^k - \omega_k P_q \nabla f(x^k), \quad (4.6)$$

où ω_k est le pas de déplacement.

4.4.2 Stratégie de pénalisation des contraintes

Soit le programme mathématique

$$\begin{cases} \text{minimiser la fonction : } f(x) \\ \text{sous les contraintes : } g_i(x) \leq b_i \quad i = 1, \dots, m. \end{cases}$$

Une façon simple de se débarrasser des contraintes est d'intégrer dans la fonction objectif une fonction qui pénalise les vecteurs qui violent les contraintes, on a

$$\min_{x \in \mathbb{R}^n} f(x) - \sum_{i=1}^m p_i(x)$$

où $p_i(\mathbf{x})$ prend une valeur élevée lorsque $g_i(\mathbf{x})$ est strictement positif, ce qui décourage l'adoption de ces vecteurs \mathbf{x} . On peut par exemple poser

$$p_i(\mathbf{x}) = \lambda \max\{0, g_i(\mathbf{x}) - b_i\}$$

où λ est un coefficient positif et grand. Si l'on souhaite que la fonction objectif soit différentiable, on peut élever l'expression précédente au carré. Pour résoudre le problème pénalisé, il suffit alors de trouver les points où le gradient de la fonction objectif s'annule.

Principe de pénalisation

Optimisation sous contraintes d'égalité

Le problème formel :

$$\begin{cases} C = \{x \in V \text{ tel que } F_j(x) = 0, j = 1, \dots, p\} \\ \forall x \in C \quad F(\bar{x}) \leq F(x) \end{cases} \quad (4.7)$$

Principe :

On supprime les contraintes mais on punit leur non-respect

Pénalisation

Définition 12 (Problème pénalisé)

$$\begin{cases} F\epsilon(x) = F(x) + \sum_i \frac{F_i(x)^2}{2\epsilon} \\ \forall x \in V \quad F\epsilon(x\epsilon) \leq F\epsilon(x) \end{cases} \quad (4.8)$$

Remarque 4 ϵ doit être petit, mais : problème des erreurs de troncature.

Interprétation

Un problème de mécanique

Équilibre d'un treillis :

Un treillis de barres, avec contact unilatéral.
La position d'équilibre pénalisée minimise

$$\min_x \frac{1}{2} \langle \mathbf{K}x, x \rangle - \langle b, x \rangle + \frac{F_1(x)^2}{2\epsilon}$$

qui est l'énergie potentielle + l'énergie interne du ressort de liaison.

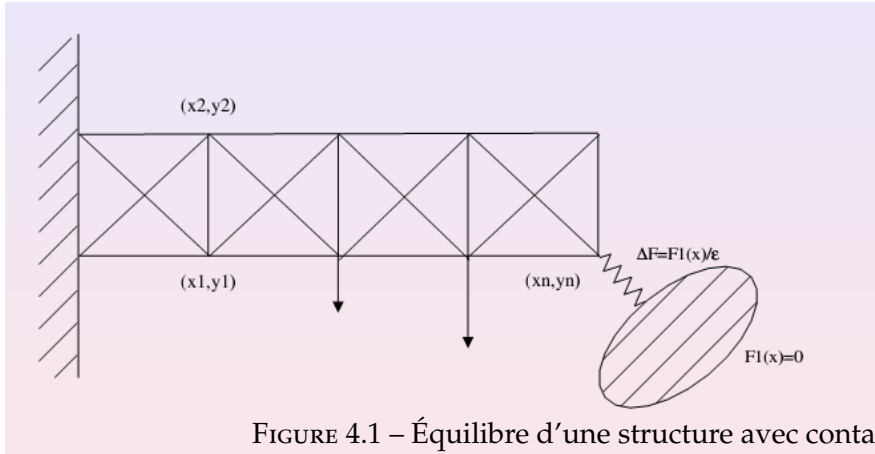


FIGURE 4.1 – Équilibre d'une structure avec contact unilatéral

Intérêt Pénalisation

Theorem 16 *La solution du problème pénalisé tend vers la solution du problème avec contraintes quand $\epsilon \rightarrow 0$. L'erreur est en $\sqrt{\epsilon}$*

Difficulté :

Le problème pénalisé est mal conditionné, la précision est médiocre