Chapitre 3

Conditions d'optimalité en optimisation sans contrainte

3.1 Introduction

On suppose dans tout ce chapitre que $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction objectif une ou deux fois différentiable, et qu'elle admet (au moins) un minimum local. On notera $x^\star$ le minimum (local) de f. Pour résoudre les problèmes d'optimisation, on s'intéresse au problème suivant : trouver le minimum d'une fonction f. D'un point de vue mathématique, les problèmes sans ou sous contrainte(s) peuvent être décrits comme

— Le problème d'optimisation sans contrainte

$$f(x^\star) = \inf_{x \in \mathbb{R}^n} f(x)$$

— Le problème d'optimisation sous contraintes

$$f(x^\star) = \inf_{x \in C} f(x)$$

où $C \subsetneq \mathbb{R}^n$. Si C est fermé et borné (donc compact) et si f est continue sur C, alors il existe un minimum global atteint en un point de C et un maximum global atteint en un point de C.

En dimension 1, si $f : \mathbb{R} \to \mathbb{R}$ est dérivable, alors tout point $x^\star$ réalisant un minimum/maximum local vérifie

$$f'(x^\star) = 0$$

De plus, un point réalisable $x^\star \in C$ est

— un minimum global de la fonction f sur le domaine C si

$$f(x^\star) \leqslant f(x), \forall x \in C$$

— un minimum local (point critique) de f sur C si

$$\exists \varepsilon > 0, \forall x \in C, \ \|x - x^\star\| \leqslant \varepsilon \ \Rightarrow f(x^\star) \leqslant f(x)$$

Dans ce chapitre, on se concentre sur le problème d'optimisation sans contrainte. Le cas sous contraintes sera étudié au Chapitre 6.

3.2 Direction réalisable et direction de descente

Définition 3.1 – Direction réalisable/*Feasible direction*/可行方向

Soit $S \subseteq \mathbb{R}^n$, $S \neq \varnothing$. Un vecteur $d \in \mathbb{R}^n$ est appelé *direction réalisable au point $x \in S$* si

$$\exists \lambda^\star > 0, \ \forall \lambda \in]0, \lambda^\star], \ x + \lambda.d \in S$$

Notons $Z(x)$ l'ensemble des directions réalisables au point x.

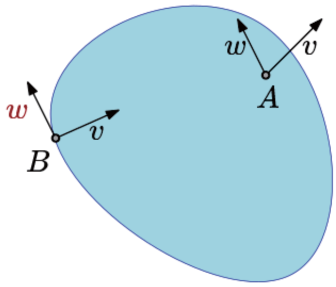

FIGURE 3.1 – Caption

Remarque 3.1

1. L'ensemble $Z(x)$ n'est jamais vide, car $0 \in Z(x)$.
2. L'ensemble $Z(x)$ n'est pas forcément fermé.

Définition 3.2 – Direction de descente/*Descent direction*/下降方向

Soit $x \in \mathbb{R}^n$, un vecteur $d \in Z(x)$ est

1. la *direction de descente au point* $x^\star$ si

$$\exists \lambda^\star > 0,\ \forall \lambda \in]0, \lambda^\star],\ f(x + \lambda.d) \leqslant f(x)$$

2. la *direction stricte de descente au point* x si

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

Notons $D(x)$ l'ensemble des directions de descente au point x.

Proposition 3.1 – Caractérisation des directions de descente

Soient $f \in \mathscr{C}^1(\mathbb{R}^n, \mathbb{R})$, $x \in \mathbb{R}^n$ *et* $d \in \mathbb{R}^n \backslash \{0\}$, *alors*

(a) si d est une direction de descente en x, alors $\langle \nabla f(x), d \rangle \leqslant 0$,

(b) si $\nabla f(x) \neq 0$ alors $d = -\nabla f(x)$ est une direction de descente stricte en x.

Démonstration 4

(a) Soit $d \in \mathbb{R}^n \backslash \{0\}$ une direction de descente en x. Alors par définition, $\exists \lambda^\star > 0$ tel que

$$f(x + \lambda.d) \leqslant f(x),\ \forall \lambda \in]0, \lambda^\star]$$

Soit $\varphi : \mathbb{R} \to \mathbb{R}$ définie par $\varphi(\lambda) = f(x + \lambda.d)$. On a $\varphi \in \mathscr{C}^1(\mathbb{R}, \mathbb{R})$, alors

$$\varphi'(\lambda) = \langle \nabla f(x + \lambda.d), d \rangle$$

Comme d est une direction de descente, on peut aussi écrire

$$\varphi(\lambda) \leqslant \varphi(0),\ \forall \lambda \in]0, \lambda^\star]$$

et donc

$$\forall \lambda \in]0, \lambda^\star[,\ \frac{\varphi(\lambda) - \varphi(0)}{\lambda} \leqslant 0$$

En passant à la limite lorsque λ tend vers 0, on déduit que $\varphi'(0) \leqslant 0$, c.à.d. $\langle \nabla f(x), d \rangle \leqslant 0$.

(b) Soit $d = -\nabla f(x) \neq 0$. On veut montrer qu'il existe $\lambda^\star > 0$ tel que si $\lambda \in]0, \lambda^\star]$ alors $f(x + \lambda.d) < f(x)$ ou encore que $\varphi(\lambda) < \varphi(0)$ où φ est la fonction définie en 1 ci-dessus. On a

$$\varphi'(0) = \langle \nabla f(x), d \rangle = -|\nabla f(x)|^2 < 0$$

Comme φ' est continue, il existe $\lambda^\star > 0$ tel que si $\lambda \in]0, \lambda^\star]$ alors $\varphi'(\lambda) < 0$. Si $\lambda \in]0, \lambda^\star]$ alors $\varphi(\lambda) - \varphi(0) = \int_0^\lambda \varphi'(t)dt < 0$, et on a donc bien $\varphi(\lambda) < \varphi(0)$ pour tout $\lambda \in]0, \lambda^\star]$, ce qui prouve que d est une direction de descente stricte en x.

L'ensemble des directions strictes de descente de f en x, $\{d \in \mathbb{R}^n, \langle \nabla f(x), d\rangle < 0\}$, forme un demi-espace ouvert de $\mathbb{R}^n$. Quelques questions à réfléchir

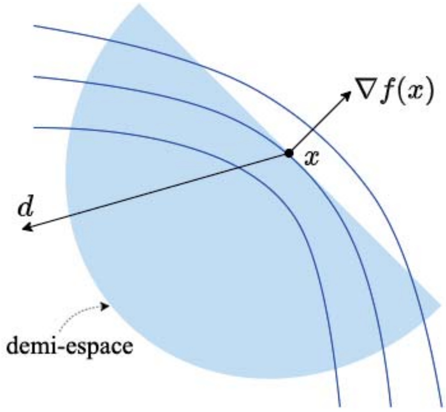

Figure 3.2 – Demi-espace des directions de descente d de f en x.

(1) Pour $f(x + \lambda.d)$, que se passe-t-il si λ est choisi avec des valeurs différentes ?

(2) Quel résultat obtiendrons-nous si nous choisissons itérativement la direction de descente ?

3.3 Résultats d'existence et d'unicité

Avant d'étudier les propriétés de la solution du problème d'optimisation sans contrainte, il faut assurer de leur existence.

Théorème 3.1 – Théorème d'existence (Weierstrass)

Soient C un ensemble compact (fermé et borné) non vide de $\mathbb{R}^n$ et $f : C \to \mathbb{R}$ une fonction continue sur C, alors f admet un minimum $x^\star$ sur C. Autrement dit, il existe un point $x^\star$ de C minimum global de f sur C

$$\forall x \in C,\ f(x) \geqslant f(x^\star)$$

Démonstration 5

Soit $(x_m)_{m\in N}$ une suite minimisante de f sur C, c'est-à-dire d'éléments de C telle que

$$x_m \in C,\ \forall m \in N \text{ et } \lim_{m\to+\infty} f(x_m) = \inf_{x\in C} f(x)$$

Comme C est borné, la suite minimisante est bornée, donc on peut extraire une sous-suite notée $(x_{\varphi(m)})_{m\in\mathbb{N}}$ qui converge vers un élément $x^\star$ de C car C est un ensemble fermé dans $\mathbb{R}^n$. Cette suite extraite vérifie

$$\lim_{m\to+\infty} f(x_{\varphi(m)}) = f(x^\star)$$

car f est continue. Par unicité de la limite

$$f(x^\star) = \inf_{x\in C} f(x)$$

Donc f réalise son minimum dans C.

Remarque 3.2

1. De la même façon, il existe un point de maximum global de f sur C.
2. Si la fonction f n'est pas continue alors elle n'admet pas nécessairement de minimum.
3. Une fonction continue sur un fermé borné n'atteint pas toujours son minimum dans un espace de dimension infinie.
4. L'existence d'une suite minimisante provient de la définition de l'inf.
5. Rappelons qu'un compact C de $\mathbb{R}^n$ est caractérisé par la propriété que toute suite de points de l'ensemble C admet une valeur d'adhérence dans l'ensemble.

Dans le cas des problèmes d'optimisation sans contraintes ($C = \mathbb{R}^n$), le théorème suivant est posé

Théorème 3.2 – Théorème d'existence

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue et coercive (i.e. infinie à l'infini : $\lim_{\|x\| \to +\infty} f(x) = +\infty$) alors f admet au moins un minimum sur $\mathbb{R}^n$. Autrement dit, il existe un point $x^\star$ de $\mathbb{R}^n$ minimum global de f sur $\mathbb{R}^n$

$$\forall x \in \mathbb{R}^n,\ f(x) \geqslant f(x^\star)$$

Démonstration 6

Soit $(x_m)_{m\in\mathbb{N}}$ une suite minimisante dans $\mathbb{R}^n$, c'est-à-dire

$$\lim_{m\to+\infty} f(x_m) = \inf_{x\in\mathbb{R}^n} f(x)$$

Comme f est coercive, la suite $(x_m)_{m\in\mathbb{N}}$ ne peut pas partir à l'infini donc elle est bornée. Donc il existe $R > 0$ tel que $(x_m)_{m\in\mathbb{N}} \subset \mathrm{BF}(0, R)$ compacte. Donc il existe une sous-suite extraite notée $(x_{\varphi(m)})_{m\in\mathbb{N}}$ de $\mathbb{R}^n$ qui converge vers un $x^\star \in \mathbb{R}^n$

$$x_{\varphi(m)} \xrightarrow[m\to+\infty]{} x^\star \in \mathbb{R}^n$$

Or f est continue, donc

$$\lim_{m\to+\infty} f(x_{\varphi(m)}) = f(x^\star)$$

Par unicité de la limite

$$f(x^\star) = \inf_{x\in\mathbb{R}^n} f(x)$$

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

Théorème 3.3 – Théorème d'unicité

Soit $C \subset \mathbb{R}^n$ un convexe. Si $f : C \to \mathbb{R}$ est strictement convexe et qu'elle admet un minimum sur C, alors il existe un unique minimum $x^\star \in C$ de f tel que

$$\forall x \in C,\ f(x) \geqslant f(x^\star)$$

Démonstration 7

Soit f strictement convexe, supposons qu'il existe $x^\star$, $\overline{x}$ dans C deux minimums de la fonction f sur $\mathbb{R}^n$ tels que

$$x^\star \neq \overline{x} \text{ et } f(x^\star) = f(\overline{x}) = \inf_{x\in\mathbb{R}^n} f(x)$$

Soit $\tilde{x} = \lambda.\overline{x} + (1-\lambda).x^\star \in C$ car C convexe avec $\lambda \in]0, 1[$ et comme f est strictement convexe

$$f(\tilde{x}) < \lambda f(\overline{x}) + (1-\lambda) f(x^\star) = f(x^\star) = \inf_{x\in\mathbb{R}^n} f(x)$$

Ceci fournit une contradiction, donc $x^\star = \overline{x}$.

3.4 Les conditions d'optimalité

Ce qui suit reste valable dans le cas où $f : C \subsetneq \mathbb{R}^n \to \mathbb{R}$, lorsque le minimum $x^\star$ se trouve à l'intérieur de l'ensemble des contraintes C.

Théorème 3.4 – Conditions nécessaires (pour des cas sans contrainte)

Soit $x^\star$ un minimum local du problème

$$\inf_{x\in\mathbb{R}^n} f(x)$$

Alors, $x^\star$ vérifie nécessairement

- ***condition nécessaire du premier ordre** : si f est différentiable en $x^\star$, on a $\nabla f(x^\star) = 0$;*
- ***condition nécessaire du second ordre** : si f est deux fois différentiable au point $x^\star$, alors la forme quadratique $H_f(x^\star)$ est semi-définie positive, c'est-à-dire*

$$\forall h \in \mathbb{R}^n,\ \langle H_f(x^\star)\cdot h, h\rangle \geqslant 0$$

où H_f est la matrice hessienne, définie par les coefficients $\frac{\partial^2 f}{\partial_{x_i}\partial_{x_j}}$.

Démonstration 8

Condition nécessaire du premier ordre
On écrit $f(x^\star) \leqslant f(x^\star + \varepsilon.h) = f(x^\star) + \varepsilon\langle\nabla f(x^\star), h\rangle + |\varepsilon.h|\varphi(\varepsilon.h)$, avec $\varphi(\varepsilon.h) \xrightarrow[\varepsilon\to 0]{} 0$, ce qui donne $0 \leqslant \varepsilon\langle\nabla f(x^\star), h\rangle + |\varepsilon.h|\varphi(\varepsilon.h)$. On divise alors par $\varepsilon > 0$ puis on fait tendre ε vers 0^+. Enfin, en choisissant dans le développement précédent $\pm h$ pour tout $h \in \mathbb{R}^n$, la conclusion s'ensuit.
Condition nécessaire du second ordre
On utilise un développement de Taylor-Young à l'ordre 2 et on utilise les mêmes notations que précédemment. On a

$$f(x^\star + h) = f(x^\star) + \langle\nabla f(x^\star), h\rangle + \frac{1}{2}\langle H_f(x^\star)\cdot h, h\rangle + \|h\|^2\varphi(h) = f(x^\star) + \frac{1}{2}\langle H_f(x^\star)\cdot h, h\rangle + \|h\|^2\varphi(h)$$

Comme précédemment, on remplace h par $\varepsilon.h$, h quelconque, ε petit, puis on divise par ε^2 et on fait tendre ε vers 0.

Remarque 3.3

La réciproque est fausse. Pourquoi ?

Exemple 3.1

Les exemples $f(x) = x^3$, $f(x) = x^4$ ...

Exemple 3.2

Trouver, s'il existe, le minimum de $f(x, y) = x^2 - xy + y^2 + 2x - y + 1$.
La première étape consiste à annuler les deux dérivées partielles afin de détecter le ou les points candidats.

$$\begin{cases} \dfrac{\partial f}{\partial x}(x, y) = 2x - y + 2 = 0 \\ \dfrac{\partial f}{\partial y}(x, y) = -x + 2y - 1 = 0 \end{cases}$$

Résolvons ce système d'équations. Les solutions sont $x = -1$ et $y = 0$. Donc, un point unique est à considérer.
La deuxième étape consiste à vérifier la matrice hessienne de la fonction f, on a

$$H_f = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

La matrice admet 1 et 3 pour valeurs propres. La fonction est donc (fortement) convexe et le point $(-1, 0)$ est bien un minimum.

Théorème 3.5 – Conditions suffisantes (pour des cas sans contrainte)

Soit f une fonction de classe $\mathscr{C}^1$ définie sur $\mathbb{R}^n$. On suppose que $\nabla f(x^\star) = 0$ et que f est deux fois différentiable en $x^\star$. Alors, $x^\star$ est un minimum (local) de f si l'une des deux conditions suivantes est vérifiée :

(a) $H_f(x^\star)$ est définie positive,

(b) $\exists r > 0$ tel que f est deux fois différentiable sur $BO(x^\star, r)$ et la forme quadratique $H_f(x)$ est semi-définie positive pour tout $x \in BO(x^\star, r)$.

Remarque 3.4

Le caractère «semi-défini positif» de la hessienne en $x^\star$ ne suffit pas pour conclure, comme en atteste l'exemple $f(x) = x^3$. En revanche, le caractère «défini-positif» de la hessienne n'est pas nécessaire, comme en témoigne l'exemple $f(x) = x^4$.

Démonstration 9

$H_f(x^\star)$ est définie positive, par conséquent, il existe $\alpha > 0$ tel que $\langle H_f(x^\star) \cdot h, h\rangle \geqslant \alpha \|h\|^2$ pour tout $h \in \mathbb{R}^n$ (rappelons que α peut être choisi égal à la plus petite valeur propre de la matrice hessienne de f en $x^\star$). On écrit alors la formule de Taylor-Young à l'ordre 2 en $x^\star$

$$f(x^\star + h) = f(x^\star) + \frac{1}{2}\langle H_f(x^\star) \cdot h, h\rangle + \|h\|^2 \varphi(h) \geqslant f(x^\star) + \left(\frac{\alpha}{2} + \varphi(h)\right) \|h\|^2 > f(x^\star)$$

pourvu que h soit choisi assez petit, puisque $\varphi(h) \xrightarrow[h \to 0]{} 0$.

Proposition 3.2 – Conditions nécessaires et suffisantes (pour des cas sans contrainte)

Soit f une fonction convexe de classe $\mathscr{C}^1$, définie sur $\mathbb{R}^n$ et $x^\star$ un point de $\mathbb{R}^n$. Alors, $x^\star$ est un point local (donc global) de f si et seulement si $\nabla f(x^\star) = 0$.

Exemple 3.3

Soit une fonction objectif avec $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, et $b = (b_1, \ldots, b_n) \in (\mathbb{R}_+)^n$ définie par

$$f(x) : \begin{cases} \mathbb{R}^n \longrightarrow \mathbb{R} \\ x \longmapsto \displaystyle\sum_{i=1}^{n} (e^{x_i} - b_i x_i) \end{cases}$$

On observe alors que f est de classe $\mathscr{C}^\infty(\mathbb{R}^n)$, son gradient est donné par

$$\nabla f(x)_i = e^{x_i} - b_i$$

et sa matrice hessienne est diagonale avec

$$H_f(x)_{ii} = e^{x_i}$$

(a) Puisque f est convexe, tout point critique est point de minimum global.

(b) Puisque f est strictement convexe, s'il existe un point critique, il est unique.

(c) Puisque f n'est pas fortement convexe, on ne peut pas utiliser la convexité pour démontrer l'existence d'un point critique.

Alors, on examine les résultats que différentes valeurs b_i apporteront.

(1) Si $b_i > 0$, $i = 1, \ldots, n$ alors il existe un unique point critique

$$\nabla f(x)_i = 0 \iff x_i = \ln(b_i)$$

C'est l'unique point de minimum global dans $\mathbb{R}^n$.

(2) Si $\exists b_i \leqslant 0$, alors il n'existe pas de point critique, donc pas de minimum.

3.5 Exercices

3.1 Soit $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x, y, z) = x^2 + y^2 + z^2 + xy + yz + xz - 3x - 4y - z + 4$.

a. Mettre f sous la forme d'une fonction quadratique $f(x) = \frac{1}{2}\langle A \cdot x, x\rangle - \langle b, x\rangle + c$ avec A est une matrice symétrique, $b \in \mathbb{R}^3$, $c \in \mathbb{R}$.

b. Soient $\lambda_{\min}$ et $\lambda_{\max}$ la plus petite et la plus grande valeur propre de A. Montrer que

$$\forall x \in \mathbb{R}^3,\ \lambda_{\min}\|x\|_2^2 \leqslant \langle A \cdot x, x\rangle \leqslant \lambda_{\max}\|x\|_2^2$$

c. Montrer que si A est définie positive alors f est « infinie à l'infini » et admet un minimum unique sur $\mathbb{R}^3$.

3.2 On considère un nuage de m points de $\mathbb{R}^2$: $M_i = (t_i, x_i)$, pour $i \in \{1, \ldots, m\}$. Ces données sont souvent le résultat de mesures et on cherche à décrire le comportement global de ce nuage. En général, ces points ne sont pas alignés. On cherche la droite approchant au mieux ces points par la méthode des moindres carrés.
Remarque : La méthode des moindres carrés consiste alors à rechercher la droite telle que la somme des carrés des distances des points du nuage à cette droite soit minimale.