REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

Université Ibn Khaldoun- Tiaret
FACULTÉ DES MATHEMATIQUES ET INFORMATIQUE
DÉPARTEMENT DES MATHEMATIQUES

MEMOIRE MASTER

Présentée par:

Boussoura Saad

Khedir Tayeb

Spécialité : *Mathématiques*

Option : *Analyse fonctionnelle et équation différentielle*

Intitulé

Méthode de la plus forte pente

Soutenue le : 20/06/ 2022 devant le jury composé de :

Président : *Laarabi Abderrahmene* *MCA. Univ. Ibn Khaldoun Tiaret*

Examinateurs : *Maazouz Kada* *MCA. Univ. Ibn Khaldoun Tiaret*

Encadreur : *Rezzoug Nadir* *MCB. Univ. Ibn Khaldoun Tiaret*

2021-2022

Remerciements

*Avant tout, on remercie **ALLAH** Tout-Puissant de nous avoir donné le courage et la volonté d'accomplir ce travail.*

*Nous tenons à exprimer toute notre reconnaissance à notre Encadreur **M. N.Rezzoug**, pour avoir dirigé ce travail avec obnégation et disponibilité. Ses conseils nous ont été d'un grand apport pour accomplir de ce mémoire.*

*Nous adressons mes plus vifs remerciements à **M. A.Larabi**, Doyen de la Faculté des mathématiques et d'informatique, pour nous avoir fait l'honneur de présider le jury de soutenance de notre mémoire.*

*Sincères remerciements vont également à **M. K.Maazouz**, qui a eu l'honneur de participer au jury de soutenance de notre mémoire en tant qu'examinateur.*

Nous ne pouvons oublier de remercier les parents pour leur soutien, leur aide et patience, tout au long de nos études.

Enfin, nous adressons nos remerciements à tous nos professeurs, collègues et tous ceux qui nous ont encouragés, à faire ce travail.

Dédicaces

je dédie ce travail :

À

Mes chèrs parents pour leur soutien, leur patience
et leurs encouragements durent mon parcours universitaire.

À

Mes fréres, ainsi ma famille BOUSSOURA et MERZOUGUI.

À

Mes Amis.

À

Tous mes professeurs
qu'il soient du primaire, du secondaire ou du supérieur.

SAAD

Dédicaces

C'est avec grand plaisir que je dédie ce modeste travail :

À

Plus chére de ma vie, ma Mère.

À

Celui qui m'a fait de moi un homme, mon Père.

À

Mes chers fréres et soeurs.

À

Tous les membres de ma famille et toute personne qui porte le nom KHEDIR et TOUIGUI.

À

Tous mes amis de promotion de 2ème année master mathématiques

À

Tous les professeurs

qu'il soient du primaire, du secondaire ou du supérieur.

TAYEB

Table des matières

Notations

$\mathbb{R}$	L'ensemble des nombres réels.
$\mathbb{R}^+$	L'ensemble des nombres réels positifs.
$\mathbb{C}$	L'ensemble des nombres complexes.
$\Vert . \Vert$	La norme.
$\langle \cdot, \cdot \rangle$	Produit scalaire.
A^T	Transposé d'une matrice A.
$\frac{\partial f}{\partial x}$	La dérivée de f suivant la variable x.
$\frac{\partial f}{\partial x \partial y}$	La dérivée de $\frac{\partial f}{\partial y}$ suivant la variable x.
$f(x) = o(g(x))$	$\lim \frac{f(x)}{g(x)} = 0$.
$\nabla f(x)$	Gradient de la fonction f.
$H(x)$	Matrice hessienne.
$\widehat{x}$	Solution optimale.
$(\mathcal{P})$	Problème d'optimisation.

INTRODUCTION

L'optimisation consiste en la recherche du minimum (ou du maximum) d'une certaine quantité, appelée cout ou objective. Si on suppose que le cout dépend de n variables réeles rassemblées en un vecteur $x = (x_1, ..., x_n) \in \mathbb{R}^n$ et que fournissent une valeur $J(x)$ où J est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, en général, les variables $x_1, ...x_n$ ne sont pas autorisées à prendre n'importe quelles valeurs, mais devront satisfaire des contraintes que l'on représent par un sous ensemble $K \subset \mathbb{R}^n$. On écrire un problème d'optimisation sous la forme générale

$$(\mathcal{P}) \qquad \min_{x \in K} J(x).$$

On dit que le problème $(\mathcal{P})$ admet une solution s'il existe un choix de variables $x_0 \in K$ tel que

$$\forall x \in K, \quad J(x_0) \leq J(x),$$

on dit alors que x_0 est un minimiseur (ou point minimum) de J sur K, et que $J(x_0)$ est un minimum de J sur K.

Pour résoudre le problème $(\mathcal{P})$ il y a plusieurs méthodes. Dans ce mémoire, on propose et étudie la méthode de la plus forte pente.

Ce travail est réparti en quatre chapitres. Le premier chapitre est consacré à quelques notions de base comme différentiabilité, gradient, matrice hessienne et fonctions de classe C^1. Un rappel sur les formules de Taylor est aussi donné dans ce chapitre.

Dans le deuxième chapitre, on introduit les ensembles et les fonctions convexes, théoreme de séparation et aprés les conditions d'optimalité d'un probleme de minimisation et quelques résultats préliminaires.

Dans le troisième chapitre, on étudie les recherches linéaires exactes et inéxactes, en particulier les recherches d'Armijo, de Goldstein et de Wolf.

Dans le derniér chapitre, on traite la méthode de la plus forte pente. On commance d'abord par définitions, théorèmes et proprietés concernant cette méthode. En fin, on propose une synthése sur les resultas de convergences.

Chapitre 1

Préliminaires

1.1 Différentiabilité, dérivée partielle, gradient, dérivée directionnelle

Définition 1.1.1. *1) Soit une fonction $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $h \in \mathbb{R}^n$. On dit que f est différentiable en $a \in U$ s'il existe une application L linéaire continue de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que*

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - L(h)}{\|h\|} = 0. \tag{1.1}$$

2) Lorsque l'application L existe, elle s'appelle la différentielle de f en a, on la note $Df(a)$, ainsi (1.1) devient

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - Df(a)(h)}{\|h\|} = 0.$$

3) Si f est différentiable en tout point de U, on dit que f est différentiable sur U.

Remarque 1.1.1. *La différentiabilité de f en a est équivalente à l'existence d'une forme linéaire L sur $\mathbb{R}^n$ telle que*

$$f(a+h) = f(a) + Lf(a)(h) + o(h). \tag{1.2}$$

Proposition 1.1.1. *Soient U un ouvert et $f : U \subset \mathbb{R}^n \to \mathbb{R}$.*

Si $f \in C^1(U)$ alors f est différentiable sur U.

Définition 1.1.2. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue. La fonction notée $\nabla_i f(x) : \mathbb{R}^n \to \mathbb{R}$, également notée $\frac{\partial f}{\partial x_i}$ est applée i^{me} dérivée partielle de f et est définie par*

$$\lim_{a \to 0} \frac{f(x_1, \ldots, x_i + a, \ldots, x_n) - f(x_1, ..., x_i, ..., x_n)}{a},$$

cette limite peut ne pas existe.

Si les dérivées partielles $\partial f / \partial x_i$ existent pour tout i, le gradient de f est défini de la façon suivante.

Définition 1.1.3. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable. La fonction notée $\nabla f(x) : \mathbb{R}^n \to \mathbb{R}$ est appelée le gradient de f au point $x = (x_1, ..., x_n)$et est défini par*

$$(\nabla f(x))^T = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

Le gradient jouera un role essentiel dans le developpement et l'analyse des algorithmes d'optimisation.

Exemple 1.1.1. *Soit $f(x_1, x_2, x_3) = e^{-3x_1} + 2x_1^2 x_3 - 4x_1 x_2 x_3$. Le gradient de f est donné par*

$$\nabla f(x_1, x_2, x_3) = \left(\begin{array}{l} -3e^{-3x_1} + 4x_1 x_3 - 4x_2 x_3 \\ -4x_1 x_3 \\ 2x_1^2 - x_1 x_2 \end{array} \right).$$

Remarque 1.1.2. *Soit f une fonction dérivable suivant $x_i, i = 1, ..., n$ et soit $e_1, e_2, ...e_n$ les élément de la base canonique de $\mathbb{R}^n$ alors*

$$\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x).$$

Où

$$(e_i)_j = \delta_{ij} = \begin{cases} 0 & Si\ j \neq i \\ 1 & Si\ j = i \end{cases} \qquad \forall i, j = 1, 2, ..., n,$$

(symboles de Kronecker)

Remarque 1.1.3. *Nous rappellons aussi la formule*

$$\frac{\partial f}{\partial h}(x) = \langle \nabla f(x), h \rangle, \quad \forall x \in \Omega, \quad \forall h \in \mathbb{R}^n.$$

Proposition 1.1.2. (Gradient de la composée)

Soit les deux ouverts $\Omega \subset \mathbb{R}^n$ et $U \subset \mathbb{R}$ et deux fonctions $f : \Omega \to \mathbb{R}$ et $g : U \to \mathbb{R}$ avec en plus $f(\Omega) \subset U$ (on peut alors definir $g \circ f : \Omega \to \mathbb{R}$). Supposons que f et g sont de classe C^1. Alors g o f est aussi de classe C^1 avec

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

Définition 1.1.4. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue. soit $x \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$. La dérivée directionnelle de f en x dans la direction d est donnée par*

$$\delta f(x, d) = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha} \text{ (si la limite existe).}$$

De plus, lorsque le gradient èxiste, la dérivée directionnelle est le produit scalaire entre le gradient de f et la direction d, i.e.

$$\delta f(x, d) = \nabla^T f(x) d.$$

Exemple 1.1.2. *Soit $f(x_1, x_2, x_3) = e^{x_1} + {x_1}^2 x_3 - x_1 x_2 x_3$ et soit*

$$d = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}.$$

La dérivée directionnelle de f dans la direction d est

$(d_1 d_2 d_3)\nabla f(x_1 x_2 x_3) = d_1(e^{x_1} + 2x_1x_3 - x_2x_3) - d_2x_1x_3 + d_3(x_1^2 - x_1x_2)$

ou $\nabla f(x_1 x_2 x_3)$ est donnée par

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} e^{x_1} + 2x_1x_3 - x_2x_3 \\ -x_1x_3 \\ x_1^2 - x_1x_2 \end{pmatrix}.$$

1.2 Taux d'accroissement

Définition 1.2.1. *Soient f une fonction définie sur un intervalle I et a, $a+h \in I$.*
On appelle taux d'accroissement de f entre a et $a+h$ le quotient

$$T_a(h) = \frac{f(a+h) - f(a)}{h}$$

Remarque 1.2.1. *Le taux d'accroissement est maximal dans la direction du gradient*

1.3 Matrice hessienne

Définition 1.3.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable. La matrice hessiene notée $H(x)$ est définie par*

$$H(x) = \nabla\left(\nabla^T f\right)(x) = \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)(x), i = 1, \ldots, n, j = 1, \ldots, n.$$

Alors

$$H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

Théorème 1.3.1. *(théorème de schwarz). Soit $\Omega \subset \mathbb{R}^n$ un ouvert.*
Si $f \in C^2(\Omega)$ alors $\nabla^2 f(x)$ est une matrice symmétrique.

Exemple 1.3.1. *Soit $f(x_1, x_2, x_3) = e^{-3x_1} + 2x_1^2 x_3 - 4x_1 x_2 x_3$.*
Alors

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} -3e^{-3x_1} + 4x_1 x_3 - 4x_2 x_3 \\ -4x_1 x_3 \\ 2x_1^2 - 4x_1 x_2 \end{pmatrix}.$$

et

$$H(x) = \begin{pmatrix} 9e^{-3x_1} + 4x_3 & -4x_3 & 4x_1 - 4x_2 \\ -4x_3 & 0 & -4x_1 \\ 4x_1 - 4x_2 & -4x_1 & 0 \end{pmatrix}.$$

Définition 1.3.2. *La matrice H est dite semi-définie positive si*

$$\forall x \in \mathbb{R}^n, x^T Hx \geq 0.$$

H est dite définie positive si

$$\forall x \in \mathbb{R}^n, x \neq 0, x^T Hx > 0.$$

1.4 Formules de Taylor

La formule de Taylor est un outil important en convexité. Nous la rappelons dans le cas général.

Soit $\Omega \subset \mathbb{R}^n$ ouvert, $f : \Omega \to \mathbb{R}$, a $\in \Omega$ et $h \in \mathbb{R}^n$ tels que [a, a+h]. Alors

1. Si $f \in C^1(\Omega)$ alors

i) Formules de Taylor avec reste intégral à l'ordre 1

$$f(a+h) = f(a) + \int_0^1 (\nabla f(a+th), h)dt.$$

ii) Formules de Taylor Maclaurin à l'ordre 1

$$f(a+h) = f(a) + < \nabla f(a+h), h > .$$

iii) Formules de Taylor-Young à l'ordre 1

$$f(a+h) = f(a) + < \nabla f(a), h > + o(\|h\|).$$

2. Si $f \in C^2(\Omega)$ alors

i) Formules de Taylor avec reste intégrale à l'ordre 2

$$f(a+h) = f(a) + < \nabla f(a), h > + \int_0^1 (1-t) < \nabla^2 f(a+th)h, h > dt.$$

ii) Formules de Taylor-Maclaurin à l'ordre 2

$$f(a+h) = f(a) + < \nabla f(a), h > + \frac{1}{2} < \nabla^2 f(a+\theta h)h, h > \text{avec}\, 0 < \theta < 1.$$

iii) Formules de Taylor-Young à l'ordre 2

$$f(a+h) = f(a) + < \nabla f(a), h > + \frac{1}{2} < \nabla^2 f(a)h, h > + o(\|h\|^2).$$

Remarque 1.4.1. *Dans la proposition précédente la notation $o(\|h\|^k)$ pour $k \in \mathbb{N}^*$ signifie une expression qui tend vers 0 plus vite que $\|h\|^k$ (c'est à dire si on la divise par $\|h\|^k$, le résultat tend vers 0 quand $\|h\|$ tend vers 0).*

Théorème 1.4.1. *Soit $f : U \subset \mathbb{R} \to \mathbb{R}$ de classe $\mathrm{C}^{n+1}(\Omega)$ et $[a, a+h] \subset U$. Alors*

$$f(a+h) = f(a) + f'(a)h + ... + \frac{1}{n!} f^{(n)}(a)h^n + \int_0^1 \frac{(1-t)^n}{n!} f^{(n+1)}(a+th)h^{n+1} dt.$$

Chapitre 2

Ensembles convexes et théorème de séparation

2.1 Ensembles convexes

Définition 2.1.1. *1) Soit S un sous ensemble de $\mathbb{R}^n$, on dit que S est convexe si et seulement si*

$$\forall x_1, x_2 \in S, \forall \lambda \in [0,1], \lambda x_1 + (1-\lambda)x_2 \in S.$$

2) Soient $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ et $\lambda_j \in \mathbb{R}$ tels que

$$\lambda_j \geq 0 \ et \ \sum_{j=1}^{k} \lambda_j = 1.$$

Toute expression de la forme suivante $\sum_{j=1}^{k} \lambda_j x_j$ s'appelle combinaison convexe des points x_j ou barycentre.

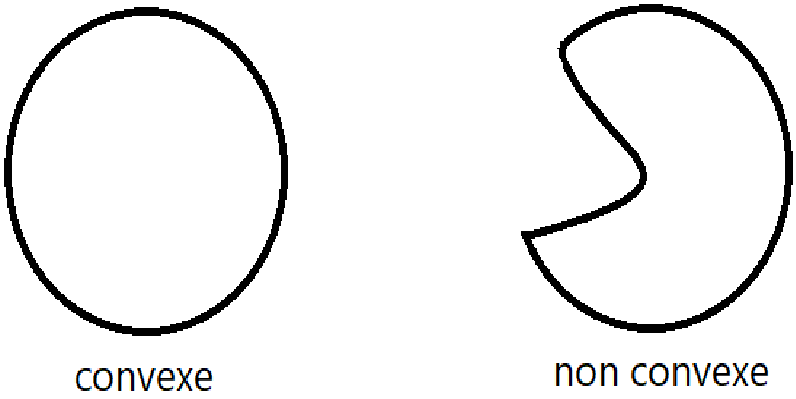


FIGURE 2.1 – ensemble convexe et non convexe

Définition 2.1.2. *Soit $S \subset \mathbb{R}^n$. On appelle envelloppe convexe de S et on le note $H(S)$, l'ensemble de toutes les combinaisons convexes de S, en d'autre termes*

$$x \in H(S) \Leftrightarrow \exists\, x_1, x_2, \dots, x_k \in S \text{ et } \lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{R}_+$$

tels que $\sum_{j=1}^{k} \lambda_j = 1$ *et*

$$x = \sum_{j=1}^{k} \lambda_j x_j \tag{2.1}$$

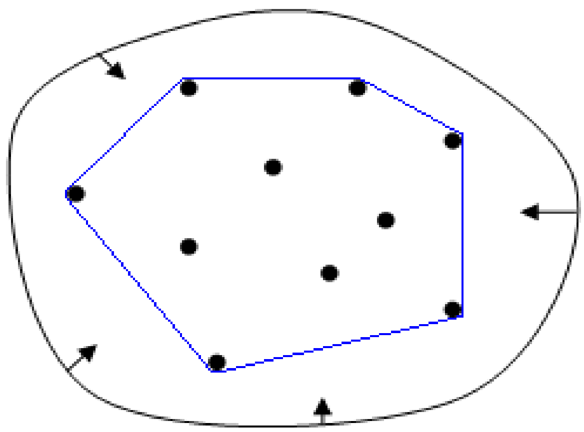


FIGURE 2.2 – envelloppe convexe

2.1.1 Propriétés de l'enveloppe convexe

Proposition 2.1.1. *Soit $S \subset \mathbb{R}^n$ quelconque, alors $H(S)$ est un convexe, c'est aussi le plus petit convexe qui contient S.*

preuve. Soient $x \in H(S)$ et $y \in H(S)$. Alors il existe $\lambda_1, \lambda_2, \ldots, \lambda_p \in \mathbb{R}, \mu_1, \mu_2, \ldots, \mu_l \in \mathbb{R}$, $x_1, \ldots, x_p \in S$ et $y_1, y_2, \ldots, y_l \in S$
tels que

$$\lambda_j \geqslant 0, \mu_j \geqslant 0, \sum_{j=1}^{p} \lambda_j = 1, \sum_{j=1}^{l} \mu_j = 1$$

et

$$x = \lambda_1 x_1 + \ldots + \lambda_p x_p, \quad y = \mu_1 y_1 + \ldots + \mu_i y_i.$$

Soit $\beta \in [0, 1]$, alors

$$\beta x + (1-\beta)y = \beta\lambda_1 x_1 + \ldots + \beta\lambda_p x_p + (1-\beta)\mu_1 y_1 + \ldots + (1-\beta)\mu_l y_l$$

avec

$$\beta(\lambda_1 + \ldots + \lambda_p) + (1-\beta)(\mu_1 + \ldots + \mu_l) = \beta + (1-\beta) = 1.$$

Donc $\beta x + (1-\beta)y$ est un barycentre des points $x_1, \ldots, x_p, y_1, y_2, \ldots, y_t$ appartenant à S. $\beta x + (1-\beta)y$ appartient donc à $H(S)$ et par suite $H(S)$ est convexe. Notons par ρ le plus petit convexe qui contient l'ensemble S. On doit démonter que $H(S) = \rho$.
1. $\rho \subset H(S)$ car ρ est le plus petit convexe qui contient S et $H(S)$ est un convexe contenant S aussi si $x \in S, x = 1.x \in H(S)$, donc $S \subset H(S)$.
2. $H(S) \subset \rho$. En effet. On a $S \subset \rho \Rightarrow H(S) \subset H(\rho)$ et $H(\rho) = \rho$ car ρ est convexe. □

Proposition 2.1.2. *Soit $S \subset \mathbb{R}^n$. L'intersection de tous les convexes contenant S est le plus petit convexe qui contient S.*

preuve. Notons par β une telle intersection ; $\beta = \bigcap_{i \in I} S_i, S_i$ convexe contenant S.
On a
a) β est un ensemble convexe .
b) $S \subset \beta$.

c) Soit S^* un convexe quelconque qui contient S. Alors $\exists i_0 \in I$ tel que $S^* = S_{i_0}$, donc $\beta \subset S_{i_0}$. $\square$

2.1.2 Intérieur et fermeture d'un convexe

Proposition 2.1.3. *Soit $S \subset \mathbb{R}^n$ un convexe non vide et $x_1 \in \bar{S}$ et $x_2 \in \mathrm{int}(S)$, alors*

$$\forall \lambda \in [0,1], \quad \lambda x_1 + (1-\lambda)x_2 \in \mathrm{int}(S).$$

preuve. $x_2 \in \mathrm{int}(S) \Rightarrow \exists \varepsilon > 0$ tel que, $B_0(x_2, \varepsilon) \subset S$, ou en d'autre terme

$$\{z : \|z - x_2\| < \varepsilon\} \subset S.$$

Considérons

$$y = \lambda x_1 + (1-\lambda)x_2 \Rightarrow x_2 = \frac{y - \lambda x_1}{1-\lambda}.$$

Montrons qu'il existe $\varepsilon_1 > 0$ tel que $B_0(y, \varepsilon_1) \subset S$. En effet choisissons $\varepsilon_1 = (1-\lambda)\varepsilon$ et prouvons que

$$\{z : \|z - y\| < \underbrace{(1-\lambda)\varepsilon}_{\varepsilon_1}\} \subset S.$$

En effet considérons z tel que $\|z - y\| < (1-\lambda)\varepsilon$

$$x_1 \in \bar{S} \Rightarrow \left\{x : \|x - x_1\| < \frac{(1-\lambda)\varepsilon - \|z-y\|}{\lambda}\right\} \cap S \neq 0.$$

Plus particulièrement, il existe $z_1 \in S$ tel que

$$\|z_1 - x_1\| < \frac{(1-\lambda)\varepsilon - \|z-y\|}{\lambda}.$$

Soit z_2 tel que

$$z_2 = \frac{z - \lambda z_1}{1-\lambda} \Rightarrow z = \lambda z_1 + (1-\lambda)z_2.$$

Démontrons que

$$z_2 \in B_0(x_2, \varepsilon)$$

ou encore $\|z_2 - x_2\| < \varepsilon$. En effet

$$\begin{aligned}
\|z_2 - x_2\| &= \left\| \frac{z - \lambda z_1}{1-\lambda} - x_2 \right\| \\
&= \left\| \frac{z - \lambda z_1 - (y - \lambda x_1)}{1-\lambda} \right\| \\
&\leq \frac{1}{1-\lambda} \left(\|z - y\| + \lambda \|x_1 - z_1\| \right) \\
&< \frac{1}{1-\lambda} \left(\|z - y\| + \frac{(1-\lambda)\varepsilon - \|z-y\|}{\lambda} \right). \\
&< \varepsilon.
\end{aligned}$$

Etant donné que ε a été choisi de sorte que $B_0(x_2, \varepsilon) \subset S$ et puisque $z_1, z_2 \in S$ et S est convexe, alors

$$\lambda z_1 + (1-\lambda) z_2 = z \in S.$$

Donc

$$\exists \varepsilon_1 > 0 : B_0(y, \varepsilon_1) \subset S.$$

Ce qui veut dire exactement que $y \in int(S)$. □

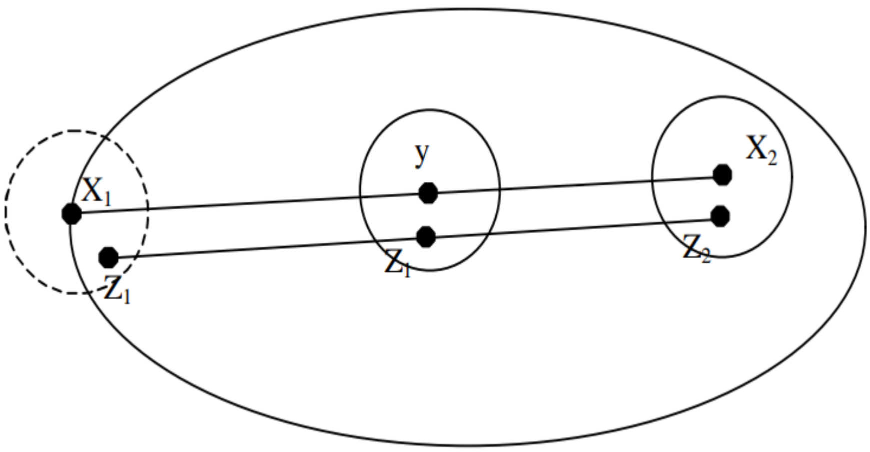


FIGURE 2.3 – Un ségement joignant un point interieur avec un point frontière de S.

Corollaire 2.1.1 Soit $S \subset \mathbb{R}^n$ tel que S est convexe et int $(S) \neq \emptyset$. Alors int(S) est convexe.

preuve. Soit $x_1 \in \text{int}(S)$, alors $x_1 \in \bar{S}$. Considérons un autre point $x_2 \in \text{int}(S)$. D'après la proposition 2.1.3 on a

$$\lambda x_1 + (1-\lambda)x_2 \in \text{int}(S).$$

Ce qui implique que $\text{int}(S)$ est convexe. □

Corollaire 2.1.2 Soit $S \subset \mathbb{R}^n$, S convexe et $\text{int}(S) \neq 0$, alors $\bar{S}$ est convexe.

preuve. Soient $x_1, x_2 \in \bar{S}$ et $z \in \text{int}(S)$, d'après la proposition 2.1.3

$$\lambda x_2 + (1-\lambda)z \in int(S).$$

Appliquons encore une fois la proposition 2.1.3 au point $x_1 \in \bar{S}$ et au point $\lambda x_2 + (1-\lambda)z \in \text{int}(S)$, on aura pour $\mu \in]0,1[$

$$\mu x_1 + (1-\mu)\left[\lambda x_2 + (1-\lambda)z\right] \in \text{int}(S) \subset S.$$

Par passage à la limite quand $\lambda \to 1$, on trouve

$$\mu x_1 + (1-\mu)x_2 \in \bar{S}.$$

Ceci imlique que $\bar{S}$ est convexe. □

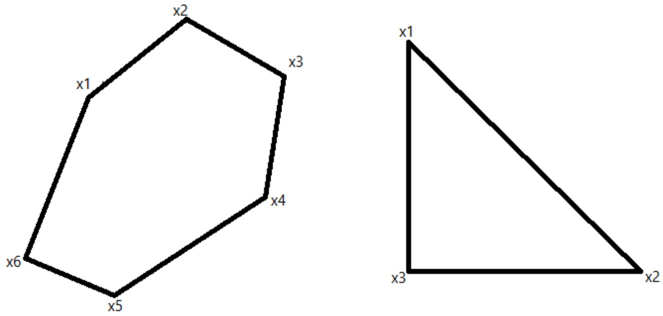
2.1.3 Polytope et simplexe

Définition 2.1.3. *L'envelloppe convexe d'un nombre fini de points* $x_1, x_2, \ldots, x_{k+1}$ *dans* $\mathbb{R}^n$ *s'appelle polytope, si*

$$x_2 - x_1, x_3 - x_1, \ldots, x_k - x_1, x_{k+1} - x_1$$

sont linéairement independants, alors $H(x_1, x_2, \ldots, x_{k-1}, x_k, x_{k+1})$ *s'appelle simplexe avec les points* $x_1 \ldots, x_k, x_{k+1}$.

Remarque 2.1.1. *Dans* $\mathbb{R}^n$, *il n'existe pas de simplexes avec plus de* $(n+1)$ *points. Dans* $\mathbb{R}^n$, *tout système forme de* $(n+1)$ *points ou plus est nécessairement lié.*



(a) Polytope **(b)** Simplexe

FIGURE 2.4 – Polytop et simplex dans $\mathbb{R}^2$

2.2 Fonction convexe

2.2.1 Définitions et propriétés générales

Définition 2.2.1. *Soient $f : S \to \mathbb{R}$, où S est un ensemble convexe non vide de $\mathbb{R}^n$.*
a) f est dite convexe dans S si et seulement si

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \tag{2.2}$$

pour tous $x_1, x_2 \in S$ et pour $\lambda \in [0,1]$.
b) f est dite strictement convexe dans S si et seulement si

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2) \tag{2.3}$$

pour tous $x_1, x_2 \in S$, $x_1 \neq x_2$ et pour $\lambda \in]0,1[$.
c) f est dite fortement convexe ou uniformement convexe dans S si et seulement s'il existe une constante $C > 0$ tel que

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{1}{2} C\lambda(1-\lambda) \|x_1 - x_2\|^2 \tag{2.4}$$

pour tous $x_1, x_2 \in S$ et pour tout $\lambda \in [0,1]$.
d) f est dite concave dans S (strictement concave dans S) si et seulement si $(-f)$ est convexe (strictement convexe) dans S.

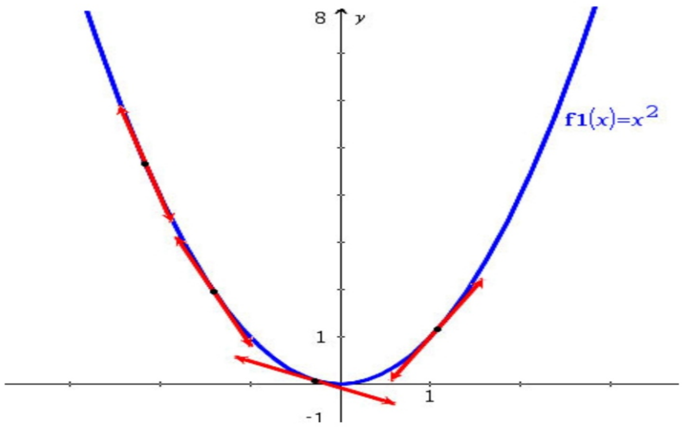


FIGURE 2.5 – Fonction convexe

Définition 2.2.2. *Soit $f : \mathbb{R}^n \to \mathbb{R}^n$ une fonction concave si -f est une fonction convexe , c'est-à-dire, si pour tout $x, y \in \mathbb{R}^n$, on a*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

$\forall x \neq y, \forall \lambda \in [0, 1]$

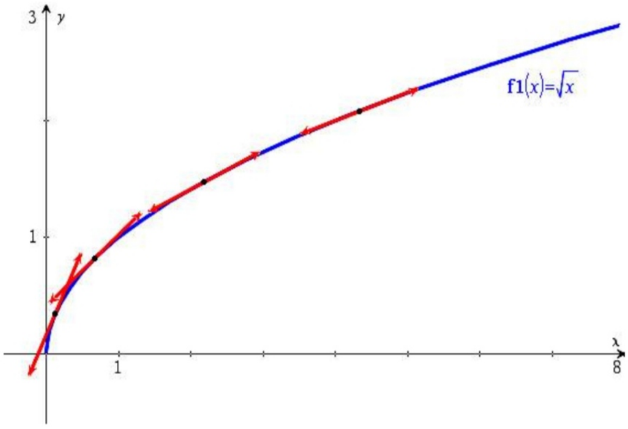


FIGURE 2.6 – Fonction concave

Proposition 2.2.1. *Soit $S \subset \mathbb{R}^n$ convexe non vide, $f : \mathbb{R}^n \to \mathbb{R}$ convexe, alors pour tout $\alpha \in \mathbb{R}$, l'ensemble*

$$S_\alpha = \{x \in S : f(x) \leq \alpha\}$$

est convexe.

preuve. Soient $x_1, x_2 \in S_\alpha$. Alors

$$f(x_1) \leq \alpha \text{ et } f(x_2) \leq \alpha. \tag{2.5}$$

Démontrons que pour $\lambda \in [0,1], \lambda x_1 + (1-\lambda)x_2 \in S_a$. En effet f étant convexe, on a

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

d'où en utilisant (2.2)

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha, \tag{2.6}$$

(2.3) implique que $\lambda x_1 + (1-\lambda)x_2 \in S_a$. Par conséquent S_α est convexe. □

Remarque 2.2.1. *La réciproque de la propsition* 2.2.1 *est fausse. D'où la definition suivante :*

Définition 2.2.3. *Soit $S \subset \mathbb{R}^n$ convexe non vide, et $f : \mathbb{R}^n \to \mathbb{R}$ une fonction. f est dite quasi convexe dans S si et seulement si pour tout $\alpha \in \mathbb{R}$, l'ensemble*

$$S_\alpha = \{x \in S : f(x) \leq \alpha\}$$

est convexe.
La définition suivante est équivalente a la définition 2.2.1 .

Définition 2.2.4. *Soit $S \subset \mathbb{R}^n$ convexe non vide, $f : \mathbb{R}^n \to \mathbb{R}$.*
f est dite quasi convexe dans S si et seulement si

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \ \textit{Maximum} \ [f(x_1), f(x_2)] \tag{2.7}$$

pour tout $x_1, x_2 \in S$ et pour tout $\lambda \in [0,1]$.

2.3 Fonctions convexes et différentiables

2.3.1 Fonctions convexes une fois différentiables

Définition 2.3.1. *Soient* $S \subset \mathbb{R}^n, f : \mathbb{R}^n \to \mathbb{R}$ *et* $\hat{x} \in \operatorname{int}(S)$.
f est dite différentiable au point $\hat{x}$, s'il existe un vecteur $A \in \mathbb{R}^n$ et une fonction $\alpha : \mathbb{R}^n \to \mathbb{R}$ tels que

$$f(x) = f(\hat{x}) + A(x - \hat{x}) + \|x - \hat{x}\| \alpha(\hat{x}, x - \hat{x})$$

où $\alpha(\hat{x}, x - \hat{x}) \underset{x \to \bar{x}}{\to} 0$ *et*

$$A = \nabla f(\hat{x}) = \left(\frac{\partial f(\hat{x})}{\partial x_1}, \ldots, \frac{\partial f(\hat{x})}{\partial x_n} \right).$$

Définition 2.3.2. *Soit $f : I \to \mathbb{R}$ une fonction à valeurs réelles définie sur un intervalle ouvert de la droite réelle. Une telle fonction peut ne pas être différentiable en tout point, par exemple, c'est le cas de la valeur absolue, f(x)=|x|.*
Cependant, la fonction f vérifie la propriété suivente : pour tout x_0 dans le domaine de définition de la fonction, il est possible de tracer une droite passant par $(x_0, f(x_0))$ et ne coupant jamais le graphe de f (il est possible d'avoir des points de tangente). L'ensemble des pentes de toutes les droites vérifiant la précédente propriété en un point x est appelé sous-gradient de f en x .

Théorème 2.3.1. *Soit $S \subset \mathbb{R}^n$ convexe et $f : \mathbb{R}^n \to \mathbb{R}$ convexe et $\hat{x} \in$ int (S). On suppose que f est differentiable au point $\hat{x}$. Alors f admet en $\hat{x}$ un seul sous gradient qui est $\nabla f(\hat{x})$. C'est- à-dire que $\partial f(\hat{x}) = \{\nabla f(\hat{x})\}$.*

preuve. Soit ξ un sous gradient appartenant à $\partial f(\hat{x})$. pour tout $d \in \mathbb{R}^n$ et pour un certain λ choisi suffisament petit ($\lambda > 0$) On a

$$f(\hat{x} + \lambda d) \geq f(\hat{x}) + \lambda \xi^\top d, \tag{2.8}$$

f est différentiable en $\hat{x}$

$$f(\hat{x} + \lambda d) = f(\hat{x}) + \lambda \nabla f(\hat{x})^\top d + \lambda \|d\| \alpha(\hat{x}, \lambda d). \tag{2.9}$$

En portant (2.9) dans (2.8), on obtient

$$\lambda\left\{d\left[(\xi-\nabla f(\hat{x})^\top\right]-[\|d\|\alpha(\hat{x},\lambda d)]\right\}\leq 0$$

divisons par $\lambda>0$ et tondons λ vers 0^+ on obtient

$$d(\xi-\nabla f(\hat{x}))^\top\leq 0,\forall d\in\mathbb{R}^n, \tag{2.10}$$

prenons $d=\xi-\nabla f(\hat{x}),(2.10)$ devient

$$\|\xi-\nabla f(\hat{x})\|\leq 0.$$

Ceci implique necessairement que

$$\xi=\nabla f(\hat{x}).$$

□

Théorème 2.3.2. *Soit $S\subset\mathbb{R}^n$ convexe, non vide, ouvert et soit $f:\mathbb{R}^n\to\mathbb{R}$ differentiable dans S. Alors*
a) f est convexe sur S si et seulement si pour tout $\hat{x}\in S$

$$f(x)\geq f(\hat{x})+\nabla f(\hat{x})^\top(x-\hat{x}),\forall x\in S. \tag{2.11}$$

b) f est strictement convexe sur S si et seulement si pour tout $\hat{x}\in S$

$$f(x)>f(\hat{x})+\nabla f(\hat{x})^\top(x-\hat{x}),\forall x\in S,\ x\neq\hat{x}. \tag{2.12}$$

preuve. **a)** Supposons que f est convexe et que S est ouvert. Puisque $int(S)=S$ et f admet en tout point $\hat{x}\in S$ un sous gradient qui est $\nabla f(\hat{x})$. Donc d'après la définition du sous gradient on a

$$f(x)\geq f(\hat{x})+\nabla f(\hat{x})^\top(x-\hat{x}),\quad\forall x\in S.$$

Réciproquement si la relation (2.11) est vérifiée pour tout $\hat{x}\in\mathrm{int}(S)=S$, alors f est convexe dans l'ensemble $\mathrm{int}(S)=S$.

b) La preuve est exactement la même que celle de a) en utilisant le corollaire 2.1.1. □

2.3.2 Fonctions convexes deux fois différentiables

Définition 2.3.3. *Soit $S \subset \mathbb{R}^n$ non vide et $f : \mathbb{R}^n \to \mathbb{R}$. f est dite deux foix differentiable au point $\hat{x} \in \operatorname{int}(S)$ s'il existe un vecteur $\nabla f(\hat{x})$ et une matrice symétrique $H(\hat{x})$ d'ordre n appellée matrice hessienne, et une fonction $\alpha : \mathbb{R}^n \to \mathbb{R}$ tels que*

$$\forall x \in S : f(x) = f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x}) + \frac{1}{2}(x - \hat{x})^\top H(\hat{x})(x - \hat{x}) + \|x - \hat{x}\|^2 \alpha(\hat{x}, x - \hat{x})$$

avec : $\alpha(\hat{x}, x - \hat{x}) \underset{x \to \hat{x}}{\to} 0,$

où

$$H(\hat{x}) = \begin{pmatrix} \frac{\partial^2 f(\hat{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 f(\hat{x})}{\partial x_1 \partial x_2} & . & \frac{\partial^2 f(\hat{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\hat{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\hat{x})}{\partial x_2 \partial x_2} & \ddots & . \\ . & . & \ddots & . \\ \frac{\partial^2 f(\hat{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\hat{x})}{\partial x_n \partial x_2} & \ddots & \frac{\partial^2 f(\hat{x})}{\partial x_n \partial x_n} \end{pmatrix}.$$

2.4 Condtions d'optimalité des problèmes d'optimisation convexe

Définition 2.4.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$, on appelle problème de minimisation sans contraintes le problème $(\mathcal{P})$ suivant*

$$\min \{f(x) x \in \mathbb{R}^n\} \qquad (\mathcal{P})$$

1) $\hat{x} \in \mathbb{R}^n$ est un minimum global de $(\mathcal{P})$ si et seulement si $f(\hat{x}) \leq f(x) \quad \forall x \in \mathbb{R}^n$.
2) $\hat{x} \in \mathbb{R}^n$ est un minimum local de $(\mathcal{P})$ si et seulement si il existe un voisinage $V_\varepsilon(\hat{x})$ tel que

$$f(\hat{x}) \leq f(x) \quad \forall x \in V_\varepsilon(\hat{x}).$$

3) $\hat{x} \in \mathbb{R}^n$ est un minimum local strict de $(\mathcal{P})$ si et seulement s'il existe un voisinage $V_\varepsilon(\hat{x})$ tel que

$$f(\hat{x}) < f(x) : \quad \forall x \in V_\varepsilon(\hat{x}), x \neq \hat{x}.$$

Définition 2.4.2. *(Optimisation avec contraintes de type égalités).*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue et K un sous-ensemble non vide de $\mathbb{R}^n$. Le problème

$$(\mathcal{P}) \qquad \min_{x \in K} f(x),$$

s'appele problème d'optimisation sous contrainte.

Remarque 2.4.1. *Soient $h_j, j = 1, ..., m$ des fonctions de classe $C^1(\mathbb{R}^n)$.*
On dit que les contraintes d'égalités si

$$k = \{x \in \mathbb{R}^n,\ h_j(x) = 0\}.$$

On dit que les contraintes d'inégalités si

$$k = \{x \in \mathbb{R}^n,\ h_j(x) \leq 0\}.$$

Théorème 2.4.1. *Soient $S \subset \mathbb{R}^n$ non vide et convexe et soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe, concidérons le problème d'optimisation convexe suivant*

$$\{\min f(x) x \in S\} \quad (\mathcal{P})$$

et $\widehat{x} \in S$. Alors $\widehat{x}$ est une solution optimale de $(\mathcal{P})$ si et seulement si f admet en $\widehat{x}$ un sous gradient ξ tel que

$$\xi^t(x - \widehat{x}) \geq 0, \quad \forall x \in S.$$

preuve. **a)** Supposons que f admet en $\widehat{x}$ un sous gradient ξ tel que $\xi^t(x-\widehat{x}) \geq 0, \forall x \in S$. ξ étant un sous gradient de f au point $\widehat{x}$, donc par définition nous avons

$$f(x) \geq f(\widehat{x}) + \xi^t(x - \widehat{x}) \geq 0, \quad \forall x \in S,$$

$\forall x \in S, \quad \xi^t(x-\widehat{x}) \geq 0$ implique que

$$f(x) \geq f(\widehat{x}) \quad \forall x \in S.$$

b) Supposons que $\widehat{x}$ est une solution optimale de (PC), et considérons les deux ensembles suivants

$$\begin{aligned} \Lambda_1 &= \left\{(x-\widehat{x}, y) : x \in \mathbb{R}^n, y > f(x) - f(\widehat{x})\right\}, \\ \Lambda_2 &= \{(x-\widehat{x}, y) : x \in S, y \leq 0\}, \end{aligned}$$

Λ_1 et Λ_2 sont convexes. Montrons que $\Lambda_1 \cap \Lambda_2 = \varnothing$. En effet car si on suppose le contraire il exitrait

$$(x^* - \widehat{x}, y^*) \in \Lambda_1 \cap \Lambda_2.$$

Donc $x^* \in S, y^* \leq 0$ et $y^* > f(x^*) - f(\widehat{x})$. Ceci implique nécessairement que

$$x^* \in S \text{ et } f(x^*) - f(\widehat{x}) < 0.$$

Ceci est contradiction avec le fait que $\widehat{x}$ est une solution optimale de $(\mathcal{P})$. Λ_1 et Λ_2 convexes et $\Lambda_1 \cap \Lambda_2 = \varnothing$, alors il existe un hyperplan dans $\mathbb{R}^{n+1}$ qui sépare simplement Λ_1 et Λ_2 . En d'autres termes

$$\exists\ \xi_0 \in \mathbb{R}^n, \mu_0 \in \mathbb{R}, \text{ et } \alpha \in \mathbb{R}, \text{ tels que } \ (\xi_0, \mu_0) \neq (\overrightarrow{0}, 0)$$

et

$$\xi_0^t(x-\widehat{x}) + \mu_0 \cdot y \leq \alpha : \quad x \in \mathbb{R}^n, y > f(x) - f(\widehat{x}) \tag{2.13}$$

$$\xi_0^t(x-\widehat{x}) + \mu_0 \cdot y \geq \alpha : \quad x \in S, y \leq 0. \tag{2.14}$$

Il n'est pas difficile de montrer en faisant un bon choix de x, que les relations précedentes impliquent nécessairement que $\alpha = 0$ et que $\mu_0 < 0$. Notons par $\xi = -\frac{\xi_0}{\mu_0}$ et divisons (2.12) et (2.13) par $-\mu_0$, on obtient

$$\xi^t(x-\widehat{x}) - y \leq 0 \quad x \in \mathbb{R}^n, y > f(x) - f(\widehat{x}) \tag{2.15}$$

et

$$\xi^t(x-\widehat{x}) - y \geq 0 \quad x \in S, y \leq 0. \tag{2.16}$$

pour $y = 0$ (2.15) donne :

$$\xi^t(x - \widehat{x}) \geq 0 \quad \forall x \in S. \tag{2.17}$$

Montrons que (2.14) implique

$$f(x) - f(\widehat{x}) \geq \xi^t(x - \widehat{x}) : \quad \forall x \in S. \tag{2.18}$$

En effet, supposons le contraire. Alors il existe $x^* \in \mathbb{R}^n$ tel que

$$f\left(x^*\right) - f(\widehat{x}) < \xi^t\left(x^* - \widehat{x}\right). \tag{2.19}$$

Donc il existe $y \in \mathbb{R}$ tel que

$$f\left(x^*\right) - f(\widehat{x}) < y < \xi^t\left(x^* - \widehat{x}\right). \tag{2.20}$$

(2.11) est contradiction avec le fait que si $y > f\left(x^*\right) - f(\widehat{x})$ alors $y \geq \xi^t\left(x^* - \widehat{x}\right)$ (voir (2.6)) De (2.5) et (2.9) on conclut qu'il existe un sous gradient ξ tel que $\xi^t(x - \widehat{x}) \geq 0 \quad \forall x \in S$. □

Théorème 2.4.2. *Soient $S \subset \mathbb{R}^n$ non vide, convexe et ouvert et $f : \mathbb{R}^n \to \mathbb{R}$ convexe et $(\mathcal{P})$ le problème d'optimisation convexe suivant*

$$\{\min f(x) : x \in S\} \quad (\mathcal{P}),$$

et $\widehat{x} \in S$. Alors $\widehat{x}$ est une solution optimale de $(\mathcal{P})$ si et seulement si f admet en $\widehat{x}$ un sous gradient nul.

preuve. Du théorème 2.4.1 on déduit que $\widehat{x}$ est une solution optimale de $(\mathcal{P})$ si et seulement si f admet en $\widehat{x}$ un sous gradient ξ tel que

$$\xi^t(x - \widehat{x}) \geq 0, \quad \forall x \in S. \tag{2.21}$$

S étant ouvert, alors il existe $\lambda > 0$ tel que $x_\lambda = \widehat{x} - \lambda\xi \in S$. Pour le point x_λ, la relation (2.12) donne

$$\xi^t(-\lambda\xi) \geq 0$$

ou encore

$$-\lambda\|\xi\| \geq 0$$

ceci implique, vu que $\lambda > 0$

$$\|\xi\| \leq 0.$$

Puisque

$$\|\xi\| \geq 0,$$

alors $\xi = 0$. □

2.5 Conditions nécessaires d'optimalité

2.5.1 Condition nécessaire d'optimalité du premier ordre

Théorème 2.5.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ différentiable au point $\hat{x} \in \mathbb{R}^n$. Si $\hat{x}$ est un minimum local de $(\mathcal{P})$ alors $\nabla f(\hat{x}) = 0$.*

preuve. :Supposons que $\nabla f(\bar{z}) \neq 0$. Puisque la direction $d = -\nabla f(\bar{\imath})$ est une direction de descente, alors il existe $\delta > 0$ tel que

$$f(\tilde{u} + \lambda b) < f(\tilde{z}) \quad \forall \lambda \in]0, \delta[.$$

Ceci est contradiction avec le fait que at est une solution optimale locale de $(\mathcal{P})$. □

2.5.2 Condition nécessaire d'optimalité du second ordre

Théorème 2.5.2. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois différentiable au point $\hat{x} \in \mathbb{R}^n$. Si $\hat{x}$ est un minimum local de $(\mathcal{P})$ alors $\nabla f(\hat{x}) = 0$ et la matrice hessienne de f au point $\hat{x}$, qu'on note $H(\hat{x})$, est semi definie positive.*

preuve. : Soit $x \in \mathbb{R}^n$ quelconque, f étant deux fois différentiable au point $\hat{x}$, on aura pour tout $\lambda \neq 0$

$$f(\hat{x} + \lambda x) = f(\hat{x}) + \frac{1}{2}\lambda^2 x^t H(\hat{x})x + \lambda^2 \left\|x^2\right\| \alpha(\hat{x}, \lambda x), \qquad \alpha(\hat{x}, \lambda x) \underset{\lambda \to 0}{\longrightarrow} 0.$$

Ceci implique

$$\frac{f(\hat{x}+\lambda x)-f(\hat{x})}{\lambda^2}=\frac{1}{2}x^tH(\hat{x})x+\left\|x^2\right\|\alpha(\hat{x},\lambda\hat{x}) \tag{2.22}$$

$\hat{x}$ est un optimum local, il existe alors $\delta>0$ tel que

$$\frac{f(\hat{x}+\lambda x)-f(\hat{x})}{\lambda^2}\geq 0, \quad \forall\lambda\in]-\delta,+\delta[.$$

Si on prend en considèration (2.22) et on passe à la limite quand $\lambda\to 0, \lambda\neq 0$, on obtient

$$x^tH(\hat{x})x\geq 0, \quad \forall x\in\mathbb{R}^n.$$

□

2.6 Conditions suffisantes d'optimalité

Théorème 2.6.1. *Soit $f:\mathbb{R}^n\to\mathbb{R}$ deux fois différentiable au point $\hat{x}\in\mathbb{R}^n$. Si $\nabla f(\hat{x})=0$ et $H(\hat{x})$ est définie positive, alors $\hat{x}$ est un minimum local strict de $(\mathcal{P})$.*

preuve. f étant deux fois différentiable au point $\hat{x}$, on aura pour tout $x\in\mathbb{R}^n$

$$f(x)=f(\hat{x})+\frac{1}{2}(x-\hat{x})^tH(\hat{x})(x-\hat{x})+\|(x-\hat{x})\|^2\alpha(\hat{x},(x-\hat{x})), \quad \alpha(\hat{x},(x-\hat{x}))\underset{x\to\hat{x}}{\to}0. \tag{2.23}$$

Supposons que $\hat{x}$ n'est pas un minimum local strict. Alors il existe une suite $\{x_k\}_{k\in\mathbb{N}}$ tel que $\forall k\in\mathbb{N}$, $x_k\neq\hat{x}$ et

$$x_k\underset{k\to\infty}{\to}\hat{x} \quad \text{et} \quad f(x_k)\leq f(\hat{x}). \tag{2.24}$$

Dans (2.23) prenons $x=x_k$, divisons le tout par $\|(x_k-\hat{x})\|^2$ et notons $d_k=\frac{(x_k-\hat{x})}{\|(x_k-\hat{x})\|}$, on obtient

$$\frac{f(x_k)-f(\hat{x})}{\|(x_k-\hat{x})\|^2}=\frac{1}{2}d_k^tH(\hat{x})d_k+\alpha(\hat{x},(x_k-\hat{x})), \quad \alpha(\hat{x},(x_k-\hat{x}))\underset{k\to\infty}{\to}0. \tag{2.25}$$

(2.24) et (2.25) impliquent

$$\frac{1}{2}d_k^t H(\hat{x})d_k + \alpha\left(\hat{x},(x_k - \hat{x})\right) \leq 0,$$

D'autre part la suite $\{d_k\}_{k\in\mathbb{N}}$ est bornée ($\|d_k\| = 1$). Donc il existe une sous suite $\{d_k\}_{k\in N_1\subset\mathbb{N}}$ telle que

$$d_k \underset{k\to\infty, k\in N_1}{\to} \tilde{d}.$$

Finalement lorsque $k \to \infty, k \in N_1$, on obtient

$$\frac{1}{2}\tilde{d}H(\hat{x})\tilde{d} \leq 0.$$

La dernière relation et le fait que $\tilde{d} \neq 0, (\|\tilde{d}\| = 1)$ impliquent que la matrice hessienne $H(\hat{x})$ n'est pas définie positive. Ceci est contradiction avec l'hypothèse. $\square$

Exemple 2.6.1. $f(x_1, x_2) = \frac{1}{2}x_1 + x_1 \cos x_2$.
Utilisons les condition d'optimalité pour identifier les minimums locaux

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_1 + \cos x_2 \\ -x_1 \sin x_2 \end{pmatrix}$$

Le gradient s'annule pour
1) $\hat{x}_k = ((-1)^{k+1}), k\pi)^T, k\in \mathbb{Z},$
2) $\bar{x_k} = (0, \frac{\pi}{2} + k\pi)^T,\ k \in \mathbb{Z}.$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 1 & -\sin x_2 \\ -\sin x_2 & -x_1 \cos x_2 \end{pmatrix}$$

$$\nabla^2 f(\hat{x}_k) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\nabla^2 f(\bar{x_k}) = \begin{pmatrix} 1 & (-1)^{k+1} \\ (-1)^{k+1} & 0 \end{pmatrix}$$

3) *$\hat{x}_k$ vérifie les conditions suffisantes d'optimalité.*

4) $\bar{x}_k$ ne vérifie pas les conditions nécessaires d'optimalité pour aucun k.

Condition suffisante d'optimalité globale

Soit une fonction $f : \mathbb{R}^n \to \mathbb{R}$ deux fois différentiable dans $\mathbb{R}^n$, et soit $\hat{x} \in \mathbb{R}^n$ un minimum local de f.

Si f est une fonction convexe, alors $\hat{x}$ est un minimum global de f .

Si de plus f est strictement convexe , $\hat{x}$ est l'unique minimum global de f.

Chapitre 3

Rechreches linéaires

3.1 Direction de descente

Définition 3.1.1. *Soit* $f : \mathbb{R}^n \to \mathbb{R}, \hat{x} \in \mathbb{R}^n$. $d \in \mathbb{R}^n$ *est dite direction de descente au point* $\hat{x}$ *si et seulement s'il existe* $\delta > 0$ *tel que*

$$f(\hat{x} + \lambda d) < f(\hat{x}) \quad \forall \lambda \in]0, \delta[.$$

Donnons maintenant une condition suffisante pour que d soit une direction de descente.

Théorème 3.1.1. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$ *differentiable au point* $\hat{x} \in \mathbb{R}^n$ *et* $d \in \mathbb{R}^n$ *une direction verifiant la condition*

$$f'(\hat{x}, d) = \nabla f(\hat{x})^t d < 0.$$

Alors d *est une direction de descente au point* $\hat{x}$.

preuve. f est différentiable au point $\hat{x}$. Donc

$$f(\hat{x} + \lambda d) = f(\hat{x}) + \lambda \nabla f(\hat{x})^t d + \lambda \|d\| \alpha(\hat{x}, \lambda d),$$

avec

$$a(\hat{x}, \lambda d) \underset{\lambda \to 0}{\to} 0,$$

ceci implique que

$$f'(\hat{x}, d) = \lim_{\lambda \to 0} \frac{f(\hat{x} + \lambda d) - f(\hat{x})}{\lambda} = \nabla f(\hat{x})^t d < 0.$$

La limite étant strictement négative, alors il existe un voisinage de zéro $V(0) =]-\delta, +\delta[$ tel que

$$\frac{f(\hat{x} + \lambda d) - f(\hat{x})}{\lambda} < 0, \quad \forall \lambda \in]-\delta, +\delta[\tag{3.1}$$

La relation (3.1) est particulièrement vraie pour tout $\lambda \in]0, +\delta[$. On obtient le résultat cherché en multipliant la relation (3.1) par $\lambda > 0$. $\square$

3.2 Schémas général des algorithmes d'optimisation sans contraintes

Supposons que d_k soit une direction de descente au point x_k. Ceci nous permet de considérer le point x_{k+1}, successeur de x_k, de la manière suivante

$$x_{k+1} = x_k + \lambda_k d_k, \quad \lambda_k \in]\, 0, +\delta[.$$

Vu la définition de direction de descente, on est assuré que

$$f(x_{k+1}) = f(x_k + \lambda_k d_k) < f(x_k).$$

Un bon choix de d_k et de λ_k permet ainsi de construire une multitude d'algorithmes d'optimisation.

3.2.1 Exemples de choix de directions de descente

Par exemple si on choisit $d_k = -\nabla f(x_k)$ et si $\nabla f(x_k) \neq 0$, on obtient la méthode du gradient. La méthode de Newton correspond à

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

Bien sur $-\nabla f(x_k)$ est une direction de descente

$$\left(\nabla f(x_k)^t d_k = -\nabla f(x_k)^t \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0\right)$$

. Pour la deuxième direction, si la matrice hessienne $H(x_k)$ est définie positive alors $d_k = -(H(x_k))^{-1} \nabla f(x_k)$ est aussi une direction de descente.

3.2.2 Exemple de choix de pas λ_k

On choisit en général λ_k de façon optimale, c'est à dire que λ_k doit vérifier

$$f(x_k + \lambda_k d_k) \leq f(x_k + \lambda d_k) : \quad \forall \lambda \in [0, +\infty[.$$

En d'autres temes on est ramené à étudier à chaque itération un problème de minimisation d'une variable réelle. C'est ce qu'on appelle recherche linéaire.

3.3 Méthode linéaire et méthodes à direction de descente

Soit $f : \mathbb{R}^n \to \mathbb{R}$. On considére le probléme de minimisation sans contraintes (PSC) suivant :

$$\min f(x), x \in \mathbb{R}^n \qquad\qquad (PSC)$$

Il existe deux grandes classes des méthodes qui contribuent résoudre le problème (PSC) .

a) les méthodes diréction de descente et rechreche linéaire.

b) les méthodes région de confiance.
Nous nous intéréssons dans cette partie aux méthodes rechreche linéaire et direction de déscente dont le principe est le suivant et à direction de déscente dont le principe suivant :
supposons qu'à l'itération k on ait un point $x_k \in \mathbb{R}^n$ et une direction de descente $d_k \in \mathbb{R}^n$, c'est à dire d_k vérifie

$$\nabla f(x_k)^T d_k < 0. \tag{3.2}$$

Le sucusseur x_{k+1} de x_k est obtenu comme suit

$$x_{k+1} = x_k + \alpha_k d_k \quad \alpha_k \in \mathbb{R}_+^* \tag{3.3}$$

a_k s'appelle le pas. Le choix de d_k et α_k caractérisent l'algorithme.
Supposons que d_n est determinée. Comment choissir le pas α_k ?
Les méthodes qui s'interessent au choix et au calcul de α_k s'appellent rechreches linéaires.
Il existe deux type de rechreches linéaires
i) les rechreches lineraire exactes.
ii) les rechreches linéaire inexactes.

3.3.1 Choix optimal du pas α_k

On peut choisir α_k de façon optimale, c'est à dire qu'en partant du point x_k et la direction d_k, f décroit le mieux possible au point $x_{n+1} = x_k + \alpha_k d_k$. Ceci est possible si on impose à α_k de vérifier la condition suivante

$$\forall \alpha \in]0, +\infty[\quad f(x_k + \alpha_k d_k) \leq f(x_k + \alpha d_k). \tag{3.4}$$

Notons $\varphi_k(\alpha)$ la fonction d'une variable réelle suivante

$$\varphi_k(\alpha) = f(x_k + \alpha d_k), \alpha \in]0, +\infty[\tag{3.5}$$

(3.4) et (3.5) donnent

$$\varphi_k(\alpha_k) \leq \varphi_k(\alpha), \quad \forall \alpha \in]0, +\infty[\tag{3.6}$$

ou encore α_n est solution optimale de

$$\varphi_k(\alpha_k) = \min\{\varphi_k(\alpha) : \alpha \in]0, +\infty[\}. \tag{3.7}$$

Trouvez α_k vérifant (3.7) s'appelle rechreche linéaire exacte. Ce travail à effectue à chaque itiration k.

Donc à chaque itiration k, on doit résoudre un problème d'optimiation dans $\mathbb{R}$.

Remarque 3.3.1. *$a > 0$, alors $H(\alpha)$ est définie positve sur $]0, +\infty[$. Alors $\varphi_n(a)$ est strictement convexe sur $[0, +\infty[$ et par conséquent le problème est un problème convexe et admet une solution optimale locale et globale unique vérifiant*

$$\varphi'(\alpha_k) = 0.$$

c'est à dire

$$a_n = -\frac{b}{2a}.$$

Remarque 3.3.2. *Effectuer une rechreche linéaire exacte à chaque itiration, est difficilement réalisable en pratique et assez couteux en temps et mémoire, c'est pour celà qu'on cherche α_k moins couteux et facile à calculer, c'est ce qu'on appelle rechrechre linéaire inéxacte.*

3.3.2 Recherche linéaire inéxacte

Si α_k est obtenue par une rechreche linéaire exacte alors f décroit le mieux en pasant du point x_k au point $x_k + \alpha_k d_k$, c'est à dire

$$f(x_k + \alpha_k d_k) \leq f(x_k + \alpha dk) \, \forall \alpha \in]0, +\infty[$$

ou encore que

$$\forall \alpha \in]0, +\infty[: f(x_k) - f(x_k + \alpha_k d_k) \geq f(x_k) - f(x_k + \alpha d_k).$$

Notons par $\tau(a)$ la quantité suivante

$$\tau(\alpha) = |f(x_k) - f(x_k + \alpha d_k)| = f(x_k) - f(x_k + \alpha d_k)$$

$\tau(a)$ représente la décroissance de f, lorsqu'on pasae du point x_k au point $x_k + \alpha d_k$. On a donc

$$\tau(\alpha_k) \geq \tau(\alpha) \forall \alpha \in]0, +\infty[.$$

Comparaisons de pas obtenus par des recherches linéaires inexactes

Comme on l'a signalé avant, α_n est optimal mais il est difficile à calculer et il est couteux en temps et en mémoire.
Il serait alors intéréssant de chrecher un autre α_n (qu'en note β_k), facile à calculer, pas couteux en temps de calcul et en mémoire et qui a'spproche le plus de α_k de sorte que la décroissance

$$\tau\left(\beta_k\right) = f\left(x_k\right) - f\left(x_k + \beta_k d_k\right)$$

soit la plus proche de

$$\tau\left(\alpha_k\right) = f\left(x_k\right) - f\left(x_k + \alpha_k d_k\right).$$

Bien sur, on a toujours

$$\tau\left(\beta_k\right) \leq \tau\left(\alpha_k\right),$$

le calcul de β_k s'appelle recherche linéaire inexacte.
Nous abordons trois type de ces recherche.

3.4 Recherche linéaire d'Armijo

3.4.1 Principe de la méthode

Il ne suffit pas de choisir à chaque itération k, α_k tel que

$$f\left(x_k + \alpha_k d_k\right) < f\left(x_k\right).$$

pour que la suite $\{x_k\}$ converge vers la solution optimale ou vers un point stationnaire. En effet si la fonction décroit trés peu en passant du point x_k vers le point $x_{k+1} = x_k + \alpha_k d_k$ la suite $\{x_k\}$ diverge ou peut converger vers un point qui n'a rien à voir avec la solution optimale.
Pour éviter cet inconvénient, il faut imposer une condition caractérisant la notion de descente suffisante de la fontion. Il faut choissir α_k de sorte que la fonction f décroit de suffisante en passant du point x_k au point passant du poin $x_{k+1} = x_k + \alpha_k d_k$.

L'idée est que la diminution jugée suffisante soit proportionnelle si la taille du pas α_k. Ainsi plus le pas sera grand, plus la diminution de la fonction f sera grande. Donc si $\gamma > 0$, on désire que

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \alpha_k \gamma$$

ou encore

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \alpha_k \gamma.$$

Le facteur γ ne peut pas être choisi de façon arbitraire. Surtout il doit varier d'une itération à l'autre. Il est plus approprié de définir γ en fonction de la pente de la fonction au point x_k dans la direction d_k. En l'occurence, nous choisirons

$$\gamma = -\beta \nabla f(x_k)^T d_k \quad 0 < \beta < 1$$

Définition (diminution suffisante condition d'Armijo)

Soient $f : \mathbb{R}^n \to \mathbb{R}$ une fonction differentiable, un point $x_k \in \mathbb{R}^n$ et une direction (de descente) $d_k \in \mathbb{R}^n$ telle que $\nabla f(x_k)^T d_k < 0$. On dit que le pas $\alpha_k > 0$ vérifie la condition d'Armijo, si on a

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta_1 \nabla f(x_k)^T d_k, \quad 0 < \beta_1 < 1.$$

Test d'Armijo,

Si

$$\varphi_k(\alpha) \leq \varphi_k(0) + \rho \phi_n'(0)\alpha$$

autrement dit ,

$$f(x_k + \alpha d_k) \leq f(x_k) + p\alpha \nabla^T f(x_k) d_k$$

alors α convient.

Si

$$\varphi_k(\alpha) > \varphi_k(0) + \rho h_k'(0)\alpha$$

autrement dit $f(x_k + \alpha d_k) > f(x_k) + \rho\alpha \nabla^T f(x_k) d_k$ alors α est trop grand.

Regle d'Armijo

Etape 0 : (initialisation)

$a_{g,1} = a_{g,1} = 0$, choisir $a_1 > 0, p \in]0, 1[$ poser $k = 1$ et aller a l'étape 1.

Etape 1 :
Si

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha\varphi_k'(0)\alpha_k : \text{STOP}\,(\alpha^* = \alpha_k)$$

si $\varphi_k(\alpha_k) > \varphi_k(0) + \rho\varphi_k'(0)\alpha_k$, alors $\alpha_{d,k+1} = \alpha_d, \alpha_{g,k+1} = \alpha_k$ et aller à létape 2 .

Etape 2 :
si $\alpha_{d,k+1} = 0$ déterminer $\alpha_{k+1} \in]\alpha_{g,k+1}, \infty[$
si $\alpha_{d,k+1} \neq 0$ déterminer $\alpha_{k+1} \in]\,\alpha_{g,k+1},\ \alpha_{d,k+1}[$ remplacer k par $k+1$ et aller d l'étape 1 .

Programme en fortran 77 d' Armijo

```
real x(2), xp(2), g(2), gp(2), f0, alpha, f, p, s, a, b, pf, t, y, phopt
integer i, k, test,n
sig = 0.3; ro = 0.1; b = 10000000; alpha = 10; test = 0; k = 0
print*, 'le dimontion n =';read*, n
do i = 1, n print *,' x0 (', i,' ) = ';read*,xp(i) end do
10 if (test= = 0.and.k.le.10) then call fonc(f0, gp,xp)
do i = 1, nx(i) = xp(i) − alpha* gp(i) end do
call fonc (f, g, x); p = 0
do i = 1, np = p + gp(i)*gp(i) end do
s = f0 − r0* alpha *p
if (f.gt.s) then b=alpha else phopt=alpha ;test = 1 end if
print*, 'alpha(,'k,') =', alpha; alpha = b/2; k = k + 1
go to 10
end if
print*,'alpha optimal=', phopt
end
subroutine fonc (f, g, x)
real f, g(2), x(2)
f = (x(1))**2 + (x(2))**4; g(1) = 2*x(1); g(2) = 4*(x(2))**3
end
```

Théorème 3.4.1. *Si $\varphi_k : \mathbb{R}_+ \to \mathbb{R}$ définie par $\varphi_k(\alpha) = f(x_k + \alpha d_n)$ est continue et bornée inférieurement, si d_n est une direction de descente en x_n $(\varphi_k'(0) < 0)$*

et si $\rho \in]0,1[$, *alors l'ensemble des pas vérifient la régle d'Armijo est non vide.*

preuve. On a

$$\varphi_k(\alpha) = f(x_k + \alpha d_k)$$
$$\varphi_p(\alpha) = f(x_k) + \rho\alpha_k \nabla^T f(x_x) d_k.$$

Le développement de Taylor-Yong en $\alpha = 0$ de φ_k est

$$\varphi_k(\alpha) = f(x_k + \alpha d_k) = f(x_k) + p\alpha\nabla^T f(x_k) d_k + \alpha\xi(\alpha) \text{ où } \xi(\alpha) \to 0, \alpha \to 0.$$

et comme $\rho \in]0,1[$ et $\varphi_k'(0) = \nabla^T f(x_k) d_k < 0$, on déduit

$$f(x_k) + \alpha\nabla^T f(x_k) d_k < f(x_k) + \rho\alpha\nabla^T f(x_k) d_k \text{ pour } \alpha > 0$$

On voit que pour $\alpha > 0$ assez petit on a

$$\varphi_k(\alpha) < \psi_\sigma(\alpha).$$

De ce qui précède et du fait que φ_k est bornée inférieurement, et $\varphi_p(\alpha) \to -\infty; \alpha \to +\infty$, on déduit que la fonction $\psi_p(\alpha) - \varphi_k(\alpha)$ a la propriété

$$\begin{cases} \psi_\rho(\alpha) - \varphi_k(\alpha) > 0 & \text{pour } \alpha \text{ assez petit} \\ \psi_\rho(\alpha) - \varphi_k(\alpha) < 0 & \text{pour } \alpha \text{ asaez grand} \end{cases}$$

donc s'sannule au moins une fois pour $\alpha > 0$.
En choisissant le plus petit de ces zéros on voit qu'il existe $\bar{a} > 0$ tel que

$$\varphi_k(\bar{\alpha}) = \psi_\rho(\bar{\alpha}) \text{ et } \varphi_k(\alpha) < \psi_p(\alpha) \quad pour \quad 0 < \alpha < \bar{\alpha}.$$

Ce qui achéve la démonstration. □

3.4.2 Contrainte des petits pas dans la règle d'Armijo

La condition d'Armijo nous permet de rejeter des pas qui, étant trop grands, ne fournissent pas une diminution suffissnte de la fonction. Ayant réglé le probléme des pas trop grands, nous allons maintenant nous interesser aux pas trop petits .

3.5 Recherche linéaire de Goldsten

La raison de l'echec de la méthode réside dans la dégénéréscence des pas α_k. Ceux ci, bien que strictement positifs, deviennent arbitrairement proches de zéro et la méthode ne peut plus progresser. On va done rajouter une de deuxiéme condition à la condition d'Armijo qui elliminerait les pas trop petits.

3.5.1 Principe de la méthode

Supposons qu'à literation k, on ait le point $x_k \in \mathbb{R}^n$, le vecteur de descente $d_k \in \mathbb{R}^n \left(\nabla f\left(x_k\right)^T d_k < 0\right)$.

On cherche le pas $\alpha_k > 0$ de Goldstein vérifant les deux conditions suivantes

$$f\left(x_k + \alpha_k d_k\right) \leq f\left(x_k\right) + \rho \alpha_k \nabla f\left(x_k\right)^T d_k, \quad \rho \in \left]0, \tfrac{1}{2}\right[\quad \text{(Goldatein1)}$$

et

$$f\left(x_k + \alpha_k d_k\right) \geq f\left(x_k\right) + (1-\rho)\alpha_k \nabla f\left(x_k\right)^T d_k, \qquad \rho \in \left]0, \frac{1}{2}\right[. \qquad \text{(Goldatein2)}$$

Si on note

$$\varphi_k(\alpha) = f\left(x_k + \alpha d_k\right),$$

alors (Goldstein 1) et (Goldstein 2) deviennent

$$\varphi_k\left(\alpha_k\right) \leq \varphi_k(0) + \rho \alpha_k \varphi_k'(0)$$

et

$$\varphi_k\left(\alpha_k\right) \geq \varphi_k(0) + (1-\rho)\alpha_k \varphi_k'(0).$$

(Goldstein1) assure une décroisance suffisante de f, les pas α_k trop petits sont nocifs pour la convergence. La suite $\{x_k\}$ pourrait converger prématurément vers un point $\bar{x}$ qui n'a rien à voir avec le problème d'optimisation.

3.5.2 Algorithme de Goldstein

L'Algorithme essaye de trouver $\alpha_k \in]\beta_1, \beta_2[$. On demarre avec un intervalle $[a_0, b_0]$ assez grand. On prend $\alpha_0 \in [a_0, b_0]$.

Si α_0 verifié Goldatein 1 et Goldstein 2 alors $\alpha_0 \in]\beta_1, \beta_2[$ et on s'arrète.
Si $\alpha_0 > \beta_2$
alors α_0 ne verifié pas Goldstein1, alors on prend $b_1 = \alpha_0$ et $a_1 = b_0$ et $\alpha_1 = \frac{a_1+b_1}{2}$
et on recommence avec α_1 .
Si $\alpha_0 < \beta_1$
alors α_0 ne vérifié pas Goldstein 2, on prend $a_1 = \alpha_0, b_1 = b_0$ et $\alpha_1 = \frac{a_1+b_1}{2}$ et on teste de nouveau α_1.
A litération k
Supposons qu'on ait $[a_k, b_k]$ et $\alpha_k = \frac{a_k+b_k}{2}$.
Si a_n verifié Goldstein l et Goldstein 2, $\alpha_k \in [\beta_1, \beta_2[$. Stop.
Si α_k ne verifié pas Goldstein 1 alors $\alpha_k > \beta_2$.
On prend $b_{k+1} = a_k, a_{n+1} = \alpha_k, a_{k+1} = \frac{a_{k+1}+b_{k+1}}{2}$.
Si α_k ne verifié pas Goldstein 2 alors $\alpha_k < \beta_1$. On prend $a_{k+1} = \alpha_k$, $b_{n+1} = b_n, \alpha_{k+1} = \frac{a_{n+1}+b_{k+1}}{2}$.
On obtient ainsi l'algorithme suivant
Algorithme de Goldstein

ETAPE 1 (Initialisation)
Choisir $\alpha_0 \in [0, 10^{100}]$ et $\rho \in]\,0, 1\,[$. Poser $a_0 = 0, b_0 = 10^{100}$. Poser $k = 0$ et aller à ETAPE2
ETAPE2 (Test Goldstein1)
Iteration k on a $[a_k, b_k]$ et α_k calculez $\varphi_k(\alpha_k)$ Si $\varphi_k(\alpha_k) \leq \varphi_k(0) + p\alpha_k\varphi_k'(0)$, allez à ETAPE 3 Sinon
Poser $b_{k+1} = \alpha_k, a_{k+1} = a_k$, et allez à ETAPE 4
ETAPE 3 (Test Goldstein2)
Si $\varphi_k(a_k) \geq \varphi_k(0)+(1-p)a_k\varphi_k'(0)$ stop. $\alpha^* = \alpha_k$. Sinon poser $a_{k+1} = a_k, b_{k+1} = b_k$ et allez à ETAPE 4
ETAPE 4
Poser $\alpha_{k+1} = \frac{a_{k+1}+b_{k+1}}{2}$
Poser $k = k + 1$ et allez à ETAPE 2.

3.6 Rechreche linéaire de Wolfe

Nous avions rencontré dans le paragraphe précédent la rechreche linéaire de Goldstien.
Reprenons les mêmes notations que celles du paragraphe précédent. A l'itération k, supposons qu'on ait le point x_k et la direction de descente d_k i.e $\left(\nabla f(x_k)^T d_k < 0\right)$.
Notons

$$\varphi_k(\alpha) = f(x_k + \alpha d_k)$$

alors

$$\varphi_k'(\alpha) = \nabla f(x_k + \alpha d_k)^T d_k$$
$$\varphi_k(0) = f(x_k)$$
$$\varphi_k'(0) = \nabla f(x_k)^T d_k < 0.$$

Notons $\Gamma(f, x_k, d_k)$ l'ensemble suivant

$$\Gamma(f, x_k, d_k) = \{\alpha > 0, f(x_k + \alpha d_k) < f(x_k)\} = \{\alpha > 0, \varphi_k(\alpha) < \varphi_k(0)\}.$$

Graphiquement $\Gamma(f, x_k, d_k)$ est l'ensemble des pointes α tel que le graphe $\varphi_k(\alpha)$ soit en dessous de la droite $y = \varphi_k(0)$.
Comme on l'a remarqué aux paragaphes précédents, plus α_k s'approche de $a^* \left(\alpha^* \text{ vérifiant } \varphi_k(\alpha^*) = \min_{\alpha \in]0,+\infty[} \varphi_k(\alpha)\right)$,
plus a_k est meilleur la descente de f s'effectue de mieux en mieux et elle sera optimale au point a^*.

Nous avions vu aussi que les deux conditions Goldstein 1 et Goldstein 2 engendrent un intervalle $[\beta_1, \beta_2]$ tel que α_k est acceptable si $a_k \in]\beta_1, \beta_1[$ L'incorvénient de la méthode de Goldstein est que α^* peut ne pas appartenir à $]\beta_1, \beta_2[$. Ceci entraine un éloignement de α^k de α^*
Les conditions de Wolfe, (Wolfe 1) et (Wolfe 2) assurent que $a^* \in]\beta\prime_1, \beta\prime_2[$ $[\beta\prime_1, \beta\prime_2]$ engendrées par les deux conditions (Wolfe 1) et (Wolfe 2).

3.6.1 Recherche linéaire de Wolfe

α_k est acceptable par la rechreche linéaire de wolfe si α_k vérifie les deux condition (Wolfe 1) et (Wolfe 2) suivantes

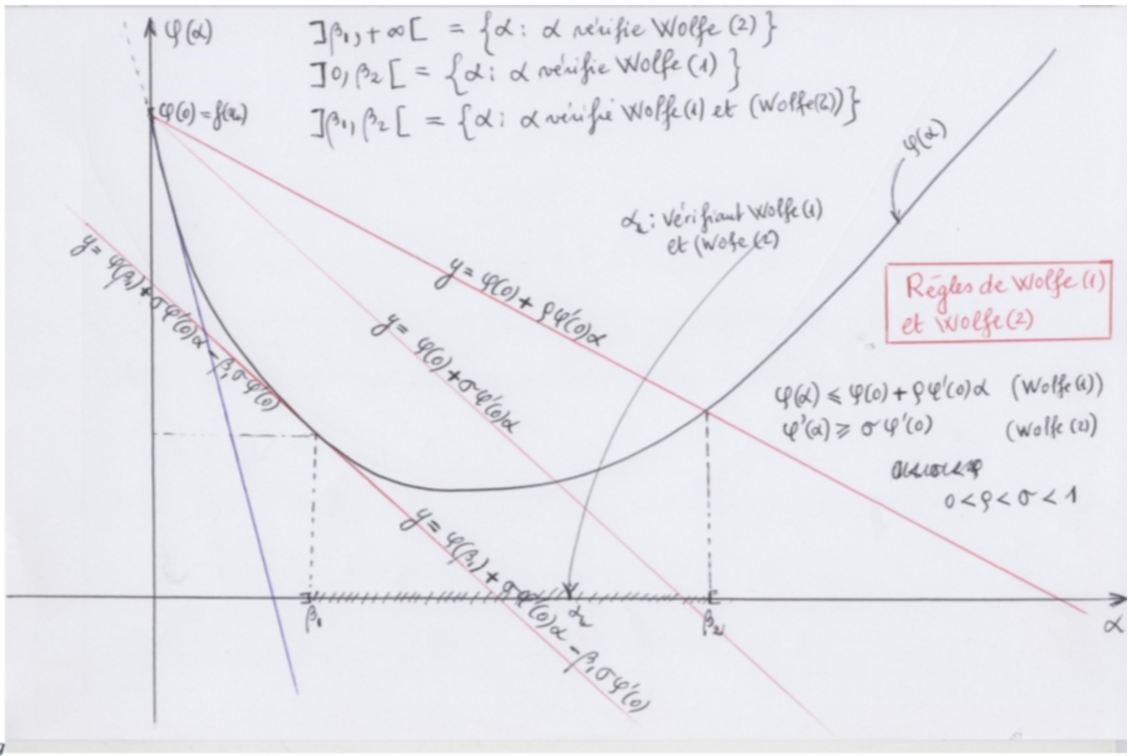
$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho\alpha_k\varphi_k'(0), \quad 0 < p < \frac{1}{2} \quad \text{(Wolfe 1)}$$

$$\varphi_k'(\alpha_k) \geq \sigma\varphi_k'(0), \quad 0 < \sigma < 1, \quad \sigma > p \quad \text{(Wolfe 2)},$$

ou en remplaçant $\varphi_k(\alpha_k), \varphi_k(0), \varphi_k'(0), \varphi_k'(\alpha_k)$ par leurs expressions, on a

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho\alpha_k \nabla f(x_k)^T d_k \quad 0 < \rho < \frac{1}{2} \quad \text{(Wolfe 3)}$$

$$\nabla f(\alpha_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k \quad 0 < \sigma < 1, \quad \sigma > \rho \quad \text{(Wolfe 4)}$$



20.jpg

FIGURE 3.1 – Rechrches linéaire de Wolfe

Commentaires sur la figure

a) $\alpha \in]0, \beta_1'[$ alors α ne vérifie pas (wolfe2) car β_1' vérifié $\varphi'(\beta_1') = \sigma\varphi'(0)$. Les droites Δ_2 et Δ_2 sont paralléles $(\Delta_2 // \Delta_3)$ et si $\alpha < \beta_1'$ alors $\varphi'(\alpha) < \varphi'(\beta_1')$.

En fait β_1 est le premier point tel que $\varphi'(\beta_1') = \delta\varphi'(0)$.
b) $\alpha \in]\beta_2', +\infty[$, α ne verifié pas (Wolfe 2), le graphe de φ est au dessus de la droite Δ_1 de coefficient directeur $p\varphi'(0)$.
c) $\alpha \in]\beta_1', \beta_2'[$ a verifié (Wolfe 1) et (Wolfe 2).

3.6.2 Rechreche linéaire de Wolfe forte

α_k verifié les condition de Wolfe forte si les deux conditions suivantes sont verifiés

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho\alpha_k\varphi_k'(0), \quad 0 < \rho < \frac{1}{2}, \qquad \text{(Wolfe forte 1)}$$
$$|\varphi_k'(\alpha_k)| \leq -\sigma\varphi_k'(0), \quad 0 < \sigma < 1, \quad \sigma > \rho. \qquad \text{(Wolfe forte 2)}$$

Théorème 3.6.1. *Si α_k vérifie (Wolfe forte 1) et (Wolfe forte 2), alors α_k vérifie (Wolfe 1) et (Wolfe 2).*

preuve. (Wolfe forte1) et (Wolfe1) sont les mêmes propositions. Reste à montrer que

$$\text{(Wolf forte 2)} \Longrightarrow \text{(Wolfe 2)} .$$

En effet. Supposons que α_k vérifie (Wolfe forte 2). Alors on a

$$\sigma\psi_k'(0) \leq \varphi_k'(a_k) \leq -\sigma\varphi_k'(0).$$

Par consequent on a

$$\varphi_k'(\alpha_k) \geq \sigma\varphi_k'(0).$$

Ceci implique que α_k verifié (Wolfe 2). □

Remarque 3.6.1. *Supposons qu'on a une suite de paramétres $\{\sigma_j\}_{j\in N}$ telle que*

$$\sigma_j \to 0, j \to \infty$$

Fixons $k \in \mathbb{N}$. Supposons aussi que pour k fixé, (Wolfe forte 1) et (Wolfe forte 2).

On obtient donc une suite $\{\alpha_{k,j}\}_{kfixe, j\in\mathbb{N}}$

$$\forall j \in \mathbb{N} \sigma_j \varphi_k'(0) \leq \varphi_k'(\alpha_{kj}) \leq -\sigma_j \varphi_k'(0).$$

Ceci implique

$$\varphi_k'(\alpha_{kj}) \underset{j\to\infty, kfixe}{\to} 0.$$

La relation précédente implique que

$$\varphi_k'(\alpha_k) \underset{k\to\infty}{\to} 0.$$

Si φ_k' *est continue et si* $\alpha_j \to \bar{\alpha}$ *alors* $\bar{\alpha} = \alpha^*$*.*

Conclusion

Dans la rechreche linéaire de Wolfe forte, si on choisit σ trés petit, on a beucoup de chances de tomber sur α_k proche de α^*. L'inconvénient est que ce choix (σ trés petit) entraine beaucoup d'itérations pour parvenir à obtenir an vérifiant (Wolfe forte 1) et (Wolfe forte 2). On pourrait choisir une suite $\{\sigma_j\}$ telle que $\sigma_j \to 0$ lentement. Ces choix influent sur la convergence et la vitesse de convergence.

Chapitre 4

Méthode de la plus forte pente

4.1 Méthode de la plus forte pente

Définition 4.1.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable. Soient $x \in \mathbb{R}^n$, et $d^* = -\nabla f\left(x_k\right)$.*
Alors pour tout $d \in \mathbb{R}^n$ tel que $\| d \| = \| \nabla f(x) \|$, on a

$$d^T \nabla f(x) \leq {d^*}^T f(x) = \nabla f(x)^T \nabla f(x),$$

et la direction du gradient est celle dans laquelle la fonction a la plus forte pente .

Théorème 4.1.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$, une fonction différentiable au point x, et supposons que $\nabla f(x) \neq 0$. Considérons le problème optimal*

$$\underset{\|d\| \leq 1}{Minimiser} f'(x, d),$$

où $f'(x, d)$ est la dérivée directionnelle de f au point a et dans la direction d. Alors la solution optimale de ce problème est donnée par

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

preuve. Puisque

$$f'(x, d) = \lim_{\lambda \to 0_+} \frac{f\left(x + \lambda d\right) - f(x)}{\lambda} = \nabla f(x)^t d$$

Notre probléme revient donc à minimiser $\nabla f(x)^t d$ dans $\{d : \|d\| \leq 1\}$. L'inégalité de Shwartz donne

$$\left|\nabla f(x)^t d\right| \leq \|\nabla f(x)\| \|d\|. \tag{4.1}$$

Si

$$\nabla f(x)^t d \geq 0$$

on a bien sur

$$\nabla f(x)^t d \geq -\|\nabla f(x)\| \|d\|.$$

Si

$$\nabla f(x)^t d \leq 0$$

(4.1) implique que

$$-\nabla f(x)^t d \leq \|\nabla f(x)\| \|d\|.$$

Par conséquent on a toujours

$$\nabla f(x)^t d \geq -\|\nabla f(x)\| \|d\|.$$

Pour $\|d\| \leq 1$, on a

$$\|\nabla f(x)\| \|d\| \leq \|\nabla f(x)\| \Rightarrow -\|\nabla f(x)\| \|d\| \geq -\|\nabla f(x)\|.$$

Donc $\forall d; \|d\| \leq 1$ on a

$$\nabla f(x)^t d \geq -\|\nabla f(x)\|.$$

D'autre part on $a\|\vec{d}\| = 1$ et $\bar{d}$ vérifie

$$\nabla f(x)^t \bar{d} = \nabla f(x)^t \left(-\frac{\nabla f(x)}{\|\nabla f(x)\|}\right) = -\|\nabla f(x)\|.$$

$\square$

Interprétation du théorème 4.1.1 : Nous allons à partir du théoreme 4.1.1 donner une idée intuitive sur l'appelation méthode de la plus forte pente. En effet d'après le théorème 4.1.1 on a

$$f'(x, d) \geq f'(x, \bar{d}) \forall d, \quad \|d\| \leq 1.$$

Soit en utilisant la définition de la dérivée directionnelle

$$\lim_{\lambda \to 0_+} \frac{f(x + \lambda d) - f(x)}{\lambda} \geq \lim_{\lambda \to 0_+} \frac{f(x + \lambda \widetilde{d}) - f(x)}{\lambda}.$$

Cette dernière inégalité implique qu' il existe $\delta > 0$ tel que

$$[f(x + \lambda d) - f(x)] - [f(x + \lambda \widetilde{d}) - f(x)] \geq 0, \forall \lambda \in] - \delta, +\delta[,$$

ou encore

$$f(x + \lambda d) \geq f(x + \lambda \widetilde{d}), \quad \forall \lambda \in] - \delta, +\delta[\text{ et } \forall d, \|d\| \leq 1.$$

4.2 L'Algorithme de la méthode de la plus forte pente

Etape intiale : Choisir un $\varepsilon > 0$. Choisir un point initial x_1. Poser $k = 1$ et aller à l'étape principale.
Etape principale : Si $\|\nabla f(x)\| < \varepsilon$ stop. Sinon poser $d_k = -\nabla f(x_k)$ et soit λ_n la solution optimale de la recherche linéaire

$$\text{Min} \{ f(x_k + \lambda d_k) ; \lambda \geq 0 \} .$$

Poser $x_{k+1} = x_k + \lambda_k d_k$. Remplacer k par $k + 1$ et répeter l'étape principale.

4.3 Inconvénients de la méthode de la plus forte pente

4.3.1 Lenteur de la méthode au voisinage des points stationnaires

Cette méthode travaille de façon performante dans les premières étapes de l'algorithme. Malheuresement, dés qu'on s'approche du point stationnaire, la méthode devient tres lente. On peut expliquer intuitivement ce phénomène par les considérations suivantes

$$f(x_k + \lambda d) = f(x_k) + \lambda \nabla f(x_k)^t d + \lambda \|d\| \alpha (x_k; \lambda d),$$

où $\alpha(x_k; \lambda d) \to 0$ quand $\lambda d \to 0$.
Si $d = -\nabla f(x_k)$, on obtient :

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

et par conséquent

$$f(x_{k+1}) - f(x_k) = \lambda \left[- \|\nabla f(x_k)\|^2 + \|\nabla f(x)\| \alpha (x_k; \lambda \nabla f(x_k)) \right].$$

D'aprés l'expression précedente, on voit que lorsque x_k s'approche d'un point stationnaire, et si f est continuement différentiable, alors $\|\nabla f(x_k)\|$ est proche de zéro. Donc le terme à droite s'approche de zéro, indépendemment de λ, et par conséquent $f(x_{k+1})$ ne s'eloigne pas beaucoup de $f(x_k)$ quand on passe du point x_k au point x_{k+1}.

4.3.2 Le phénomène de Zigzaguing

Il n'est pas facile de vérifier que pour la méthode du gradient on a toujours

$$d_k^t \cdot d_{k+1} = 0,$$

c'est à dire que la suite $\{x_k\}$ engendrée par l'algorithme de la méthode du gradient, zigzague. Ceci crée un phénomène de ralentissement dans l'acheminement des points x_k vers la solution optimale.

4.4 Quelques remèdes

4.4.1 Changement de direction

Au lieu de prendre comme direction de descente, la direction :

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

on prend des directions de la forme

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

où D est une matrice choisie convenablement (D pourrait être par exemple l'inverse de la matrice hessienne au point x_k c'est à dire $\left(H(x_k)\right)^{-1}$). Un autre choix pourrait s'operer de la façon suivante

$$d_k = -\nabla f(x_k) + g_k$$

où g_k est un vecteur approprié.

4.4.2 Accéleration de la convergence

On peut aussi accelerer ls comergence de la méthode du gradient. Pour celà on transforme grace à un algorithme d'acceleration de la convergence la suite $\{x_k\}$ en une suite $\{y_k\}$ qui convergerait vers la même limite que la suite $\{x_k\}$, mais convrgerait plus rapidement. Si on note par x^* cette limite commune on exprime cette rapidité par la limite suivante

$$\lim_{k\to\infty} \frac{y_k - x^*}{x_k - x^*} = 0$$

4.4.3 Tests numériques

Nous allons tester l'algorithme "steepest descent" avec diffréntes recherches linéaires, comme fonction test on prend la vallée - banane - de Rosenbrock.
Il sagit d'une fonction deux variables du quatriéme degré qui se présente comme

une vallée en U dont le fond assez plat est incurvé suivant une parabole.

$$f(x,y) = 100.(x^2 - y)^2 + (x-1)^2$$

Tableau 2.1 : Steepest descent avec recherche linéaire exacte

k	d_k	x_k	$f(x_k)$	$\|g(x_k)\|$
1	$-$	$(-1.20, 1.000)$	24.200	232.867
10	4.777×10^{-03}	$(-1.009, 1.019)$	4.037	2.928
100	3.811×10^{-03}	$(-0.771, 0.596)$	3.138	3.092
200	8.999×10^{-03}	$(0.147, 0.018)$	0.727	1.642
500	1.634×10^{-03}	$(0.807, 0.650)$	3.726×10^{-02}	0.220
1000	1.243×10^{-03}	$(0.944, 0.891)$	3.095×10^{-03}	5.547×10^{-02}
1500	1.161×10^{-03}	$(0.981, 0.962)$	3.590×10^{-04}	1.826×10^{-02}
2000	1.136×10^{-03}	$(0.993, 0.986)$	4.538×10^{-05}	6.4229×10^{-03}

Tableau 2.2 : Steepest descent avec recherche linéaire d'Armijo

k	d_k	x_k	$f(x_k)$	$\|g(x_k)\|$
1	$-$	$(-1.200, 1.000)$	24.200	232.867
10	2.499×10^{-03}	$(-1.016, 1.035)$	4.067	3.239
100	4.999×10^{-03}	$(0.360, 0.123)$	0.413	1.399
200	2.499×10^{-03}	$(0.683, 0.466)$	0.100	0.409
500	2.499×10^{-03}	$(0.930, 0.865)$	4.841×10^{-03}	0.124
1000	1.250×10^{-03}	$(0.976, 0.953)$	5.501×10^{-04}	2.304×10^{-02}
1500	2.499×10^{-03}	$(0.988, 0.977)$	1.225×10^{-04}	1.390×10^{-02}
2000	4.999×10^{-03}	$(0.994, 0.989)$	2.838×10^{-05}	1.037×10^{-02}

4.5 Convergence de la méthode de la plus forte pente

4.5.1 Cas des recherches linéaires exactes

Avant de donner le théorème principal concernant la convergence de la méthode de la plus forte pente avec des recherches linéaires exactes, démontrons d'abord ce réaultat intermédiaire.

Théorème 4.5.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ et $L \subset \mathbb{R}$ un intervalle fermé et soit M l'application multivoque suivante*

$$\begin{aligned} M : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ (x, d) &\longrightarrow M(x, d) = \{y : y = x + \bar{\lambda} d\} \\ \text{où} \qquad f\left(x + \bar{\lambda} d\right) &= \min\{f(x + \lambda d);\ \lambda \in L\}. \end{aligned}$$

Si f est continue en x et si $d \neq 0$, alors M est férmée au point (x, d).

preuve. Soit $\{x_k d_k\}_{k\in\mathbb{N}}$ et $\{y_k\}_{k\in\mathbb{N}}$ telles que

$$(x_k, d_k) \underset{k\to\infty}{\longrightarrow} (x, d)\,, y_k \in M\,(x_k, d_k)\,, y_k \underset{k\to\infty}{\longrightarrow} y. \tag{4.2}$$

Démontrons que

$$y \in M(x, d).$$

En effet $y_k \in M\,(x_k, d_k)$. Alors

$$y_k = x_k + \overline{\lambda_k} \cdot d_k, \tag{4.3}$$

$\overline{\lambda_k}$ est solution optimale de la recherche linéaire

$$f\left(x_k + \overline{\lambda_k} \cdot d_k\right) = \min\left\{f\left(x_k + \lambda \cdot d_k\right); \lambda \in L\right\}. \tag{4.4}$$

Puisque

$$d_k \underset{n\to\infty}{\longrightarrow} d \neq 0,$$

alors

$$d_k \neq 0, \quad \forall k \geq k_0, \quad k_0 \in \mathbb{N}, \tag{4.5}$$

(4.2) et (4.4) impliquent

$$\overline{\lambda_k} = \frac{\|y_k - x_k\|}{\|d_k\|}, \tag{4.6}$$

en remarquant que L est férmé, on a

$$\overline{\lambda_k} \underset{k\to\infty}{\longrightarrow} \bar{\lambda} = \frac{\|y - x\|}{\|d\|} \in L. \tag{4.7}$$

$(4.1), (4.2), (4.5)$ et (4.6) donnent

$$y = x + \bar{\lambda}d,$$

(4.3) implique

$$f\left(x_k + \overline{\lambda_k}d_k\right) \leq f\left(x_k + \lambda d_k\right).$$

Faisons tendre k vers l'infini et prenons en considération le fait que f est continue au point (x, d), on obtient

$$f\left(x + \bar{\lambda}d\right) \leq f\left(x + \lambda d\right), \quad \forall \lambda \in L.$$

Ceci veut dire exactement que le point $y = x + \bar{\lambda}d \in M(x, d)$. $\square$

4.6 Théorème de Zangwill

Définition 4.6.1. *Soit $h : X \to R$ est une fonction de descente (relativement une application) si elle est continue et possède les propriétés suivantes*
$x \notin \Omega \Longrightarrow h(y) < h(x)\ \forall y \in A(x)$,
$x \in \Omega \Longrightarrow h(y) < h(x)\ \forall y \in A(x)$.

Théorème 4.6.1. *Soit $(\mathcal{P})$ le probléme d'optimisation suivant*

$$(\mathcal{P}) \qquad \min_{x \in X} f(x)$$

et Ω l'ensemble des solutions.
Soit $A : X \overset{m}{\to} X$ une application mutivoque et considérons une suite $\{x_k\}$ engendrée par l'algorithme, c'est à dire vériant $x_{k+1} \in A(x_k)$
Supposons que les trois conditions suivantes soient vérifiées
***C1-** Les points (x) sont tous contenus dans un ensemble compact $K \subset X$.*
***C2-** Il existe une fonction de descente h.*
***C3-** L'application multivoque A est fermée dans $X \backslash \Omega$, $\forall\ x \in X \backslash \Omega$, $A(x) \neq \emptyset$.*
Alors pour tout x limite d'une sous-suite convergente de la suite $\{x_k\}$, on a $x \in \Omega$ et $h(x_k) \underset{k \to \infty}{\overset{k \in \mathbb{N}}{\to}} h(x)$.
Exemple la condition C1 :
Si f est une fonction de descente, on suppose que l'ensemble

$$X_{f(x_0)} = \{x/x \in X : f(x) \leq f(x_0)\}$$

est bornée.

Exemple de la condition C2 :

On peut prendre la fonction de descente égale à $f(x)$ ou $|\nabla f(x)|$.

preuve. Soit $\{x_k\}_{k\in\mathbb{N}}$ une suite engendrée par l'algorithme telle que $\{x_k\}_{k\in\mathbb{N}} \subset K$, il existe alors une sous-suite $\{x_k\}_{\substack{k\in N_1 \\ N_1\subset\mathbb{N}}}$ convergente

$$x_k \underset{k\to\infty}{\overset{k\in N_1}{\longrightarrow}} x, x \in K.$$

La continuité de h donne

$$h(x_k) \underset{k\to\infty}{\overset{k\in N_1}{\longrightarrow}} h(x).$$

$\forall\varepsilon > 0, \exists\nu(\varepsilon) : \forall k \geq \nu(\varepsilon), k \in N_1$ on a $h(x_k) - h(x) \leq \varepsilon$.

Pour tout $k \geq \nu(\varepsilon), k \in \mathbb{N}$ on a

$$h(x_k) - h(x) = h(x_k) - h(x_\nu) + h(x_\nu) - h(x) \leq \varepsilon,$$

car $h(x_k) - h(x_\nu) \leq 0$ pour tout $k \geq \nu(\varepsilon)$ et $h(x_\nu) - h(x) \leq \varepsilon$.

Montrons que $x \in \Omega$

Pour cela considérons la suite $\{x_{k+1}\}_{\substack{k\in N_1 \\ N_1\subset\mathbb{N}}}$. Comme $\{x_{k+1}\}_{\substack{k\in N_1 \\ N_1\subset\mathbb{N}}} \subset K$, il existe alors une sous-suite $\{x_{k+1}\}_{\substack{k\in N_2 \\ N_2\subset\mathbb{N}}}$ telle que

$$x_{k+1} \underset{k\to\infty}{\overset{k\in N_2}{\longrightarrow}} x'$$

Donc

$$x_k \underset{k\to\infty}{\overset{k\in N_2}{\longrightarrow}} x$$

$$x_{k+1} \underset{k\to\infty}{\overset{k\in N_2}{\longrightarrow}} x'.$$

Supposons que $x \notin \Omega$. A étant fermée en x, on en déduit que $x' \in A(x)$. Comme $h(x_k) \underset{k\to\infty}{\overset{k\in N}{\rightarrow}} h(x)$ et en particulier $h(x_{k+1}) \underset{k\to\infty}{\overset{k\in N}{\longrightarrow}} \mathrm{h}(x')$, l'inégalité stricte $h(y) < h(x)$ pour $y = x'$ n'est pas vérifiée. Ceci est en contradiction avec la définition de la fonction de descente donc $x \in \Omega$. $\square$

Théorème 4.6.2. *(convergence de la méthode de la plus forte pente avec des recherches linéaines exactes) Soit* $f : \mathbb{R}^n \to \mathbb{R}$*, telle que* $f \in C^1(\mathbb{R}^n)$ *et* $(\mathcal{P})$ *le problème de minimisation sans contraintes suivant*

$$\min\{f(x) : x \in \mathbb{R}^n\} \qquad (\mathcal{P})$$

On suppose que l'ensemble $\delta(\alpha_0)$ *suivant*

$$\delta(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}, \quad x_0 \in \mathbb{R}^n,$$

est borné. Soit $\{x_k\}$ *une suite génerée par l'algorithme de la plus forte pente, c'est à dire*
Initialisation : $x_0 \in \mathbb{R}^n$*, point initial, poser* $k = 0$ *et aller à Etape principale*
Etape principale : *Si* $\nabla f(x_k) = 0$ *Stop.*
Sinon
Calculer λ_k *verifiant* $f(x_k + \lambda_k d_k) = \text{Min}\{f(x_k + \lambda d_k); \lambda \geq 0\}$
Calculer $x_{k+1} = x_k - \lambda_k \nabla f(x_k)$
Poserk $= k + 1$ *et aller d Etape principale.*
Fin
Soit x^* *une limite quelconque d'une sous suite convergente de* $\{x_k\}$*. Alors*

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

preuve. c'est une application du théorème de Zangwill , en prenant en considération le théorème 4.5.1. En prenant en considération les notations du théorème de Zangwill, on considère dans notre cas

$$X = \mathbb{R}^n, \Omega = \{x : \nabla f(x) = 0\},$$
$$A : \mathbb{R}^n \longrightarrow \mathbb{R}^n : A = M \circ D,$$

avec

$$\begin{array}{ll} D : \mathbb{R}^n & \longrightarrow \mathbb{R}^n \times \mathbb{R}^n \\ x & \longrightarrow D(x) = (x, -\nabla f(x)) \end{array}$$

et

$$M : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$(x, d) \longrightarrow M(x, d) = x + \bar{\lambda} d,$$

$\bar{\lambda}$ vérifiant

$$f\left(x + \bar{\lambda} d\right) = Min\left\{f(x + \lambda d) ; \lambda \geq 0\right\}.$$

Ceci étant l'application multivoque A s'écrire sous la forme

$$\begin{aligned} A : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x \quad &\longrightarrow A(x) = x - \bar{\lambda} \nabla f(x) \end{aligned}$$

$\bar{\lambda}$ vérifiant

$$f(x - \bar{\lambda} \nabla f(x)) \leq f(x - \lambda \nabla f(x)) : \forall \lambda \in [0, +\infty[.$$

M est fermée au point (x, d). A étant la composé d'une fonction univoque continue (D) et d'une fonction multivoque fermée (M) est fermée. La suite $\{x_k\}_{k \in N}$ vérifie

$$x_0 \in \mathbb{R}^n, \quad x_{k+1} \in \mathcal{A}(x_k), \quad k = 0, 1, \ldots$$

Pour la fonction de descente on prend la fonction objectif f (notée h dans lénoncé du théoreme de Zangwill). Vérifians que c'est effectivement une fonction de descente. En effet

i) Si $x \in \Omega$, alors $\nabla f(x) \neq 0$ et par conséquent $-\nabla f(x)$ est une direction de descente. Soit

$$y \in A(x) : y = x - \bar{\lambda} \nabla f(x).$$

Si on prend en considération la définition de direction de descente et le fait que $\bar{\lambda}$ est solution optimale dans tout $[0, +\infty[$ on aura nécessairement

$$f(y) < f(x).$$

ii) Si $x \in \Omega$, alors $\nabla f(x) = 0$ et par conséquent vu que

$$y = A(x) = x - \bar{\lambda} \nabla f(x)$$

on aura

$$y = x.$$

Ceci implique

$$f(y) = f(x).$$

Toutes les conditions du théorème de Zangwill sant vérifiées. Par conséquent si la suite $\{x_k\}_{k\in N}$ est infinie, alors toute limite x^* d'une sous suite convergente de la suite $\{x_k\}_{k\in N}$ vérifie

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

Si $\{x_k\}_{k\in\mathbb{N}}$ s'arrète à l'itération k_0, on a aussi

$$\nabla f\left(x_{k_0}\right) = 0.$$

□

4.6.1 Cas des recherches linéaires inexactes

Algorithme

Initialisation Fixons $\bar{\lambda} > 0$, $0 < \epsilon < 1$ et $x_0 \in \mathbb{R}^n$ initial, poser $K = 0$ et aller à Etape principale.

Etape principale A l'iteration k défnissons la direction $d_k = -\nabla f(x_k)$ et considérons la fonction d'Armijo

$$\hat{\theta}(\lambda) = \theta(0) + \lambda\epsilon\theta'(0)$$

où $\theta(\lambda)$ est la fonction suivante

$$\theta(\lambda) = f(x_k + \lambda d_k) = f[x_k - \lambda\nabla f(x_k)] : \quad \lambda \geq 0.$$

Si $\nabla f(x_k) = 0$ stop.

Sinon

Trouver le plus petit entier $t \geq 0$ tel que

$$(\bar{\theta}\frac{\hat{\lambda}}{2^t}) \leq (\hat{\theta}\frac{\hat{\lambda}}{2^t})$$

et défnir le successeur x_{k+1} dex_k comme suite

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

avec

$$\lambda_k = \frac{\bar{\lambda}}{2^t}$$

Poser k = k + 1 et aller à Etape principale.
Fin.

Théorème 4.6.3. *(convergence de la methode de la plus forte pente avec la recherche linéaire d'Armijo) . Soit* $f : \mathbb{R}^n \to \mathbb{R}$*, telle que* $\nabla f(x)$ *est continuement Lipschitzien de constante G dans l'ensemble* $\delta(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}, x_0 \in \mathbb{R}^n$ *quelconque. Considérons l'algorithme de la méthode de la plus forte pente avec la recherche linéaire d'Armijo, c'est à dire l'algorithme défini précédent.*
Soit x_k *une suite générée par cet algorithme. Alors ou bien il s'arrete aprés un nombre fini d'itérations en un point* x_{k_0} *tel que* $\nabla f(x_{k_0})$=*O, ou bien*

$$\nabla f(x_k) \to 0$$

quand $k \to +\infty$*.*

Conclusion

Une idée naturelle consiste à suivre la direction de plus forte descente et à faire un pas qui rende la fonction à minimiser la plus petite possible dans cette direction. Cette méthode est appelée **méthode de la plus forte pente**.
La méthode de la plus forte pente est une sorte d'idéalisation d'une part, nous ne avons pas en pratique calculer de façon exacte un point minimum α_k de l'objectif dans une direction donnée et le problème n'est en général pas trivial. Pour ces raisons, on peut lui préférer parfois la méthode de la plus forte pente.

Bibliographie

[1] L. Armijo, Minimzation of function having lipschitz continous first partial derivatives, Pacific Journal of Mathematics, Vol. 16(1966), pp. $1-3$.

[2] L. Andrieu, Optimisation sous contrainte en probabilité, THESE présentée pour l'obtention du titre de Docteur de l'Ecole Nationale des Ponts ct Chaussées, décembre (2004)

[3] M. Belloufi, R. Benzine, and Y. Laskri, Modification of the Armijo line search to satisfy the convergence properties of HS method, An International Journal of Optimization and Control : Theories & Applications $(IJOCTA), 141(2013), 145-152$.

[4] A. Cauchy, Analyse mathématique, Méthode générale pour la résolution des systémes d'équations simultanées, Comptes Rendus de l'académie des Sciences de Paris, (1847) t-25, pp. 536-538.

[5] Y. H. Dai and Y. Yuan, Convergence properties of the conjugate descent method, Adv. Math. (China), 26(1996), pp. $552-562$.

[6] Y. H. Dai and Y. Yuan, Convergence of the Fletcher-Reeves method under a generalized Wolfe wearch, J. Comput. Math., 2 (1996), pp. $142-148$.

[7] A. A. Goldstein, On steepest descent, SIAM J. on Control A, Vol. 3, No. 1 (1965), pp. 147-151.

[8] Michel Bierlaire, Introduction à l'optimisation différentiable.

[9] Pr Rachid Benzine 2007, Optimisation convexe, optimisation sans contrainte.

[10] M. S. Bazaraa and H. D. Sherali, et C. M. Shetty, Nonlinear Programming. Theory and Algorithms, Wiley-Interscience(1993).

[11] H. P. Crowder and P. Wolfe, Linear convergence of the conjugate gradient method, IBM J. Res. Dev., 16(1969), pp. $431 - 433$.

[12] X.D. Chen and J. Sun, Global convergence of a two-parameter family of conjugate gradient methods without line search, Journal of Computational and Applied Mathematics 146(2002)$37 - 45$.

[13] Y. H. Dai, New properties of a nonlinear conjugate gradient method, Numer. Math. 89(2001), pp. $83 - 98$.

[14] Y. H. Dai, A nonmonotone conjugate gradient algorithm for unconstrained optimization, J. Syst. Sci. Complex., 15(2002), pp. $139 - 145$.

[15] Y. H. Dai, J. Y. Han, G. H. Liu. D. F. Sun, H. X. Yin, and Y. Yuan. Convergence properties of nonlinear conjugate gradient methods, SIAM J. Optim., 10 (1999). pp. $345 - 358$.

[16] Y. H. Dai and L. Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, Appl. Math. Optim., 43(2001), pp. $87 - 101$

[17] Y. H. Dai and Q. Ni, Testing different conjugate gradient methods for large-scale unconstrained optimization, J. Comput. Math., 21 (2003), pp. $311 - 320$.