الجمهوريـة الجزائـريـة الديمقراطيـة الشعبيـة
République Algérienne Démocratique et Populaire
وزارة التعليـم العالـي والبحـث العلمـي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

N° Réf :……………

Centre Universitaire

Abd elhafid boussouf Mila

Institut des sciences et de la technologie **Département de Mathématiques etInformatique**

Mémoire préparé En vue de l'obtention du diplôme de Licence

en: - Filière mathématiques

Algorithme d'optimisation avec et sans contrainte

Préparé par :
- **Medjdoub Abd elmounaim**
- **Sekfali Fatima**
- **Rachedi Asma**

Encadrer par :
Benaouicha Loubna

Année universitaire :2014/2015

Remerciements

Nous tenons à remercier tout d'abord dieu, le tout puissant pour la volonté et le courage qu'il m'a donné pour mener à terme ce travail.

Nous tenons à remercier notre encadreur madam Benaouicha Loubna, qui nous en couragé en nous faisant part d'observations constructives et pour ses préscieux conseils.

Nous remercions nos parents, pour tous les sacrifices qu'ils ont consentis pour nous permettre de suivre nos études dans les meilleures conditions et de

nous avoir encouragé tout au long de ces années.

En fin, Nous tenons à remercier tout les amis au département de mathématiques et information, et ailleurs, pour leur disponibilité et leur soutien.

Table des matières

Introduction

L'optimisation est une branche des mathématique cherchant à analyse et à résoudre des problèmes de décision qui consiste généralement à déterminer le meilleur élément décision d'un ensemble afin de maximiser ou de minimiser un critère qualitatif qui mesure la qualité de la décision.

Le mot optimisation vient de latin « optimon » qui signifie la meilleur, l'optimisation joue un rôle important en économie financier, en analyse numérique, en statistique (l'estimation de vraisemblance d'une distribution), la recherche de stratégie dans le cadre du théorème des jeux et aussi en théorème de control.

Ce mémoire est structuré en l'introduction générale, trois chapitres et conclusion générale.

Dans le premier chapitre, nous avons donné des définitions d'un problème d'optimisation, rappels et notion de calcul différentiel.

Dans le deuxième chapitre, nous avons présenté l'algorithme d'optimisation sous contrainte, ces définitions les conditions d'optimalité sans contrainte, les résultats d'existence et d'unicité, et méthode de descente.

Dans le dernière chapitre, nous avons présenté algorithme d'optimisation avec (sous) contrainte, ces définition condition d'optimalité simple et d'obtenir les résultats d'existence et d'unicité ensuite la méthode du gradient avec projection.

Notation

- $\mathbb{R}$: l'ensemble des nombres réels.
- ∂f : les dérivés partiele de la fonction f.
- H : l'espace de Hilbert .
- $\|\cdot\|$: la norme.
- $\langle \ , \ \rangle$: le produit scalaire .
- $B(x^{*}, \alpha)$: la boule ouverte de centre x^{*} et de rayon α.
- $D(A)$ domaine de A.
- K^{0} intérieur de K.

Chapitre 1

Optimisation

1.1 Définitions et rappels de calcul différentiel

1.1.1 Définitions et résultats préliminaires

Définition 1.1.1 *(La norme)*

Soit E un espace vectoriel sur un corps K $(K = \mathbb{R}$ ou $K = C)$
une norme sur E est une application $N : E \longrightarrow \mathbb{R}_+$ vérifiant les propriétés suivantes :
1. $\forall x \in E,\ N(x) = 0 \Longleftrightarrow x = 0.$
2. $\forall x \in E, \forall \lambda \in K,\ \ N(\lambda x) = |\lambda|\, N(x).$
3. $\forall x,\ y \in E,\ \ N(x + y) \leq N(x) + N(y).$

Pour $x \in E$, $N(x)$ est appelé norme sur E et notée $\|x\|$.

Définition 1.1.2 *(Fonction linéaire)*

Soit $f : E \longrightarrow F$ une fonction, on dit que f est un fonction linéaire, vérifiant les propriétés suivantes :
1. $\forall \lambda \in K,\ f(\lambda x) = \lambda f(x).$
2. $\forall\ x,\ y \in E,\ f(x + y) = f(x) + f(y).$

Définition 1.1.3 *(espace de Hilbert)*

. un espace préhilbertien complet pour la norme associe au produit scalaire est appelé espace de hilbert.

. On appelle espace de Hilbert (où Hilbertien) tout espace préhilertien complet. Donc espace de Hilbert et un espace vectoriel muni d'une norme, cet espace est complet pour la norme $N(x) = \sqrt{\langle x, x\rangle}$ et on not H.

Définition 1.1.4 *(espace préhilbertien)*

un espace préhilbertien est un espace vectoriel d'un produit scalaire , c'est-à-dire, d'une application :

$$h : E \cdot E \longrightarrow C$$

verifiant les propriétes suivants :

pour tout $u, v, w \in E$;$\lambda, \mu \in \mathbb{C}$, on a :

1. $h\left[\lambda u+\mu v, w\right]=\lambda h\left[u, w\right]+\mu h\left[v, w\right].$
2. $h\left[u, \lambda v+\mu w\right]=\overline{\lambda} h\left[u, w\right]+\overline{\mu} h\left[v, w\right].$
3. $h\left[u, v\right]=\overline{h\left[v, u\right]}.$
4. $h\left[u, u\right] \geq 0.$
5. $h\left[u, u\right]=0 \Leftrightarrow u=0.$

les propriétes (1) et (2) caractérisent les forme ses qui-lineaires, si de plus h verifie (3), on dit qu'elle est hermetienne.

Définition 1.1.5 *(fonction convexe)*

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est dite covexe si pour tout x, $y \in \mathbb{R}^n$, et pour tout $\lambda \in [0,1]$, on a :

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

Définition 1.1.6 *(fonction strictement convexe)*

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est dite strictement convexe si pour tout x, $y \in \mathbb{R}^n$, $x \neq y$, et pour tout $\lambda \in \left]0,1\right[$, on a :

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

Définition 1.1.7 *(matrice hessinne)*

Soit : $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction deux fois différentiable. La fonction notée

$$\nabla^2 f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^{n*n}$$

est appelée matrice hessienne de f, définie par :

$$\nabla^2 f(x) = \begin{pmatrix} \partial^2 f(x) / \partial x_1^2 & \partial^2 f(x) / \partial x_1 \partial x_2 & \cdots & \partial^2 f(x) / \partial x_1 \partial x_n \\ \partial^2 f(x) / \partial x_2 \partial x_1 & \partial^2 f(x) / \partial x_2^2 & \cdots & \partial^2 f(x) / \partial x_2 \partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 f(x) / \partial x_n \partial x_1 & \partial^2 f(x) / \partial x_n \partial x_2 & \ldots & \partial^2 f(x) / \partial x_n^2 \end{pmatrix}$$

La matrice hessienne est toujours symétrique.

Définition 1.1.8 *La matrice hessienne $Hess(f)_x$ de f au point x est la matrice*

$$Hess(f)_x = (\partial_{ij}^2 f(x)) = \begin{pmatrix} \partial_{11}^2 f(x) & \cdots & \partial_{1n}^2 f(x) \\ \vdots & & \vdots \\ \partial_{n1}^2 f(x) & \cdots & \partial nn^2 f(x) \end{pmatrix}, \quad \begin{matrix} i = 1, .., n \\ j = 1, .., n \end{matrix}$$

Il lui est associé la forme quadratique.

Définition 1.1.9

$$h = (h_1, \cdots, h_n) \in \mathbb{R}^n \longmapsto hess(f)_x(h) = \sum_{i,j=1}^{n} \partial_{ij}^2 f(x) h_i h_j$$

Soit $g = \nabla f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, et $x \in \mathbb{R}^n$, alors $Dg(x) \in M_n(\mathbb{R})$ et on peut définir la matrice hessienne de f, qu'on note H_f, Par :

$$H_{f(x)} = Dg(x) = D(Df)(x) = (b_{i,j})_{i,j=1\ldots n} \in M_n(\mathbb{R}) \textit{où } b_{i,j} = \partial_{i,j}^2 \ f(x) \ \textit{ où } \partial_{i,j}^2$$

désinge la dérivée partielle par rapport à la variable i de la dérivée partielle par rapport à la variable j. Notons que par définition, $Dg(x)$ est la matrice jacobienne de g en x.

Exemple 1.1.10

La hessienne de la fonction polynomiale

$$P(x,y) = ax^2 + 2bxy + cy^2 + dx + ey + f$$

est la matrice (indépendante de x, y)

$$Hess(f)_{(x,y)} = 2\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Soit $f(x,\, y) = xy - ln(x^2 + y^2)$. Déterminer la hessienne au point $(1,2)$.

on suppose que vous connaissez suffisamment bien les opérations sur dérivées pour rendre inutiles les détails de calcul :

$$\frac{\partial f}{\partial x} = y - \frac{2x}{x^2 + y^2}$$

$$\frac{\partial f}{\partial y} = x - \frac{2y}{x^2 + y^2}$$

Les dérivées secondes étant dans cet exemple un peu plus compliquées à établir, je détaille leur calcul. Celui-ci nécessite en effet un petit entraînement quand on est en présence d'une forme $(u/v)\prime = (u\prime v - v\prime u)/v^2)$ car la confusion guette...

$$H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

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

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

Et enfin, dérivons $f\prime(x)$ par rapport à y :

$$\frac{\partial^2 f}{\partial x \partial y} = 1 + \frac{4xy}{(x^2 + y^2)^2} = \frac{\partial^2 f}{\partial x \partial y}$$

Il nous reste à remplacer x par 1 et y par 2, ce qui donne ce spectaculaire résultat :

$$H = \begin{pmatrix} -6/25 & 33/25 \\ 33/25 & 6/25 \end{pmatrix}$$

Définition 1.1.11 *(le gradient)*

Soit $a \in U$. On appelle gradient de f en le vecteur :

$$\nabla f(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}$$

Définition 1.1.12 *(le gradient)*

La forme différentielle totale d'une fonction $f(x, y, z)$, où x, y et z sont les trois variables de l'espace, est :

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz,$$

qui peut s'écrire sous la forme d'un produit scalaire de deux vecteurs $\left(\overrightarrow{u}, \overrightarrow{dl}\right)$ avec :

$$\overrightarrow{u} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix}, \overrightarrow{dl} = \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}$$

Le champ vectoriel $\overrightarrow{u}$ s'exprime par un opérateur nommé gradient que l'on note :

$$\overrightarrow{grad}(f) = \overrightarrow{\nabla} f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}.$$

Définition 1.1.13 *(fonction continue C^0)*

1. Soit f une fonction définie sur un intervalle ouvert I de $\mathbb{R}$ et soit $a \in I$.

2. Si tout intervalle $]\ f(a) - \varepsilon,\ f(a) + \varepsilon[$ contient toutes les valeurs de $f(x)$ dès que x est assez proche de a, (ce qui signifie que $f(x)$ tend vers $f(a)$ quand x tend vers a), on dit que la fonction f est continue au point a, ce qui se traduit par la définition mathématique suivante :

3. $\forall \varepsilon > 0$, $\exists \alpha(\varepsilon) > 0$, $\forall x$; $x \in]a - \alpha, a + \alpha[\Rightarrow f(x) \in]f(a) - \varepsilon,\ f(a) + \varepsilon[$.

Si la fonction f est continue en tout point a de l'intervalle I , on dit qu'elle est continue sur I .

Exemple 1.1.14

Presque toutes les fonction sont continues sur tout intervalle continues dans leur domaine de définition :

1. les fonction valeur absolue est continue sur $\mathbb{R}$.
2. les fonction polynômes sont continues sur $\mathbb{R}$.
3. les fonction rationnelles sont continues sur tout intervalle contenu dans leur domaine de définition.
4. les fonction exponentielles sont continues sur $\mathbb{R}$.
5. les fonction logarithme népérien et logarithme décimal sont continues sur $]0, +\infty[$.
6. les fonction racines sont continues sur $[0, +\infty[$.
7. les fonction sinus et cosinus sont continues sur $\mathbb{R}$, la fonction tangente est continue sur tout intervalle ne contenant pas un nombre de la forme $\frac{\pi}{2} + k\pi$, $k \in \mathbb{Z}$.

Définition 1.1.15 *(fonctions de classe C^1)*

On dit que f est de classe C^1 sur U si toutes ses dérivées partielles sont définies et continues sur U .

Soit V un ouvert de $\mathbb{R}^p$. On dit que f est un C^1 de U dans V si f est une bijection de U dans V, est de classe C^1 sur U, et si sa réciproque f^{-1} est C^1 sur V.

Remarque 1.1.16 *Si f est un C^1-difféomorphisme de U dans V et $W \subset U$ est ouvert, alors $f(W)$ est ouvert comme image réciproque de W par l'application continue f^{-1}.*

Définition 1.1.17 *(problème d'optimisation)*

soit $f \in C\ (\mathbb{R}^n, \mathbb{R})$

Le problème d'optimisation sans conrainte s'écrit :

$$\begin{cases} \text{Trouver } x^* \in \mathbb{R}^n \text{ tel que} \\ f(x^*) \leq f(y),\ \forall y \in \mathbb{R}^n \ . \end{cases} \tag{1}$$

Le problème d'optimisation avec containte s'écrit :

$$\begin{cases} \text{Trouver } x^* \in K \quad \text{tel que} \\ f(x^*) \leq f(y),\ \forall y \in K. \end{cases} \tag{2}$$

Où $K \subset \mathbb{R}^n$ et $K \neq \mathbb{R}^n$

Si x^* est solution du problème (1), on dit que $x^* \in \arg\min_{\mathbb{R}^n} f$, et si x^* est solution du problème (2), on dit que $x^* \in \arg\min_{K} f$.

1.1.2 Définition de problèms standard et canonique

Forme canonique :

Un problème d'optimisation sous la forme canonique se présente sous la forme suivante :

$$\begin{cases} \min \langle c, x\rangle = c^t x \\ Ax \leq b \\ \text{où } x \in \mathbb{R}^n,\ c \in \mathbb{R}^n \\ b \in \mathbb{R}^m \\ A \subset M_{m.n}(\mathbb{R}) \\ x \geq 0 \end{cases}$$

maximiser $c^t x$ sous les contraintes $Ax \leq b$ et $x \geq 0$.

Conduit au même type de problème (en changeant c en $-c$).

Forme standard :

Par contre si le domaine des solutions realisables est défini par un systeme d'equations linéaire on dit que le programme est écrit sous forme standard :

$$\begin{cases} \min c^t x \\ Ax = b \\ x \geq 0 \end{cases}$$

· On passe de la forme canonique (avec A, b, c) à la forme standard de la manière suivante :

$$\begin{pmatrix} \langle a_i, x\rangle \leq b_i \\ \text{dans } \mathbb{R}^n \end{pmatrix} \quad \text{devient} \quad \begin{pmatrix} \text{il existe } U_i \geq 0 \text{ tel que} \\ \langle a_i, x\rangle + U_i = b_i \end{pmatrix}$$

d'ou

$$\begin{pmatrix} Ax \leq b \\ x \geq 0 \\ \text{dans } \mathbb{R}^n \end{pmatrix} \quad \text{devient} \quad \begin{pmatrix} Ax + \operatorname{Im} u = b \\ x \geq 0, u \geq 0 \end{pmatrix}$$

1.1.3 Rappels et notations de calcul différentiel

Soient E et F des espaces vectoriels normés, f une application de E dans F et $x \in E$. On dit que f est différentiable en x s'il existe $T \in \mathcal{L}(E,\, F)$ (ou $\mathcal{L}(E,\, F)$ est l'ensemble des applications linéaires de E dans F) telle que :

$$f(x+h) = f(x) + T(h) + \|h\|\, \varepsilon(h) \text{ avec } \varepsilon(h) \to 0 \text{ quand } h \to 0$$

L'application T est alors unique et on note $Df(x) = T \in \mathcal{L}(E,\, F)$.

On peut remarquer qu'en dimension infinie, T dépend des normes associées à E et F. Voyons maintenant quelques cas particuliers d'espaces E et F :

Cas où $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$ soit $f : \mathbb{R}^n \to \mathbb{R}^p$, $x \in \mathbb{R}^n$ et supposons que f est différentaible en x, alors :

$$\underbrace{Df(x)(y)}_{\in\, \mathbb{R}^P} = \underbrace{Ay}_{\in\, \mathbb{R}^P}, \forall y \in \mathbb{R}^n.$$

On confond alors l'application linéaire $Df(x) \in \mathcal{L}(\mathbb{R}^n,\, \mathbb{R}^p)$ et la matrice $A(x) \in M_{p,n}(\mathbb{R})$ qui la représente. On écrit donc :

$$A(x) = Df(x) = (a_{i,j})_{1 \le i \le p, 1 \le j \le n} \text{ ou } a_{i,j} = \partial_j f_i(x),$$

∂_j désignant la dérivée partielle par rapport à la j-éme variable.

Prenons $n = 3$ et $p = 2$, notons $x = (x_1, x_2, x_3)^t$ et considérons $f : \mathbb{R}^3 \to \mathbb{R}^3$définie par :

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

On vérifiera par le calcul que pour $h = (h_1, h_2, h_3)^t$ on a :

$$D\ f(x)h = \begin{pmatrix} 2x_1h_1 + 3x_2^2h_2 + x_3^3h_3 \\ 2h_1 - h_2 \end{pmatrix}$$

et donc d'apres les notations précédentes :

$$A(x) = \begin{pmatrix} 2x_1 & 3x_2^2 & 4x_3^3 \\ 2 & -1 & 0 \end{pmatrix}$$

Cas où $E = \mathbb{R}^n$, $F = \mathbb{R}$ C'est un sous-cas du paragraphe précédent, puisqu'on est ici dans le cas $p = 1$. Soit $x \in \mathbb{R}^n$ et f une fonction de E dans F différentiable en x ;

on a donc (avec l'abus de notation singualé dans le paraghraphe précédent) $Df(x)^t \in \mathbb{R}^n$. Pour $(x,y) \in (\mathbb{R}^n)^2$, on a donc :

$$Df(x)y = \sum_{j=i}^{n} \partial_j \ f(x) \ y_j = \nabla f(x) \cdot y \text{ où } \nabla f(x) = \begin{bmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{bmatrix} \in \mathbb{R}^n$$

Cas où E est un espace de Hilbert et $F = \mathbb{R}$ On généralise ici le cas présenté au paragraphe précédent. Soit $f : E \to \mathbb{R}$ différentiable en $x \in E$.

Alors $Df(x) \in \mathcal{L}(E, \mathbb{R}) = \acute{E}$, où $\acute{E}$ désigne le dual topologique de E, c.à.d. l'ensemble des formes linéaires continues sur E.

Par le théorème de représentation de Riesz, il existe un unique $u \in E$ tel que $df(x)(y) = (u \mid y)_E$ pour tout $y \in E$, où $(. \mid .)_E$ désigne le produit scalaire sur E. On appelle encore gradient de f en x ce vecteur u. On a donc $u = \nabla f(x) \in E$ et pour $y \in E$,

$$Df(x)(y) = (f(x) \mid y)_E \ .$$

Différentielle d'ordre 2, matrice hessienne Renvenons maintenanr au cas général de deux espaces vectoriels normés E et F, et supposons maintenant que $f \in C^2(E, F)$. Le fait que $f \in C^2(E, F)$ signifie que $Df \in C^1(E, \mathcal{L}(E, F))$.

Par définition, on a $D^2 f(x) \in \mathcal{L}(E, \mathcal{L}(E, F))$ et donc pour $y \in E$,

$D^2 f(x)(y) \in \mathcal{L}(E, F)$; en particulier, pour $z \in E$, $D^2 f(x)(y)(z) \in F$.

Considérons maintenant le cas particulier $E = \mathbb{R}^n$ et $F = \mathbb{R}$. On a :

$$f \in C^2(\mathbb{R}^n, \mathbb{R}) \Longleftrightarrow \left[f \in C^1(\mathbb{R}^n, \mathbb{R}) \ \ et \ \nabla f \in C^1(\mathbb{R}^n, \mathbb{R}^n) \right] .$$

Chapitre 2

Algorithmes d'optimisation sans contrainte (Méthodes de gradient)

2.1 Optimisation sans contrainte

2.1.1 Définition

Soit $f \in C(E, \mathbb{R})$ et E un espace vectoriel normé. On cherche soit un minimum global de f, c.à.d :

$$x^* \in E \quad \text{telque} : f(x^*) \leq f(y), \ \forall y \in E. \tag{3}$$

ou un minimum local, c.à.d. :

$$x^* \in E \text{ tel que } \exists \alpha > 0 \ , \ \ f(x^*) \leq f(y), \forall y \in B(x^*, \alpha). \tag{4}$$

2.1.2 Condition d'optimalité

Proposition 2.1.1 *(Condition nécessaire d'optimalité)*

Soit E un espace vectoriel normé, et soient $f \in C(E, \mathbb{R})$, et $x^* \in E$ tel que f est différentiable en x^*.

Si x^* est solution de (4) alors :

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

Preuve. Supposons qu'il existe $\alpha > 0$ tel que $f(x^*) \leq f(y)$ pour tout $y \in B(x^*, \alpha)$.

Soit $z \in E \setminus \{0\}$, alors si $|t| < \frac{\alpha}{\|z\|}$, on a $x^* + tz \in B(x^*, \alpha)$ (où $B(x^*, \alpha)$ désigne la boule ouverte de centre x^* et de rayon α) et on a donc :

$$f(x^*) \leq f(x^* + tz).$$

Comme f est diférentiable en x^*, on a :

$$f(x^* + tz) = f(x^*) + \nabla f(x^*)(tz) + |t|\varepsilon_z(t).$$

où $\varepsilon_z(t) \to 0$ lorsque $t \to 0$. On a donc $f(x^*) + t\nabla f(x^*)(z) + |t|\varepsilon_z(t) \geq f(x^*)$. Et pour $\frac{\alpha}{\|z\|} > t > 0$, on a $\nabla f(x^*)(z) + \varepsilon_z(t) \geq 0$. En faisant tendre t vers 0, on obtient que :

$$\nabla f(x^*)(z) \geq 0, \forall z \in E.$$

On a aussi $\nabla f(x^*)(-z) \geq 0, \forall z \in E \setminus \{0\}$, et donc : $-\nabla f(x^*)(z) \geq 0, \forall z \in E$.On en conclut que :

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

■

Exemple 2.1.2

Soit l'application suivante :

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x_1, x_2) &\longmapsto f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \end{aligned}$$

on a :

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f(x_1,x_2)}{\partial x_1} \\ \frac{\partial f(x_1,x_2)}{\partial x_2} \end{pmatrix}$$

$$\Longrightarrow \nabla f(x_1, x_2) = \begin{pmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{pmatrix}$$

$$\nabla f(x_1, x_2) = 0 \Longrightarrow \begin{cases} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 = 0......(*) \\ 200x_2 - 200x_1^2 = 0................(**) \end{cases}$$

$$(**) \Longrightarrow 200x_2 = 200x_1^2$$

donc

$$x_2 = x_1^2..........(***)$$

on remplace $(***)$ dans $(*)$

$$\Longrightarrow 400x_1^3 - 400x_1x_1^2 + 2x_1 - 2 = 0$$

$$\Longrightarrow 400x_1^3 - 400x_1^3 + 2x_1 - 2 = 0$$

$$\Longrightarrow 2x_1 - 2 = 0$$

$$\Longrightarrow 2x_1 = 2$$

$$\Longrightarrow x_1 = 1$$

$$(***) \Longrightarrow x_2 = 1$$

donc

$$x^* = (1, 1)$$

Remarque 2.1.3 *Attention, la proposition précédente donne une condition nécessaire mais non suffsante. En effet, $\nabla f(x^*) = 0$ n'entraîne pas que f atteigne un minimum (ou un maximum) même local, en x^*. Prendre par exemple $E = \mathbb{R}$, $x^* = 0$ et la fonction f définie par : $f(x) = x^3$ pour s'en convaincre.*

Proposition 2.1.4 *(Conditions suffsante d'optimalite)*

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ deux fois différentiable dans un sous ensemble ouvert V de $\mathbb{R}^n$ et soit $x^* \in V$ qui vérifie les conditions :

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

et

$$\nabla^2 f(x^*) \text{ est définie positive}$$

i.e : $\forall x \in \mathbb{R}^n, \|x\| \neq 0,\ x^t \nabla^2 f(x^*) > 0$ Alors x^* est point minimum local strict de f .

Exemple 2.1.5

Soit f une application :

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) = 1 - \exp(-x^2) \end{aligned}$$

$$\nabla f(x) = \left(\frac{\partial f\ (x)}{\partial x} \right)$$

$$\Longrightarrow \nabla f(x) = \left(2x \exp\left(-x^2\right)\right)$$

$$\nabla f(x) = 0 \Longrightarrow 2x \exp\left(-x^2\right) = 0$$

$$\Longrightarrow x = 0 \quad et \quad \exp(-x^2) \neq 0$$

Alors

$$x^* = 0$$

$$\nabla^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial^2 x}\right)$$

$$\nabla^2 f(x) = 2\exp(-x^2) - 4x^2\exp(-x^2) = \exp(-x^2)(2 - 4x^2)$$

$$\nabla^2 f(0) = 2 > 0$$

Alors x^* est point minimum local.

Exemple 2.1.6

Soit f une application :

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x_1, x_2) &\longrightarrow f(x_1, x_2) = x_1^2 - x_2^2 - 2x_1x_2 + x_1 \end{aligned}$$

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f(x_1,x_2)}{\partial x_1} \\ \frac{\partial f(x_1,x_2)}{\partial x_2} \end{pmatrix}$$

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2x_1 - 2x_2 + 1 \\ -2x_2 - 2x_2 \end{pmatrix}$$

$$\nabla f(x_1, x_2) = 0 \Longrightarrow \begin{cases} 2x_1 - 2x_2 + 1 = 0 \ldots\ldots\ldots\ldots\ldots (*) \\ -2x_2 - 2x_2 = 0 \ldots\ldots\ldots\ldots (**) \end{cases}$$

$(**) \Longrightarrow x_1 = -x_2 \ldots\ldots\ldots\ldots (***)$

On remplace $(***)$ dans $(*)$ donc :

$$2x_1 - 2x_2 + 1 = 0$$

$$\Longrightarrow x_1 = \frac{1}{4}$$

On a :

$$x_1 = -x_2 \Longrightarrow x_2 = -\frac{1}{4}$$

Donc :

$$x^* = \left(\frac{1}{4}, -\frac{1}{4}\right)$$

$$\nabla^2 f(x_1,x_2) = \begin{pmatrix} \frac{\partial^2 f(x_1,x_2)}{\partial^2 x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_1,x_2)}{\partial^2 x_1} \end{pmatrix}$$

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

Donc :

$$\begin{aligned} x^t \nabla^2 f\left(\frac{1}{4}, -\frac{1}{4}\right) x &= (x_1, x_2) \begin{pmatrix} 2 & -2 \\ -2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} 2x_1 - 2x_2 & -2x_1 - 2x_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= 2x_1^2 - 2x_2x_1 - 2x_1x_2 - 2x_2^2 \end{aligned}$$

donc :

$$x^t \nabla^2 f(\frac{1}{4}, -\frac{1}{4}) x = 2x_1^2 - 4x_1x_2 - 2x_2^2.$$

On peut rien dire .

alors $x^* = \left(\frac{1}{4}, -\frac{1}{4}\right)$ n'est pas un point minimum local .

Proposition 2.1.7 *(Conditions suffisants d'optimalite globale)*

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ deux fois différentiable dans $\mathbb{R}^n$; et $x \in \mathbb{R}^n$ minimum local de f.

Si f est une fonction convexe alors x^* est un minimum global de f .

Si de plus f est structement convexe x^* est unique minimum global de f .

2.1.3 Résultats d'existence et d'unicité :

Théorème 2.1.8 *(Existence)*

Soit $E = \mathbb{R}^n$ et $f : E \to \mathbb{R}$ une application telle que :

(i) f est continue.

(ii) $f(x) \to +\infty$ quand $\|x\| \to +\infty$.

Alors il existe $x^* \in \mathbb{R}^n$ tel que $f(x^*) \leq f(y)$ pour tout $y \in \mathbb{R}^n$.

Remarque 2.1.9

1. Le théorème est faux si E est de dimension infinie (i.e. si E est espace de Banach au lieu de $E = \mathbb{R}^n$), car si E est de dimension infinie, B_R n'est pas compacte.

2. L'hypothèse (ii) du théorème peut être remplacée par :

$$(ii)\prime \quad \exists\, b \in \mathbb{R}^n, \exists\, R > 0,\ telque\ : \|x\| \geq R \Rightarrow f(x) \geq f(b).$$

3. Sous les hypothèses du théorème il n'y a pas toujours unicité de x^* même dans le cas $n = 1$, prendre pour s'en convaincre la fonction f définie de $\mathbb{R}$ dans $\mathbb{R}$ par :

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

Théorème 2.1.10 *(Unicité)*

Soit E un espace vectoriel normé est $f : E \longrightarrow \mathbb{R}$ strictement convexe alors il existe au plus un $x^* \in E$ tel que :

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

Preuve. Soit f strictement convexe, supposont qu'il existe x^*et $x^{**} \in E$ tel que : $f(x^*) = f(x^{**}) = \min_{R^n} f$ Comme f est strictement convexe si $x^* \neq x$ alors :

$$f(\frac{1}{2}x^* + \frac{1}{2}x^{**}) < \frac{1}{2}f(x^*) + \frac{1}{2}f(x^{**}) = \frac{1}{2}f(x^*) + \frac{1}{2}f(x^*) = f(x^*)$$

$$\min_{x \in R^n} f(x) \Longrightarrow f(x^*) < f(\frac{1}{2}x^* + \frac{1}{2}x^{**})(\text{contradiction})$$

Ce qui impossible car $(f(\frac{1}{2}x^* + \frac{1}{2}x^{**}) > \min_{x\in R^n} f(x) = f(x^*))$

Donc :

$$x^* = x^{**}.$$

■

Remarque 2.1.11

1. Ce théorème ne donne pas, l'existence. Par exemeple dans le cas $n = 1$ la fonction f définie par $f(x) = \exp(x)$: n'atteint pas son minimum, car $\min_{\mathbb{R}^n} f = 0$ et $f(x) \neq 0$ pour tout $x \in \mathbb{R}$; pourtant f est strictement convexe.

2. Par contre, si on réunit les hypothèses des théoréme (2.1.8) et (2.1.10), on obtient le résultat d'existense et unicité suivant :

Théorème 2.1.12 *(Existence et Unicité)*

Soit $E = \mathbb{R}^n$ et soit $f : E \longrightarrow \mathbb{R}$. On supposeque :
(i) f continue.
(ii) $f(x) \longrightarrow +\infty$.
(iii) f et strictement convexe.
Alors il existe un unique $x^* \in \mathbb{R}^n$ tel que $f(x^*) = \min_{\mathbb{R}^n} f$.

Remarque 2.1.13 *Ce théorème reste vrai si E est un espace de Hilbert , on a besoin dans ce cas pour la partie existence des hypothèses (i), (ii) et de la convexité de f.*

2.2 Méthodes de descente

2.2.1 Définition

Soient $f \in C(E, \mathbb{R})$ et $E = \mathbb{R}^n$.

1. Soit $x \in E$, on dit que $w \in E \backslash \{0\}$ est une direction de descente en x s'il existe $\rho_0 > 0$ tel que :

$$f(x + \rho w) \leq f(x), \forall \rho \in [0, \rho_0].$$

2. Soit $x \in E$, on dit que $w \in E \backslash \{0\}$ est une direction de descente stricte en x s'il existe $\rho_0 > 0$ tel que :

$$f(x + \rho w) < f(x), \forall \rho \in \ ]0, \rho_0].$$

3. Une “méthode de descente" pour la recherche de x^* tel que $f(x^*) = \min\limits_E f$ consiste à construire une suite $(x_n)_n$ de la manière suivante :
(a) Initialisation $x_0 \in E$.
(b) Itération n : on suppose $x_0 ... x_n$ connus $(n \geq 0)$.
i). On cherche w_n direction de descente stricte de x_n.
ii). On prend $x_{n+1} = x_n + \rho_n w_n$ avec $\rho_n > 0$“bien choisi".

Proposition 2.2.1 *Soient $E = \mathbb{R}^n$, $f \in C^1(E, \mathbb{R})$, $x \in E$ et $w \in E \backslash \{0\}$; alors :*

1. Si w direction de descente en x alors : $w \cdot \nabla f(x) \leq 0$.
2. Si $\nabla f(x) \neq 0$ alors $w = -\nabla f(x)$ est une direction de descente stricte en x.

2.2.2 Algorithme du gradient à pas fixe

Soient $f \in C^1(E, \mathbb{R})$, et E $= \mathbb{R}^n$

On se donne $\rho > 0$

$$\left\{\begin{array}{c} \text{Initialisation : } x_0 \in E, \\ \text{Itérationn : } x_n \text{ connu}, (n \geq 0) \\ w_n = -\nabla f(x_n), \\ x_{n+1} = x_n + \rho w_n. \end{array}\right. \tag{5}$$

Théorème 2.2.2 *(Convergence du gradient à pas fixe)*

Soient $E = \mathbb{R}^n$ et $f \in C^1(E, \mathbb{R})$ On suppose que :

1.$\exists\ \alpha > 0$ tel que $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha \left\|x - y\right\|^2, \forall (x, y) \in E^2$.

2.$\exists\ M > 0$ tel que $\left\|\nabla f(x) - \nabla f(y)\right\| \leq M \left\|x - y\right\|, \forall (x, y) \in E^2$.

alors :

1. f est strictement convexe.
2. $f(x) \to +\infty$ quand $|x| \to +\infty$.
3. Il existe un et un seul $x^* \in E$ tel que $f(x^*) = \min\limits_E f$ (conséquence de 1. et 2.).
4. Si $0 < \rho < \frac{2\alpha}{M^2}$ alors la suite $(x_n)_{n \in \mathbb{N}}$ construite par (5) convergen vers x^* lorsque $n \to +\infty$.

2.2.3 Algorithme du gradient à pas optimal

L'idée de l'algorithme du gradient à pas optimal est d'essayer de calculer à chaque itération le paramètre qui minimise la fonction dans la direction de descente donnée par le gradient.

Soients $f \in C^1(E, \mathbb{R})$ et $E = \mathbb{R}^n$, cet algorithme s'écrit :

$$\left\{\begin{array}{c} \text{Initialisation} \quad : \ x_0 \in \mathbb{R}^n \\ \text{Itération :} \qquad x_n \text{ connu} \\ \text{On calcule } w_n = -\nabla f(x_n) \\ \text{On choisit } \rho_n \geq 0 \text{ tel que} \\ f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n)\ \forall\ \rho \geq 0. \\ \text{On pose} \quad x_{n+1} = x_n + \rho_n w_n. \end{array}\right. \tag{6}$$

Les questions aux quelles on doit répondre pour s'assurer du bien fondé de ce nouvel algorithme sont les suivantes :

1. Existe_t_il ρ_n tel que $f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n), \forall \rho \geq 0$?
2. Comment calcule_t'on ρ_n ?

3. La suite $(x_n)_{n\in\mathbb{N}}$ construite par l'algorithme converge-t-elle ?

La réponse aux questions 1. et 3. est apportée par le théorème suivant :

Théorème 2.2.3 *(Convergence du gradient à pas optimal)*

Soit $f \in C^1(\mathbb{R}^n,\ \mathbb{R})$ telle que $f(x) \to +\infty$ quand $|x| \to +\infty$.

Alors :

1. La suite $(x_n)_{n\in\mathbb{N}}$ est bien définie par (6). On choisit $n > 0$ tel que :

$$f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n), \forall \rho \geq 0.$$

(ρ_n existe mais n'est pas nécessairement unique).

2. La suite $(x_n)_{n\in\mathbb{N}}$ est bornée et si $(x_{nk})_{k\in\mathbb{N}}$ est une sous suite convergente, i.e. $x_{nk} \to x$ lorsque $k \to +\infty$, on a nécessairement $\nabla f(x) = 0$.

De plus si f est convexe on a $f(x) = \min\limits_{\mathbb{R}^n} f$.

3. Si f est strictement convexe on a alors $x_n \to x^*$ quand $n \to +\infty$, avec $f(x^*) = \min\limits_{\mathbb{R}^n} f$.

On en donne ici les idées principales.

i . On utilise l'hypothèse $f(x) \to +\infty$ quand $|x| \to +\infty$ pour montrer que la suite $(x_n)_{n\in\mathbb{N}}$ construite par (6) existe : en effet, à x_n connu,

1er cas : si $\nabla f(x_n) = 0$, alors $x_{n+1} = x_n$ et donc $x_p = x_n\ \forall p \geq n$.

2ème cas : si $\nabla f(x_n) \neq 0$, alors $w_n = \nabla f(x_n)$ est une direction de descente stricte.

Dans ce deuxième cas, il existe donc ρ_0 tel que :

$$f(x_n + \rho w_n) < f(x_n), \forall \rho \in]0, \rho_0]. \tag{7}$$

De plus, comme $w_n \neq 0$, $|x_n + \rho w_n| \to +\infty$ quand $\rho \to +\infty$ et donc $f(x_n + \rho w_n) \to +\infty$ quand $\rho \to +\infty$. Il existe donc $M > 0$ tel que si $\rho > M$ alors $f(x_n + \rho w_n) \geq f(x_n)$. On a donc :

$$\min_{\rho\in\mathbb{R}_+^*} f(x_n + \rho w_n) = \min_{\rho\in[0,M]} f(x_n + \rho w_n).$$

Comme $[0, M]$ est compact, il existe $\rho_n \in [0, M]$ tel que :

$$f(x_n + \rho_n w_n) = \min_{\rho\in[0,M]} f(x_n + \rho w_n).$$

De plus on a grâce à (7) que $\rho_n > 0$.

Chapitre 3

Algorithmes d'optimisation avec contrainte

3.1 Optimisation avec contraintes

3.1.1 Définition

Soit $E = \mathbb{R}^n$, soit $f \in C(E, \mathbb{R})$, et soit K un sous ensemble de E. On s'intéresse à la recherche de $x^* \in K$ tel que :

$$\begin{cases} x^* \in K \\ f(x^*) = \min\limits_K f \end{cases} \tag{8}$$

Ce problème est un problème de minimisation avec contrainte (ou"sous contrainte") au sens où, l'on cherche de qui minimise f en astreignant x a être dans K : voyons quelques exemples de ces contrainte (définie par, l'ensemble K), qu'on va expliciter à, laide des p fonctions continues, $g_i \in C(E, \mathbb{R})$, $i = 1, \ldots.., p$.

Contraintes égalités

On pose

$$K = \{x \in E,\ g_i(x) = 0,\ i = 1...p\}.$$

On verra plus loin que le problème de minimisation de f peut alors être résolu grâce au théorème des multiplicateurs de Lagrange (voir théorème 3.1.5).

Contraintes inégalités

On pose :

$$K = \{x \in E,\ g_i(x) \leq 0,\ i = 1...p\}.$$

On verra plus loin que le problème de minimisation de f peut alors être résolu grâce au théorème de Kuhn–Tucker (voir théorème 3.1.8).

Contraintes d'égalités et d'inégalité

L'ensemble des contrainte est donné par :

$$K = \{\ x \in E,\ g_i(x) = 0,\ \forall i = 1, ..., p,\ \text{et } g_i(x) \leq 0,\ \forall i = 1, ..., p\}.$$

3.1.2 Existence – Unicité – Conditions d'optimalité simple

Théorème 3.1.1 *(Existence)*

Soit $E = \mathbb{R}^n$ et $f \in C\ (E, \mathbb{R})$.

1. Si K est un sous-ensemble fermé borné de E, alors il existe $x^* \in K$ tel que :

$$f(x^*) = \min_K f.$$

2. Si K est un sous-ensemble fermé de E, et si f est croissante à l'infini, c'est–à–dire que $f(x) \to +\infty$ quand $|x| \to +\infty$, alors
$\exists\ x^* \in K$ tel que :

$$f(x^*) = \min_K f.$$

Théorème 3.1.2 *(Unicité)*

Soit $E = \mathbb{R}^n$ et $f \in C\ (E, \mathbb{R})$. On suppose que f est strictement convexe et que K est convexe. Alors il existe au plus un élément x^* de K tel que :

$$f(x^*) = \min_K f.$$

Preuve. Supposons que x^*et x^{**} soient deux solutions du problème (10), avec $x^* \neq x^{**}$Alors :

$$f(\frac{1}{2}x^* + \frac{1}{2}x^{**}) < \frac{1}{2}f(x^*) + \frac{1}{2}f(x^{**}) \Longrightarrow f(\frac{1}{2}x^* + \frac{1}{2}x^{**}) < f(x^*) = \min_{x\in\mathbb{R}^n} f(x)$$

$$\Longrightarrow f(\frac{1}{2}x^* + \frac{1}{2}x^{**}) < \min_{x\in\mathbb{R}^n} f(x^*)(contradiction).$$

Ce qui est impossible car $((f\frac{1}{2}x^* + \frac{1}{2}x^{**}) < \min_{x\in\mathbb{R}^n} f(x^*))$.

Donc $x^* = x^{**}$. ■

•Les théorèmes, d'existence et d'unicuté, on déduit immédiatement le théorème, d'existence et d'uncité suivant :

Théorème 3.1.3 *(Existence et unicité)*

Soient $E = \mathbb{R}^n$, $f \in C$ $(E, \mathbb{R}^n)$ une fonction strictement convexe et K un sous ensemble convexe fermé de E.

1. Si K est borné ou si f est croissante à l'infini, c'est–à–dire si $f(x) \to +\infty$ quand $\|x\| \to +\infty$, alors il existe un unique élément x^* de K solution du problème de minimisation (8), i.e. tel que :

$$f(x^*) = \min_K f.$$

2. On peut remplacer $E = \mathbb{R}^n$ par E espace de Hilbert de dimension infnie dans le dernier théorème, mais on a besoin dans ce cas de l'hypothèse de convexité de f pour assurer l'existence de la solution.

Condition d'optimalité simple

Proposition 3.1.4

Soient $E = \mathbb{R}^n$, $f \in C(E, \mathbb{R}^n)$ et $x^* \in K$ tel que $f(x^*) = \min_K f$. On suppose que f est différentiable en x^*.

(i) Si $x^* \in \overset{0}{K}$ alors $\nabla f(x^*) = 0$.

(ii) Si K est convexe, alors $\nabla f(x) \cdot (x - x^*) \geq 0$ pour tout $x \in K$.

Condition d'optimalité égalités

Dans tout ce paragraphe, on considèrera les hypothèses et notations suivantes :

$$\begin{aligned} f &\in C(\mathbb{R}^n,\mathbb{R}), g_i \in C^1(\mathbb{R}^n,\mathbb{R}), i=1,....,p; \\ K &= \{x \in \mathbb{R}^n,\ g_i(x)=0,\ \forall i=1,......,p\}\,; \\ g &= (g_1......g_p)^t \in C^1(\mathbb{R}^n,\mathbb{R}^p). \end{aligned} \tag{10}$$

Théorème 3.1.5 *(Multiplicateurs de lagrange)*

Soit $x^* \in K$ tel que $f(x^*) = \min\limits_K f$. On suppose que f est différentiable en x et $\dim(\mathrm{Im}(Dg(x^*)) = p$ (ou $rang(Dg(x^*) = p)$, alors : il existe $(\lambda_1,....,\lambda_p)^t \in \mathbb{R}^p$ tel que :

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) = 0.$$

(cette derniére égalité a lieu dans $\mathbb{R}^n$) .

Remarque 3.1.6 *(Utilisation pratique du théorème de Lagrange)*

Soit $f \in C^1(\mathbb{R}^n,\mathbb{R})$, $g = (g_1,...,g_p)^t$ avec $g_i \in C(\mathbb{R}^n,\mathbb{R})$ pour $i=1,...,p$, et soit $K = \{u \in \mathbb{R}^n, g_i(u) = 0, i = 1,...,p\}$. Le problème qu'on cherche à résoudre est le problème de minimisation (8) qu'on rappelle ici :

$$\begin{cases} u^* \in K \\ f(u^*) = \min\limits_K f \end{cases}$$

D'après le théorème des multiplicateurs de Lagrange, si u^* est solution de (8) et $Im(Dg(u^*) = \mathbb{R}^p$, alors il existe $(\lambda_1,..,\lambda_p) \in \mathbb{R}^p$ tel que u^* est solution du problème

$$\begin{cases} \frac{\partial f}{\partial x_j}(u^*) + \sum \lambda_i \frac{\partial g_i}{\partial x_j} = 0,\ j=1,....,n \\ g_i(u) = 0\ ,\ i=1,...,p \end{cases} \tag{11}$$

Le système (11) est un système non linéaire de $(n+p)$ équations et à $(n+p)$ inconnues $(x^*,...,x_n^*,\lambda_i...\lambda_p)$. Ce système sera résolu par une méthode de résolution de système non linéaire.

Condition d'optimalité inégalité

Soit $f \in C(\mathbb{R}^n,\mathbb{R})$ et $g_i \in C^1(\mathbb{R}^n,\mathbb{R})$, $i=1,........,p$, on considère maintenant un ensemble K de la forme : $K = \{x \in \mathbb{R}^n,\ g_i(x) \le 0\ \forall i = 1,.......,p\}$, et on cherche à résoudre le problème de minimisation (8) qui sécrit :

$$\begin{cases} x^* \in K \\ f(x^*) \leq f(x), \forall x \in K \end{cases}$$

Remarque 3.1.7 *Soit x^* une solution de (8) et supposons que $g_i(x^*) < 0$, pour tout $i \in \{1, ..., p\}$. il existe alors $\varepsilon > 0$ tel que si $x \in B(x^*, \varepsilon)$ alors : $g_i < 0$ pour tout $i = 1, ..., p$.*

On a donc $f(x^*) \leq f(x)$, $\forall\ x \in B(x^*, \varepsilon)$. On est alors ramené à un problème de minimisation sans contraintes, et si f est différentiable en x^*, on a donc : $\nabla f(x^*) = 0$.

Théorème 3.1.8 *(Kuhn - Tucker)*

Soit $f \in C(\mathbb{R}^n, \mathbb{R})$, soit $g_i \in C^1(\mathbb{R}^n, \mathbb{R})$, pour $i = 1, ..., p$; et soit

$$K = \{x \in \mathbb{R}^n;\ g_i(x) \leq 0\ ,\ \forall i = 1, ..., p\}\,.$$

On suppose qu'il existe x^* solution de (8), et on pose $I(x^*) = \{i \in \{1, ..., p\}\ ;\ |g_i(x^*)| = 0\}$.

Et On suppose que f est différentiable en x^* et que la famille (de $\mathbb{R}^n$) $\{\ \nabla g_i(x^*),\ i \in I(x^*)\}$ est libre.

alors il existe une famille $(\lambda_i)_{i \in I}(x^*) \subset \mathbb{R}_+$ telle que :

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*) = 0.$$

1. Le théorème de Kuhn -Tucker s'applique pour des ensembles de contraintes de type inégalité. Si on a une contrainte de type égalité, on peut évidemment se ramener à deux contraintes de type inégalité en remarquant que :

$$\{h(x) = 0\} = \{h(x) \leq)\} \cap \{-h(x) \leq 0\}\,.\,\{\nabla g_1(x), \nabla g_2(x)\} = \{\ \nabla h(x^*), \nabla h(x^*)\}\,.$$

n'est pas libre. On ne peut donc pas appliquer le théorème de Kuhn -Tucker sous la forme donnée précédemment dans ce cas .

2. Dans la pratique, on a intérêt à écrire la conclusion du théoréme de Kuhn -Tucker (i.e. l'exstence de. la famille $(\lambda_i)_{i \in I(x^*)}$) sous la forme du système de $n + p$ équations et $2p$ inéquations à résoudre suivant :

$$\begin{cases} \nabla f(x^*) + \sum\limits_{i=1} \lambda_i \nabla g_i\ (x^*) = 0 \\ \lambda_i g_i\ (x^*) = 0\ \ \forall i = 1, ..., p, \\ g_i\ (x^*) \leq 0\ \ \forall i = 1, ..., p, \\ \lambda_i \geq 0\ \ \forall i = 1, ..., p, \end{cases}$$

$\lambda_i \geq 0$, $\quad g_i\ (x^*) \leq 0, \quad i = 1, ..., p$.

3.2 Méthodes de gradient avec projection

On rappelle le résultat suivant de projection sur un convexe fermé :

Proposition 3.2.1 *(Projection sur un convexe fermé)*

Soit E un espace de Hilbert, muni d'une norme $\|.\|$ induite par un produit scalaire $(.,.)$, et soit K un convexe fermé non vide de E. Alors, pour tout $x \in E$, il existe un unique $x_0 \in K$ tel que $\|x - x_0\| \leq \|x - y\|$ pour tout $y \in K$. On note $x_0 = p_K(x)$ la projection orthogonale de x sur K. On a également :

$$x_0 = P_K(x) \text{ si et seulement si } (x - x_0, x_0 - y) \geq 0, \forall y \in K.$$

Dans le cadre des algorithmes de minimisation avec contraintes que nous allons développer maintenant, nous considèrerons $E = \mathbb{R}^n$, $f \in C^1(\mathbb{R}^n, \mathbb{R})$ une fonction convexe, et K fermé convexe non vide. On cherche à calculer une solution approchée de x^*, solution du problème (8).

3.2.1 Algorithme du gradient à pas fixe avec projection sur K (GPFK)

La méthode est comme suit :

1. Initialisation : on part de $x^* \in K$, on fait $K = 0$.
2. On calcule $y_k = proj_k(x_k - \rho \nabla f(x_k))$.
3. Test d'arrêt : $y_k = x_k$.
4. Sinon on prend pour direction de descente $d_k = y_k - x_k$.
5. On fait une recherche linéaire à partire de x_k dans la direction d_k. On obtient t_k.
6. On fait $x_{k+1} = x_k + \rho t_k d_k$; $k = k + 1$ et on retourne en 2.

Si $d_k = 0$ alors x_k est un point critique et solution optimale dans le cas où f est convexe.

Vérifions dans le cas contraire que d_k est une direction de descente.

Par définition de la projection on a :

$$\langle y_k - x_k + \alpha \nabla f(x),\ x - y_k \rangle \geq 0, \forall x \in K.$$

Prenons $x = x_k$ alors :

$$0 > -\left\| d_k \right\|^2 \geq \langle\ \alpha \nabla f(x_k),\ d_k \rangle\,.$$

D'autre part la convexité de k implique :

$$x_k + t(y_k - x_k) \in K \text{ pour tout } t \in]0,1[\ .$$

Lemme 3.2.2 *Soit $(x_n)_n$ construite par l'algorithme (G P F K). On suppose que : $x_n \longrightarrow x$ quand $n \to \infty$. Alors x est solution de (8).*

Soit $P_k : \mathbb{R}^n \to K \subset \mathbb{R}^n$ la projection sur K définie par la proposition (3.2.1). Alors P_k est continue.

Donc si $x_n \to x$ quand $n \to +\infty$ alors $x = P_k(x - \rho \nabla f(x))$ et $x \in K$ (car $x_n \in K$ et K est fermé).

La caractérisation de $P_k(x - \rho \nabla f(x))$ donnée dans la proposition (3.2.1) donne alors :
$(x - \rho \nabla f(x) - x / x - y) \geq 0$ pour tout $y \in K$, et comme $\rho > 0$, ceci entraîne
$(\nabla f(x) / x - y)$ pour tout $y \in K$.

Or f est convexe donc $f\ (y) \geq f(x) + \nabla f(x)(y - x)$ pour tout $y \in K$, et donc $f(y) \geq f(x)$ pour tout $y \in K$, ce qui termine la démonstration.

Théorème 3.2.3 *(convergence de l'algorithme G P F K)*

Soit $f \in C^1(\mathbb{R}^n, \mathbb{R})$, et K convexe fermé non vide. On suppose que :
1. Il existe $\alpha > 0$ tel que :

$$(\nabla f(x) - \nabla f(y) \mid x - y) \geq \alpha \left| x - y \right|^2, \text{ pour tout } (x, y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

2. Il existe $M > 0$ tel que :

$$\left| \nabla f(x) - \nabla f(y) \right| \leq M \left| x - y \right|, \text{ pour tout } \ (x, y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

alors :
1. Il existe un unique élément $x \in K$ solution du probléme (8).
2. Si $0 < \rho < \frac{2\alpha}{M^2}$, la suite$(x_n)$ définie par l'algorithme (P G F K) converge vers x^* lorsque $n \longrightarrow \infty$.

1. La condition 1. donne que f est strictement convexe et que $f(x) \to +\infty$ quand $|x| \to +\infty$. Comme K est convexe fermé non vide, il existe donc un unique x^* solution de (8).

2. On pose, pour $x \in \mathbb{R}^n$, $h(x) = P_k(x - \rho\nabla f(x))$. On a donc $x_{n+1} = h(x_n)$.

Pour montrer que la suite $(x_n)_{n\in\mathbb{N}}$ converge, il suffit donc de montrer que h est strictement contractante dès que

$$0 < \rho < \frac{2\alpha}{M^2}. \tag{12}$$

Grâce au lemme (3.2.6) démontré plus loin, on sait que p_K est contractante. Or h est définie par :

$$h(x) = P_k(h^*(x)) \text{ où } h^*(x) = x - \rho\nabla f(x).$$

On a déjà vu que h^* est strictement contractante si la condition (12) est vérifiée (voir théorème (2.2.2), et plus précisément :

$$|h^*(x) - h^*(y)| \leq (1 - 2\alpha\rho + M^2\rho^2)|x - y|^2.$$

On en déduit que :

$$|h(x)-h(y)|^2 \leq |P_K(h^*(x))- \ P_K(h^*(y))|2 \leq |h^*(x)-h^*(y))|^2 \leq (1-2\alpha\rho+\rho^2M^2)|x-y|^2.$$

L'application h est donc strictement contractante dès que $0 < \frac{2\alpha}{M^2}$. La suite $(x_n)_{n\in\mathbb{N}}$ converge donc bien vers $x = x^*$

Lemme 3.2.4 *(Propriétè de contraction de la projection orthogonale)*

Soit E un espace de Hilbert, $\|.\|$ la norme et $(.,.)$ le produit scalaire, K un convexe fermé non vide de E et P_k la projection orthgonale sur K définie par la proposition (3.2.1), alors :

$$\|P_K(x) - P_K(y)\| \leq \|x - y\| \text{ pour tout } (x, y) \in E^2.$$

Comme E est un espace de Hilbert,
$\|P_K(x) - P_K(y)\|^2 = (P_K(x) - P_K(y)|P_K(x) - P_K(y))$.
On a donc :
$\|P_K(x) - P_K(y)\|^2 = (P_K(x) - x + x - y + y - P_K(y)|P_K(x) - P_K(y))$
$= (P_K(x) - x|P_K(x) - P_K(y))_E + (x - y|P_K(x) - P_K(y)) +$
$(y - P_K(y)|P_K(x) - P_K(y))$.

Or $(P_K(x) - x | P_K(x) - P_K(y)) \geq 0$ et $(y - P_K(y) | P_K x) - P_K(y))$, d'où :
$\| P_K(x) - P_K(y) \| \leq (x - y | P_K(x) - P_K(y))$,
et donc, grâce à l'inégalité de Cauchy-Schwarz,
$\| P_K(x) - P_K(y) \| \leq \| x - y \| \, \| P_K(x) - P_K(y) \| \leq \| x - y \|$.

Conclusion

En conclusion, on peut dire que la méthode de projection est une méthode itératif pour résoudre un problème d'optimisation avec contrainte qui permit les plus performantes pour résoudre des problèmes d'optimisation sous contrainte.

l'objectif de méthode de gradient avec projection est de vous apprendre à coder des algorithmes gènèreux de minimisation sans et avec contrainte et rèsolution de problème par l'algorithme de descente de gradient.

Ce mémoire a pour but de résolution de deux problèmes d'optimisation sans contrainte et sous contrainte.

Bibliographie

[1] Cour d'optimisation 3eme année. (2014 /2015).

[2] H.boumaza-B.collas-S.collion-M.dellinger-Z.faget-L.lazzarini-F.schaffhhauser, mathématique Analyse L3 : paris, (juin 2009).

[3] Cour Opératours Linéaires Hilbert 3eme année. (2014 /2015).

[4] Cours d'Analyse numérique, Raphaèle Herbin, 19 octobre 2008, Université Aix Marseille 1.

[5] Optimisation sous contrainte, Laurent Guillopé, Université de Nantes.

[6] L2 Parcours Spécial, Calcul différentiel et intégral, J. Royer, Université Toulouse3.

[7] Rapport Les Multiplicateurs de Lagrange Optimisation et Signal, M. HITTA Amara, Univ. 8 Mai 1945, Guelma, (2007).

[8] Mechel Bierlaire, Introduction a l'optimisation différentiable, Presses polytechniques et Universitaires Romandes.