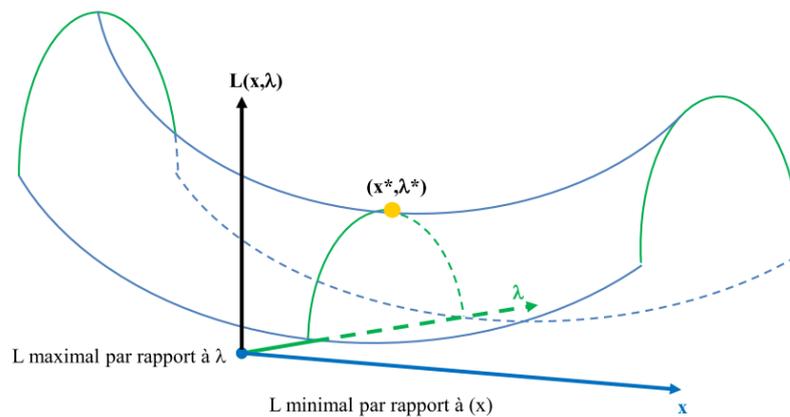

Optimisation sans contrainte

Cours et exercices

Point col (ou point selle)



Bekkouche Noria

Décembre 2022

TABLE DES MATIÈRES

0.1	Problématique	3
0.1.1	Cadre	3
0.2	Différents types d'optimisation	4
0.2.1	Classification des problèmes d'optimisation	4
1	Rappels et compléments de calcul différentiel et analyse convexe	6
1.1	Espace vectoriels normés	6
1.2	Espaces euclidiens. Espaces de Hilbert	8
1.3	Calcul différentiel	8
1.3.1	Fonctions numériques d'une variable réelle	8
1.3.2	Fonctions vectorielles d'une variable réelle	9
1.3.3	Fonctions vectorielles de plusieurs variables :	9
1.3.4	Dérivée directionnelle	12
1.3.5	Direction de descente	13
1.3.6	Développement de Taylor	14
1.4	Convexité	15
1.4.1	Ensembles convexes	15
1.4.2	Fonction convexe	17
2	Problème d'optimisation sans contraintes	28
2.1	Existence et unicité de la solution du problème	29
2.1.1	Dans la pratique	29
2.1.2	Existence d'une solution	30
2.1.3	Condition du premier ordre	33
2.1.4	Conditions du seconde ordre	34
3	Algorithmes	40
3.0.1	Schémas général des algorithmes d'optimalité	42
3.0.2	Méthode à directions de descente- une itération	42
3.0.3	Méthode de Gradient	43
3.0.4	Méthodes de gradient à pas optimal	43
3.0.5	Méthode de Gradient à pas fixe	45
3.0.6	Méthode du gradient conjugué	46
3.0.7	Méthode de Newton	47

3.0.8	Méthode de Quasi-Newton	50
-------	-----------------------------------	----

L'optimisation est une branche des mathématiques cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes qui consistent à minimiser ou maximiser une fonction sur un ensemble.

L'optimisation joue un rôle important en recherche opérationnelle (domaine à la frontière entre l'informatique, les mathématiques et l'économie), dans les mathématiques appliquées (fondamentales pour l'industrie et l'ingénierie), en analyse numérique, en statistique pour l'estimation du maximum de vraisemblance d'une distribution, pour la recherche de stratégies dans le cadre de la théorie des jeux, ou encore en théorie du contrôle et de la commande.

L'optimisation intervient dans de nombreux domaines :

1. En recherche opérationnelle (problème de transport, économie, gestion de stocks...)
2. En analyse numérique (approximation/résolution de systèmes linéaires, non linéaires...)
3. En automatique (modélisation de systèmes, filtrage...)
4. En ingénierie (dimensionnement de structures, conception optimale de systèmes (réseaux, ordinateurs...))

Les premiers problèmes d'optimisation auraient été formulés par Euclide, au III^e siècle avant notre

0.1 Problématique

0.1.1 Cadre

Soit V est un espace vectoriel normé, muni de la norme $\|\cdot\|$. Dans ce cours, on s'intéresse au problème suivant

$$\left\{ \inf_{x \in K} f(x) \right. \quad (1)$$

un point $x_0 \in K$ qui réalise ce minimum (resp. maximum) i.e. $f(x_0) = v$.

Vocabulaire

Où $K \subset V$ et $f : K \rightarrow \mathbb{R}$ est une fonction, appelée fonction cot ou critère.

1. v est la valeur optimale.
2. x_0 est la solution optimale.

- (a) Si $K = V$, on dit que (1) est un problème d'optimisation sans contrainte.
- (b) Si $K \subsetneq V$, on dit que (1) est un problème d'optimisation sous contrainte.
- (c) Si $\dim K < +\infty$ (resp. $\dim K = +\infty$), on dit que (1) est un problème d'optimisation en dimension finie (resp. infinie). Dans le cadre de ce cours, on étudiera essentiellement l'optimisation sans contrainte en dimension finie.
- (d) écriture du problème : $\min_{x \in K} f(x)$ resp. $\max_{x \in K} f(x)$.

0.2 Différents types d'optimisation

0.2.1 Classification des problèmes d'optimisation

1. **Optimisation linéaire**

f est une fonction linéaire : $f(x) = \langle c, x \rangle$

K est défini par des fonctions affines : $ax + b \geq 0$

2. **Optimisation linéaire quadratique**

f est une fonction convexe quadratique : $f(x) = \frac{1}{2} \langle Ax, x \rangle + \langle b, x \rangle$

A est une matrice symétrique semi-définie positive

K est défini par des fonctions affines : $ax + b \geq 0$

3. **Optimisation convexe**

f est une fonction convexe et K un domaine convexe

4. **Optimisation différentiable**

f est une fonction différentiable

K est défini par des fonction (=contraintes) différentiables

5. **Optimisation non différentiable**

ex : $f(x) = \max_{x \in K} \{f_1(x), \dots, f_m(x)\}$

6. **Optimisation en dimension infinie**

ex : problèmes variationnels

$$J(u) = \int_0^T L(t, u(t), \dot{u}(t)) dt$$

avec u dans un ensemble de fonctions X , $u(0) = u_0$ et $u(T) = u_T$

Nous étudierons, dans ce cours, uniquement des problèmes d'optimisation non linéaire.

Pour résoudre le problème (1), il faut se poser les questions suivantes :

- Est-ce que $\inf_{x \in K} f(x)$ existe ? c.à.d. est-ce que f est bornée inférieurement ?
- Est-ce que l'infimum est atteint dans K ? c.à.d. est-ce qu'il existe $x_0 \in K$ vérifiant $f(x_0) = \min_{x \in K} f(x)$?
- Est-ce que x_0 est unique ? Sinon, quelle est la taille de l'ensemble des solutions ?
- Est-ce que l'on peut caractériser x_0 ? c.à.d. peut-on trouver des conditions nécessaires pour caractériser un minimum : $\langle \text{si } u \text{ vérifie (1), alors } x \text{ vérifie la propriété } N(x) \rangle$ et / ou trouver des conditions suffisantes pour tre un point optimal : $\langle \text{si } x \text{ vérifie la propriété } S(x), \text{ alors } x \text{ vérifie (1)} \rangle$.
- Trouver un algorithme d'optimisation pour déterminer les solutions de (1).

Pour répondre à ces questions on est en particulier amené à étudier :

-la structure de V : espace vectoriel, muni d'une norme, d'un produit scalaire, de dimension finie ou infinie

- les propriétés de $K \subset V$: fermé, borné, convexe, . . .

- les propriétés de $f : V \rightarrow \mathbb{R}$: continuité, différentiabilité, convexité, . . .

Dans ce cours on va donner des éléments de réponse à ces questions.

CHAPITRE 1

RAPPELS ET COMPLÉMENTS DE CALCUL DIFFÉRENTIEL ET ANALYSE CONVEXE

1.1 Espace vectoriels normés

Soit V un espace vectoriel sur le corps \mathbb{R} .

Définition 1.1.1

Un produit scalaire sur V , si

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$$

est une application bilinéaire, symétrique, et définie positive, c'est-à-dire qui vérifie :

1. $\langle u, \cdot \rangle : V \rightarrow \mathbb{R}$ est linéaire pour tout $u \in V$,
2. $\langle \cdot, v \rangle : V \rightarrow \mathbb{R}$ est linéaire pour tout $v \in V$,
3. $\langle u, v \rangle = \langle v, u \rangle$ pour tout $u, v \in V$,
4. $\langle v, v \rangle = 0 \Leftrightarrow v = 0$, et $\langle v, v \rangle \geq 0$ pour tout $v \in V$.

Définition 1.1.2

L'application $u \mapsto \|u\|$ de V dans \mathbb{R}_+ définie par $\|v\| = \sqrt{\langle v, v \rangle}$ pour tout $v \in V$,
une norme sur l'espace V telle que

1. $\forall u \in V : \|u\| = 0 \Leftrightarrow u = 0_E$
2. $\forall u \in V; \forall \lambda \in \mathbb{R} : \|\lambda u\| = |\lambda| \|u\|$
3. $\forall (u, v) \in V \times V : \|u + v\| \leq \|u\| + \|v\|$.

Proposition 1.1.1

Les applications

$$\begin{array}{lll} V \times V & \rightarrow & V; & \mathbb{R} \times V & \rightarrow & V; & V & \rightarrow & \mathbb{R}_+ \\ (u, v) & \mapsto & u + v & (\lambda, u) & \mapsto & \lambda u & u & \mapsto & \|u\| \end{array}$$

sont continues.

Proposition 1.1.2

Soient $(V, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ des e.v.n, $\Phi : V \rightarrow F$ une application linéaire, alors

$$\Phi \text{ est continue} \Leftrightarrow \exists c > 0, \quad \forall u \in V : \|\Phi(u)\|_F \leq c\|u\|_V.$$

Proposition 1.1.3

Soient $(V, \|\cdot\|_V)$ et $(F, \|\cdot\|_F)$ des e.v.n. On note (V, F) l'espace vectoriel des applications linéaires continues de V dans F .

Pour $f \in (V, F)$ on pose $\|f\| = \sup_{\|u\|_E \leq 1} \|f(u)\|_F$. Alors :

1. $f \mapsto \|f\|$ est une norme sur (V, F) ;
2. $\forall x \in E : \|f(u)\|_F \leq \|f\| \|u\|_E$;
3. $\|f\| = \sup_{\|u\|_E \leq 1} \|f(u)\|_F = \sup_{\|u\|_V = 1} \|f(u)\|_F = \sup_{u \in V, u \neq 0_V} \frac{\|f(u)\|_F}{\|u\|_V}$

Proposition 1.1.4: A

plication : Si $\dim V = n$ et $\dim F = m$ alors $f : V \rightarrow F$ linéaire est représenté par une matrice A de m lignes et n colonnes, on peut alors définir une *norme de matrice subordonnée aux normes vectorielles* par

$$\|A\| = \sup_{\|u\|_E = 1} \|Au\|_F = \sup_{u \in E, u \neq 0_V} \frac{\|Au\|_F}{\|u\|_V}$$

Proposition 1.1.5

Soit $(V, \|\cdot\|_V)$ un e.v.n. On note $V' = (V, \mathbb{R})$ l'e.v.n. des formes linéaires continues sur V , alors : Une forme linéaire $J \in V'$ si et seulement si $\ker J$ est fermé dans V .

On désigne par $(V; F)$ ou simplement (V) si $V = F$, l'espace vectoriel formé par les applications linéaires continues de V dans F . C'est un espace vectoriel norme par

$$\|A\| = \sup_{\substack{u \in V \\ u \neq 0}} \frac{\|Au\|_F}{\|u\|_V}$$

et c'est un espace vectoriel complet si l'espace F est complet. Dans le cas où $V = F = \mathbb{R}^n$, la norme définie ci-dessus n'est pas autre chose qu'une norme matricielle subordonnée (à une norme vectorielle sur \mathbb{R}^n). Lorsque $F = \mathbb{R}$, on écrit généralement

$$(V; \mathbb{R}) = V',$$

et on appelle l'espace V' le dual de V .

1.2 Espaces euclidiens. Espaces de Hilbert

On appelle espace préhilbertien un espace vectoriel normé muni d'un produit scalaire. S'il est complet pour cette norme, c'est un espace de Hilbert.

Tout espace vectoriel norme de dimension finie étant complet, l'espace \mathbb{R}^n muni du produit scalaire euclidien est un exemple d'espace de Hilbert. Notons au passage l'inégalité de Schwarz :

$$|\langle u, v \rangle| \leq \|u\| \|v\| \quad \text{pour tout } u, v \in V,$$

qui sert notamment à démontrer l'inégalité triangulaire pour la norme associée au produit scalaire. L'inégalité de Cauchy-Schwarz pour le produit scalaire euclidien ou l'inégalité de Cauchy-Schwarz pour les fonctions :

$$\left| \int_I uv dx \right| \leq \left(\int_I |u|^2 dx \right)^{\frac{1}{2}} \left(\int_I |v|^2 dx \right)^{\frac{1}{2}}, \quad I \subset \mathbb{R},$$

en sont des cas particuliers. On remarquera que l'inégalité de Schwarz entraine la continuité du produit scalaire, considéré comme application du produit $V \times V$ dans \mathbb{R}^n .

Enfin on rappelle que cette inégalité devient une égalité si, et seulement si, les deux vecteurs qui y figurent sont linéairement dépendants.

1.3 Calcul différentiel

L'objectif du présent les principales notations et définitions et les résultats fondamentaux du calcul différentiel dans les espaces vectoriels normes (dérivabilité d'une fonction composée, théorème des accroissements finis, théorème des fonctions implicites, formules de Taylor). On s'est volontairement limité aux dérivées premières et secondes, puisque nous n'aurons pas l'usage de dérivées d'ordre supérieur.

1.3.1 Fonctions numériques d'une variable réelle

Soit K un intervalle ouvert de \mathbb{R} et $f : K \rightarrow \mathbb{R}$ une application.

Définition 1.3.1

f est dérivable sur K ssi pour tout $x \in K$,

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

existe. On note alors cette limite $f'(x)$.

Définition 1.3.2

f est de classe C^1 sur K – on note $f \in C^1(K, \mathbb{R})$ – ssi f est dérivable sur K et l'application $x \rightarrow f'(x)$ est continue sur K .

Exemples 1.3.1. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par

$$f(x) = e^{2x} - x^2$$

. La fonction f est de classe C^1 sur \mathbb{R} car pour tout réel x ,

$$f'(x) = 2e^{2x} - 2x$$

qui définit une fonction continue sur \mathbb{R} .

La fonction f définie sur \mathbb{R} par

$$f(x) = x^2 \sin \frac{1}{x}$$

pour $x > 0$ et $f(0) = 0$. Alors f est dérivable sur \mathbb{R}^+ :

$$\forall x > 0, f'(x) = 2x \sin \frac{1}{x} - \sin \frac{1}{x^2} \text{ et } f'(0) = 0.$$

Mais f n'est pas de classe C^1 sur \mathbb{R}^+ car f' n'est pas continue en 0.

1.3.2 Fonctions vectorielles d'une variable réelle

Soit K un intervalle de \mathbb{R} et $K : V \rightarrow \mathbb{R}^p$ une application. Pour tout $x \in K$, on note $f(x) = (f_1(x), \dots, f_p(x))^T$.

Exemples 1.3.2. Soit $f : \mathbb{R}_+^* \rightarrow \mathbb{R}^2$ définie par

$$\forall x \in \mathbb{R}_+^*, f(x) = (\sqrt{x} - x, e^x - \ln x)^T$$

$f \in C^1(\mathbb{R}_+^*)$ car

$$\forall x \in \mathbb{R}_+^*, f'(x) = \left(\frac{1}{2\sqrt{x}} - 1, e^x - \frac{1}{x} \right)^T$$

qui est continue sur \mathbb{R}_+^* .

1.3.3 Fonctions vectorielles de plusieurs variables :

Soit K un ouvert de \mathbb{R}^n (pour simplifier, on peut considérer que K est un produit d'intervalles ouverts $(I_1 \times \dots \times I_n)$) et $f : K \rightarrow \mathbb{R}^n$ une application. On notera $x = (x_1, \dots, x_n)^T$ pour $x \in K$.

Définition 1.3.3

La fonction f est différentiable en $x \in K$ ssi l'application

$$t \in I_i \rightarrow f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n) \in \mathbb{R}^n$$

est dérivable en $t = x_i$. On note alors sa dérivée

$$\frac{\partial f}{\partial x_i}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)$$

(appelée dérivée partielle de f selon la i^{e} direction).

En outre, la fonction f est dite de classe C^1 sur K si elle est différentiable en chaque x de K , et chacune des ses dérivées partielles est continue sur K .

Exemples 1.3.3. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par

$$f(x_1, x_2) = (f_1(x_1, x_2), f_2(x_1, x_2)) = (x_1^2 - 3x_2^3 + 2x_2 - 1, e^{2x_1} - x_2)^\top$$

Alors f est de classe C^1 sur \mathbb{R}^2 et

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = (2x_1, 2e^{2x_1})^\top \quad \text{et} \quad \frac{\partial f}{\partial x_2}(x_1, x_2) = (-9x_2^2 + 2, -1)^\top.$$

Définition 1.3.4

Soit $f \in C^1(K, \mathbb{R}^p)$. On appelle matrice jacobienne de f en x la matrice $Jf(x)$ de taille $p \times n$, telle que

$$[Jf(x)]_{ij} = \frac{\partial f_i}{\partial x_j}(x) \quad (1 \leq i \leq p, 1 \leq j \leq n).$$

Remarque 1.3.1. On appelle aussi cette matrice la différentielle de f en x – on note alors $df(x)$ ou $f'(x)$. Il s'agit en fait de la définition intrinsèque de la différentiabilité : la dérivée est le terme de degré un dans le développement de Taylor de f autour de x . La fonction f est différentiable en $x \in K$ ssi il existe une application linéaire G de \mathbb{R}^n dans \mathbb{R}^p telle que

$$f(x+h) = f(x) + G.h + O(\|h\|^2).$$

Cette application G est appelée différentielle de f en x .

Exemples 1.3.4. On reprend la fonction de l'exemple précédent. La matrice jacobienne est donnée par :

$$Jf(x) = \begin{pmatrix} 2x_1 & -9x_2^2 \\ 2e^{2x_1} & -1 \end{pmatrix}$$

On rappelle aussi le résultat de dérivation composée :

Théorème 1.3.1

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ et $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$ de classe C^1 . Alors $g \circ f : \mathbb{R}^n \rightarrow \mathbb{R}^q$ est de classe C^1 et

$$J[g \circ f](x) = Jg(f(x)) \cdot Jf(x).$$

Cas particulier $p = 1$: gradient, hessienne

Dans toute la suite, on considère $f : K \rightarrow \mathbb{R}$ une application.

Différentiation à l'ordre 1

Définition 1.3.5

Si $f \in C^1(K, \mathbb{R})$, alors on appelle gradient de f en x le vecteur de \mathbb{R}^n

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

Remarquons que $f(x)$ est un nombre alors que $\nabla f(x)$ est un vecteur.

Exemples 1.3.5. Soit $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ définie par $f(x) = \|x\|^2$. Si on écrit $x = (x_1, x_2, x_3)^\top$, alors

$$f(x) = x_1^2 + x_2^2 + x_3^2$$

Alors le gradient est donné par

$$\nabla f(x) = (2x_1, 2x_2, 2x_3)^\top = 2x$$

Définition 1.3.6

On dit que x est un point stationnaire de f si $\nabla f(x) = 0$.

Différentiation à l'ordre 2

Définition 1.3.7

f est de classe C^2 sur K ssi $f \in C^1(K, \mathbb{R})$ et $\nabla f \in C^1(K, \mathbb{R})$. On note alors Hf ou $\nabla^2 f$ la matrice jacobienne de ∇f ; elle est appelée hessienne de f .

Cette matrice carrée de taille n est donnée par

$$[Hf(x)]_{ij} = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

Exemples 1.3.6. On reprend l'exemple précédent. Alors $f \in C^2(\mathbb{R}^3, \mathbb{R})$ et la hessienne est donnée par

$$Hf(x) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Théorème 1.3.2: (Schwartz)

Soit $f \in C^2(K, \mathbb{R})$, alors

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

Proposition 1.3.1: (Lien entre ∇ et ∇^2)

1. La i -ème ligne de $\nabla^2 f(x)$ Jacobienne du i -ème élément de ∇f .
2. On a

$$\nabla^2 f(x)h = \nabla \langle \nabla f(x), h \rangle ; \forall x \in K, \forall h \in \mathbb{R}^n$$

Preuve :

1. évidente
2. On a

$$\begin{aligned} \frac{\partial}{\partial x_i} \langle \nabla f(x), h \rangle &= \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n \frac{\partial f}{\partial x_j}(x) h_j \right) \\ &= \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x) h_j \\ &= (\nabla^2 f(x)h)_i. \end{aligned}$$

Exemples 1.3.7. Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction constante alors $\nabla f = \nabla^2 f = 0$:

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$f(x) = \langle a, x \rangle \quad \forall x \in \mathbb{R}^n,$$

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

$$\nabla f = a$$

(le gradient est constant).

Ceci nous donne

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

1.3.4 Dérivée directionnelle

Définition 1.3.8

On appelle dérivée directionnelle de f dans la direction d au point x , notée $\delta f(x, d)$, la limite (ventuellement $\pm\infty$) du rapport :

$$\frac{f(x, hd) - f(x)}{h} \text{ lorsque } h \text{ tend vers } 0.$$

Autrement dit :

$$\delta f(x, d) = \lim_{h \rightarrow 0} \frac{f(x, hd) - f(x)}{h} = \nabla^\top f(x) d.$$

Remarque 1.3.2. 1. Si $\|d\| = 1$: la dérivée directionnelle est le taux d'accroissement de f dans la direction d au point x .

2. Pour tout $x \in K$ et $h \in \mathbb{R}^n$ on note

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

(c'est la dérivée directionnelle de f en x de direction h) où on a noté $g(t) = f(x + th)$.

Exemples 1.3.8. Soit $f(x_1, x_2, x_3) = x_1 + x_1^3 x_3 - x_1 x_2 x_3$ et soit

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

La dérivée directionnelle de f dans la direction d est

$$(d_1 d_2 d_3) \nabla f(x_1, x_2, x_3) = d_1 (1 + 3x_1^2 x_3 - x_2 x_3) - d_2 (x_1 x_3) + d_3 (x_1^3 - x_1 x_2)$$

où $\nabla f(x_1, x_2, x_3)$ est donné par

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} 1 + 3x_1^2 x_3 - x_2 x_3 \\ x_1 x_3 \\ x_1^3 - x_1 x_2 \end{pmatrix}.$$

Définition 1.3.9: (Fonction différentiable)

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

Si, pour tout $d \in \mathbb{R}^n$; la dérivée directionnelle de f dans la direction d existe, alors la fonction f est dite différentiable.

1.3.5 Direction de descente

une direction de descente est une direction le long de laquelle la fonction à minimiser a une dérivée directionnelle strictement négative. Ces directions sont utilisées par les méthodes à directions de descente.

Définition 1.3.10

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable. Soient $x, d \in \mathbb{R}^n$. La direction d est une direction de descente en x si

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

Théorème 1.3.3

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable au point $x \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ tel que

$$f'(x, d) = (\nabla f(x), d) < 0$$

Alors, d est une direction de descente de f en $x \in \mathbb{R}^n$.

Preuve :

Soit f une fonction différentiable au point $x \in \mathbb{R}^n$, donc

$$f(x + \lambda d) = f(x) + \lambda \nabla f(x)^\top d + \lambda \|d\| \alpha(x, \lambda d), \quad \forall x \in \mathbb{R}^n \text{ où } \alpha(x, \lambda d) \text{ quand } \lambda \rightarrow 0.$$

Donc

$$\frac{f(x + \lambda d) - f(x)}{\lambda} = \nabla f(x)^\top d + \|d\| \alpha(x, \lambda d)$$

et

$$\lim_{\lambda \rightarrow 0} \frac{f(x + \lambda d) - f(x)}{\lambda} = \nabla f(x)^\top d \Rightarrow f'(x, d) = (\nabla f(x), d)$$

et comme

$$(\nabla f(x), d) < 0 \Rightarrow f'(x, d) < 0$$

donc

$$\exists \delta > 0, \text{ tq } \frac{f(x + \lambda d) - f(x)}{\lambda} < 0, \forall \lambda \in]0, \delta[$$

$$\exists \delta > 0, \forall \lambda \in]0, \delta[, f(x + \lambda d) < f(x).$$

Donc le vecteur d est une direction de descente.

1.3.6 Développement 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, $f : \Omega \rightarrow \mathbb{R}$; $a \in \Omega$ et $h \in \mathbb{R}^n$ tels que $[a; a + h]$. Alors :

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

(a) **Formule de Taylor à l'ordre 1 avec reste intégral**

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

(b) **Formule de Taylor - Maclaurin à l'ordre 1**

$$f(a + h) = f(a) + \langle \nabla f(a + \theta h), h \rangle.$$

(c) **Formule de Taylor - Young à l'ordre 1**

$$f(a + h) = f(a) + \langle \nabla f(a + \theta h), h \rangle + o(\|h\|).$$

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

(a) **Formule de Taylor à l'ordre 2 avec reste intégral**

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

(b) **Formule de Taylor - Maclaurin à l'ordre 2**

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

(c) Formule de Taylor - Young à l'ordre 2

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

Remarque 1.3.3. 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 $\|h\|^k$ tend vers 0).

1.4 Convexité

La convexité joue un rôle très important dans la théorie classique de l'optimisation. Elle est un outil indispensable pour la recherche des conditions à la fois nécessaires et suffisantes d'optimalité.

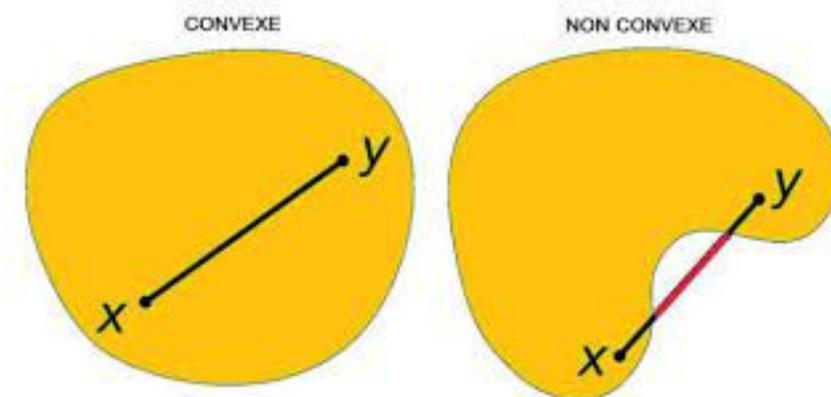
Dans ce chapitre nous introduisons la notion d'ensemble et fonction convexe, démontrons leurs principales propriétés géométriques et topologiques.

1.4.1 Ensembles convexes

Définition 1.4.1

Un ensemble $C \subset \mathbb{R}^n$ est dit convexe si pour tout couple $(x, y) \in C^2$ et $\forall \lambda \in [0, 1]$ on a ;

$$\lambda x + (1 - \lambda)y \in C.$$



Définition 1.4.2

On rappelle qu'étant donné deux points $a, b \in \mathbb{R}^n$, on définit le segment $[a, b]$ par

$$[a, b] = \{x \in \mathbb{R}^n : x = (1-t)a + tb \text{ pour un certain } t \in [0, 1]\}.$$

Une partie $K \subset \mathbb{R}^n$ est convexe si, pour tous points $a, b \in K$, le segment $[a, b]$ est compris dans K

Propriétés des ensembles convexes

1. La définition d'ensemble convexe peut s'interpréter en disant que le segment reliant x et y doit être dans C .
2. Soit $x_1, x_2, \dots, x_k \in \mathbb{R}^n$ et t_j telle que $t_j \geq 0$ et $\sum_{j=1}^k t_j = 1$. Toute expression de la forme

$$\sum_{j=1}^k t_j x_j.$$

S'appelle combinaison convexe des points x_j ou barycentre.

3. Si C_1 et C_2 sont deux ensembles convexes de \mathbb{R}^n , alors $K = C_1 \cap C_2$ est convexe et

$$K = \{x \mid x = x_1 + x_2, \quad x_1 \in C_1 \text{ et } x_2 \in C_2\},$$

est convexe.

Exemples 1.4.1. 1. Une intersection de parties convexe est convexe. Cela suit directement de la définition.

2. Un sous-espace affine K de \mathbb{R}^n est convexe. On rappelle que $K \subset \mathbb{R}^n$ est un sous-espace affine s'il existe un sous-espace vectoriel $V \subset \mathbb{R}^n$ et un vecteur $v \in \mathbb{R}^n$, tel que $x \in K$ si et seulement si $x - v \in V$. Montrons que K est convexe. Soient $x, y \in K$, et soit $z = (1-t)x + ty$ avec $t \in [0, 1]$. On a $x - v, y - v \in V$, et donc $z - v = (1-t)(x - v) + t(y - v) \in V$ et donc $z \in K$.

1.4.2 Fonction convexe

Définition 1.4.3

une application réelle $J : K \subset V \rightarrow \mathbb{R}$ est définie sur un ensemble convexe K d'un espace vectoriel V est dite convexe sur l'ensemble K si pour tous les points $u, v \in K$ et pour tout nombre réel positif ou nul t tel que $0 \leq t \leq 1$ l'inégalité suivante est vérifiée :

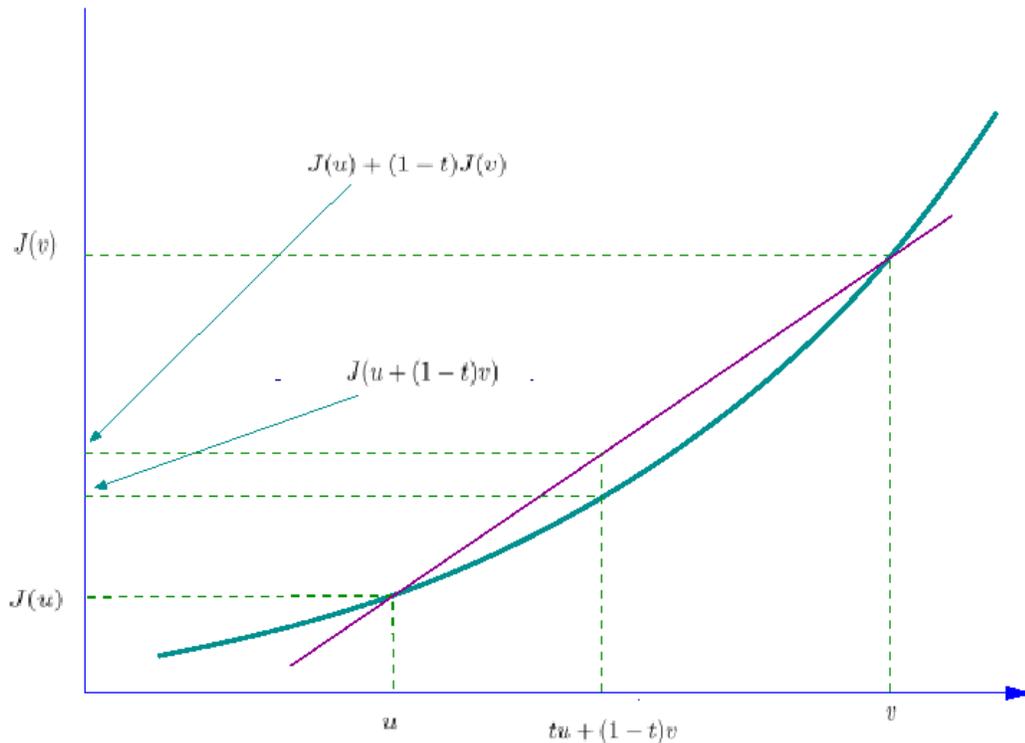
$$J(tu + (1-t)v) \leq tJ(u) + (1-t)J(v)$$

et strictement convexe sur l'ensemble K si

$$u, v \in K, u \neq v \text{ et } t \in]0, 1[$$

implique

$$J(tu + (1-t)v) < tJ(u) + (1-t)J(v).$$



Définition 1.4.4

f est fortement convexe de constante $\alpha > 0$ si :

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) - \frac{\alpha}{2}t(1-t)\|x_1 - x_2\|^2$$

On montre facilement qu'une fonction fortement convexe est strictement convexe.

Remarque 1.4.1.

convexité forte \implies convexité stricte \implies convexité.

Définition 1.4.5

Une fonction $f : K \subset V \rightarrow \mathbb{R}$ définie sur une partie convexe K d'un espace vectoriel V est dite (strictement) concave si la fonction $(-f)$ est (strictement) convexe.

Exemples 1.4.2. 1. Une fonction affine est convexe sur \mathbb{R}^n , mais non-strictement convexe (une fonction affine est donnée pour tout $x \in \mathbb{R}^n$ par $f(x) = \langle a, x \rangle + b$ pour un certain $a \in \mathbb{R}^n$ et $b \in \mathbb{R}$).

2. Une norme est convexe sur \mathbb{R}^n mais non strictement convexe. La convexité résulte de l'inégalité triangulaire : pour tout $x, y \in \mathbb{R}^n$ et pour tout $t \in]0, 1[$,

$$\|(1-t)x + ty\| \leq (1-t)\|x\| + t\|y\|.$$

Que l'inégalité ne soit pas stricte résulte de ce que, si on prend $y = \lambda x$ pour un certain $\lambda \geq 0$, on a l'égalité dans l'expression précédente.

Opérations sur les fonctions convexe

Nous donnons quelques propriétés de stabilité des fonctions convexes. Soit $K \subset \mathbb{R}^n$ un ensemble convexe.

1. La somme de deux fonctions convexes sur un ensemble K est convexe.
2. Si $f : K \rightarrow \mathbb{R}$ est convexe, et si $\lambda \geq 0$, alors λf est convexe sur K .
3. Le maximum de deux fonctions convexes sur K est convexe.
4. Si $f : K \rightarrow \mathbb{R}$ est convexe, et que $L : \mathbb{R}^m \rightarrow \mathbb{R}^n$ est linéaire, alors $f \circ L$ est convexe sur le convexe

$$V = \{x \in \mathbb{R}^m : L(x) \in K\}.$$

Définitions et propriétés d'une fonction convexe

Epigraphe d'une fonction convexe

Définition 1.4.6

Soient K un ensemble de \mathbb{R}^n non vide et $f : K \rightarrow \mathbb{R}$. L'épigraphe de f qu'on note $\text{epi}(f)$ est un sous ensemble de \mathbb{R}^{n+1} défini par :

$$\text{epi}(f) = \{(x, y) \in \mathbb{R}^{n+1} / x \in K, y \in \mathbb{R}, y \geq f(x)\}$$

qui est équivalent à

$$\text{epi}(f) = \{(x, y) \in K \times \mathbb{R}, y \geq f(x)\}$$

C'est à dire l'ensemble des points qui sont au-dessus du graphe de f .

Théorème 1.4.1

Soient K un ensemble convexe de \mathbb{R}^n non vide et $f : K \rightarrow \mathbb{R}$, alors

La fonction est convexe \iff L'ensemble $(\text{epi}(f))$ est convexe.

Preuve :
Exercice

Convexité et continuité

Commençons par énoncer et démontrer une version simple de l'inégalité de Jensen. Soit un ensemble convexe $K \subset \mathbb{R}^n$ et soit une fonction $f : K \rightarrow \mathbb{R}^n$.

Proposition 1.4.1

Supposons que f soit convexe sur K , et soient $m_1 \geq 0$ et $p_1, \dots, p_m \geq 0$ tels que $p_1 + \dots + p_m = 1$. Alors

$$f(p_1x_1 + \dots + p_mx_m) \leq p_1f(x_1) + \dots + p_mf(x_m)$$

pour tout $x_1, \dots, x_m \in K$

Preuve :

C'est immédiat pour $m = 1$, c'est équivalent à la définition de convexité pour $m = 2$; le résultat s'obtient par récurrence sur m pour $m > 2$.

Théorème 1.4.2

Si K est ouvert et si f est convexe sur K , alors f est continue sur K .

Preuve :

Pour simplifier, nous faisons la preuve dans le cas particulier $K = \mathbb{R}^n$ et $n = 2$.

De plus, pour alléger les notations, nous démontrons la continuité en l'origine. On munit \mathbb{R}^2 de la norme $\|\cdot\|_\infty$. Nous procédons en deux étapes :

1^{ère} étape : Nous montrons qu'il existe $C \in \mathbb{R}$ tel que, pour tout $r > 0$ dans un voisinage de 0, et pour tout x tel que $\|x\| \leq r$, on ait

$$f(x) - f(0) \leq Cr.$$

Pour voir à, supposons x dans le 1er quadrant (les trois autres cas sont analogues). On a

$$x = p_10 + p_2e_1 + p_3e_2$$

où $p_1 = 1 - x_1 - x_2$, $p_2 = x_1$, $p_3 = x_2$.

(et $0 = (0, 0)$, $e_1 = (1, 0)$ et $e_2 = (0, 1)$). Pour r assez petit on a

$$p_1, p_2, p_3 > 0 \text{ et } p_1 + p_2 + p_3 = 1.$$

Dès lors, par convexité de f (inégalité de Jensen), on trouve

$$f(x) \leq p_1f(0) + p_2f(e_1) + p_3f(e_2)$$

et donc

$$f(x) - f(0) \leq |p_1 - 1||f(0)| + p_2|f(e_1)| + p_3|f(e_2)|.$$

L'affirmation suit alors de ce que

$$|p_1 - 1| \leq 2r, p_2 \leq r, p_3 \leq r.$$

2^{ème} étape : Nous montrons qu'il existe $C_0 \in \mathbb{R}$ tel que, pour tout $r > 0$ dans un voisinage de 0, et pour tout x tel que $\|x\| \leq r$, on ait

$$f(0) - f(x) \leq C'r.$$

Pour voir a, on écrit cette fois

$$0 = p_1x + p_2(-e_1) + p_3(-e_2)$$

avec

$$p_1, p_2, p_3 > 0 \text{ et } p_1 + p_2 + p_3 = 1;$$

cette décomposition est rendue possible en prenant

$$p_1 = \frac{1}{1 + x_1 + x_2}, p_2 = \frac{x_1}{1 + x_1 + x_2}, p_3 = \frac{x_2}{1 + x_1 + x_2}$$

(on suppose toujours x dans le 1er quadrant). Par convexité de f , on trouve

$$f(0) - f(x) \leq |p_1 - 1||f(x)| + p_2|f(-e_1)| + p_3|f(-e_2)|.$$

Prenant r dans un voisinage de l'origine, observant qu'encore une fois

$$|p_1 - 1| \leq 2r, p_2 \leq r, p_3 \leq r$$

et se servant de la première étape pour borner $|f(x)|$, on obtient l'affirmation souhaitée.

La continuité en l'origine est une conséquence directe des deux affirmations ci-dessus.

Remarque 1.4.2. : *La proposition est fausse si K n'est pas ouvert. En effet, prenons par exemple $K = [0, 1] \subset \mathbb{R}$ et considérons $f : K \rightarrow \mathbb{R}$ telle que $f(x) = 0$ si $x \in]0, 1[$ et $f(x) = 1$ si $x \in \{0, 1\}$.*

Définition 1.4.7

Soit f une fonction définie de K dans $\overline{\mathbb{R}} = \mathbb{R} \cup +\infty$, on appelle domaine de f et on note

$$\text{dom}(f) = \{x \in K \text{ tels que } f(x) \neq +\infty\}.$$

Définition 1.4.8

Une fonction f définie sur K est dite coercive si

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

Définition 1.4.9

On dit qu'une fonction f définie de K dans $\mathbb{R} \cup +\infty$ est semicontinue inférieurement (on notera sci) si pour tout $x \in K$,

$$\liminf_{y \rightarrow x} f(y) \geq f(x).$$

Remarque 1.4.3.

1. La semi-continuité inférieure est stable par somme.
2. Toute fonction continue est sci.
3. L'indicatrice d'un convexe fermé est sci (c'est à dire la fonction qui vaut 0 en tout point du convexe et $+\infty$ en dehors)
4. Si f est sci, pour tout $\alpha \in \mathbb{R}$, les ensembles

$$\{x \in K, f(x) \leq \alpha\} \text{ et } \{(x, \alpha) \in K \times \mathbb{R} \text{ tels que } f(x) \leq \alpha\}$$

sont fermés.

Définition 1.4.10

Une fonction de K dans $\mathbb{R} = \mathbb{R} \cup \pm\infty$ est dite propre si elle n'est pas identiquement égale à $+\infty$ et si elle ne prend pas la valeur $-\infty$.

Convexité et différentiabilité au premier ordre Voici un premier résultat permettant de reconnaître la convexité d'une fonction au moyen de ses dérivées premières.

Proposition 1.4.2

Soit f une fonction dérivable sur un intervalle I de \mathbb{R} . On dit que f est convexe si et seulement si sa dérivée est croissante sur I , c-à-d :

$$\text{La fonction } f \text{ est convexe} \iff f' \text{ est croissante sur } I.$$

Preuve :

1. Supposons que f convexe et montrons que f' est croissante. Soient $0 \leq \lambda_1 \leq \lambda_2$. Alors, par la convexité de f

$$f(x + \lambda_1 d) = f\left[\left(1 - \frac{\lambda_1}{\lambda_2}\right)x + \frac{\lambda_1}{\lambda_2}(x + \lambda_2 d)\right] \leq f\left(1 - \frac{\lambda_1}{\lambda_2}\right)x + \frac{\lambda_1}{\lambda_2}f(x + \lambda_2 d)$$

donc

$$\frac{f(x + \lambda_1 d) - f(x)}{\lambda_1} \leq \frac{f(x + \lambda_2 d) - f(x)}{\lambda_2}.$$

D'où le résultat.

2. Supposons que f' croissante, et montrons que f est convexe. Soient donc x, y tel que $y > x$. Montrons que l'application

$$\phi : t \rightarrow tf(x) + (1-t)f(y) - f(tx + (1-t)y)$$

est positive sur $[0, 1]$. On constate que $\phi(0) = \phi(1) = 0$. De plus,

$$\phi'(t) = f(x) - f(y) + (y-x)f'((x-y)t + y).$$

Or $y - x > 0$, donc on a

$$\phi'(t) = k + af'(y - at)$$

, avec $a > 0$. Cette fonction est décroissante, puisque f' est croissante. Si $\phi'(1) > 0$, alors quelque soit $t \in [0, 1]$,

$$\phi'(t) > 0 \text{ et } \phi(1) = \phi(0) + \int_0^1 \phi'(t)dt > \phi(0) = \phi(1),$$

ce qui est absurde. D'où $\phi'(1) \leq 0$.

Par un raisonnement analogue, on montre que $\phi'(0) \geq 0$. De là, on déduit facilement que : $\forall t \in [0, 1], \phi(t) \geq 0$ (sinon, en raisonnant sur les valeurs de la dérivée notamment, on trouve rapidement une contradiction). Donc f est convexe.

Corollaire 1.4.1

Soit f une fonction deux fois dérivable sur un intervalle I de \mathbb{R} . La fonction f est convexe si et seulement si sa dérivée seconde f'' est à valeurs positives ou nulles.

La fonction f est convexe $\iff f'' \geq 0$ sur I .

La fonction f est concave $\iff f'' \leq 0$ sur I .

1. la fonction $f(x) = x^2$ est convexe car $f''(x) = 2 > 0, \forall x \in \mathbb{R}$.
2. la fonction $f(x) = \exp x$ est convexe car $f''(x) = \exp x > 0, \forall x \in \mathbb{R}$.
3. la fonction $f(x) = \ln x$ est concave car $f''(x) = -\frac{1}{x^2} < 0, \forall x \in]0, +\infty[$.

Théorème 1.4.3

Soient E un espace normé, K un ouvert convexe de E et $f : K \rightarrow \mathbb{R}$ une fonction différentiable. Alors, les propriétés suivantes sont équivalentes :

1. f est convexe sur K ;
2. $\forall x, y \in K, f(y) \geq f(x) + f'(x) \cdot (y - x)$;
3. $\forall x, y \in K, (f'(y) - f'(x)) \cdot (y - x) \geq 0$.

Un résultat analogue permet de caractériser la stricte convexité d'une fonction. Il suffit

de remplacer les inégalités ci-dessus par des inégalités strictes et de supposer que les points d'évaluation x et y diffèrent.

Théorème 1.4.4

Soient E un espace normé, K un ouvert convexe de E et $f : K \rightarrow \mathbb{R}$ une fonction différentiable. Alors, les propriétés suivantes sont équivalentes :

1. f est strictement convexe sur K ;
2. $\forall x, y \in K, x \neq y : f(y) > f(x) + f'(x) \cdot (y - x)$;
3. $\forall x, y \in K, x \neq y : (f'(y) - f'(x)) \cdot (y - x) > 0$.

On peut enfin caractériser la forte convexité au moyen des dérivées premières.

Théorème 1.4.5

Soient E un espace euclidien, K un ouvert convexe de E et $f : K \rightarrow \mathbb{R}$ une fonction différentiable. Alors, les propriétés suivantes sont équivalentes :

1. f est fortement convexe sur K ;
2. $\exists \alpha > 0, \forall x, y \in K : f(y) \geq f(x) + f'(x) \cdot (y - x) + \frac{\alpha}{2} \|y - x\|^2$;
3. $\exists \alpha > 0, \forall x, y \in K : (f'(y) - f'(x)) \cdot (y - x) > \alpha \|y - x\|^2$.

Convexité et différentiabilité de seconde ordre

Définition 1.4.11

Soit $H(x_*)$ la matrice Hessienne de la fonction f au point x_* , alors

1. La matrice $H(x_*)$ est dite semi-définie positive (SDP) et on note $H(x_*) \geq 0$ ssi :

$$\forall x \in \mathbb{R}^n, x^t H(x_*) x \geq 0.$$

Donc , toutes les valeurs propres de la matrice $H(x_*)$ sont positives.

2. La matrice $H(x)$ est dite définie positive (DP) et on note $H(x_*) > 0$ ssi :

$$\forall x \in \mathbb{R}^n, x^t H(x_*) x > 0.$$

Donc, toutes les valeurs propres de la matrice $H(x)$ sont strictement positives.

Théorème 1.4.6

Soient $K \subset \mathbb{R}^n$ ouvert convexe non vide et $f : K \rightarrow \mathbb{R}$ une fonction réelle deux fois différentiable dans K , alors

1. La fonction f est convexe dans K si et seulement si la matrice hessienne est semi définie positive (s.d.p) en tout point de K .
2. La fonction est strictement convexe dans K si et seulement si la matrice hessienne est définie positive (d.p) en tout point de K .

Preuve :

1. Supposons que f est convexe dans K et montrons que $H(x_*)$ semi définie pour tout x_*
2. La fonction f est convexe sur K si et seulement si

$$\forall x_* \in K; f(x_* + \lambda x) \geq f(x_*) + \lambda \cdot (\nabla f(x_*))^t x, \forall x \in K. \quad (1.1)$$

Et comme la fonction f est deux fois différentiable dans K en utilise la formule de Taylor Maclaurin d'ordre 2, alors

$$f(x_* + \lambda x) = f(x_*) + \lambda \cdot (\nabla f(x_*))^t x + \frac{1}{2} \cdot \lambda^2 x^t H(x_*) x + \lambda^2 \|x\|^2 \alpha(x_*, \lambda x), \forall x \in K. \quad (1.2)$$

De (1.1) et (1.2), on a

$$\frac{1}{2} \cdot \lambda^2 x^t H(x_*) x + \lambda^2 \|x\|^2 \alpha(x_*, \lambda x) \geq 0$$

On divise par λ^2 et lorsque $\lambda \rightarrow 0$, on obtient $x^t H(x_*) x \geq 0$.

3. Supposons que $H(x_*)$ semi définie pour tout x_* et montrons que f est convexe dans K c-à-d montrons que

$$\forall x_* \in K; f(x) \geq f(x_*) + (\nabla f(x_*))^t (x - x_*), \forall x \in K.$$

En effet, soient x, x_* quelconques de K

$$f(x) = f(x_*) + (\nabla f(x_*))^t (x - x_*) + \frac{1}{2} (x - x_*)^t H(\zeta) (x - x_*); \text{ où } \zeta \in [x, x_*]$$

$$\text{donc } \zeta = tx + (1-t)x_*, t \in]0, 1[$$

$$f(x) - [f(x_*) + (\nabla f(x_*))^t (x - x_*)] = \frac{1}{2} (x - x_*)^t H(\zeta) (x - x_*) \geq 0; \text{ car } H \text{ est s.d.p}$$

D'où la convexité de f .

Exemples 1.4.3. Soit la fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ définie par

$$f(x, y, z) = (x - 2)^2 + (y - 3)^2 + z^2$$

f est-elle convexe ?

On sait que la fonction f est 2 fois différentiable sur \mathbb{R}^3 , et

$$f \text{ est convexe} \iff H(x, y, z) \geq 0, \forall (x, y, z) \in \mathbb{R}^3$$

1. Calculons le gradient

$$\nabla f(x, y, z) = \begin{pmatrix} 2x - 4 \\ 2y - 6 \\ 2z \end{pmatrix}$$

2. Calculons la matrice hessienne

$$H(x, y, z) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} = 2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \geq 0, \forall (x, y, z) \in \mathbb{R}^3.$$

La matrice hessienne a trois valeurs propres : $\lambda_1, \lambda_2, \lambda_3$ Donc, elle est semi définie positive, on conclut que la fonction f est convexe.

Remarque 1.4.4. Dans le cas où f est convexe, alors tout minimum local est aussi global. De plus si f est strictement convexe, alors tout minimum local devient non seulement global mais aussi unique

Exercice 1.4.1. Montrer d'après la définition que la fonction :

$$f(x, y) = x^2 + y^2$$

est différentiable dans \mathbb{R}^2 . Calculer la différentielle.

Exercice 1.4.2. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = \begin{cases} \frac{x^2 y^3}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon} \end{cases}$$

1. Est-elle continue dans \mathbb{R}^2 ?
2. Est-elle dérivable dans \mathbb{R}^2 ?
3. Est-elle de classe C^1 dans \mathbb{R}^2 ?
4. Est-elle différentiable dans \mathbb{R}^2 ?

Exercice 1.4.3. Montrer qu'une norme est convexe.

Exercice 1.4.4. Montrer que la fonction indicatrice d'un ensemble U définie par

$$1_U = \begin{cases} 0 & \text{si } x \in U \\ +\infty & \text{sinon} \end{cases}$$

est convexe si et seulement si U est convexe.

Exercice 1.4.5. Soit C_1, C_2 deux parties convexes d'un espace vectoriel réel E et soit $s \in [0, 1]$. On pose

$$C = sC_1 + (1 - s)C_2 = \{sx + (1 - s)y; x \in C_1, y \in C_2\}.$$

Démontrer que C est convexe.

Exercice 1.4.6. Soit C_1 et C_2 deux ensembles convexes de \mathbb{R}^n et

$$C_1 + C_2 = \{x + y; x \in C_1, y \in C_2\}.$$

Démontrer que $C_1 + C_2$ est convexe.

Exercice 1.4.7. Soient $a, b \in \mathbb{R}$.

1. Montrer que $e^{\frac{a+b}{2}} \leq \frac{e^a + e^b}{2}$.
2. Montrer que $f(x) = \ln(\ln(x))$ est concave sur $]1, +\infty[$.
3. En déduire que $\forall a, b > 1, \ln\left(\frac{a+b}{2}\right) \geq \sqrt{\ln a \cdot \ln b}$.

Exercice 1.4.8. Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}$ telles que f et g soient convexes, et g est croissante. Démontrer que $g \circ f$ est convexe.

Exercice 1.4.9. Soit a une forme bilinéaire symétrique de \mathbb{R}^n dans \mathbb{R} .

1. Montrer que l'on peut trouver une matrice symétrique A d'ordre n telle que :

$$\forall u, v \in \mathbb{R}^n \quad a(u, v) = (Au, v).$$

2. Calculer le gradient et la dérivée seconde (hessien) de la fonctionnelle J définie sur \mathbb{R}^n par :

$$J(v) = \frac{1}{2}(Av, v) - (b, v), \text{ où } b \in \mathbb{R}^n \text{ est fixé.}$$

3. A quelle condition sur A , la fonction J est-elle convexe ? strictement convexe ?

Exercice 1.4.10. Soient a et b deux nombres réels et soit f la fonction définie sur

$$\Omega = \{(x, y) \in \mathbb{R}^2; x > 0, y > 0\}$$

par

$$f(x, y) = x^a + y^b$$

1- Etudier la convexité, stricte convexité, concavité, stricte concavité de f sur Ω suivant les valeurs de a et b .

Exercice 1.4.11. Soit g la fonction à deux variables définie par

$$g(x, y) = \frac{\exp(x+y)}{\sqrt{x+y}}.$$

- (a) Déterminer son domaine de définition Dg . On admet qu'il est ouvert et que g est de classe C^1 sur Dg .
- (b) Dg est-il convexe ?
- (c) Etudier la convexité de g sur son domaine de définition.
- (d) Calculer les dérivées partielles d'ordre 1 de g sur Dg .

Exercice 1.4.12. 1. Que dire de la somme de deux fonctions convexes ?

2. Soient f et g convexes sur I , que dire de $\sup(f, g)$ et de $\inf(f, g)$?
3. Soient f et g convexes sur \mathbb{R} avec g croissante, montrer que $g \circ f$ est convexe. En déduire que si $h : \mathbb{R} \rightarrow \mathbb{R}^+$ est telle que $\ln h$ est convexe, alors h est convexe.

CHAPITRE 2

PROBLÈME D'OPTIMISATION SANS CONTRAINTES

Soit $f : K \rightarrow \mathbb{R}$; On appelle problème de minimisation sans contraintes le problème noté :

$$\min_{x \in \mathbb{R}^n} f(x) \quad (2.1)$$

On veut résoudre $\min_{x \in K \subset \mathbb{R}^n} f(x)$ ou $\max_{x \in K \subset \mathbb{R}^n} f(x)$ i.e. on cherche v valeur optimale et x_* tel que $f(x_*) = v$. Les minima locaux et globaux de f sur \mathbb{R}^n sont définis de la manière suivante :

Définition 2.0.1

On dit que x_* est un minimum local de

$$\exists V \in V(x_0) \text{ tel que } \forall x \in V, f(x) \geq f(x_*).$$

Définition 2.0.2

On dit que x_* est un minimum local strict de

$$\exists V \in V(x_0) \text{ tel que } \forall x \in V, f(x) > f(x_*).$$

Définition 2.0.3

On dit que x_* est un minimum global si

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

Définition 2.0.4

On dit que x_* est un minimum global strict si

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

Définition 2.0.5

Une suite $(x_n)_{n \in \mathbb{N}}$ de points de \mathbb{R}^n est une *suite minimisante* si

$$\lim_{n \rightarrow +\infty} f(x_n) = \inf_{x \in \mathbb{R}^n} f(x).$$

Définition 2.0.6

Un point stationnaire qui n'est ni un minimum ni un maximum est un point singulier.

Définition 2.0.7

x_0 est un point stationnaire ssi $\nabla f(x_0) = 0$.

Remarque 2.0.1. 1. Un minimum global est clairement un minimum local

2. - Si on dit simplement minimum on comprend minimum global.

3. Si on connaît tous les minima locaux alors le plus petit est le minimum.

4. - Toute solution optimale isolée est une solution optimale locale stricte. La réciproque n'est pas toujours vraie.

2.1 Existence et unicité de la solution du problème

Examinons maintenant les questions d'existence et d'unicité de la solution du problème (1). en dimension finie, l'unicité d'une solution éventuelle est en général établie indépendamment de l'existence, le plus souvent à partir de la convexité de l'ensemble K et de la stricte convexité de la fonctionnelle f .

2.1.1 Dans la pratique

Nous noterons (1) le problème $\min_{x \in K} f(x)$ et (2.1) le problème $\min_{x \in \mathbb{R}^n} f(x)$.

Lemme 2.1.1: S

Soit $K \subset \mathbb{R}^n$ un ouvert. Alors

$$x_* \text{ est solution de (1)} \Leftrightarrow \begin{cases} x_* \text{ solution de (2.1)} \\ x_* \in K \end{cases}$$

En pratique, pour résoudre (1), on résout d'abord (2.1), c'est plus simple puis

1. si (2.1) a une solution $x_* \in K$ alors x_* est solution de (1). Sinon, si $x_* \notin K$ alors (1) n'a pas de solution.
2. si (2.1) n'a pas de solution, on ne sait pas pour (1).

2.1.2 Existence d'une solution

Théorème 2.1.1

d'existence 1 : (Théorème de Weierstrass) Soient K un ensemble compact (fermé et borné) non vide de \mathbb{R}^n et $f : K \rightarrow \mathbb{R}$ une fonction continue sur K , alors f admet au moins un min x_* sur K . Autrement dit, il existe un point x_* de K minimum global de f sur K i.e.

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

Preuve :

Soit (x_n) une suite minimisante de f sur K c'est-à-dire d'éléments de K telle que

$$x_n \in K, \forall n \in \mathbb{N} \text{ et } \lim_{n \rightarrow +\infty} f(x_n) = \inf_{x \in K} f(x)$$

Comme K est borné, la suite minimisante soit bornée, donc on peut extraire une sous-suite notée (x_n) qui converge vers un élément x_* de K car K est un ensemble fermé. Cette suite extraite vérifie

$$\lim_{n \rightarrow +\infty} f(x_n) = f(x_*)$$

car f est continue, d'où par unicité de la limite

$$f(x_*) = \inf_{x \in K} f(x).$$

Donc f réalise son minimum dans K .

Théorème 2.1.2

Soit $f : \mathbb{R} \rightarrow \mathbb{R}^n$ une fonction continue et coercive alors f admet au moins un minimum sur \mathbb{R}^n . Autrement dit, il existe un point x_* de \mathbb{R}^n minimum global de f sur \mathbb{R}^n

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

Preuve :

Soit (x_n) une suite minimisante dans \mathbb{R}^n c'est à dire

$$\lim_{n \rightarrow +\infty} f(x_n) = f(x_*)$$

d'où par unicité de la limite

$$f(x_*) = \inf_{x \in \mathbb{R}^n} f(x).$$

Donc ; f réalise son minimum dans \mathbb{R}^n .

Théorème 2.1.3: d'unicité

Si de plus $f : \mathbb{R}^n \rightarrow \mathbb{R}$ strictement convexe, alors il existe un unique minimum $x_* \in \mathbb{R}^n$ de f tel que :

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

Preuve :

Soit f strictement convexe, supposons qu'il existe $x_*, x'_* \in \mathbb{R}^n$ deux min de la fonction f sur \mathbb{R}^n tels que

$$x_* \neq x'_* \text{ et } f(x_*) = f(x'_*) = \inf_{x \in \mathbb{R}^n} f(x).$$

Soit $x''_* = \lambda x'_* + (1 - \lambda)x_* \in \mathbb{R}^n$ avec $\lambda \in (0, 1)$ et comme f est strictement convexe

$$f(x''_*) < \lambda f(x'_*) + (1 - \lambda)f(x_*) = f(x_*) = \min_{x \in \mathbb{R}^n} f(x).$$

Ceci fournit une contradiction, ce qui est impossible ; donc $x_* = x'_*$.

Définition 2.1.1

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$. On dit que f est une fonction elliptique si elle est de classe C^1 et si $\exists b_1 > 0$ telle que

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 3b_1 \|x - y\|^2, \forall x, y \in \mathbb{R}^n.$$

Corollaire 2.1.1

Si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est différentiable, on a les équivalences

1. f est α -elliptique ;
2. $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2} \|y - x\|^2, \forall (x, y) \in \mathbb{R}^2;$
3. $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \alpha \|y - x\|^2, \forall (x, y) \in \mathbb{R}^2.$

Théorème 2.1.4: d'existence et d'unicité)

(Soit f une fonction $C^1(\mathbb{R}^n, \mathbb{R})$ on suppose que f est α -elliptique c'est dire qu'il existe $\alpha > 0$ (appelée la constante d'ellipticité) tel que,

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

Alors, f est strictement convexe et coercive. En particulier le problème (2.1) admet une solution unique.

Preuve :

Il est tout à fait clair que l'ellipticité implique la stricte convexité qui implique elle-même la convexité. Si f est α -elliptique et différentiable, en utilisant la caractérisation précédente, on obtient aisément

$$\langle \nabla f(x) - \nabla f(0), x - 0 \rangle \geq \alpha \|x\|^2$$

ce qui implique que f est coercive.

Proposition 2.1.1: (Caractérisation de l'ellipticité avec ∇^2) :

Soit f une fonction de classe C^2 . Alors f est elliptique si et seulement si $\exists \beta > 0$ tel que

$$\langle \nabla^2 f(x)h, h \rangle \geq \beta, \forall x, h \in \mathbb{R}^n \quad (2.2)$$

Preuve :

Supposons que f est elliptique et montrons (2.2). Soit $h \in \mathbb{R}^n$ fixé et notons $g : \mathbb{R}^n \rightarrow \mathbb{R}$ la fonction donnée par

$$g(x) = \langle \nabla f(x), h \rangle, \forall x \in \mathbb{R}^n$$

Nous avons

$$\langle \nabla^2 f(x)h, h \rangle = \langle \nabla g(x), h \rangle = \frac{\partial g}{\partial h}(x) = \lim_{t \rightarrow 0} \frac{\langle \nabla f(x + th), h \rangle - \langle \nabla f(x), h \rangle}{t}$$

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

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

ce qui nous donne (2.2) avec $\beta = \alpha$.

-Supposons maintenant que (2.2) est satisfaite et montrons que f est elliptique. Soient $x, y \in \mathbb{R}^n$ fixées arbitraires, et considérons la fonction $g_1 : \mathbb{R}^n \rightarrow \mathbb{R}$ donnée par

$$g_1(z) = \langle \nabla f(z), x - y \rangle, \forall z \in \mathbb{R}^n. \text{ Alors}$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = g_1(x) - g_1(y) = \langle \nabla g_1(y + \theta(x - y)), x - y \rangle$$

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

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

et ceci nous permet de déduire, en utilisant aussi (2.2) :

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = \langle \nabla^2 f(y + \theta(x - y))(x - y), x - y \rangle \geq \beta \|x - y\|^2$$

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

Exemples des fonctions elliptiques

1. Si $n = 1$; Toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ de classe C^2 et satisfaisant $:\exists \alpha > 0, f''(x) \geq \alpha, \forall x \in \mathbb{R}$ est une fonction elliptique.

Exemples 2.1.1. (a) $f(x) = ax^2 + bx + c$ avec $a > 0$.

(b) $f(x) = x^2 + \sin(x)$ (car $f''(x) = 2 - \sin(x) \geq 1, \forall x \in \mathbb{R}$).

2. **Le cas général** ($n \in \mathbb{N}^*$); Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ donnée par

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c, \forall x \in \mathbb{R}^n;$$

avec $A \in M_n(\mathbb{R})$ une matrice carrée réelle et symétrique de taille n , avec $b \in \mathbb{R}^n$ un vecteur et $c \in \mathbb{R}$ un scalaire (on appelle fonction (ou forme) quadratique une fonction de ce type). On calcule facilement :

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

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

donc la hessienne de f est constante.

Comme A est symétrique alors on sait que toutes les valeurs propres de A sont réelles et que

$$\langle Ah, h \rangle \geq \lambda_{\min} \|h\|^2, \forall h \in \mathbb{R}^n$$

où $\lambda_{\min} \in \mathbb{R}$ est la plus petite valeur propre de A .

Rappelons aussi que A est une matrice définie positive si et seulement si $\lambda_{\min} > 0$.

(par définition une matrice carrée réelle $B \in M_n(\mathbb{R})$ est définie positive si $\langle Bx, x \rangle > 0$; $\forall x \in \mathbb{R}^n$; $x \neq 0$).

En utilisant la Proposition précédente on déduit que f est une fonction elliptique si et seulement si A est une matrice définie positive.

2.1.3 Condition du premier ordre

Dans cette section, nous allons chercher à obtenir des conditions nécessaires et parfois suffisantes de minimalité.

L'objectif est d'une certaine manière beaucoup plus pratique puisque ces conditions d'optimalité seront le plus souvent utilisées pour tenter de calculer un minimum ou le maximum. Ces conditions sont utiles pour :

1. Vérifier l'optimalité éventuelle d'un point $x \in \mathbb{R}^n$, voir si c'est un minimum, un maximum ou un point stationnaire.
2. Calculer les solutions de (2.1).

Condition d'optimalité nécessaires du premier ordre :CON1

Etant donné un point x_* , la propriété de différentiabilité continue de la fonction f fournit une première manière de caractériser une solution optimale.

Théorème 2.1.5

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $x_* \in \mathbb{R}^n$, si x_* est un minimum local de (2.1) alors $\nabla f(x_*) = 0$.

Preuve :

On démontre par l'absurde, on suppose que $\nabla f(x_*) \neq 0$. Donc on a un vecteur $d = -\nabla f(x_*)$ qui est une direction de descente i.e.

$$\nabla f(x_*)^t \cdot d = -\|\nabla f(x_*)\|^2 < 0$$

et par le théorème il existe $\delta > 0$ tel que

$$f(x_* + \alpha d) < f(x_*), \quad \forall \alpha \in]0, \delta[.$$

Ce qui donne une contradiction avec le fait que x_* est un minimum local, d'où $\nabla f(x_*) = 0$.

Condition d'optimalité nécessaire et suffisante du premier ordre :CONS1

comme celle de la convexité de f permettant en effet de préciser la nature (maximum ou minimum) des extremums considérés. Bien entendu, on pourrait énoncer des résultats analogues pour les maximums relatifs. Si f est convexe, la condition nécessaire du premier ordre est également suffisante.

Théorème 2.1.6

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction convexe différentiable au point $x_* \in \mathbb{R}^n$, un x_* est un minimum de (2.1) si et seulement si

$$\nabla f(x_*) = 0.$$

Preuve :

On a vu que la condition est toujours nécessaire, montrons qu'elle est suffisante.

Soit $x_* \in \mathbb{R}^n$ tel que $\nabla f(x_*) = 0$ et comme f est convexe donc

$$\forall x \in \mathbb{R}^n, f(x) \geq f(x_*) + (\nabla f(x_*))^t (x - x_*) \Rightarrow f(x) \geq f(x_*),$$

On a donc immédiatement le fait que x_* réalise un minimum de f sur \mathbb{R}^n .

Les deux résultats suivants donnent des conditions nécessaires puis des conditions suffisantes du second ordre. Cette condition va faire intervenir la dérivée seconde de f .

2.1.4 Conditions du second ordre

Les résultats qui suivent seront énoncés pour des minimums relatifs, la prise en compte des dérivées secondes

Condition d'optimalité nécessaire du seconde ordre :CON2

Théorème 2.1.7

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction deux fois différentiable au point $x_* \in \mathbb{R}^n$. Si f possède un minimum local en x_* alors :

1. $\nabla f(x_*) = 0$,
2. La matrice Hessienne $H(x_*)$ soit semi définie positive.

Preuve :

Soit f une fonction deux fois différentiable au point $x_* \in \mathbb{R}^n$ et d un vecteur de \mathbb{R}^n , donc pour λ assez et en utilisé la formule de Taylor-Maclaurin d'ordre 2

$$f(x_* + \lambda d) = f(x_*) + \lambda(\nabla f(x_*), d) + \frac{1}{2}\lambda^2 d^t H(x_*) d + \lambda^2 \|d\|^2 \alpha(x_*, \lambda d), \forall \lambda \in \mathbb{R}^n$$

où $(x_*, \lambda d) \rightarrow 0$ quand $\lambda \rightarrow 0$.

le point $x_* \in \mathbb{R}^n$ est le minimum de f c'est-à-dire $\nabla f(x_*) = 0$
donc

$$\frac{f(x_* + \lambda d) - f(x_*)}{\lambda^2} \geq 0, \forall \lambda \in]0, \delta[,$$

$$\frac{1}{2}d^t H(x_*) d + \|d\|^2 \alpha(x_*, \lambda d) \geq 0, \forall \lambda \in]0, \delta[$$

Par passage à la limite lorsque $\lambda \rightarrow 0$ on obtient

$$\frac{1}{2}d^t H(x_*) d \geq 0, \forall d \in \mathbb{R}^n.$$

Donc la matrice hessienne $H(x_*)$ est semi définie positive.

Condition d'optimalité suffisante du seconde ordre :COS2

Le théorème suivant établit une condition suffisante pour qu'un vecteur soit un minimum local, si f est deux fois différentiable.

Théorème 2.1.8

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction deux fois différentiable au point $x_* \in \mathbb{R}^n$. Si $\nabla f(x_*) = 0$, La matrice Hessienne $H(x_*)$ soit semi définie positive alors f possède un minimum local en x_* .

Preuve :

Soit f une fonction deux fois différentiable au point $x_* \in \mathbb{R}^n$, donc f s'écrit sous la forme

$$f(x) = f(x_*) + (\nabla f(x_*))^t (x - x_*) + \frac{1}{2}(x - x_*)^t H(x_*) (x - x_*) + \|x - x_*\|^2 \alpha(x_*, x - x_*),$$

$\forall x \in \mathbb{R}^n$, où $\alpha(x_*, x - x_*) \rightarrow 0$, quand $x \rightarrow x_*$

Supposons que $x_* \in \mathbb{R}^n$ n'est pas un min local stricte c'est-à-dire

$$\forall V_\epsilon(x_*), \exists x_k \in V_\epsilon(x_*), f(x_k) \leq f(x_*).$$

Notons

$$d_k = \frac{x_k - x_*}{\|x_k - x_*\|}, \|d_k\|^2 = 1,$$

d'après le théorème de Bolzano Weierstrass

$$\exists N_1 \in \mathbb{N}, d_k \xrightarrow[k \rightarrow \infty]{} \tilde{d}, k \in N_1$$

donc

$$f(x_k) - f(x_*) = \frac{1}{2}(x - x_*)^t H(x_*)(x - x_*) + \|x - x_*\|^2 \alpha(x_*, x_k - x_*),$$

$$\frac{f(x_k) - f(x_*)}{\|x - x_*\|^2} = \frac{1}{2} d_k^t H(x_*) d_k + \alpha(x_*, \lambda d),$$

or

$$f(x_k) \leq f(x_*) \Rightarrow (x_k) - f(x_*) \leq 0,$$

$$\forall k, \frac{1}{2} d_k^t H(x_*) d_k + \alpha(x_*, x_k - x_*) \leq 0,$$

passant à la limite quand $k \rightarrow +\infty$

$$\tilde{d}^t H(x_*) \tilde{d} \leq 0, \tilde{d} \neq 0 \text{ car } d_k \rightarrow \tilde{d} \text{ et } \|d_k\| = 1.$$

Donc la matrice hessienne $H(x_*)$ n'est pas définie positive.

Exemples 2.1.2. *Considérons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$*

$$f(x) = (x - 5)^2 + (y - 2)^2$$

Etape 1 : Identification des points critiques nous allons utiliser la première condition d'optimalité la fonction f est différentiable et sont gradient est

$$\text{on a } \begin{cases} f'_x = 2(x - 5) \\ f'_y = 2(y - 2) \end{cases} \text{ Les coordonnées } (x, y) \text{ d'un point critique sont solution de } \begin{cases} f'_x = 0 \\ f'_y = 0 \end{cases}$$

$$\text{soit de } \begin{cases} f'_x = 2(x - 5) = 0 \\ f'_y = 2(y - 2) = 0 \end{cases} \text{ . On trouve donc } \begin{cases} x = 5 \\ y = 2 \end{cases}$$

Etape 2 : La nature de point critique nous allons utiliser la deuxième condition d'optimalité, nous permettent de classer ces points critiques.

La matrice hessienne de cette fonction

$$H(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

On a $\det H(5, 2) = 4 > 0$ et c'est valeurs propres sont positives alors la matrice hessienne est définie positives, et donc le point $(x, y) = (5, 2)$ est un minimum local.

Exercice 2.1.1. Les fonctions f suivantes sont-elles *coercive* ?

1. $f : \mathbb{R} \rightarrow \mathbb{R}; f(x) = -3x^3 + 2x^2 + 1.$
2. $f : \mathbb{R} \rightarrow \mathbb{R}; f(x) = 2x^3 + 3x.$
3. $f : \mathbb{R}^2 \rightarrow \mathbb{R}; f(x_1, x_2) = 2x_1^2 + x_2 - 5.$
4. $f : \mathbb{R}^2 \rightarrow \mathbb{R}; f(x_1, x_2) = 2x_1^2 + x_2^2.$
5. $f : \mathbb{R}^n \rightarrow \mathbb{R}; f(x) = (a, x) + b$ avec $a \in \mathbb{R}^n$ et $b \in \mathbb{R}$
6. $f : \mathbb{R}^2 \rightarrow \mathbb{R}; f(x_1, x_2) = 2x_1^2 + x_2^3 + 2x_2^2.$

Exercice 2.1.2. Rechercher les points critiques et déterminer leur nature (*maximum local, minimum local, col*) pour les fonctions f définies ci-dessous :

1. $f(x, y) = (x - 5)^2 + (y - 2)^2.$
2. $f(x, y) = 2x^2 + 6y^2 - 5x + 4y.$
3. $f(x, y) = 4x^2 - 12xy + y^2.$
4. $f(x, y) = ax^2 + by^2$ ($a, b \in \mathbb{R}$).
5. $f(x, y) = x^2y^2 + x^2 + y^2 + 2axy$ ($a \geq 0$).
6. $f(x, y) = \frac{xy}{(1+x^2)(1+y^2)}.$

Exercice 2.1.3. On considère la fonction f suivante :

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

- a) Calculer les dérivées partielles de f et les dérivées secondes de f .
- b) évaluez les dérivées, les dérivées partielles, les dérivées secondes de f aux points

$$A = \begin{pmatrix} x = 1 \\ y = -1 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} x = -1 \\ y = 1 \end{pmatrix}$$

A et B sont-ils des points critiques ? si oui déterminer la nature de ces points critiques ?

c) Déterminer la nature de ces points critiques.

Exercice 2.1.4. Soit f la fonction numérique à variable réelle définie par :

$$\forall x \in \mathbb{R}, f(x) = \sqrt{x^4 + 1} - x + 3.$$

- (a) Montrer que f admet un minimum sur \mathbb{R} .
- (b) Déterminer les points critiques de f .
- (c) Parmi les points critiques, lesquels correspondent à des minima ?

Exercice 2.1.5. Soient $A \in M_n(\mathbb{R})$ une matrice symétrique définie positive et b un vecteur de \mathbb{R}^n . On définit la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ par

$$\forall x \in \mathbb{R}^n, f(x) = \frac{1}{2}x^T A x - b^T x.$$

Montrer que f admet un minimum absolu sur \mathbb{R}^n .

2. Déterminer ce minimum.

Exercice 2.1.6. Soit la fonction f définie par

$$f(x, y) = (y - 1) \ln(y - 1) - \ln(x) + x^2 - xy + 2y^2 - 7y - \frac{3}{2}x + 3.$$

- (a) Donner le domaine de définition D_f de f et faire un dessin de cet ensemble.
- (b) L'ensemble D_f est-il convexe ? Est-il ouvert.
- (c) Montrer que la fonction $\varphi : u \rightarrow \ln(u)$ est convexe sur son ensemble de définition.
- (d) En déduire la convexité de f sur D_f .
- (e) Montrer que $(2, 2)$ est un point critique.
- (f) En déduire la nature de $(2, 2)$.
- (g) La fonction f admet-elle un maximum global sur D_f .

Exercice 2.1.7. Trouver les points critiques et discuter leur nature pour $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

- (a) $f(x, y) = (x - 1)^2 + 2y^2$
- (b) $f(x, y) = 2x^3 - 6xy + 3y^2$
- (c) $f(x, y) = e^{x-y}(x^2 - 2y^2)$
- (d) $f(x, y) = (x^2 + y^2)e^{-(x^2+y^2)}$.
- (e) $f(x, y) = x^3 + y^3 - 3cxy$, ($c \in \mathbb{R}$)
- (f) $f(x, y) = x^2 - \cos(y)$,
- (g) $f(x, y) = x^2y^3$.

Exercice 2.1.8. 1- Soit $J(u) = \frac{1}{2}(Au, v) - (b, v)$ où A est une matrice symétrique de \mathbb{R}^n dans \mathbb{R}^n et $v \in \mathbb{R}^n$, une fonctionnelle quadratique de \mathbb{R}^n dans \mathbb{R} . Démontrer les propositions suivantes :

- (a) J est convexe si et seulement si A est semi-définie positive.
- (b) J est strictement convexe si et seulement si A est définie positive.
- (c) $\exists c \in \mathbb{R}^n$ tel que $\forall v \in \mathbb{R}^n - \{u\}$ $J(u) < J(v)$ si et seulement si A est définie positive.
- (d) $\exists c \in \mathbb{R}^n$ tel que $\forall v \in \mathbb{R}^n$ $J(u) \leq J(v)$ si et seulement si A est semi-définie positive.

2-Soit x^* , un minimum local pour le problème

$$\begin{cases} \min f(x) \\ x \in \mathbb{R}^n. \end{cases} \quad (2.3)$$

i . Montrer que si f est différentiable en x^* , alors $\nabla f(x^*) = 0$. On dit que x^* est un point stationnaire ou critique.

ii . Montrer que si f est deux fois différentiable en x^* , alors $\text{Hess}f(x^*)$ est semi-définie positive.

Exercice 2.1.9. On considère la fonction f définie sur \mathbb{R}^2 par

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

1. tels que pour tous $(x, y) \in \mathbb{R}^2$

$$f(x, y) \geq 2\|(x, y)\|^2 - 162$$

où la notation $\|\cdot\|$ désigne la norme euclidienne de \mathbb{R}^2 . En déduire que le problème

$$\inf_{(x,y) \in \mathbb{R}^2} f(x, y) \quad (2.4)$$

possède au moins une solution.

2. La fonction f est-elle convexe sur \mathbb{R}^2 ?

3. Déterminer les points critiques de f , et préciser leur nature (minimum local, maximum local, point-selle, ...). Résoudre alors le problème (2.4).

CHAPITRE 3

ALGORITHMES

Dans ce chapitre, nous allons présenter quelques algorithmes permettant de calculer (de manière approchée) la ou les solutions du problème (2.1) de départ.

Les algorithmes d'optimisation est une procédure mathématique qui permet d'obtenir et cherchent à déterminer le jeu de paramètres d'entrée d'une fonction donnant à cette fonction la valeur maximale ou minimale, d'une fonction réelle f (que l'on appelle fonction objective)

$$\inf_{x \in \mathbb{R}} f(x) \quad (3.1)$$

Avec $X_* = (x_{*1}, x_{*2}, x_{*3}, \dots)$ sont les coordonnées du point critique. On fait appel aux algorithmes d'optimisation afin de résoudre des problèmes de différente nature, comme par exemple, trouver les zéros de fonctions non-linéaires, ajustement de données expérimentales selon le critère des moindres carrés linéaire et non-linéaire, résolution de systèmes d'équations à une ou plusieurs variables ... etc. En général, la recherche des extremums est atteinte en procédant au calcul des dérivées premières (gradient de la fonction) et des dérivées secondes (Hessien de la fonction).

Les métaheuristiques sont une classe d'algorithmes d'optimisation qui tentent d'obtenir une valeur approchée de l'optimum global dans le cas de problèmes d'optimisation difficile. Elles ne donnent cependant aucune garantie sur la fiabilité du résultat.

On supposera que x_* existe (éventuellement qu'il est unique) et on se propose de trouver une **approximation numérique** de x , en construisant une suite

$$\{x^{(k)}\}_{k \in \mathbb{N}} \subset \mathbb{R}^n \text{ telle que } x^{(k)} \rightarrow x_* \text{ pour } k \rightarrow +\infty;$$

Le principe est de construire un algorithme itératif de la forme

$$x^{k+1} = x^k - \rho_k d^k \quad (3.2)$$

d^k est la direction de descente, ρ_k est le pas. Il est, soit fixé, éventuellement le même pour toutes les étapes (on parle alors de méthode à pas variable), soit calculé à chaque étape de façon à minimiser f dans la direction d^k (on parle alors de méthode à pas optimal).

Pour s'approcher de la solution optimale du problème (3.2) (dans le cas général, c'est un point en lequel ont lieu peut être avec une certaine précision les conditions nécessaires d'optimalité de f), on se déplace naturellement à partir du point x^k dans la direction de la décroissance de la fonction f .

Définition 3.0.1

Algorithme Un algorithme est défini par une application A de \mathbb{R}^n dans \mathbb{R}^n permettant la génération d'une suite d'éléments de \mathbb{R}^n par la formule :

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné, } k = 0 \text{ étape d'initialisation,} \\ x_{k+1} = A(x_k); k = k + 1 \text{ itération } k : \end{cases}$$

Ecrire un algorithme n'est ni plus ni moins que se donner une suite $\{x_k\}_{k \in \mathbb{N}}$ de \mathbb{R}^n ; étudier la convergence de l'algorithme, c'est étudier la convergence de la suite $\{x_k\}_{k \in \mathbb{N}}$.

Convergence globale

On dit qu'un algorithme est globalement convergent (où encore, possède la propriété de la convergence globale) si, quelque soit le point de départ x_0 choisi, la suite $\{x_k\}$ générée par cet algorithme (où une sous suite) converge vers un point satisfant une condition nécessaire d'optimalité.

Vitesse de convergence

La convergence globale d'une algorithme ayant été établie, nous nous intéressons maintenant à l'évaluation de son efficacité d'un point de vue pratique, l'efficacité d'un algorithme dépend du nombre d'itérations nécessaires pour obtenir un approximation à ϵ près (ϵ fixé à l'avance) de l'optimum x^* .

Ceci conduit à attribuer à chaque algorithme une indice d'efficacité appelé se vitesse de convergence.

Nous introduisons ici les principales définitions de base qui seront abondamment utilisées par la suite.

Plaçons nous dans \mathbb{R}^n , où $\|\cdot\|$ désigne la norme euclidienne et considérons une suite $\{x_k\}$ convergeant vers x^* .

- Si $\limsup \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \lambda < 1$

On dit que la convergence est linéaire et λ est le taux de convergence associé.

- Si $\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \rightarrow 0$ quand $k \rightarrow \infty$

on dit que la convergence est superlinéaire.

- Plus précisément si $\exists p > 1$ tel que :

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} < +\infty$$

on dit que la convergence est superlinéaire d'ordre p .

- En particulier si

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} < +\infty$$

on dit que la convergence est quadratique (superlinéaire d'ordre 2).

3.0.1 Schémas général des algorithmes d'optimalité

Il convient de souligner que la plupart des algorithmes d'optimisation, avec contrainte ou non,

fonctionnent selon un schéma général consistant, à chaque itération, à se rapprocher du minimum par la résolution d'un sous problème de minimisation. Les méthodes (ou les algorithmes) itératives d'optimalité fait partie d'une classe plus grande de méthodes numériques appelées méthodes de descente.

Un algorithme à directions de descente est un algorithme d'optimisation différentiable, destiné à minimiser une fonction réelle différentiable définie sur un espace euclidien (par exemple \mathbb{R}^n , muni d'un produit scalaire) ou, plus généralement, sur un espace hilbertien. L'algorithme est itératif et procède donc par améliorations successives. Au point courant, un déplacement est effectué le long d'une direction de descente, de manière à faire décroître la fonction.

3.0.2 Méthode à directions de descente- une itération

Etape 0 :(initialisation)

On suppose qu'au début de l'itération k , on dispose d'un itéré $x_k \in \mathbb{R}^n$

Etape1 :

Test d'arrêt : si $\nabla f(x_k) \simeq 0$, arrêt de l'algorithme ;

Etape2 :

Choix d'une direction de descente $d_k \in \mathbb{R}^n$;

Etape3 :

Recherche linéaire : déterminer un pas $\rho_k > 0$ le long de d_k de manière à "faire décroître f suffisamment" ;

Etape4 :

Si la recherche linéaire réussie

$$x_{k+1} = x_k + \rho_k d_k;$$

remplacer k par $k + 1$ et aller à l'étape 1.

Exemple de choix de direction de descente :

1. Par exemple s'on choisit

$$d_k = -\nabla f(x_k) \text{ et si } \nabla f(x_k) \neq 0,$$

on obtient la méthode du gradient.

2. Dans le cas

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

on obtient la méthode de Newton, où la matrice Hessienne $H(x_k)$ est définie positive.

Exemple de choix de pas λ_k :

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

$$f(x_k + \rho_k d_k) \leq f(x_k + \rho d_k) : \forall \rho \in [0, +\infty[.$$

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

3.0.3 Méthode de Gradient

La méthode (ou algorithme) du Gradient fait partie d'une classe plus grande de méthodes numériques appelées méthodes de descente.

Ici on choisit à chaque étape $d_k = -\nabla f(x_k)$.

L'algorithme de Gradient

1. Initialisation $K = 0$
choix de x_0 , $\rho_0 > 0$ et ϵ ,
2. Itération k

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

3. Critère d'arrêt si

$$\|x_{k+1} - x_k\| < \epsilon \text{ où si } \|\nabla f(x_k)\| < \epsilon \text{ stop}$$

Sinon, on pose $k = k + 1$, et on retourne à 2.

avec ϵ est un réel positif (petit) donné qui représente la précision désirée (la valeur que nous devons fixer pour ϵ dépend du problème considéré).

3.0.4 Méthodes de gradient à pas optimal

la méthode est la suivante :

$$\begin{cases} d_k = -\nabla f(x_k) \\ x_{k+1} = x_k + \rho_k d_k, \rho_k \end{cases} \quad (3.3)$$

où ρ_k est choisi par la règle de minimisation, il consiste à choisir, à chaque itération ρ_k comme étant

la solution optimale du problème de minimisation unidimensionnelle de f le long de la demi-droite

définie par le point x_k et la direction d_k . Donc, ρ_k est choisi de manière à ce que :

$$f(x_k + \rho_k d_k) = \min_{\rho \in \mathbb{R}} f(x_k + \rho d_k), \forall \rho > 0 \quad (3.4)$$

en supposons qu'un tel minimum existe.

Remarque 3.0.1. Nous faisons l'itération (3.3) dans le cas où $\nabla f(x_k) \neq 0$, donc le problème de minimisation (3.4) a un sens.

On a le résultat suivant :

Théorème 3.0.1

Soit f une fonction $C^1(\mathbb{R}^n, \mathbb{R}^n)$, coercive et strictement convexe. On suppose qu'il existe une constante $M > 0$ tel que

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

Alors, si on choisit le pas ρ_k dans un intervalle $[\beta_1, \beta_2]$ tel que $0 < \beta_1 < \beta_2 < \frac{2}{M}$, la méthode du gradient converge vers le minimum de f .

Preuve :

La fonction f admet un minimum x_0 unique sur \mathbb{R}^n caractérisé par $\nabla f(x_0) = 0$ puisque f est strictement convexe. Montrons que la suite (x_k) engendrée par l'algorithme converge vers x_0 , on a :

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

On applique cette relation à $y = x_{k+1}$, $x = x_k$

$$f(x_{k+1}) = f(x_k) + (\nabla f(x_k), x_{k+1} - x_k) + \int_0^1 (\nabla f(x_k + t(x_{k+1} - x_k)) - \nabla f(x_k), x_{k+1} - x_k) dt$$

Comme $x_{k+1} = x_k - \rho_k \nabla f(x_k)$, on obtient

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -\frac{1}{\rho_k} \|x_{k+1} - x_k\|^2 + \int_0^1 \|\nabla f(x_k + t(x_{k+1} - x_k)) - \nabla f(x_k)\| \cdot \|x_{k+1} - x_k\| dt \\ &\leq -\frac{1}{\rho_k} \|x_{k+1} - x_k\|^2 + \frac{M}{2} \|x_{k+1} - x_k\|^2 = \left(\frac{M}{2} - \frac{1}{\rho_k}\right) \|x_{k+1} - x_k\|^2 \end{aligned}$$

Si on choisit le pas ρ_k dans un intervalle $[\beta_1, \beta_2]$ tel que $0 < \beta_1 < \beta_2 < \frac{2}{M}$, nous obtenons alors

$$f(x_{k+1}) - f(x_k) \leq \left(\frac{M}{2} - \frac{1}{\rho_k}\right) \|x_{k+1} - x_k\|^2$$

La suite $f(x_k)$ est alors strictement décroissante, comme elle est minorée car

$$f(x_k) \geq f(x_0), \forall k$$

elle est convergente. Cela entraîne d'une part que $(f(x_{k+1}) - f(x_k))$ tend vers 0 et d'autre part, la suite (x_k) est bornée (car f est coercive). On peut donc extraire une sous suite convergente vers x . De plus, comme

$$\|x_{k+1} - x_k\|^2 \leq \left(\frac{M}{2} - \frac{1}{\rho_k}\right)^{-1} (f(x_{k+1}) - f(x_k)).$$

La suite $(x_{k+1} - x_k)$ tend également vers 0.

Par conséquent

$$\nabla f(x_k) = \frac{x_{k+1} - x_k}{\rho_k} \rightarrow 0$$

Par continuité de $\nabla f(\bar{x}) = 0$, donc \bar{x} est l'unique minimum x_0 de f . Ceci étant vrai pour toute valeur d'adhérence de la suite (x_k) cela prouve que toute la suite (x_k) converge vers x_0 .

3.0.5 Méthode de Gradient à pas fixe

On peut utiliser un pas fixé a priori $\rho > 0$, $\forall k$ on obtient alors la méthode de gradient simple

$$\begin{cases} d_k = -\nabla f(x_k) \\ x_{k+1} = x_k + \rho d_k \end{cases}$$

Pour $f \in C^1$, cette méthode converge si ρ est choisi assez petit.

Le choix de pas :

- Un pas bien choisi donne des résultats à ceux obtenus par la plus profonde descente.
- Un pas plus petit atténue les zigzags des itérés mais augmente significativement le nombre d'itérations.
- Un pas trop grand fait diverger la méthode.

Cas particulier : fonction quadratiques

Dans ce paragraphe nous supposons que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est donnée par

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c \quad (3.5)$$

avec $A \in M_n(\mathbb{R})$ matrice SDP (c'est à dire symétrique et définie positive), $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ donc c'est une forme quadratique associée à une matrice SDP.

Nous devons calculer $\rho_k \in \mathbb{R}$ qui minimise la fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ donnée par

$$g(\rho) = f(x_k - \rho \nabla f(x_k)).$$

Alors ρ_k satisfait nécessairement

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

Un calcul simple nous donne

$$g'(\rho) = -\langle \nabla f(x_k - \rho \nabla f(x_k)), \nabla f(x_k) \rangle$$

c'est à dire, comme $\nabla f(x_k) = Ax_k - b$

$$\begin{aligned} g'(\rho) &= -\langle \nabla(x_k - \rho \nabla f(x_k)) - b, Ax_k - b \rangle \\ &= -\|Ax_k - b\|^2 + \rho \langle A(Ax_k - b), Ax_k - b \rangle \end{aligned}$$

nous obtenons alors

$$\rho_k = \frac{\|Ax_k - b\|^2}{\langle A(Ax_k - b), Ax_k - b \rangle} \quad (3.6)$$

Remarquons qu'on a $\langle A(Ax_k - b), Ax_k - b \rangle > 0$ (car A est SDP et $Ax_k - b = \nabla f(x_k) \neq 0$).
Donc la méthode de gradient à pas optimal dans le cas f quadratique est :

$$x_{k+1} = x_k - \rho_k (Ax_k - b)$$

avec ρ_k donnée par (3.6) valable uniquement pour $Ax_k - b \neq 0$.

3.0.6 Méthode du gradient conjugué

Algorithme de la méthode du gradient conjugué

Les méthodes du gradient conjugué sont utilisées pour résoudre des systèmes d'équations linéaires dont la matrice est symétrique définie positives, on l'utilise aussi pour résoudre les grands systèmes linéaires est une méthode itérative qui converge en un nombre fini d'itérations (au plus égal à la dimension du système linéaire).

Définition 3.0.2

Soit A une matrice symétrique $n \times n$, définie positive. On dit que deux vecteurs x et y de \mathbb{R}^n sont A -conjugués (ou conjugués par rapport à A) s'ils vérifient

$$x^T A y = 0.$$

Principe de la méthode

Soit $\{d_0, d_1, \dots, d_n\}$ une famille de vecteurs A -conjugués. On appelle alors méthode de directions conjuguées toute méthode itérative appliquée à une fonction quadratique strictement convexe de n variables : $f(x) = \frac{1}{2}xAx^T + b^T x + c$, avec $x \in \mathbb{R}^n$ et $A \in M_{n \times n}$ est symétrique et définie positive, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$, conduisant à l'optimum en n étapes au plus.

Cette méthode est de la forme suivante :

Elles reposent sur le concept des directions conjuguées parce que les gradients successifs sont orthogonaux entre eux et aux directions précédentes.

L'idée de la méthode étant donnés un point initial x_0 de \mathbb{R}^n et n -directions A -conjugués est de construire itérativement des directions d_1, \dots, d_k mutuellement conjuguées on définit le -schéma suivant :

$$x_{k+1} = x_k + \rho_k d_k, \text{ avec} \tag{3.7}$$

où ρ_k est le scalaire minimisant $f(x)$ selon la direction $x_k + \rho_k d_k$ et qui est défini par

$$\rho_k = -\frac{r_k^T d_k}{d_k^T A d_k} \tag{3.8}$$

$$r_k = -\nabla f(x_k) = b - Ax_k \tag{3.9}$$

L'algorithme ci-dessous résout $Ax = b$, où A est une matrice réelle, symétrique, et définit positive. Le vecteur d'entré x_0 peut tre une approximation de la solution initiale ou 0.

Algorithme de la méthode de Gradient Conjuguée

1. **Initialisation** $k = 0$,

Choix de x_0 dans \mathbb{R}^n ; $r_0 = -\nabla f(x_0) = b - Ax_0$, $p_0 = r_0$

2. **Itération** k

$$\alpha_k = -\frac{r_k^\top r_k}{p_k^\top A p_k}$$

$$x_{k+1} = x_k + \alpha_k p_k, r_{k+1} = r_k - \alpha_k A p_k,$$

3. **Critère d'arrêt**

Si r_{k+1} est suffisamment petit, alors on sort de la boucle

$$\beta_k = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}, p_{k+1} = r_{k+1} + \beta_k p_k,$$

Sinon, on pose $k = k + 1$, et on retourne à 2.

Théorème 3.0.2

Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est quadratique et elliptique la méthode de gradient conjugué converge en n -itérations au plus où n est l'ordre de A .

3.0.7 Méthode de Newton

La méthode de Newton n'est pas une méthode d'optimisation à proprement parler.

C'est en réalité une méthode utilisée pour résoudre des équations non linéaires de la forme $F(x) = 0$ où F est une fonction de \mathbb{R}^n dans \mathbb{R}^n . Nous allons d'abord la décrire puis montrer comment on peut l'appliquer à la recherche de minimum.

Dans le cas $n = 1$, cette méthode de Newton s'appelle également la méthode de la tangente.

L'idée est de remplacer le point x_k obtenu à l'itération k , par le point d'intersection de la tangente en $(x_k; f(x_k))$ à la courbe représentative de f avec l'axe des abscisses.

Considérons maintenant le cas général, i.e. on dispose d'une fonction $F : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe (au moins) C^1 telle que l'équation 0 ($\nabla F(x) = 0$) admette au moins une solution x et la matrice $H(x)$ est définie positive.

Principe de la méthode

Considérons le problème d'optimisation sans contraintes (P)

$$(P) : \min_{x \in \mathbb{R}^n} f(x)$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$

Le principe de la méthode de Newton consiste à minimiser successivement les approximations du second ordre de f , plus précisément si

$$f(x) = f(x_k) + \nabla f(x_k)^t (x - x_k) + (x - x_k)^t H(x_k) (x - x_k) + o(\|x - x_k\|^2);$$

posons

$$q(x) = f(x) + \nabla f(x_k)^t (x - x_k) + (x - x_k)^t H(x_k) (x - x_k)$$

Soit x_{k+1} l'optimum de q , alors il vérifie $\nabla q(x_{k+1}) = 0$; soit en remplaçant :

$$\nabla f(x_k) + H(x_k)(x_{k+1} - x_k) = 0;$$

ou encore

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

donc

$$x_{k+1} = x_k - [H(x_k)]^{-1}\nabla f(x_k).$$

Théorème 3.0.3

Soit F est une fonction de classe C^2 de \mathbb{R}^n dans \mathbb{R}^n et x un zéro de F (c'est-à-dire \mathbb{R}^n). On suppose en outre que ce zéro est isolé et que $DF(x)$ est inversible (DF désigne la dérivée première de F).

Alors il existe une boule fermée B centrée en x_* telle que, pour tout point $x_0 \in B$, la suite (x_k) définie par la méthode de Newton est entièrement contenue dans B et converge vers x qui est le seul zéro de F dans B .

Enfin la convergence est géométrique : il existe $B \in]0; 1[$ tel que

$$\forall k \geq 0; \|x_k - x_*\| \leq B \|x_0 - x_*\|.$$

En d'autres termes, si on choisit le point de départ x_0 "assez près" de x_* , alors ; l'algorithme converge vers x_* .

Preuve :

Comme F est C^1 et $DF(x)$ est inversible, il existe une boule centrée en x_* : $B(x, r_0)$ sur laquelle $DF(\cdot)$ est inversible et $DF(\cdot)^{-1}$ est uniformément bornée par m .

Appliquons la formule de Taylor avec reste intégral entre x_* et un itéré x_k en supposant que $x_k \in B(x, r_0)$

$$F(x_*) - F(x_k) = DF(x_k)(x_* - x_k) + \int_0^1 D^2F(x_* + t(x_k - x_*)) \cdot (x_* - x_k)^2 t dt \quad (3.10)$$

Comme F est C^2 , $D \in F$ est continue et uniformément bornée sur $B(x_*, r_0)$ par un réel $M > 0$, et nous avons

$$\left| \int_0^1 D^2F(x_* + t(x_k - x_*)) \cdot (x_* - x_k)^2 t dt \right| \leq \frac{M}{2} \|x_* - x_k\|^2.$$

Par conséquent, comme

$$\begin{aligned} x_{k+1} - x_* &= x_k - x_* - DF^{-1}(x_k) [F(x_k) - F(x_*)] \\ &= DF(x_k)^{-1} [F(x_*) - F(x_k) - DF(x_k)(x_* - x_k)] \\ &= DF(x_k)^{-1} \left[\int_0^1 F(x_*) - F(x_k) - DF(x_k)(x_* - x_k)^2 t dt \right]; \end{aligned}$$

on obtient

$$\|x_{k+1} - x_*\| \leq m \frac{M}{2} \|x_* - x_k\|^2. \quad (3.11)$$

Posons $r = \min(r_0, \frac{2}{mM}, 1)$; il est alors facile de voir par récurrence que si $x_0 \in B(x_*, r)$ alors pour tout k , $x_k \in B(x, r)$: la suite des itérés est bien définie et reste dans la boule. Posons alors $e_k = \frac{mM}{2} \|x_k - x\|$. La relation (3.11) donne

$$e_{k+1} = e_k^2.$$

La suite e_k converge donc vers 0 si $e_0 = \frac{mM}{2} \|x_0 - x_*\| < 1$, cad si x_0 est dans la boule $B(x, r)$ où $r = \min(r, \frac{1}{mM})$ par exemple.

Remarque 3.0.2. - Cette méthode est bien définie à chaque itération si la matrice hessienne $H(x_k)$ est définie positive : ceci est vrai en particulier au voisinage de la solution x_* cherchée si on suppose que $H(x_*)$ est définie positive (par continuité de H).

- La méthode de Newton est un algorithme de descente à pas fixe égal à 1.

- Si la fonctionnelle f est quadratique, strictement convexe, alors l'algorithme converge en une seule itération.

Algorithme de la méthode de Newton

1. Initialisation $k = 0$,

choix de x_0 dans un voisinage de x_* , ϵ

2. Itération k

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

3. Critère d'arrêt

si $\|x_{k+1} - x_k\| < \epsilon$ stop

Sinon, on pose $k = k + 1$, et on retourne à 2.

Inconvénients de la méthode de Newton :

1. Cette méthode fonctionne très bien pour les petites dimensions ($1 \leq n \leq 10$) lorsque on peut calculer facilement $H(x)$ et $(H(x))^{-1}$, ce calcul nécessite des opérations plus nombreuses et coûteuses dans les problèmes des grandes tailles.

2. Comme $x_{k+1} = x_k - (H(x_k))^{-1} \nabla f(x_k)$; le point (x_{k+1}) n'est pas toujours bien défini c-à-d

possible que $(H(x_k))^{-1}$ n'existe pas (cela intervient typiquement lorsque la méthode atteint un région où f est linéaire donc ses secondes dérivées partielles valent zéro), et même elle est

existe, la direction $d_k = -(H(x_k))^{-1} \nabla f(x_k)$ n'est pas toujours une direction de descente (si

$(H(x_k))$ est définie positive alors d_k est une direction de descente).

3- L'inconvénient majeur de la méthode est sa sensibilité au choix du point de départ x_0 : Si ce point est mal choisi (trop loin de la solution) la méthode peut soit diverger, soit converger vers une autre solution. Pour choisir le point de départ x_0 "assez près" de x_* on essaie de s'approcher de x_* par une méthode de type gradient par exemple, puis on applique la méthode de Newton.

3.0.8 Méthode de Quasi-Newton

Les méthodes de Quasi-Newton sont élaborées pour l'optimisation et pour pallier aux inconvénients de la méthode de Newton elle :

- 1- garde la rapidité de la méthode de Newton,
- 2- évite le calcul (coteux) de la matrice $[H(x_k)]$ à chaque itération.
- 3- plus robustes par rapport au du point de départ. On y trouve les méthodes dites "région de

confiance" qui s'attachent à rendre la méthode robuste (i.e. peu sensible) par rapport x_0 .

- 4- $(H(x))^{-1}$ n'est pas nécessairement connue, peut tre très chère à calculer, et $H(x_k)$ peut tre très difficile à inverser. On remplace alors et $(H(x_k))^{-1}$ par une matrice D_k , éventuellement constante, qui est censée approcher $H(x_k)$ ou bien son inverse. Parfois mme, on remplace $(H(x_k))^{-1} \nabla f(x_k)$ par un vecteur y_k facile à calculer.

Donc, l'algorithme de cette méthode est donné comme suit :

Algorithme de la méthode de Quasi-Newton

1. Initialisation $k = 0$,

choix de x_0, α_0, ϵ

2. Itération k

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

3. Critère d'arrt

si $\|x_{k+1} - x_k\| < \epsilon$ stop

Sinon, on pose $k = k + 1$, et on retourne à 2.

Remarque 3.0.3. Dans cet algorithme, on va utiliser une approximation D_k de $(H(x_k))^{-1}$ et puis on va trouver α_k (par une recherche linéaire) qui minimise la fonction $\phi(\alpha) = f(x_k + \alpha_k d_k)$.

EXERCICES ET TRAVEAUX PRATIQUES

TP 01

On commence par la dimension 2 car on peut représenter les fonctions $\mathbb{R}^2 \rightarrow \mathbb{R}$ en 3D ou grce aux lignes de niveau. On s'intéresse à un problème de minimisation d'une fonction quadratique, car on sait l'étudier théoriquement : on pourra donc comparer les résultats numériques aux résultats attendus.

Soit

$$A = \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}, b = (1, 1)^t \text{ et } J(x) = \frac{1}{2}Ax.x - b.x$$

1. Expliquer pourquoi le problème de minimisation de J sur \mathbb{R}^2 a une solution unique.
2. En déterminer la solution.
3. Représenter J sur le pavé $[-10, 10]^2$ grce aux commandes meshgrid et mesh.

TP 02

1. Programmer et tester les algorithmes suivante :
 - Gradient à pas constant.
 - Algorithme de Newton.
 - Relaxation avec sous-programme Newton (pour \mathbb{R}).
2. Pour chacun d'entre eux, une étude de sensibilité sur le point de départ (initialisation) et le pas éventuel sera menée le plus rigoureusement possible. On fera une comparaison numérique des trois méthodes surtout en termes de :
 - Vitesse de convergence - nombre d'itérations - temps CPU
 - Robustesse et domaine de validité en particulier sur les exemples suivants (f de \mathbb{R}^2 dans \mathbb{R}).
 - a $f(x, y) = x^2 - 5xy + y^4 - 25x - 8y$
 - b $f(x, y) = (x^4 - 3) + y^4$
 - c $f(x, y) = x^4 - 4y^3 + 6(x^2 + y^2) - 4(x + y)$.

TP 03

Programmer l'algorithme du gradient conjugué pour résoudre $Ax = b$; comparer avec les procédures de MATLAB en terme de précision et de temps CPU.

On fera des tests sur des matrices définies positives de grande taille et sur l'exemple suivant

$$A = \begin{pmatrix} 0.78 & -0.02 & -0.12 & -0.14 \\ -0.02 & 0.86 & -0.04 & 0.06 \\ -0.12 & -0.04 & 0.72 & -0.08 \\ -0.14 & 0.06 & -0.08 & 0.74 \end{pmatrix} \text{ et } b = \begin{pmatrix} 0.76 \\ 0.08 \\ 1.12 \\ 0.68 \end{pmatrix}.$$

Exercice 3.0.1. Soit $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle$ où A est une matrice symétrique définie positive $N \times N$ de valeurs propres $0 < \lambda_1 < \lambda_2 < \dots < \lambda_N$ et $b \in \mathbb{R}^n$. Notons x_* le minimum de f sur \mathbb{R}^n . On considère la suite (u_k) d'éléments de \mathbb{R}^n telle que $x_0 \in \mathbb{R}^n$ quelconque et

$$\forall k \in N, u_{k+1} = u_k - \rho \nabla f(u_k)$$

où ρ est un réel positif fixé.

- Montrer que $u_{k+1} - u_* = (I_n - \rho A)(u_k - u_*)$, où I_n désigne la matrice identité de taille N

- Montrer que l'algorithme de gradient à pas fixe converge pour $0 < \rho < \frac{2}{\lambda_N}$.

Exercice 3.0.2. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}.$$

On fixe $x_1, \dots, x_n \in \mathbb{R}$. Résoudre le problème suivant :

$$\mu^* = \arg \max_{\mu \in \mathbb{R}} \left[\prod_{i=1}^n f(x_i) \right].$$

[Indication : on pourra passer au logarithme.]

Exercice 3.0.3. Soient $n \geq 2$ et $f : \mathbb{R} \rightarrow \mathbb{R}$ la fonction définie par

$$f(x) = (1 + x_n)^3 \prod_{i=1}^{n-1} x_i^2 + x_n^2.$$

Montrer que 0 est le seul point critique de f , que f y atteint un minimum local strict, mais pas global.

Exercice 3.0.4. On considère la fonction

$$f(x, y) = \frac{1}{2}x^2 + \frac{1}{2}y^2$$

En partant du point initial $(x_0, y_0) = (1, 1)$ et en appliquant la méthode du gradient avec k optimale, calculez (x_1, y_1) ; (x_2, y_2) et (x_3, y_3) .

Exercice 3.0.5. Vérifier que le calcul de l'inverse d'un scalaire par la méthode de Newton correspond à la méthode itérative :

$$x_{k+1} = x_k(2 - \alpha x_k); k \geq 0$$

Construire, par analogie, une méthode itérative d'approximation de l'inverse d'une matrice inversible A , de la forme :

$$\begin{cases} B_0 \text{ matrice arbitraire,} \\ B_{k+1} = \text{fonction}(B_k, A); k \geq 0 \end{cases}$$

Démontrer qu'une CNS de convergence de cette méthode est $\rho(I - AB_0) < 1$ où $\rho(I - AB_0)$ désigne le rayon spectral de la matrice $I - AB_0$.

Supposant la matrice A symétrique, définie, positive et supposant connu son rayon spectral, comment choisir simplement la matrice B_0 pour vérifier la condition précédente ?

Exercice 3.0.6. Soit la fonction de Rosenbrock définie comme suit :

$$f(X) = f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

1- Calculer le gradient et la matrice Hessienne de la fonction f .

2- Vérifier que $x_* = [1, 1]^t$ est un minimum local de f .

3- Calculer les 5 premiers itérés de la méthode de Newton pour minimiser f en commençant par $x_0 = [-1, -2]^t$. Calculer la norme de l'erreur $\|x - x_*\|$ à chaque itération et déterminer si le taux de convergence est quadratique

- [1] G. Allaire, Analyse numérique et optimisation, Edition 2002.
- [2] M. Belloufi, Cours d'optimisation sans contraintes, 2015
- [3] M. Bergounioux, Optimisation et contrôle des systèmes linéaires, Dunod, Paris, 2001.
- [4] J-F. Bonnans, J-C. Gilbert, C. Lemaréchal, C. Sagastizàbal, Optimisation Numérique, Aspects théoriques et pratiques, Springer M&A 27, 1997.
- [5] R. Fletcher (1987), Practical methods of optimization, John Wiley&Sons, Chichster.
- [6] R. Fletcher, Practical Methods of Optimization vol. 1 : Unconstrained Optimization, John Wiley & Sons, New York, 1987.
- [7] J. C. Gilbert , Eléments d'optimisation différentiable : théorie et algorithmes, Notes de cours, École Nationale Supérieure de Techniques Avancées, Paris, (2007).
- [8] R.T. Picard, Convexité et applications, Notes de cours. Université de Rennes 1.
- [9] G.R.Walsh, Methods of optimization, A wiley- Interscience Publication, 1975.

Sites Internet

- 1- <http://dumas.perso.math.cnrs.fr/LSMA651-2011.html>
- 2- <http://dumas.perso.math.cnrs.fr/>
- 3- <http://aeropedia.enac-aerospace.eu/index.php?title=Optimisation>