



DÉPARTEMENT STPI

3ÈME ANNÉE MIC

Introduction à l'Optimisation Numérique

Frédéric de Gournay & Aude Rondepierre

Table des matières

Introduction	5
Rappels de topologie dans \mathbb{R}^n	7
0.1 Ouverts et fermés de \mathbb{R}^n	7
0.2 Notions de minimum, maximum, infimum, supremum	8
1 Formulation et analyse d'un problème d'optimisation	13
1.1 Description d'un problème d'optimisation	13
1.2 Condition suffisante d'existence d'un point minimum	15
1.2.1 Contre-exemples à l'existence d'un minimum	15
1.2.2 Cas où l'ensemble X des contraintes est borné	16
1.2.3 Cas où l'ensemble X des contraintes est non borné	16
1.2.4 Cas des contraintes d'égalité	18
1.3 Convexité et optimisation	19
1.3.1 Résultats d'existence et d'unicité en optimisation convexe	20
1.3.2 Caractérisation différentielle de la convexité	21
2 Optimisation numérique sans contraintes	23
2.1 Conditions d'optimalité	23
2.2 Généralités sur les algorithmes de descente	25
2.2.1 Notion de direction de descente	25
2.2.2 Algorithme de descente modèle	28
2.2.3 Convergence et vitesse de convergence	29
2.3 Premiers algorithmes de descente	31
2.3.1 Algorithmes de gradient à pas fixe/pas optimal	31
2.3.2 Méthode de Newton locale	35
2.3.3 Méthode de Gauss-Newton	37
3 Introduction à l'optimisation sous contraintes	39
3.1 Conditions d'optimalité	40
3.1.1 Cas d'une contrainte d'inégalité	40
3.1.2 Cas de contraintes d'inégalité et d'égalité	43
3.2 Lagrangien du problème	45
3.3 Algorithme du gradient projeté	50

A Compléments	55
A.1 Rappels de calcul différentiel	55
A.2 Quelques démonstrations	57
A.2.1 Hessienne et convexité : démonstration du Théorème 1.5	57
A.2.2 Méthode de Gauss-Newton pour la résolution des problèmes de moindres carrés	59

Introduction

L'optimisation et particulièrement l'optimisation numérique a connu un essor important ces dernières années avec l'avènement de l'ordinateur. Elle est souvent l'ultime étape de l'analyse numérique où, après avoir étudié un phénomène physique, l'avoir mis en équation, avoir étudié ces équations et avoir montré que l'on pouvait calculer les solutions avec un ordinateur, on commence à optimiser le système en changeant certains paramètres pour changer la solution dans un sens désiré.

Nous ne nous intéresserons qu'aux notions de bases de l'optimisation et pour se fixer les idées, imaginons-nous un étudiant qui doit réviser ses examens. Nous supposons qu'il a trois matières à passer et qu'il révise x_i heures sur la $i^{\text{ème}}$ matière, pour $i = 1, 2, 3$. Nous supposons que la note n_i de la matière i dépend uniquement du temps passé x_i . Pour se fixer les idées supposons que

$$n_1(x_1) = 3x_1, \quad n_2(x_2) = x_2^2, \quad n_3(x_3) = x_3(100 - x_3)$$

Cet étudiant veut passer au plus 10 heures au total à réviser et veut optimiser sa moyenne totale. On note $x = (x_1, x_2, x_3)$ le triplet qui correspond aux heures passées à travailler, et on note $f(x) = \frac{1}{3}(3x_1 + x_2^2 + x_3(100 - x_3))$ sa moyenne. D'après les données du problème x ne peut pas prendre n'importe quelles valeurs réelles, on dit que le problème est "contraint". En effet x doit appartenir à X , l'ensemble des **contraintes** :

$$X = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \text{ tel que } \forall i, x_i \geq 0 \text{ et } x_1 + x_2 + x_3 \leq 10\}$$

On note le problème de l'étudiant

$$\max_{x \in X} f(x)$$

En généralisant notre exemple à la modélisation mathématique de problème d'optimisation, on distingue trois étapes :

1. Identification des variables de décisions (désignées par le vecteur $x \in \mathbb{R}^3$ dans notre exemple) : ce sont les paramètres sur lesquels l'utilisateur peut agir pour faire évoluer le système considéré.
2. Définition d'une fonction coût ou fonction objectif (la fonction f dans notre exemple) permettant d'évaluer l'état du système (ex : rendement, performance, ...).
3. Description des contraintes imposées aux variables de décision (l'ensemble X dans notre exemple).

Le problème d'optimisation consiste alors à déterminer les variables de décision conduisant aux meilleures conditions de fonctionnement du système (ce qui revient à minimiser ou maximiser

la fonction coût), tout en respectant les contraintes d'utilisation définies à l'étape 3.

Dans ce cours, nous nous intéressons aux méthodes numériques pour l'optimisation continue, différentiable et non linéaire.

Le premier chapitre traite des notions mathématiques fondamentales à maîtriser avant de s'intéresser à la résolution à proprement parler de tout problème d'optimisation : la description mathématique d'un problème d'optimisation, la notion de solution locale et quelques éléments d'analyse convexe.

Dans les chapitres suivants, nous entrons dans le vif du sujet en nous intéressant à la résolution de problèmes d'optimisation sans contrainte (cf chapitre 2), puis avec contraintes (cf chapitre 3). Pour chacun de ces problèmes nous tenterons de répondre aux questions suivantes :

1. Existe-t-il une solution (même locale) du problème considéré ? si oui, a-t-on unicité ?
2. Comment la caractériser ? (cf conditions d'optimalité).
3. Comment la calculer ? Quel type d'algorithme choisir ?

Rappels de topologie dans \mathbb{R}^n

0.1 Ouverts et fermés de \mathbb{R}^n

Soient $x \in \mathbb{R}^n$ et $r > 0$. On appelle *boule ouverte de centre x et de rayon r* l'ensemble :

$$B(x, r) = \{y \in \mathbb{R}^n \text{ tel que } \|y - x\| < r\}.$$

Définition 0.1 (Ouvert, fermé) *Un ensemble O de \mathbb{R}^n est dit ouvert si pour tout $x \in O$, il existe une boule ouverte $B(x, r)$ de centre x et de rayon r incluse dans O . Un ensemble F de \mathbb{R}^n est dit fermé si son complémentaire est un ouvert.*

Exemple 0.1.1 1. Les ensembles \emptyset et \mathbb{R}^n sont ouverts, mais ils sont aussi fermés car leur complémentaires respectifs sont \mathbb{R}^n et \emptyset . Ce sont d'ailleurs les seuls exemples d'ensemble à la fois ouverts et fermés.

2. Les boules ouvertes sont des ouverts.

3. Les intervalles de la forme $]a, b[$, $-\infty \leq a < b \leq +\infty$, sont des ouverts de \mathbb{R} .

4. Les intervalles de la forme $[a, b]$, $-\infty < a < b < +\infty$, sont des fermés de \mathbb{R} .

L'interprétation géométrique de "ouvert" et "fermé" est qu'un ensemble est ouvert si et seulement si il ne contient aucun point de sa frontière. Un ensemble est fermé si et seulement si il contient tous les points de sa frontière. Les ensembles qui contiennent une partie de leur frontière mais pas l'intégralité ne sont ni ouverts ni fermés.

Aller plus loin : Il faut faire un peu attention à la définition que l'on vient de donner des ensembles ouverts et fermés car il en existe d'autres, qui donnent des ouverts et des fermés différents. Se donner une définition des ouverts s'appelle, en mathématique, "se donner une topologie". La définition que nous avons donné est la définition classique qui construit la "topologie usuelle".



Proposition 0.1 (Caractérisation des fermés) *Soit $F \subset \mathbb{R}^n$. F est un fermé si et seulement si toute suite convergente d'éléments de F a sa limite dans F .*

Preuve. (\Rightarrow) Supposons F fermé et notons O son complémentaire. Soit $(x_n)_{n \in \mathbb{N}}$ une suite d'éléments de F qui converge vers $x \in \mathbb{R}^n$. Supposons par l'absurde que $x \in O$.

Comme O est un ouvert de \mathbb{R}^n , il existe $r > 0$ tel que : $B(x, r) \subset O$. Or $(x_n)_{n \in \mathbb{N}}$ converge vers x d'où :

$$\exists N \in \mathbb{N}, \forall n \geq N, \|x_n - x\| < r,$$

Ceci implique $x_n \in B(x, r) \subset O$ et donc $x_n \in O$ partir d'un certain rang, ce qui est impossible car $x_n \in F$ quel que soit $n \in \mathbb{N}$ par définition.

(\Leftarrow) Supposons que toute suite convergente de F admet une limite dans F . Montrons que O , le complémentaire de F , est ouvert.

Par l'absurde, supposons que O ne soit pas ouvert, i.e. : il existe $x \in O$ tel que : $\forall r > 0, B(x, r) \not\subset O$. Autrement dit :

$$\forall r > 0 \quad B(x, r) \cap F \neq \emptyset.$$

Pour $r = \frac{1}{n+1}$, nous construisons une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de F telle que

$$x_n \in B(x, \frac{1}{n+1}).$$

Ceci peut être traduit en $\|x_n - x\| \leq \frac{1}{n+1}$. Il suit que x_n converge vers x . Comme F est fermé, $x \in F$, ce qui est impossible car $x \in O$. \square

0.2 Notions de minimum, maximum, infimum, supremum

On distinguera les notions de minimum et de maximum des notions d'infimum et de supremum. Ces notions sont des prérequis pour les démonstrations des résultats d'existence et d'unicité d'extrema d'une fonction donnée.

Définition 0.2 (Minorant/Majorant) Soit E un sous-ensemble de \mathbb{R} , $m \in \mathbb{R} \cup \{-\infty, +\infty\}$ est un minorant de E ssi m est inférieur ou égal à tous les éléments de E tandis que $M \in \mathbb{R} \cup \{-\infty, +\infty\}$ est un majorant de E ssi il est supérieur ou égal à tous les éléments de E . Ainsi

$$(m \text{ minorant} \iff \forall x \in E \quad m \leq x) \quad \text{et} \quad (M \text{ majorant} \iff \forall x \in E \quad M \geq x).$$

Si E admet un minorant (resp. majorant) fini alors il est dit minoré (resp. majoré).

Définition 0.3 (Infimum/Supremum) Soit $E \subset \mathbb{R}$. L'infimum $\inf(E) \in \mathbb{R} \cup \{-\infty, +\infty\}$ de E est le plus grand des minorants. Le supremum $\sup(E) \in \mathbb{R} \cup \{-\infty, +\infty\}$ de E est le plus petit des majorants. On les note respectivement

$$\inf(E) = \inf_{x \in E} (x) \quad \text{et} \quad \sup(E) = \sup_{x \in E} (x).$$

Aller plus loin : La définition 0.3 n'est pas une vraie définition. En effet rien ne prouve que l'on puisse trouver "le plus grand des minorants". Nous sommes en train d'essayer d'expliquer ce qu'est le plus grand ou le plus petit élément d'un ensemble et nous utilisons pour cela le fait que l'on puisse trouver "le plus grand des minorants". Si on réfléchit bien, nous sommes en train de tourner en rond. Le fait qu'il existe toujours un infimum et un supremum

s'appelle l'"axiome de la borne supérieure" et vient de la construction de \mathbb{R} (qui est complètement hors programme). C'est d'ailleurs l'unique propriété qui fait la spécificité de \mathbb{R} . En effet les autres propriétés importantes de \mathbb{R} sont qu'il est un corps archimédien (caractéristiques qu'il partage avec \mathbb{Q}).

Par définition, on autorise les minorants ou les majorants, à être infinis. Se pose la question de savoir si l'infimum (resp. le supremum) est infini. La réponse à cette question est donnée par la proposition suivante.

Proposition 0.2 *Soit $E \subset \mathbb{R}$, alors $\inf(E) \in \mathbb{R}$ (resp. $\sup(E) \in \mathbb{R}$) si et seulement si E est minoré (resp. majoré).*

Nous avons parlé d'infimum et de supremum, nous les relierons maintenant aux définitions classiques de minimum et de maximum.

Définition 0.4 (Minimum, maximum) *Soit $E \subset \mathbb{R}$. L'infimum de E est appelé minimum ssi $\inf(E) \in E$. Le supremum de E est appelé maximum ssi $\sup(E) \in E$. Dans ce cas, on les note respectivement $\min(E)$ et $\max(E)$.*

Exemple 0.2.1 *Si on se donne $E =]0, 1]$, alors l'ensemble des minorants de E est $[-\infty, 0]$ et l'ensemble des majorants est $[1, +\infty]$. On en déduit que le l'infimum vaut 0 et le supremum vaut 1. Comme 1 appartient à E alors c'est aussi le maximum de E . Cependant 0 n'appartient pas à E et donc E n'a pas de minimum. On dit souvent que l'infimum de E n'est pas atteint.*

Nous démontrons maintenant une propriété importante des fermés.

Proposition 0.3 *Soit $F \subset \mathbb{R}$ un fermé non vide de \mathbb{R} .*

- Si F est minoré alors F admet un minimum et ainsi $\inf(F) = \min(F) \in \mathbb{R}$.
- Si F est majoré alors F admet un maximum et ainsi $\sup(F) = \max(F) \in \mathbb{R}$.

Pour démontrer la proposition 0.3, on introduit le concept de suites minimisante et de suite maximisante :

Définition-Proposition 0.1 (Suite minimisante/Suite maximisante) *Soit $E \subset \mathbb{R}$, $E \neq \emptyset$. Il existe $(x_n)_n$ et $(y_n)_n$ deux suites d'éléments de E telle que*

$$\lim_{n \rightarrow +\infty} x_n = \inf(E), \quad \lim_{n \rightarrow +\infty} y_n = \sup(E).$$

La suite $(x_n)_n$ est appelée "suite minimisante" et $(y_n)_n$ une "suite maximisante" de E .

Preuve. Comme $E \neq \emptyset$, nécessairement : $\inf(E) \in \mathbb{R} \cup \{-\infty\}$ et $\sup(E) \in \mathbb{R} \cup \{+\infty\}$. *1er cas :* $\inf(E) \in \mathbb{R}$. Comme $\inf(E)$ est le plus grand des minorants de E , alors quel que soit $n \in \mathbb{N}$, $\inf(E) + 1/n$ n'est pas un minorant de E . Il existe donc un élément $x_n \in E$ tel que

$$x_n \leq \inf(E) + 1/n.$$

Cette suite (x_n) d'éléments de E vérifie $\inf(E) \leq x_n \leq \inf(E) + 1/n$ et admet donc $\inf(E)$ comme limite.

2nd cas : $\inf(E) = -\infty$. E admet seulement $-\infty$ comme minorant. Par conséquent pour tout $n \in \mathbb{N}$, il existe $x_n \in E$ tel que

$$x_n \leq -n.$$

La suite x_n ainsi construite converge vers $-\infty$. La construction de la suite $(y_n)_n$ est similaire à celle de $(x_n)_n$ et est laissée en exercice. \square

Preuve de la proposition 0.3. Soit $(x_n)_n$ une suite minimisante

$$\lim_{n \rightarrow +\infty} x_n = \inf(F).$$

Comme F est fermé, $\inf(F) \in F$ et, par conséquent, $\inf(F)$ est le minimum de F . La démonstration concernant le maximum est similaire et est laissée en exercice. \square

Théorème 0.1 (Bolzano-Weierstrass) *De toute suite bornée $(x_n)_n$ de \mathbb{R}^p , il existe une sous-suite qui converge.*

Preuve. Soit $(x_n)_n$, une suite bornée de \mathbb{R}^p , il nous faut montrer qu'il existe une fonction $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ strictement croissante telle que la suite $(x_{\sigma(n)})_n$ converge vers une limite que nous noterons x^* . Nous posons $\sigma(0) = 0$ et nous construisons σ par le procédé récurent suivant :

Comme $(x_n)_n$ est bornée, il existe un $A > 0$ tel que pour tout n , x_n soit inclus dans le pavé $[-A, A]^p$. Nous coupons chaque intervalle $[-A, A]$ en deux, ce qui donne deux intervalles $[-A, 0]$ et $[0, A]$ de sorte que le pavé $[-A, A]^p$ soit découpé en 2^p pavés de taille A . Comme il existe une infinité d'éléments de la suite, il y a forcément un de ces 2^p petits pavés qui contient une infinité de terme. On choisit un tel pavé et on prend $\sigma(1)$ comme le plus petit des indices j tel que x_j appartienne à ce pavé et soit strictement plus grand que $\sigma(0)$. On continue à construire σ ainsi de suite en coupant chaque pavé en 2^p pavés plus petits à chaque fois et en choisissant un petit pavé qui contient une infinité de termes de la suite. On a ainsi construit σ telle que σ est strictement croissante et que pour tout $n_0 \in \mathbb{N}$ il existe un pavé de taille $2^{-n_0}A$ tel que pour tout $n > n_0$, $x_{\sigma(n)}$ soit contenu dans ce pavé. Une telle suite est dite : *suite de Cauchy*. Pour prouver que cette suite converge, notons x^i la $i^{\text{ème}}$ coordonnée de x et introduisons les suites $(y_n)_n$ et $(z_n)_n$ définies par :

$$y_n^i = \sup_{k \geq n} x_{\sigma(k)}^i \quad \text{et} \quad z_n^i = \inf_{k \geq n} x_{\sigma(k)}^i$$

La suite $(y_n^i)_n$ (resp. $(z_n^i)_n$) est décroissante (resp. croissante) et minorée (resp majorée) donc converge vers une limite notée x_+^i (resp x_-^i). Or comme pour tout $n \geq n_0$ la suite $x_{\sigma(n)}$ est contenue dans un pavé de taille $2^{-n_0}A$, alors

$$\forall n_0 \geq 0, \forall n > n_0, \quad |y_n^i - z_n^i| \leq 2^{-n_0}A$$

En faisant tendre d'abord n vers $+\infty$ puis ensuite n_0 vers $+\infty$, on voit que les limites x_+^i et x_-^i sont égales. On utilise ensuite le théorème des gendarmes avec

$$\forall n \in \mathbb{N}, \quad z_n^i \leq x_{\sigma(n)}^i \leq y_n^i,$$

pour montrer que la suite numérique $x_{\sigma(n)}^i$ tend bien vers une limite pour tout i et par ainsi la suite de vecteurs $x_{\sigma(n)}$ tend bien vers une limite. \square

Aller plus loin : Quand vous verrez les notions de *compact*, *complet*, *suite de Cauchy*, de *lim inf et de lim sup*, il est utile de revoir la démonstration du théorème 0.1 qui montre que les ensemble fermés bornés de \mathbb{R}^n sont compact. Par exemple, on démontre dans la démonstration du dessus que les suites de Cauchy convergent (ce qui veut dire que \mathbb{R}^n est complet). La démonstration de ce fait est basée sur l'introduction des suites x_n^+ et x_n^- qui sont à la base de la notion de *lim inf* et *lim sup* et sur l'axiome de la borne supérieure. C'est pourquoi, il est sans doute important de comprendre cette démonstration qui est essentiellement la même que d'autres plus complexes que vous verrez plus tard.

Chapitre 1

Formulation et analyse d'un problème d'optimisation

1.1 Description d'un problème d'optimisation

Comme nous l'avons vu en introduction, tous les problèmes d'optimisation que nous considérerons peuvent être exprimés de la façon suivante :

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{sous la contrainte : } x \in X. \quad (1.1)$$

où X est un sous-ensemble de \mathbb{R}^n . On pourra écrire aussi

$$\min_{x \in X} f(x)$$

Les variables $x = (x_1, \dots, x_n)$ sont appelées "variables d'optimisation" ou *variables de décision*. La fonction $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est appelée *fonction objectif* et l'ensemble X *ensemble des contraintes*. Le problème (1.1) est dit *réalisable* si $X \neq \emptyset$.

Résoudre le problème (1.1) revient chercher des points de minimum local (ou global, c'est encore mieux !) au sens de la définition suivante :

Définition 1.1 (Minimum local/Minimum global) Soit $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction.

- $x \in \mathbb{R}^n$ est un point de minimum local de f sur X si

$$x \in X \quad \text{et} \quad \exists r > 0 \mid \forall y \in X \cap B(x, r), \quad f(x) \leq f(y). \quad (1.2)$$

On dit alors que $f(x)$ est un minimum local de f sur X .

- $x \in \mathbb{R}^n$ est un point de minimum global de f sur X si

$$x \in X \quad \text{et} \quad \forall y \in X, \quad f(x) \leq f(y). \quad (1.3)$$

On dit alors que $f(x)$ est un minimum global de f sur X .

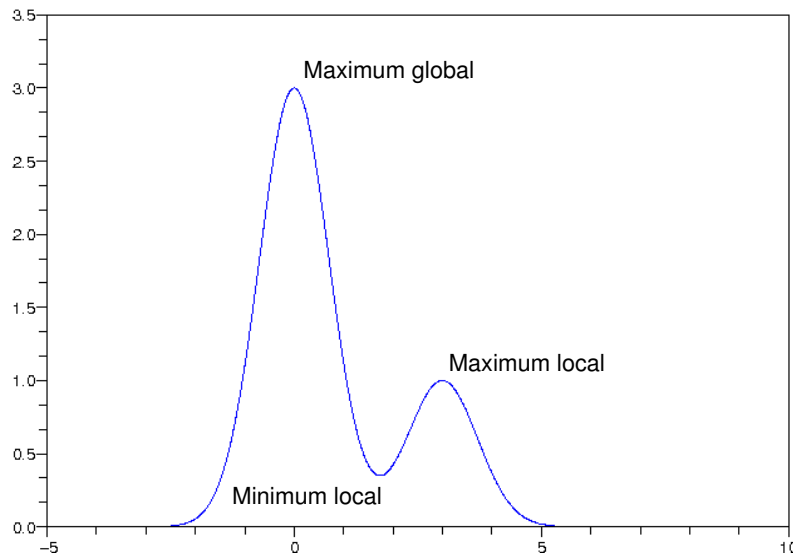
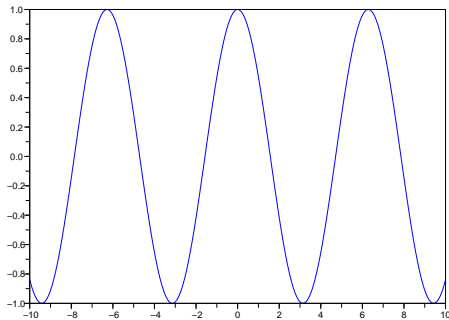
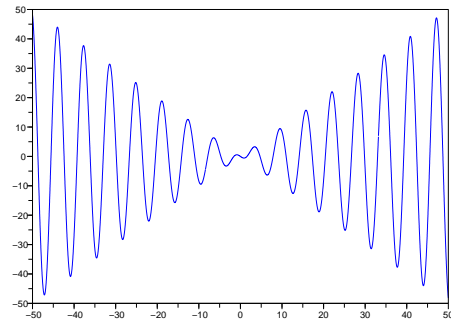


FIGURE 1.1 – Exemples de minima et maxima locaux et globaux pour la fonction $f : x \mapsto 3e^{-x^2} + e^{-(x-3)^2}$.



Pour la fonction $f : x \mapsto \cos(x)$, il existe une infinité de minima et maxima globaux.



Pour la fonction $f : x \mapsto x \cos(x)$, il existe une infinité de minima et maxima locaux mais aucun minimum ou maximum global.

Les notions de maximum local et global sont définies de façon tout à fait similaire. En fait, on peut facilement démontrer que les problèmes (avec ou sans contraintes) :

$$\min_{x \in X} f(x) \quad \text{et} \quad \max_{x \in X} -f(x)$$

sont équivalents dans le sens où ils ont même ensemble de solutions et :

$$\min_{x \in X} f(x) = -\max_{x \in X} -f(x) \quad \text{ou encore} \quad \max_{x \in X} f(x) = -\min_{x \in X} -f(x).$$

Ainsi la recherche d'un maximum pouvant se ramener la recherche d'un minimum, nous ne nous intéresserons qu'à la recherche du minimum.

Aller plus loin : On peut se demander s'il existe d'autres problèmes d'optimisation que nous n'aborderons pas. L'hypothèse simplificatrice la plus forte que nous ayons fait est de supposer que X soit un sous-ensemble de \mathbb{R}^n , donc un sous-ensemble d'un espace vectoriel *fini*. En fait beaucoup de problèmes d'optimisation se posent dans des espaces de dimension infini. Par exemple le problème des géodésiques dans S^2 (la sphère de dimension 2, par exemple la terre) revient à trouver le chemin le plus court du point A vers le point B , c'est-à-dire de minimiser la longueur des fonctions de $[0, 1]$ dans S^2 qui valent A en 0 et B en 1. L'ensemble sur lequel on minimise est l'ensemble des fonctions de $[0, 1]$ dans \mathbb{R}^3 qui est un ensemble vectoriel de dimension 3 et les contraintes sont le point de départ, le point d'arrivée et le fait que tous les chemins doivent être inscrits sur la terre S^2 . Quand les problèmes sont posés sur des espaces vectoriels de dimension infinie, des problématiques se posent que nous n'aborderons pas ici.

Considérons à nouveau le problème (1.1) mais d'un point de vue ensembliste : résoudre

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{sous la contrainte : } x \in X.$$

revient à chercher le minimum de l'image directe $f(X) = \{f(x) : x \in X\}$ de X par f .

D'après les résultats du paragraphe 0.2, on sait qu'il existe un infimum fini dès que $f(X)$ est minoré, et que cet infimum est un minimum (i.e. est atteint) si $f(X)$ est fermé. Il s'agit donc maintenant d'étudier les propriétés respectives de f et X et d'en déduire des résultats d'existence de solution du problème (1.1).

1.2 Condition suffisante d'existence d'un point minimum

Dans ce paragraphe, nous nous intéressons aux propriétés de la fonction objectif f et de l'ensemble des contraintes X afin d'en déduire des conditions suffisantes d'existence d'un minimum de l'ensemble $f(X)$, i.e. d'une solution du problème (1.1).

Il existe principalement deux théorèmes donnant des conditions suffisantes d'existence d'un point de minimum : le premier dans le cas où l'ensemble des contraintes est fermé borné, le second pour un ensemble de contraintes fermé mais non borné.

1.2.1 Contre-exemples à l'existence d'un minimum

Avant de s'intéresser au théorème qui garantit l'existence d'un minimum, il est bon de connaître des exemples typiques où il n'y a pas de minimum. Le problème $\min_X f(x)$ n'a pas de solution dans les cas suivants :

- **Fonction non continue :** Si $X = \mathbb{R}$ et $f(x) = x^2 + 1$ sur \mathbb{R}^* et $f(0) = 3$, l'infimum de f est 1 mais $f(0) = 3$, il n'existe pas de points de $x \in X$ tels que $f(x) = 1$.
- **Ensemble non fermé :** Si $X =]0, 1]$ et $f(x) = x^2 + 1$. L'infimum de f sur X vaut 1 mais le point $x = 0$ n'appartient pas à X .
- **Minimum à l'infini :** Si $X = \mathbb{R}$ et $f(x) = -x^2$, alors l'infimum de f vaut $-\infty$ et est atteint en $x = \pm\infty$. Cependant il n'y a pas de minimum.

1.2.2 Cas où l'ensemble X des contraintes est borné

Nous nous intéressons d'abord au cas où X est borné. Le premier résultat est que l'image d'un fermé borné par une application continue est un fermé borné : si f est continue et X fermé borné, alors $f(X)$ est fermé borné.



Théorème 1.1 (Théorème de Weierstrass) Soit X un ensemble fermé borné non vide de \mathbb{R}^n et $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une application continue sur X .

Alors f est bornée et atteint ses bornes. Autrement dit, il existe $x \in X$ point de minimum global de f sur X i.e. :

$$\forall y \in X, f(x) \leq f(y).$$

De la même façon, il existe un point de maximum global de f sur X .

Preuve. Soit $(x_n)_{n \in \mathbb{N}}$ une suite minimisante dans $f(X)$, i.e. d'éléments de X telle que

$$\lim_{n \rightarrow +\infty} f(x_n) = \inf f(X).$$

Comme X est fermé borné, il existe une sous-suite extraite $(x_{\sigma(n)})_{n \in \mathbb{N}}$ qui converge vers un $x \in X$. Cette suite extraite vérifie

$$x_{\sigma(n)} \rightarrow x \quad \text{et} \quad f(x_{\sigma(n)}) \rightarrow \inf_{y \in X} f(y).$$

Or f est continue, d'où par unicité de la limite, il suit

$$f(x) = \inf_{y \in X} f(y) \text{ avec } x \in X,$$

et f réalise son minimum sur X . □

1.2.3 Cas où l'ensemble X des contraintes est non borné

Dans cette section nous nous intéressons au cas où l'ensemble des contraintes est non borné. C'est le cas tout particulièrement où il n'y a pas de contraintes, c'est-à-dire $X = \mathbb{R}^n$. L'idée générale est que le seul problème est qu'il existe un infimum mais qu'il n'est pas atteint car celui-ci se trouve à l'infini. Une fonction qui illustre ce problème est la fonction $f : x \mapsto x^{-1}$ qui admet 0 comme infimum sur \mathbb{R}_+ pour $x = +\infty$.

Nous introduisons la notion de *coercivité* d'une fonction qui empêche que l'infimum se trouve à l'infini.

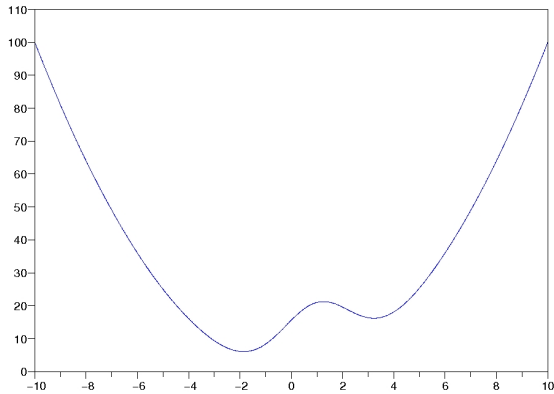
Définition 1.2 Une application $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *infinie à l'infini* (ou *coercive*) sur X ssi

$$\forall A \in \mathbb{R}, \quad \exists R > 0 \mid \forall x \in X, \quad [\|x\| \geq R \implies f(x) \geq A] \quad (1.4)$$

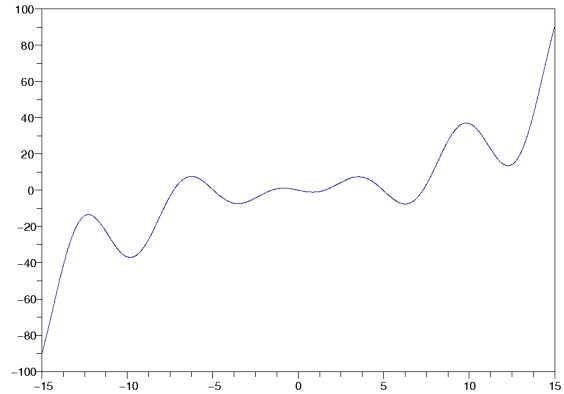
On note : $\lim_{\substack{\|x\| \rightarrow +\infty \\ x \in X}} f(x) = +\infty.$

Exemple 1.2.1 1. $f_1(x) = \|x\|_2$ est coercive.

2. $f_2(x) = x_1^2 - x_2^2$ n'est pas coercive : en effet, la suite de terme général $x_n = (0, n)$, $n \in \mathbb{N}$, est telle que : $\lim_{n \rightarrow +\infty} \|x_n\| = \lim_{n \rightarrow +\infty} n = +\infty$ mais : $\lim_{n \rightarrow +\infty} f_2(x_n) = \lim_{n \rightarrow +\infty} -n^2 = -\infty$.



Exemple de fonction coercive (f_1)



Exemple de fonction non coercive (f_2).

Comme la définition 1.2 n'est pas facile à manier en pratique, on utilise souvent la proposition suivante, qui est une hypothèse un peu plus forte, pour montrer que la fonction est infinie à l'infini.

Proposition 1.1 Soit $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une application et $g : \mathbb{R} \rightarrow \mathbb{R}$ vérifiant

$$f(x) \geq g(\|x\|) \quad \text{avec} \quad \lim_{t \rightarrow +\infty} g(t) = +\infty.$$

Alors, f est infinie à l'infini.

Preuve. Comme g tend vers $+\infty$ en $+\infty$

$$\forall A \in \mathbb{R}, \quad \exists R > 0 \mid \forall t \in \mathbb{R} \quad t \geq R \implies g(t) \geq A.$$

Avec $t = \|x\|$ et comme $g(x) \geq f(\|x\|)$, nous obtenons (1.4). □



Théorème 1.2 Soient F un fermé non vide de \mathbb{R}^n et $f : F \rightarrow \mathbb{R}$ une application continue infinie à l'infini sur F . Alors f admet un point de minimum global sur F , i.e. il existe $x \in F$ tel que

$$\forall y \in F, \quad f(y) \geq f(x).$$

Preuve. L'idée de la preuve est d'utiliser le théorème de Weierstrass (Théorème 1.1) en minimisant f sur un fermé borné K que nous construirons et en montrant que minimiser f sur K revient à minimiser f sur F . **(i)** Définissons un K fermé borné. Comme $F \neq \emptyset$, on a :

$\inf(F) \in \mathbb{R} \cup \{-\infty\}$. Soit $A \in \mathbb{R}$, tel que $A > \inf_{y \in F} f(y)$. Comme f est infinie à l'infini, il existe $R_1 > 0$ tel que pour $y \in \mathbb{R}^n$

$$\|y\| > R_1 \implies f(y) > A.$$

De plus F est non vide : il existe donc $R_2 > 0$ tel que

$$\overline{B}(0, R_2) \cap F \neq \emptyset.$$

Choisissons $R = \max(R_1, R_2)$

$$\|y\| > R \implies f(y) > A \quad \text{et} \quad \overline{B}(0, R) \cap F \neq \emptyset.$$

On introduit : $K = \overline{B}(0, R) \cap F$ non vide, borné ($\|y\| \leq R$) et fermé (intersection de 2 fermés).

(ii) Minimisons f sur K . Comme f est continue et K est fermé borné, f atteint son minimum sur K , i.e. :

$$\exists x \in K \quad | \quad f(x) = \inf_{y \in K} f(y), \quad (1.5)$$

(iii) Montrons que minimiser sur K revient à minimiser sur F . D'une part, nous avons

$$\inf_{y \in F} f(y) = \inf \left(\inf_{y \in K} f(y); \inf_{y \in F \setminus K} f(y) \right).$$

D'autre part, pour $z \in F$ et $z \notin K$, on a $\|z\| \geq R$, soit :

$$f(z) > A > \inf_{y \in F} f(y)$$

Par conséquent : $\inf_{y \in F} f(y) < \inf_{y \in F \setminus K} f(y)$. Il suit

$$\inf_{y \in F} f(y) = \inf_{y \in K} f(y)$$

et d'après (1.5) il existe $x \in K \subset F \quad | \quad f(x) = \inf_{y \in F} f(y)$. □

1.2.4 Cas des contraintes d'égalité

L'hypothèse X fermé est assez difficile à montrer en pratique sauf dans le cas (fréquent en optimisation) où X est défini par des égalités et des inégalités :

$$X = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$$

où $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^q$. L'écriture " $h(x) = 0$ " représente en fait p contraintes d'égalité :

$$h_i(x) = 0, \quad i = 1, \dots, p,$$

et de même " $g(x) \leq 0$ " représente q contraintes d'inégalité :

$$g_i(x) \leq 0, \quad i = 1, \dots, q.$$

Dans le cas où les fonctions contraintes g et h sont continues, on a le résultat suivant :



Proposition 1.2 Soient $g : \mathbb{R}^n \rightarrow \mathbb{R}^q$ et $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ deux fonctions continues.

- $X = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ est un ensemble fermé de \mathbb{R}^n .
- $X = \{x \in \mathbb{R}^n : g(x) < 0\}$ est un ouvert de \mathbb{R}^n .

Ainsi, on peut conclure directement que si f , g et h sont continues et soit

- l'ensemble des contraintes $X = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ est borné,
- Ou f est infinie à l'infini,

alors le problème :

$$\begin{aligned} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t. :} & g(x) \leq 0, h(x) = 0. \end{aligned}$$

admet au moins une solution globale. Autrement dit, f admet au moins un point de minimum global sur X .

Aller plus loin : Si l'ensemble des contraintes X n'est plus un sous-ensemble de \mathbb{R}^n mais un sous-ensemble d'un espace vectoriel métrique complet (un Banach) de dimension **infinie** alors le théorème de Weierstrass n'est plus vrai. Il faut dans ce cas une hypothèse supplémentaire pour assurer l'existence d'un minimum. Il faut en effet supposer que X est compact ce qui assure que de toute suite on peut extraire une sous-suite qui converge et utiliser le même genre d'argument que la démonstration du théorème de Weierstrass.

Aller plus loin : La condition de continuité de f n'est pas nécessaire, on peut la remplacer dans la preuve du théorème de Weierstrass par la condition plus faible de "semi-continuité inférieure" qui dit essentiellement que $f(\lim x_n) \leq \lim f(x_n)$, alors que la continuité impose l'égalité.

Aller plus loin : En dimension infinie, toutes les normes ne sont pas équivalentes, donc on peut choisir les normes que l'on veut. Tout le jeu en dimension infinie est de trouver des normes (en fait on n'a besoin que d'une topologie) pour que X soit compact tout en ayant f semi-continue inférieure.

1.3 Convexité et optimisation

Les résultats précédents ne nous donnent aucune information quant à l'unicité éventuelle d'un point de minimum global, ni sur le lien entre possibles minima locaux et minimum global. Nous introduirons donc la notion de *convexité* qui permet de garantir que les minimums locaux sont en fait des minimums globaux. Attention la notion de convexité est suffisante pour garantir que nous avons trouvé un minimum global mais aucunement nécessaire.

Les problèmes dont les données sont convexes, constituent une classe importante en optimisation, car fréquemment rencontrés dans les applications et à la base de nombreuses méthodes développées pour des problèmes plus généraux.

Définition 1.3 (Ensemble convexe) Soit $X \subset \mathbb{R}^n$. L'ensemble X est convexe ssi

$$\forall (x, y) \in X^2, \quad \forall \lambda \in]0, 1[, \quad \lambda x + (1 - \lambda)y \in X,$$

c'est-à-dire, si x et y sont deux éléments de X alors le segment qui relie x à y est inclus dans X .

Exemple 1.3.1 – \mathbb{R}^n est convexe.

– Les boules ouvertes et les boules fermées sont convexes

Définition 1.4 (Fonction convexe/strictement convexe) Soit $X \subset \mathbb{R}^n$ convexe et $f : X \rightarrow \mathbb{R}$.

- f est convexe ssi

$$\forall (x, y) \in X^2, \quad \forall \lambda \in]0, 1[, \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

- f est strictement convexe ssi

$$\forall (x, y) \in X^2, \quad x \neq y, \quad \forall \lambda \in]0, 1[, \quad f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y).$$

On interprète géométriquement le fait qu'une fonction soit convexe en disant qu'elle est située sous ses cordes.

La convexité est une notion globale, qui va donner des informations sur le caractère global d'un point de minimum.

Définition 1.5 (Problème convexe) Soit $X \subset \mathbb{R}^n$ convexe et $f : X \rightarrow \mathbb{R}$ une fonction.

Le problème $\min_{x \in X} f(x)$ est dit convexe si et seulement si la fonction objectif f est convexe et si l'ensemble X des contraintes est convexe.

Si f est strictement convexe et l'ensemble X est convexe, alors le problème $\min_{x \in X} f(x)$ est dit strictement convexe.

1.3.1 Résultats d'existence et d'unicité en optimisation convexe



Théorème 1.3 (Condition suffisante d'optimalité globale) Soient $X \subset \mathbb{R}^n$ un ensemble convexe et $f : X \rightarrow \mathbb{R}$ une fonction. Soit x^* un point de minimum local de f sur X .

- Si f est convexe, alors x^* est un point de minimum global de f sur X .
- Si f est strictement convexe, alors x^* est l'unique point de minimum global de f sur X .

Preuve. i. Par l'absurde : soit $x^* \in X$ un point qui réalise un minimum local de f sur X , i.e. :

$$\exists r > 0 \mid \forall y \in X \text{ avec } \|y - x^*\| < r, \quad f(y) \geq f(x^*). \quad (1.6)$$

tel que x^* ne réalise pas un minimum global, i.e. qu'il existe un point x^+ de X tel que :

$$f(x^+) < f(x^*). \quad (1.7)$$

On introduit le point : $y_{r/2} = \alpha x^+ + (1 - \alpha)x^*$, avec : $\alpha = \frac{r}{2\|x^+ - x^*\|} \in]0, 1[$.

D'après les hypothèses (1.6) et (1.7), on a : $x^+ \notin B(x^*, r)$, ce qui implique : $r < \|x^+ - x^*\|$, et : $\alpha \in]0, 1[$. Par convexité de X , le point $y_{r/2}$ appartient donc au segment $[x^+, x^*]$ lui-même contenu dans X . De plus :

$$\begin{aligned} f(y_{r/2}) &\leq \alpha f(x^+) + (1 - \alpha)f(x^*) && \text{par convexité de } f \\ &< f(x^*) && \text{d'après (1.6).} \end{aligned}$$

Ceci contredit (1.6) car : $\|y_{r/2} - x^*\| = \frac{r}{2}$. Le point $y_{r/2}$ appartient donc également à la boule ouverte $B(x^*, r)$ i.e. : $f(y_{r/2}) \geq f(x^*)$.

ii. Raisonnons à nouveau par l'absurde. Soient x_1 et x_2 deux éléments de X réalisant le minimum de f . Par convexité de X , $\frac{x_1 + x_2}{2} \in X$, et comme f est strictement convexe, il suit

$$\begin{aligned} f\left(\frac{x_1 + x_2}{2}\right) &< \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) \\ &< \frac{1}{2}\min_{y \in X} f(y) + \frac{1}{2}\min_{y \in X} f(y) = \min_{y \in X} f(y), \end{aligned}$$

Ce qui est impossible. □

Exemple 1.3.2 Nous allons vérifier que chaque hypothèse du théorème est nécessaire.

- Soit la fonction $f : x \rightarrow x^{-1}$, montrer qu'elle est convexe. Si $X = [1, 2] \cup [3, 4]$. Montrer que X n'est pas convexe mais est fermé borné. Montrer que 2 est un point de minimum local mais n'est pas un point de minimum global.
- Soit la fonction $f : x \rightarrow x \cos(x)$ et $X = [0, 8\pi]$. Montrer que X est convexe mais pas f et montrer qu'il existe des points de minimum locaux qui ne sont pas des minimums globaux.

En conclusion, nous obtenons le résultat suivant :

Corollaire 1.1 Soit $X \subset \mathbb{R}^n$ un ensemble fermé non vide, et convexe. Soit $f : X \rightarrow \mathbb{R}$ supposée continue et strictement convexe. Si une des deux conditions suivantes est vérifiée :

- soit X est borné,
- soit f est infinie à l'infini

Alors f admet un unique point de minimum global sur X .

1.3.2 Caractérisation différentielle de la convexité

Dans le cas où la fonction f est différentiable (à l'ordre 1 ou à l'ordre 2) en tout point, on sait caractériser la convexité d'une fonction via son gradient ou sa hessienne. Ces caractérisations sont la plupart du temps plus faciles à manipuler que la définition 1.4.

Théorème 1.4 Soit $X \subset \mathbb{R}^n$ convexe et $f : X \rightarrow \mathbb{R}$ différentiable. La fonction f est convexe ssi :

$$\forall (x, y) \in X^2, \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad (1.8)$$

ou de façon équivalente, ssi :

$$\forall (x, y) \in X^2, \quad \langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0. \quad (1.9)$$

Preuve. Soit $(x, y) \in X^2$. Par convexité de f , on a donc pour tout $t \in]0, 1[$:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) = f(x) + t(f(y) - f(x)),$$

soit : $\frac{f(x + t(y-x)) - f(x)}{t} \leq f(y) - f(x)$. En passant à la limite pour $t \rightarrow 0^+$, il suit (1.8).

Réciproquement, on applique (1.8) $tx + (1-t)y$ et x , puis $tx + (1-t)y$ et y , d'où :

$$\begin{aligned} f(x) &\geq f(tx + (1-t)y) + (1-t) \langle \nabla f(tx + (1-t)y), y - x \rangle \\ f(y) &\geq f(tx + (1-t)y) - t \langle \nabla f(tx + (1-t)y), y - x \rangle \end{aligned}$$

En combinant ces deux inégalités, on obtient : $tf(x) + (1-t)f(y) \geq f(tx + (1-t)y)$, et donc la convexité de f .

En changeant les rôles de x et y dans (1.8), puis en sommant les deux inégalités obtenues, on démontre sans problème que (1.8) implique (1.9). Pour montrer la réciproque, on introduit :

$$\varphi : t \in [0, 1] \mapsto tf(x) + (1-t)f(y) - f(tx + (1-t)y),$$

et on montre que φ est positive sur $[0, 1]$. Cette démonstration est laissée en exercice (on pourra tracer le tableau de variation de φ sur $[0, 1]$). \square

Si de plus, la fonction f est deux fois différentiable, on a alors une caractérisation d'ordre deux de la convexité via la Hessienne. On rappelle (théorème de Schwarz) que si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est de classe \mathcal{C}^2 , alors sa matrice hessienne $H[f](x)$ est symétrique.



Théorème 1.5 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 . On note $H[f](x)$ sa hessienne en x .

- Si $H[f](x)$ est semidéfinie positive pour tout $x \in \mathbb{R}^n$, alors f est convexe.
- Si $H[f](x)$ est définie positive pour tout $x \in \mathbb{R}^n$, alors f est strictement convexe.

La preuve de ce résultat est hors programme et faite en annexe.

Chapitre 2

Optimisation numérique sans contraintes

Nous nous intéressons dans ce chapitre à la conception de méthodes numériques pour la recherche des points $x \in \mathbb{R}^n$ qui réalisent le minimum d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$:

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

où f est supposée au moins différentiable. On parle d'optimisation sans contrainte.

2.1 Conditions d'optimalité

Nous commençons dans cette section à étudier ce que l'on appelle les *conditions d'optimalité* qui sont des caractérisations des points de minimum.



Théorème 2.1 (Conditions nécessaires d'optimalité locale) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application différentiable. Si $x^* \in \mathbb{R}^n$ réalise un minimum local (resp. maximum local) de f , alors :

$$\nabla f(x^*) = 0 \quad (\text{CN d'optimalité du 1}^{er} \text{ ordre})$$

Si, de plus, f est deux fois différentiable dans un voisinage ouvert de x^* , alors :

$$\begin{aligned} H[f](x^*) \text{ est semidéfinie positive} & \quad (\text{CN d'optimalité du 2}^{nd} \text{ ordre}) \\ (\text{resp. } H[f](x^*) \text{ est semidéfinie négative}) \end{aligned}$$

Preuve. Soit $h \in \mathbb{R}^n$, $h \neq 0$. Pour s assez petit, on définit $\varphi : s \in \mathbb{R} \mapsto f(x^* + sh)$. φ admet donc un minimum local en $s = 0$, d'où : $\varphi'(0) = \nabla f(x^*)^\top h = 0$. Ceci étant vrai pour tout h , on en déduit : $\nabla f(x^*) = 0$.

Supposons maintenant f deux fois différentiable. On écrit le développement de Taylor d'ordre 2 de la fonction φ . Comme $\nabla f(x^*) = 0$, on obtient :

$$f(x^* + sh) - f(x^*) = \frac{s^2}{2} h^\top H[f](x^*) h + o(s^2).$$

soit : $\frac{s^2}{2} h^\top H[f](x^*)h + o(s^2) \geq 0$ puisque x^* est un point de minimum local de f . Après division par s^2 , on fait tendre s vers 0 et on obtient : $h^\top H[f](x^*)h \geq 0$. \square

La condition d'optimalité du premier ordre montre que les points où le gradient s'annule sont des points importants. Ils sont appelés *points critiques* et sont définis comme suit :

Définition 2.1 (Points critiques) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application différentiable. Tout point $x \in \mathbb{R}^n$ vérifiant :

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

est appelé *point critique* (ou *point stationnaire*) de f .

Mais attention ! Les conditions du théorème 2.1 ne sont que nécessaires : tout point où le gradient est nul n'est pas nécessairement un extremum. Les exemples suivants montrent les différents types de cas qui peuvent apparaître.

Exemple 2.1.1 – La fonction $x \mapsto x^2$ admet un point critique en $x = 0$ qui est aussi minimum local.

– La fonction $x \mapsto -x^2$ admet un point critique en $x = 0$ qui est aussi maximum local.

– La fonction $x \mapsto x^3$ admet un point critique en $x = 0$ qui n'est ni minimum local ni maximum local, c'est ici un point d'inflexion.

– La fonction $x \mapsto x^2 - y^2$ admet un point critique en $(x, y) = (0, 0)$ qui n'est ni minimum local ni maximum local, c'est ici un point-selle.

Cependant, la condition du premier ordre joue un rôle central en optimisation numérique : elle permet de sélectionner un certain nombre de points candidats à être des extrema locaux, même s'il faut vérifier que les points critiques que l'on a sélectionné sont bien des minima locaux. Par exemple dans l'exemple suivant

Exemple 2.1.2 Trouver le(s) minima globaux de $f : x \mapsto ax^2 + bx + c$ avec $a, b, c \in \mathbb{R}$.

Si $a > 0$ alors la fonction est infinie à l'infini et admet un minimum global. Si $a < 0$ la fonction vaut $-\infty$ en $\pm\infty$ et si $a = 0$ et $b \neq 0$ la fonction vaut $-\infty$ en $+\infty$ ou $-\infty$ selon le signe de b et dans tous ces cas n'admet pas de minimum global. Si $a = 0$ et $b = 0$, la fonction est constante et tout point est minimum global. Nous nous intéressons donc au seul cas $a > 0$. L'équation des points critiques est $2ax + b = 0$, il existe un seul point critique $x = -b/(2a)$ et dans le cas $a > 0$ nous savons qu'il existe un minimum global qui est donc point critique donc c'est $-b/(2a)$.

Nous avons vu qu'être un point fixe est une condition nécessaire pour être un extremum, nous voyons maintenant une condition suffisante.



Théorème 2.2 (Condition suffisante d'optimalité locale) Soit O un ouvert de \mathbb{R}^n . Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application supposée de classe C^2 sur O . Si $\bar{x} \in O$ vérifie :

$$\nabla f(\bar{x}) = 0 \quad \text{et} \quad H[f](\bar{x}) \text{ symétrique, définie positive (resp. définie négative)}$$

Alors \bar{x} est un point de minimum local (resp. maximum local) de f .

Remarque 2.1 D'un point de vue géométrique, la condition du second ordre : " $H[f](\bar{x})$ définie positive", revient à dire que f est localement convexe en x^* , i.e. convexe dans un voisinage ouvert de x^* . En pratique, elle est difficile à vérifier systématiquement car elle nécessite de calculer les dérivées secondes et d'étudier les valeurs propres de la matrice hessienne.

Exemple 2.1.3 L'exemple qui suit montre que la condition du Théorème 2.2 est suffisante mais non nécessaire :

Montrer que 0 est un minimum global de $f : x \mapsto x^4$, que c'est bien un point critique de f mais que la Hessienne de f en ce point est semidéfinie positive mais n'est pas définie positive (elle est nulle).

Si de plus la fonctionnelle à optimiser est convexe ou strictement convexe, en appliquant le Théorème 1.3 au convexe $X = \mathbb{R}^n$, on obtient :

Théorème 2.3 (Condition Suffisante d'optimalité globale) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application différentiable et \bar{x} un point critique de f .

- i. Si f est convexe, alors \bar{x} est un point de minimum global de f .
- ii. Si f est strictement convexe, alors \bar{x} est l'unique point de minimum global de f .

2.2 Généralités sur les algorithmes de descente

Nous allons maintenant nous intéresser aux algorithmes de calcul de minimum et plus particulièrement aux algorithmes de descente. Partant d'un point x_0 arbitrairement choisi, un algorithme de descente va chercher à générer une suite d'itérés $(x_k)_{k \in \mathbb{N}}$ telle que :

$$\forall k \in \mathbb{N}, \quad f(x_{k+1}) \leq f(x_k).$$

Commençons par définir plus précisément la notion de descente.

2.2.1 Notion de direction de descente

Le gradient joue un rôle essentiel en optimisation. Dans le cadre des méthodes d'optimisation, il sera également important d'analyser le comportement de la fonction objectif dans certaines directions. Commençons pour cela par rappeler le concept de dérivée directionnelle :

Définition 2.2 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application continue. Soit $x \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$.

La dérivée directionnelle de f en x dans la direction d est définie par :

$$df(x; d) := \lim_{t \rightarrow 0^+} \frac{f(x + td) - f(x)}{t},$$

si cette limite existe.

Proposition 2.1 Si f est différentiable en un point $x \in \mathbb{R}^n$, alors pour tout $d \neq 0$, f admet une dérivée dans la direction d en x et :

$$df(x; d) = Df(x)(d) = \nabla f(x)^\top d.$$

On rappelle que la réciproque est fautive ! La dérivabilité selon tout vecteur en x n'implique pas nécessairement la différentiabilité de f en x .

La dérivée directionnelle donne des informations sur la pente de la fonction dans la direction d , tout comme la dérivée donne des informations sur la pente des fonctions à une variable. En particulier,

- si $df(x; d) > 0$ alors f est croissante dans la direction d .
- si $df(x; d) < 0$ alors f est décroissante dans la direction d .

Dans ce dernier cas, on dira que d est une direction de descente de f .

Définition 2.3 (Direction de descente) Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $x \in \mathbb{R}^n$. Le vecteur $d \in \mathbb{R}^n$ est une direction de descente pour f à partir du point x si $t \mapsto f(x + td)$ est décroissante en $t = 0$, c'est-à-dire s'il existe $\eta > 0$ tel que :

$$\forall t \in]0, \eta], f(x + td) < f(x). \quad (2.1)$$

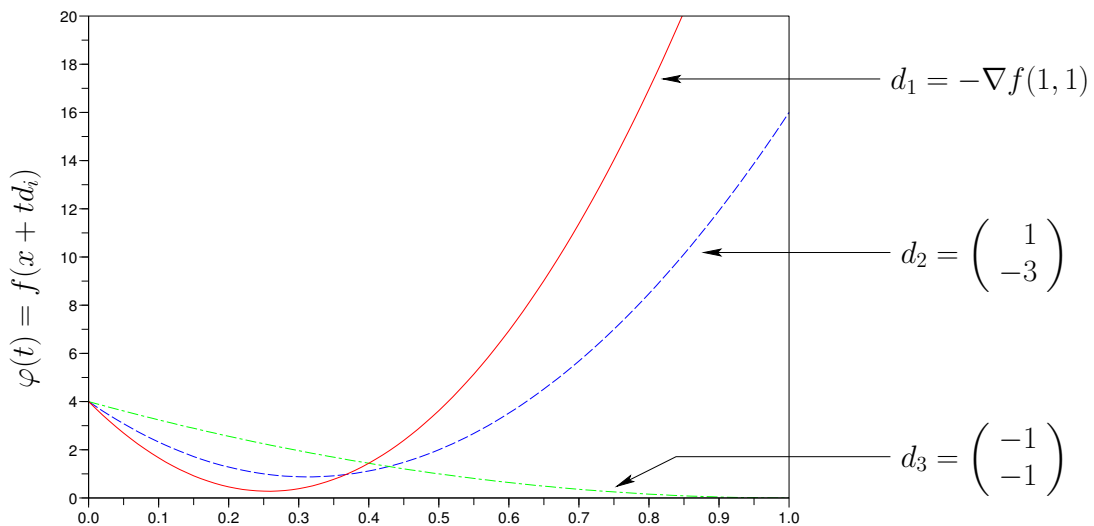


FIGURE 2.1 – Allure de la fonction $f : x \mapsto \frac{1}{2}x_1^2 + 2x_2^2$ au point $x = (1, 1)^\top$ dans plusieurs directions.



Proposition 2.2 Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable et $x \in \mathbb{R}^n$ tel que : $\nabla f(x) \neq 0$. Soit $d \in \mathbb{R}^n$ un vecteur

- Si $\nabla f(x)^\top d < 0$ alors d est une direction de descente.
- Si d est une direction de descente alors $\nabla f(x)^\top d \leq 0$.

Preuve de la proposition 2.2. Soit d telle $\nabla f(x)^\top d < 0$. On écrit le développement de Taylor-Young de $t \mapsto f(x + td)$: pour t assez petit,

$$f(x + td) = f(x) + t\nabla f(x)^\top d + t\epsilon(t), \text{ avec } \epsilon(t) \xrightarrow{t \rightarrow 0} 0.$$

Comme $-\nabla f(x)^\top d > 0$, il existe $\eta > 0$ tel que si $|t| \leq \eta$, alors $|\epsilon(t)| < -\nabla f(x)^\top d$. D'où :

$$\forall t \in]0, \eta], \quad f(x + td) - f(x) = t [\nabla f(x)^\top d + \epsilon(t)] < 0.$$

Ce qui montre la première partie de la proposition. Soit maintenant d une direction de descente, supposons que $\nabla f(x)^\top d > 0$, en appliquant le même raisonnement que ci-dessus, on montre qu'il existe un η tel que pour tout t tel que $t \leq \eta$, on a $\epsilon(t) > -\nabla f(x)^\top d$ et ainsi $f(x + td) - f(x) > 0$, ce qui contredit le fait que d soit une direction de descente. \square

Parmi toutes les directions de descente existantes en un point x donné, il est naturel de s'intéresser à celle où la pente est la plus forte. Un résultat remarquable montre que cette direction est donnée par le gradient (ou plus exactement son opposé). Pour le démontrer, il suffit de comparer les dérivées directionnelles :

Théorème 2.4 (Direction de plus forte descente) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable. Soit $x \in \mathbb{R}^n$. Alors pour toute direction d de norme constante égale à $\|d\| = \|\nabla f(x)\|$, on a :

$$(-\nabla f(x))^\top \nabla f(x) \leq d^\top \nabla f(x), \quad (2.2)$$

Ainsi, la direction $d^* = -\nabla f(x)$ est appelée "direction de plus forte descente".

Preuve. Soit $d \in \mathbb{R}^n$ une direction quelconque de norme : $\|d\| = \|\nabla f(x)\|$. On a alors :

$$\begin{aligned} (-d)^\top \nabla f(x) &\leq \| -d \| \|\nabla f(x)\| && \text{d'après l'inégalité de Cauchy-Schwarz} \\ &\leq \|\nabla f(x)\|^2 = \nabla f(x)^\top \nabla f(x) && \text{puisque } \|d\| = \|\nabla f(x)\|. \end{aligned}$$

On en déduit donc (2.2). \square

Sans surprise, si l'opposé du gradient correspond à la plus forte descente, le gradient correspond, lui, à la plus forte pente de montée :

Corollaire 2.1 Le vecteur $\nabla f(x)$ est appelé direction de plus forte pente de f au point $x \in \mathbb{R}^n$ et si $d^* = \nabla f(x)$ alors

$$(d^*)^\top \nabla f(x) = \max_{d \in \mathbb{R}^n} (d^\top \nabla f(x)) \text{ sous la contrainte } \|d\| = \|\nabla f(x)\|$$

Exercice 2.2.1 Retrouver les conditions nécessaires d'optimalité du théorème 2.1 en utilisant la notion de direction de descente.

2.2.2 Algorithme de descente modèle

Partant d'un point x_0 arbitrairement choisi, un algorithme de descente va chercher à générer une suite d'itérés $(x_k)_{k \in \mathbb{N}}$ telle que :

$$\forall k \in \mathbb{N}, \quad f(x_{k+1}) \leq f(x_k)$$

D'après la caractérisation de la descente (cf proposition 2.2), il s'agit donc à chaque itération k , de trouver un point x_{k+1} dans une direction d vérifiant : $\nabla f(x_k)^\top d < 0$.

Le schéma général d'un algorithme de descente est le suivant :

ALGORITHME DE DESCENTE MODÈLE.

Données: $f : \mathbb{R}^n \rightarrow \mathbb{R}$ supposée au moins différentiable, x_0 point initial arbitrairement choisi

Sortie: une approximation de la solution du problème : $\min_{x \in \mathbb{R}^n} f(x)$

1. $k := 0$

2. Tant que "test de convergence" non satisfait,

(a) Trouver une direction de descente d_k telle que : $\nabla f(x_k)^\top d_k < 0$.

(b) *Recherche linéaire* : Choisir un pas $s_k > 0$ à faire dans la direction d_k et tel que :

$$f(x_k + s_k d_k) < f(x_k).$$

(c) Mise à jour : $x_{k+1} = x_k + s_k d_k$; $k := k + 1$;

3. Retourner x_k .

Oracle/Boîte noire. Pour obtenir le prochain itéré, l'algorithme aura besoin d'informations sur la fonction objectif f : la valeur numérique de f en un point donné x , et souvent également du gradient $\nabla f(x)$. Ces informations sont fournies en "boîte noire", i.e. par un sous-programme indépendant de l'algorithme d'optimisation choisi : routine de calcul du gradient par différences finies lorsque celui-ci n'est pas calculable explicitement, ou simulateur renvoyant les valeurs numériques $f(x)$ et $\nabla f(x)$ sans formule mathématique explicite par exemple.

Test de convergence/Test d'arrêt. Soit x^* un point de minimum local du critère f à optimiser. Supposons que l'on choisisse comme test d'arrêt dans l'algorithme de descente modèle, le critère idéal : " $x_k = x^*$ ". Dans un monde idéal (i.e. en supposant tous les calculs exacts et la capacité de calcul illimitée), soit l'algorithme s'arrête après un nombre fini d'itérations, soit il construit (théoriquement) une suite infinie $x_1, x_2, \dots, x_k, \dots$ de points de \mathbb{R}^n qui converge vers x^* .

En pratique, un test d'arrêt devra être choisi pour garantir que l'algorithme s'arrête toujours après un nombre fini d'itérations et que le dernier point calcul soit suffisamment proche de x^* .

Soit $\varepsilon > 0$ la précision demandée. Plusieurs critères sont à notre disposition : tout d'abord (et c'est le plus naturel), un critère d'optimalité basé sur les conditions nécessaires d'optimalité du premier ordre présentées dans la section 2.1 : on teste si

$$\|\nabla f(x_k)\| < \varepsilon, \tag{2.3}$$

auquel cas l'algorithme s'arrête et fournit l'itéré courant x_k comme solution.

En pratique, le test d'optimalité n'est pas toujours satisfait et on devra faire appel à d'autres critères (fondés sur l'expérience du numérique) :

- Stagnation de la solution : $\|x_{k+1} - x_k\| < \varepsilon \|x_k\|$.
- Stagnation de la valeur courante : $|f(x_{k+1}) - f(x_k)| < \varepsilon |f(x_k)|$.
- Nombre d'itérations dépassant un seuil fixé à l'avance : $k < \text{IterMax}$.

et généralement une combinaison de ces critères :

Critère d'arrêt = Test d'optimalité satisfait
 OU (Stagnation de la valeur courante & Stagnation de la solution)
 OU Nombre maximum d'itérations autorisées dépassé.

En pratique, on préférera travailler avec les erreurs relatives plutôt qu'avec les erreurs absolues, trop dépendantes de l'échelle.

2.2.3 Convergence et vitesse de convergence

Étudier la convergence d'un algorithme, c'est étudier la convergence de la suite des itérés générés par l'algorithme. Un algorithme de descente selon le modèle précédent, est dit *convergent* si la suite de ses itérés $(x_k)_{k \in \mathbb{N}}$ converge vers un point limite x^* , solution du problème :

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

De plus, la convergence est dite *locale* si elle n'a lieu que pour des points initiaux x_0 dans un voisinage de x^* . Sinon elle est dite *globale*.

En pratique, le but d'un algorithme d'optimisation ne sera que de trouver un point critique (i.e. un point vérifiant la condition d'optimalité du premier ordre : $\nabla f(x^*) = 0$). On introduit alors la notion de convergence globale d'un algorithme d'optimisation :

Définition 2.4 Soit un algorithme itératif qui génère une suite $(x_k)_{k \in \mathbb{N}}$ dans \mathbb{R}^n afin de résoudre le problème :

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

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une application de classe C^1 . L'algorithme est dit *globalement convergent* si quel que soit le point initial $x_0 \in \mathbb{R}^n$,

$$\lim_{k \rightarrow +\infty} \|\nabla f(x_k)\| = 0.$$

Cette propriété garantit que le critère d'arrêt " $\|\nabla f(x_k)\| \leq \varepsilon$ " sera satisfait à partir d'un certain rang quelle que soit la précision $\varepsilon > 0$ demandée.

Remarque 2.2 Attention, la notion de convergence que l'on a vu ne suppose pas que l'algorithme converge vers un minimum, même un minimum local. Prenons par exemple la fonction

$f : (x, y) \mapsto x^2 - y^2 + y^4$ qui est infinie à l'infini. Son minimum global est obtenu pour les points $(x, y) = (0, \pm 1/\sqrt{2})$. on peut montrer cependant que l'algorithme qui consiste à poser

$$d_k = (-2x_k, 2y_k - 3y_k^3), \quad s_k \ll 1$$

est bien un algorithme de descente. Il converge au sens de la définition 2.4, cependant, si on part du point $(1, 0)$, l'algorithme converge vers le point $(0, 0)$ qui est bien un point de gradient nul mais pas un point de minimum (même local).

Il est bien entendu très important de garantir la convergence d'un algorithme sous certaines hypothèses, mais la vitesse de convergence et la complexité sont également des facteurs à prendre en compte lors de la conception ou de l'utilisation d'un algorithme ; en effet, on a tout intérêt à ce que la méthode choisie soit à la fois rapide, précise et stable. Pour cela, on introduit les notions de vitesse (ou taux) de convergence qui mesurent l'évolution de l'erreur commise $\|x_k - x^*\|$.

Définition 2.5 Soit $(x_k)_{k \in \mathbb{N}}$ une suite d'itérés générées par un algorithme convergent donné. On note x^* la limite de la suite $(x_k)_{k \in \mathbb{N}}$ et on suppose : $\forall k \in \mathbb{N}, x_k \neq x^*$ (sinon l'algorithme convergerait en un nombre fini d'itérations). La convergence de l'algorithme est dite :

– linéaire si l'erreur $e_k = \|x_k - x^*\|$ décroît linéairement i.e. s'il existe $\tau \in]0, 1[$ tel que :

$$\lim_{k \rightarrow +\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \tau.$$

– superlinéaire si

$$\lim_{k \rightarrow +\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

– d'ordre p s'il existe $\tau \geq 0$ tel que :

$$\lim_{k \rightarrow +\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} = \tau.$$

En particulier, si $p = 2$, la convergence est dite quadratique (grosso modo partir d'un certain rang, le nombre de chiffres significatifs exacts double à chaque itération).

Bien entendu, on a intérêt à ce que la convergence d'un algorithme soit la plus élevée possible afin de converger vers la solution en un minimum d'itérations pour une précision donnée.

Exemple 2.2.1 La fonction $f : x \mapsto x^3 - 6x + 1$ admet un minimum local sur \mathbb{R} en $x^* = \sqrt{2}$. Partant d'une approximation grossière $x_0 = 2$ de x^* , comparons plusieurs algorithmes de calcul approché de x^* avec 5 chiffres significatifs exacts :

- Soit l'algorithme $x_{k+1} = x_k - \alpha(x_k^2 - 2)$. Vérifier que pour $0 < \alpha < \frac{1}{\sqrt{2}}$, cet algorithme converge linéairement avec un taux $\tau = |2\alpha\sqrt{2} - 1|$.

α	$\frac{2}{3}$	0.5	$\frac{1}{3}$	$\frac{1}{2\sqrt{2}}$
$\tau = 2\alpha\sqrt{2} - 1 $	0.885	0.414	0.057	0
Nb d'itérations	105	15	6	4
Nb chiffres sign. exacts	5	5	7	10

Si $\alpha = \frac{1}{2\sqrt{2}}$, la convergence est dite superlinéaire et c'est la meilleure convergence possible de l'algorithme en question.

- Soit l'algorithme : $x_{k+1} = \frac{1}{2}(x_k + \frac{2}{x_k})$ dont la convergence est quadratique. On peut montrer que 4 itérations suffisent pour calculer une valeur approche de x^* avec 5 chiffres significatifs exacts ; en réalité, on a même 11 chiffres significatifs exacts dès la quatrième itération.

2.3 Premiers algorithmes de descente

Un algorithme de descente est déterminé par les stratégies de choix des directions de descente successives, puis par le pas qui sera effectué dans la direction choisie. Concentrons nous dans cette partie sur le choix de la direction de descente : l'idée est de remplacer f par un modèle local plus simple, dont la minimisation nous donnera une direction de descente de f .

2.3.1 Algorithmes de gradient à pas fixe/pas optimal

Soit $x_k \in \mathbb{R}^n$ l'itéré courant. Étant donnés la valeur $f(x_k)$ et le gradient $\nabla f(x_k)$ (notre "oracle"), on remplace f au voisinage de x_k par son développement de Taylor au premier ordre :

$$f(x_k + d) \sim f(x_k) + \nabla f(x_k)^\top d.$$

On voudrait que la dérivée directionnelle $\nabla f(x_k)^\top d$ soit la plus petite possible dans un voisinage de $d = 0$. On cherche donc à résoudre :

$$\min_{d \in \mathbb{R}^n} \nabla f(x_k)^\top d \quad \text{s.c.} \quad \|d\| = \|\nabla f(x_k)\|,$$

dont la solution nous est donnée par le théorème 2.4, par :

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

Le choix de la direction de plus forte descente définit une famille d'algorithmes appelés algorithmes de descente de gradient dont le schéma est le suivant :

ALGORITHME DE DESCENTE DE GRADIENT.

Données: f , x_0 première approximation de la solution cherchée, $\varepsilon > 0$ précision demandée

Sortie: une approximation x^* de la solution de : $\nabla f(x) = 0$

1. $k := 0$;
2. Tant que critère d'arrêt non satisfait,
 - (a) *Direction de descente* : $d_k = -\nabla f(x_k)$.
 - (b) *Recherche linéaire* : trouver un pas s_k tel que : $f(x_k + s_k d_k) < f(x_k)$.
 - (c) $x_{k+1} = x_k - s_k \nabla f(x_k)$; $k := k + 1$;
3. Retourner x_k .

Il reste maintenant à définir une stratégie de recherche linéaire pour le calcul du pas. Nous étudions ici en première approche une méthode à pas optimal, puis une à pas fixe.

Méthode de plus profonde descente (“Steepest descent”)

Une idée naturelle consiste à suivre la direction de plus forte descente et à faire un pas qui rende la fonction à minimiser la plus petite possible dans cette direction. Cette méthode est appelée méthode de gradient à pas optimal ou encore **méthode de plus profonde descente**. L'étape 2(a) de l'algorithme de descente de gradient est alors remplacée par :

RECHERCHE LINÉAIRE EXACTE.

2. (a) Calculer un pas optimal s_k solution de : $\min_{s>0} f(x_k + s d_k)$.

La méthode de plus profonde descente est une sorte d'idéalisation : d'une part, nous ne savons pas en pratique calculer de façon exacte un point minimum s_k de l'objectif dans une direction donnée et le problème n'est en général pas trivial. D'autre part, la résolution du problème de minimisation unidimensionnel de l'étape 2 (a), même de façon approchée, coûte cher en temps de calcul. Pour ces raisons, on peut lui préférer parfois l'algorithme de gradient à pas constant (ou à pas fixe).

Algorithme de gradient à pas fixe

L'idée est très simple : on impose une fois pour toutes la taille du pas effectué selon la direction de descente calculée à chaque itération. Les itérations 2 (b) et (c) de l'algorithme de descente de gradient sont alors remplacées par :

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

La question est alors : comment choisir un pas qui garantisse la convergence de l'algorithme ?

Quelques observations numériques. On souhaite minimiser $f : (x, y) \in \mathbb{R}^2 \mapsto \frac{1}{2}x^2 + \frac{7}{2}y^2$, en utilisant les algorithmes de descente de gradient à pas fixe et pas optimal.

Commençons par analyser le problème de minimisation : d'une part, la fonction f est deux fois différentiable sur \mathbb{R}^2 et strictement convexe. D'autre part, le point $(0, 0)$ vérifie les conditions suffisantes d'optimalité du théorème 2.3. Donc $(0, 0)$ est l'unique point de minimum global de f .

Soit $X_k = (x_k, y_k) \in \mathbb{R}^2$ l'itéré courant tel que : $\nabla f(x_k, y_k) \neq 0$. Calculons par la méthode de plus profonde descente, l'itéré suivant :

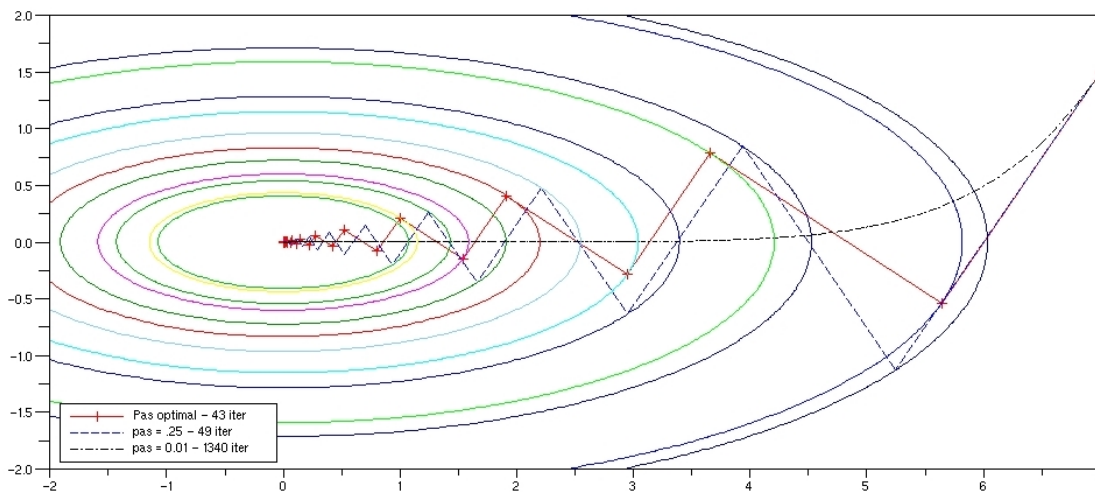


FIGURE 2.2 – Itérations des algos de gradient à pas fixe et optimal, générées à partir du point (7, 1.5).

- Direction de plus forte descente : $d_k = -\nabla f(X_k) = \begin{pmatrix} -x_k \\ -7y_k \end{pmatrix}$.
- Calcul du pas optimal s_k solution, si elle existe, du problème à une dimension :

$$\min_{s>0} f(X_k + sd_k) = \min_{s>0} \frac{1}{2}x_k^2(1 - s)^2 + \frac{7}{2}y_k^2(1 - 7s)^2.$$

La solution se calcule de façon immédiate : $s_k = (x_k^2 + 7^2y_k^2)/(x_k^2 + 7^3y_k^2)$.

A chaque itération, la méthode génère donc le point : $x_{k+1} = x_k + \frac{x_k^2 + 7^2y_k^2}{x_k^2 + 7^3y_k^2} \begin{pmatrix} -x_k \\ -7y_k \end{pmatrix}$.

Appliquons maintenant ces deux méthodes à partir du point $x_0 = (7, 1.5)$. Leurs comportements sont illustrés par la figure 2.2 et les itérations sont décrites dans les tableaux 2.1 et 2.2

Cet exemple met en évidence la lenteur de la méthode de plus profonde descente, caractérisée par le comportement en zigzag des itérés. Essayons de comprendre d'où vient ce phénomène.

A l'itération $k + 1$, l'algorithme de plus profonde descente minimise $\varphi : s \in \mathbb{R} \mapsto f(x_k - s\nabla f(x_k))$. L'objectif f étant supposé différentiable, la fonction φ est dérivable sur \mathbb{R} de dérivée :

$$\varphi'(s) = -\langle \nabla f(x_k), \nabla f(x_k - s\nabla f(x_k)) \rangle.$$

Soit s_k le pas optimal calculé ; nécessairement s_k vérifie : $\varphi'(s_k) = 0$, soit :

$$\langle \nabla f(x_k), \nabla f(x_k - s_k \nabla f(x_k)) \rangle = 0.$$

Le point $x_{k+1} = x_k - s_k \nabla f(x_k)$ vérifie donc : $\langle \nabla f(x_k), \nabla f(x_{k+1}) \rangle = 0$.

|| Deux directions de descente successives calculées par l'algorithme de plus profonde descente sont **orthogonales** ce que traduisent les zigzags des itérés, observés sur la figure 2.2.

k	$f(x_k, y_k)$	$\ \nabla f(x_k, y_k)\ _2$	s_k	x_k	y_k
0	32.375	10.547512	—	7	1.5
1	16.925373	7.9786973	0.1940299	5.641791	-0.5373134
2	8.8484403	6.5973298	0.3513514	3.6595401	0.7841872
3	4.6258889	3.5448339	0.1940299	2.9494801	-0.2809029
4	2.4183752	3.4490276	0.3513514	1.9131763	0.4099663
5	1.2643059	1.8532089	0.1940299	1.541963	-0.1468536
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
40	$1.751e-10$	2.9343653×10^{-5}	0.3513514	1.63×10^{-5}	0.35×10^{-5}
41	$9.155e-11$	1.5725775×10^{-5}	0.1940299	1.31×10^{-5}	-0.12×10^{-5}
42	$4.786e-11$	1.536522×10^{-5}	0.3513514	0.85×10^{-5}	0.18×10^{-5}
43	$2.502e-11$	0.8292768×10^{-5}	0.1940299	0.69×10^{-5}	0.07×10^{-5}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
76	$1.268e-20$	0.2523886×10^{-9}	0.3513514	0.14×10^{-9}	0.03×10^{-9}
77	$6.630e-21$	0.1303840×10^{-9}	0.1940299	0.11×10^{-9}	-0.01×10^{-9}
78	$3.466e-21$	0.1303840×10^{-9}	0.3513514	0.72×10^{-10}	0.16×10^{-10}
79	$1.812e-21$	$0.6989278 \times 10^{-10}$	0.1940299	0.58×10^{-10}	-0.05×10^{-10}

TABLE 2.1 – Itérations de la méthode de plus profonde descente. Le critère d’optimalité est satisfait en 43 itérations pour une précision $\varepsilon = 10^{-5}$ et en 79 itérations si $\varepsilon = 10^{-10}$.

pas	0.325	0.25	0.125	0.05	0.01
Nb d’itérations	DV	49	101	263	1340

TABLE 2.2 – Nombres d’itérations de l’algorithme de gradient à pas fixe pour approcher l’unique argument minimum de f à 10^{-5} près, en fonction du pas choisi - Point initial : $x_0 = (7, 1.5)$.

Enfin, les données du tableau 2.2 illustrent l’importance du choix du pas dans l’algorithme de pas fixe : un pas “bien choisi” donne des résultats comparables à ceux obtenus par la plus profonde descente, un pas plus petit atténue les zigzag des itérés mais augmente significativement le nombre d’itérations et enfin, un pas trop grand fait diverger la méthode.

2.3.2 Méthode de Newton locale

Pour construire les méthodes de gradient, nous avons remplacé f par son approximation linéaire au voisinage de l’itéré courant. Nous avons vu que ces méthodes ne sont pas très performantes, en partie parce qu’elles ne tiennent pas compte de la courbure (ou de la Hessienne) qui est une information de second ordre.

Principe Supposons maintenant que f est de classe C^2 et remplaçons f au voisinage de l’itéré courant x_k par son développement de Taylor de second ordre :

$$f(y) \sim q(y) = f(x_k) + \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2} \langle H[f](x_k)(y - x_k), y - x_k \rangle,$$

où la valeur $f(x_k)$, le gradient $\nabla f(x_k)$ et la matrice hessienne $H[f](x_k)$ sont donnés par notre oracle (boite noire).

On choisit alors comme point x_{k+1} le minimum de la quadratique q lorsqu'il existe et est unique, ce qui n'est le cas que si $H[f](x_k)$ est définie positive. Or le minimum de q est réalisé par x_{k+1} solution de : $\nabla q(x_{k+1}) = 0$, soit :

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

ou encore, en supposant que $H[f](x_k)$ est définie positive :

$$x_{k+1} = x_k - H[f](x_k)^{-1} \nabla f(x_k). \quad (2.4)$$

On reconnaît dans la formule (2.4) les itérations de la méthode de Newton vue en cours d'analyse numérique, appliquée ici à la résolution de l'équation : $\nabla f(x) = 0$. La méthode ne doit cependant jamais être appliquée en utilisant une inversion de la matrice Hessienne (qui peut être de très grande taille et mal conditionnée) mais plutôt en utilisant :

$$x_{k+1} = x_k + d_k$$

où d_k est l'unique solution du système linéaire :

$$H[f](x_k)d_k = -\nabla f(x_k).$$

d_k est appelée **direction de Newton**.

Cette méthode est bien définie si à chaque itération, la matrice hessienne $H[f](x_k)$ est définie positive : ceci est vrai en particulier au voisinage de la solution x^* cherchée si on suppose que $H[f](x^*)$ est définie positive (par continuité de $H[f]$).

Algorithme

MÉTHODE DE NEWTON LOCALE.

Données: $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 , x_0 première approximation de la solution cherchée,
 $\varepsilon > 0$ précision demandée

Sortie: une approximation x^* de la solution

1. $k := 0$;
 2. Tant que $\|\nabla f(x_k)\| > \varepsilon$,
 - (a) Calculer d_k solution du système : $H[f](x_k)d_k = -\nabla f(x_k)$;
 - (b) $x_{k+1} = x_k + d_k$;
 - (c) $k := k + 1$;
 3. Retourner x_k ;
-

Remarque 2.3 1. La méthode de Newton est un algorithme de descente à pas fixe égal à 1.

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

Exercice 2.3.1 Démontrer les deux assertions de la remarque 2.3.

Convergence de la méthode de Newton locale

L'algorithme hérite des propriétés de l'algorithme de Newton vu en cours d'analyse numérique pour la résolution des équations non-linéaires :

Proposition 2.3 Soit f de classe C^3 et x^* un point de minimum local de f . On suppose en outre que la matrice Hessienne $H[f](x^*)$ est définie positive.

1. Il existe un voisinage \mathcal{V}^* de x^* tel que si $x_0 \in \mathcal{V}^*$, alors la suite des itérés $(x_k)_{k \in \mathbb{N}}$ générés à partir de x_0 par la méthode de Newton locale, converge vers x^* .
2. La convergence est au moins quadratique.

La méthode peut diverger si le point initial n'est pas suffisamment proche d'un point de minimum local, et elle n'est pas définie si les matrices $H[f](x_k)$ ne sont pas définies positives. Utilisée dans le cadre de l'optimisation, la méthode de Newton locale présente un autre inconvénient : la solution identifiée à la fin de l'algorithme n'est pas forcément un point de minimum local, mais uniquement un point critique de f .

2.3.3 Méthode de Gauss-Newton

Si maintenant F désigne une application de \mathbb{R}^n dans \mathbb{R}^m , avec par exemple $m > n$, le système d'équations $F(x) = 0$ n'a généralement pas de solutions. Le problème de moindres carrés associé à F consiste à rechercher x^* tel que

$$r(x^*) = \min \left\{ r(x) = \frac{1}{2} \sum_{i=1}^m F_i(x)^2 = \frac{1}{2} \|F(x)\|_2^2, x \in \mathbb{R}^n \right\}. \quad (2.5)$$

De tels problèmes se rencontrent fréquemment dans le cadre de l'identification de paramètres. Les variables x_i sont les n paramètres d'un modèle physique non linéaire. On effectue $m > n$ mesures, et on cherche les x_i qui permettent d'ajuster au mieux ce modèle aux mesures.

La solution de (2.5) est caractérisée par $\nabla r(x^*) = 0$. Pour appliquer la méthode de Newton, on doit résoudre des systèmes de la forme

$$H_r(x) d = -\nabla r(x) \iff \left(J_F(x)^T J_F(x) + \sum_{i=1}^m F_i(x) H_{F_i}(x) \right) d = -J_F(x)^T F(x), \quad (2.6)$$

où $J_F(x) = \nabla F(x)^\top = [\nabla F_1(x) \dots \nabla F_m(x)]^\top$ désigne la matrice Jacobienne de F en x . Les calculs permettant d'obtenir la relation (2.6) sont détaillés en annexe.

La matrice Hessienne $H_r(x)$ de $r(x)$ a une expression assez compliquée. Cependant le terme $\sum_{i=1}^m F_i(x) H_{F_i}(x)$ est tel que lorsque le résidu $\|F(x)\|$ devient petit, c'est-à-dire lorsque l'on se rapproche de la solution, il devient lui même négligeable.

La méthode de Gauss-Newton consiste à remplacer (2.6) par :

$$J_F(x)^\top J_F(x) d = -J_F(x)^\top F(x). \quad (2.7)$$

Une autre façon de voir les choses consiste à remplacer dans l'expression de r , la fonction F par une approximation linéaire au voisinage du point courant x . Autrement dit, on s'intéresse au problème approché :

$$\min_y \tilde{r}(y) = \frac{1}{2} \|F(x) + J_F(x)(y - x)\|^2.$$

La solution de ce problème est caractérisée par la relation : $\nabla \tilde{r}(y) = 0$ i.e. :

$$J_F(x)^\top (F(x) + J_F(x)(y - x)) = 0 \quad \text{soit :} \quad J_F(x)^\top J_F(x)(y - x) = -J_F(x)^\top F(x)$$

On retrouve ainsi la direction de Gauss-Newton trouvée précédemment.

ALGORITHME DE GAUSS-NEWTON.

Données: F fonction différentiable, x_0 point initial, $\varepsilon > 0$ précision demandée

Sortie: une approximation de la solution du problème de moindres carrés :

$$\min_{x \in \mathbb{R}^n} r(x) = \frac{1}{2} F(x)^\top F(x).$$

1. $k := 0$;

2. Tant que ,

(a) *Calcul d'une direction de recherche :* calculer d_{k+1} solution de :

$$J_F(x_k)^\top J_F(x_k) d = -J_F(x_k)^\top F(x_k).$$

(b) $x_{k+1} = x_k + d_{k+1}$;

(c) $k := k + 1$;

3. Retourner s_k .

Application aux moindres carrés linéaires. Dans le cas où la fonction F est linéaire, i.e. :

$$F(x) = Ax - b, \quad \text{avec} \quad A \in M_{n,p}(\mathbb{R}),$$

on obtient : $J_F(x) = A$, et l'équation de Gauss-Newton (2.7) devient : $A^\top A d_{k+1} = -A^\top (Ax_k - b)$. Comme $d_{k+1} = x_{k+1} - x_k$, on obtient :

$$A^\top A x_{k+1} = A^\top b$$

et ceci quel que soit x_k . On reconnaît ici le système d'équations normales du problème de moindres carrés linéaire :

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2, \quad (2.8)$$

On rappelle que d'après le théorème 2.4 du cours d'analyse numérique, x^* est solution du problème (2.8) si et seulement si x^* vérifie les équations normales : $A^\top Ax = A^\top b$. De plus si A est de rang plein, x^* est l'unique solution de (2.8).

Ceci signifie donc que la méthode de Gauss-Newton identifie la solution en une seule itération lorsque la fonction F est linéaire.

Exercice 2.3.2 Soit $F : \mathbb{R}^3 \rightarrow \mathbb{R}^n$ une application définie par :

$$F_i(x_0, x_1, x_2) = c_i - x_0 - x_1 e^{-x_2 t_i}, \quad i = 1, \dots, n.$$

1. Calculer ∇F_i et $H[F_i]$.
2. En déduire : J_r et H_r où $r = \frac{1}{2} F(x)^\top F(x)$.

Chapitre 3

Introduction à l'optimisation sous contraintes

Ce chapitre est une courte introduction à l'optimisation sous contraintes. On s'intéresse à la résolution de problèmes d'optimisation de la forme :

$$\min_{x \in X} f(x), \quad (3.1)$$

où X est un sous-ensemble non vide de \mathbb{R}^n défini par des contraintes d'égalité ou d'inégalités de fonctions :

$$X = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\} \quad (3.2)$$

où $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^q$ sont **continues**. Ici les écritures $h(x) = 0$ et $g(x) \leq 0$ signifient :

$$\begin{cases} \forall i = 1, \dots, p, h_i(x) = 0 \\ \forall j = 1, \dots, q, g_j(x) \leq 0. \end{cases}$$

L'ensemble X est appelé *ensemble ou domaine des contraintes*. Tout point $x \in \mathbb{R}^n$ vérifiant : $x \in X$, est appelé *point admissible* du problème (3.1).

D'après les résultats du chapitre 1, rappelons que si g et h sont continues alors X est un ensemble fermé (mais non nécessairement borné). Dans la suite nous travaillerons toujours avec un problème sous "forme standard" c'est-à-dire un problème écrit sous la forme suivante :

$$(P) \left| \begin{array}{l} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.c.} \quad h_i(x) = 0, \quad i = 1, \dots, p \\ \quad \quad g_j(x) \leq 0, \quad j = 1, \dots, q. \end{array} \right.$$

Par exemple, la forme standard du problème

$$\max_{(x,y) \in \mathbb{R}^2} f(x,y) \text{ s.t. : } x^2 + y^2 \geq 1 \text{ et } x + y = 5,$$

est :

$$\min_{(x,y) \in \mathbb{R}^2} -f(x,y) \text{ s.t. : } 1 - x^2 - y^2 \leq 0 \text{ et } x + y - 5 = 0.$$

Proposition 3.1 *Le problème (P), écrit sous forme standard, est dit convexe si h est affine, g convexe et si la fonction objectif f est convexe sur X .*

Concernant les contraintes d'inégalités, nous aurons également besoin de la notion de contrainte active :

Définition 3.1 (Contraintes actives/inactives) *Une contrainte d'inégalité $g_j(x) \leq 0$, $j \in \{1, \dots, q\}$, est dite active (ou "saturée") en \bar{x} si :*

$$g_j(\bar{x}) = 0,$$

et inactive en \bar{x} si :

$$g_j(\bar{x}) < 0.$$

Ce qui rend la notion de contrainte active attrayante, c'est le fait qu'au point de minimum local $x^* \in X$, les contraintes actives peuvent être remplacées par des contraintes d'égalité et les contraintes inactives peuvent être ignorées. Si l'intérêt de cette simplification reste essentiellement théorique (on ne connaît pas x^* à l'avance !), un intérêt pratique est que si l'on trouve un point de minimum local du problème sans contrainte, c'est également un point de minimum local du problème contraint. En un point $x \in X$, l'ensemble des contraintes **actives** est l'ensemble des j tel que $g_j(x) = 0$.

Remarquons également que si en un point $x \in X$, la contrainte $g_k(x)$ n'est pas active, alors $g_k(x) < 0$ et ainsi, par continuité de g , pour tout x' dans un voisinage de x , on aura encore $g_k(x') < 0$. Donc localement, on est sûr que la contrainte est vérifiée.

3.1 Conditions d'optimalité

Les conditions d'optimalité dans le cadre de l'optimisation sous contrainte sont de la même forme que les conditions d'Euler du premier ordre vues au Chapitre 2, à savoir des conditions sur le gradient de f aux minima locaux. Il est à noter que la condition d'Euler (le gradient s'annule au points minima) n'est plus valable telle quelle, comme le montre l'exemple suivant :

Exemple 3.1.1 *Soit le problème :*

$$\text{Minimiser } f(x) = x^2, x \in \mathbb{R}^n, \text{ sous la contrainte : } x \geq 1.$$

La solution de ce problème est : $x = 1$, et pourtant : $f'(1) = 2 \neq 0$.

3.1.1 Cas d'une contrainte d'inégalité

Le cas le plus simple que nous saurons prouver (et le seul que nous ayons le temps d'aborder en profondeur dans le cadre de notre cours) est celui d'une seule contrainte d'inégalité, c'est-à-dire un problème du type

$$\min_{g(x) \leq 0} f(x),$$

où f et g sont des fonctions de \mathbb{R}^n dans \mathbb{R} . La condition d'optimalité du premier ordre est :

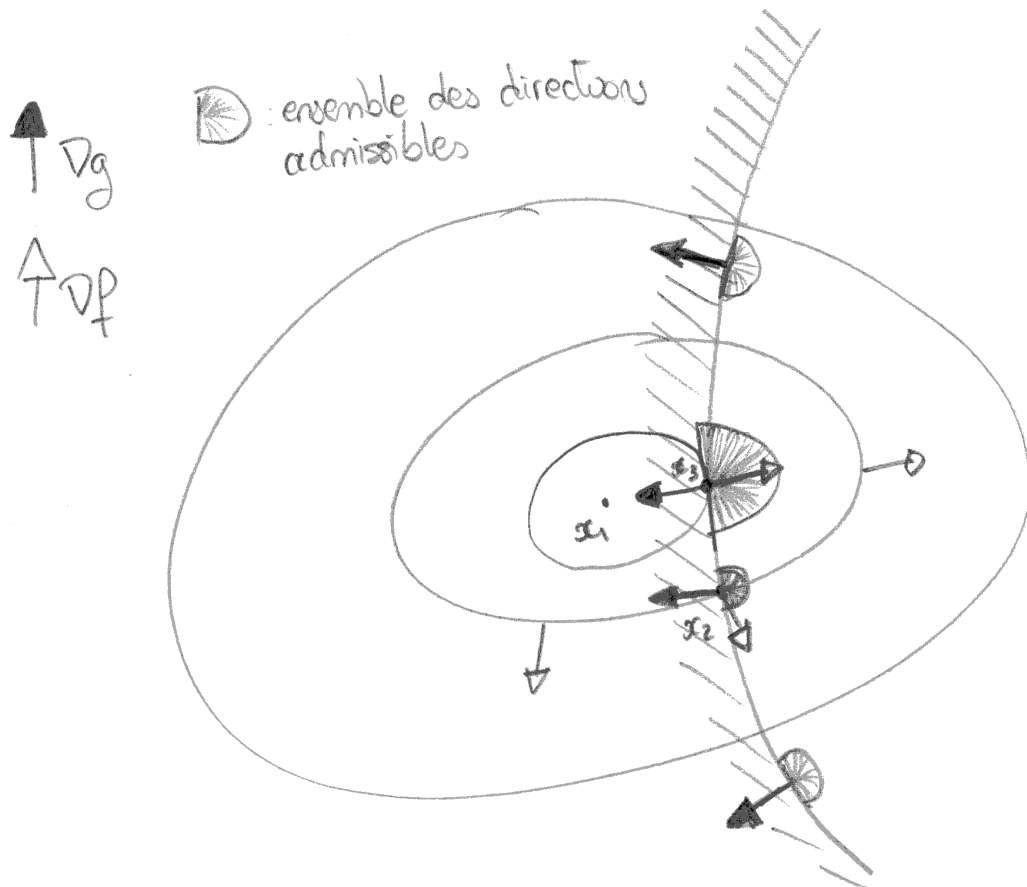
Théorème 3.1 Si f et g sont des fonctions différentiables de \mathbb{R}^n dans \mathbb{R} et si x^* est un minimum local de f sur l'ensemble $X = \{x | g(x) \leq 0\}$, si on suppose de plus que soit $\nabla g(x^*) \neq 0$ soit g est non-active en x^* . Alors il existe $\gamma \geq 0$ tel que

$$\nabla f(x^*) + \gamma \nabla g(x^*) = 0 \quad \text{et} \quad \gamma g(x^*) = 0 \quad (3.3)$$

De plus si f et g sont convexes alors l'équation (3.3) est suffisante pour assurer que x^* est un minimum local (et même global) de f sur l'ensemble X .

On remarque que l'on a besoin d'une hypothèse qui est $\nabla g(x^*) \neq 0$ quand g est active en x^* , cette hypothèse s'appelle "qualification des contraintes".

Exemple 3.1.2 Dans l'exemple ci-dessous, le point x_1 est minimum sans contrainte et le minimum global sous contrainte est le point x_3 .



Preuve. La preuve se fait dans le même ordre d'idée que dans le cas sans contraintes, c'est-à-dire que les directions admissibles ne peuvent pas être directions de descente. On rappelle qu'une

direction d est "admissible en un point x " si il existe $\varepsilon > 0$ tel que pour tout α dans $]0, \varepsilon[$, $x + \alpha d$ appartient à X . De même une direction est dite "de descente au point x " si il existe $\varepsilon > 0$ tel que pour tout α dans $]0, \varepsilon[$, $f(x + \alpha d) < f(x)$. Dire que x^* est un minimum local sur X entraîne donc qu'une direction admissible ne peut pas être une direction de descente. On rappelle que si d est telle que $\langle \nabla f(x), d \rangle < 0$, alors un DL de Taylor montre que d est une direction de descente au point x .

Nous pouvons distinguer deux cas, si la contrainte est active ou non en x^* . Premièrement, si la contrainte est non-active alors $g(x^*) < 0$ et, par continuité, toute direction est direction admissible et en particulier ne peut être direction de descente. On a ainsi $\langle \nabla f(x^*), d \rangle \geq 0$ pour tout $d \in \mathbb{R}^n$ et on peut conclure comme dans le cas sans contrainte que $\nabla f(x^*) = 0$.

Si la contrainte est active en x^* , alors $g(x^*) = 0$, toutes les directions ne sont pas forcément admissibles (effectivement pour certaines directions, g peut devenir positif), cependant tous les d tel que $\langle \nabla g(x^*), d \rangle < 0$ sont directions admissibles (un DL de Taylor montre que dans ces directions, g devient strictement négatif). Comme une direction admissible ne peut pas être direction de descente, on en conclut que

$$\langle \nabla f(x^*), d \rangle \geq 0 \quad \text{pour tout } d \text{ tel que } \langle \nabla g(x^*), d \rangle < 0$$

Un lemme fameux (Lemme de Farkas) permet de conclure que si $\nabla g(x^*) \neq 0$, alors $\nabla f(x^*)$ doit être colinéaire et de sens opposé à g , ce qui se traduit par l'existence d'un $\gamma \geq 0$ tel que :

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

Le lemme de Farkas peut être démontré graphiquement assez simplement en dimension 2. Si on suppose (quitte à effectuer un changement de coordonnées) que $\nabla g(x^*)$ est le vecteur $(1, 0)$, alors l'ensemble des directions admissibles est l'ensemble des directions pointant vers le demi-plan $x < 0$ et si toutes ces directions doivent avoir un produit scalaire positif avec $\nabla f(x^*)$, on voit graphiquement que $\nabla f(x^*)$ doit être colinéaire et de sens opposé à g .

Finalement la formulation de (3.3) n'est qu'une formulation subtile des deux cas traités (soit $\gamma = 0$ dans le cas de la contrainte non-active, soit $g(x^*) = 0$ dans le cas de la contrainte active).

□

On peut se servir de ce Théorème pour chercher les minima locaux d'une fonction

Exemple 3.1.3 Soit le problème :

$$\text{Minimiser } f(x) = x^2, x \in \mathbb{R}^n, \text{ sous la contrainte : } x \geq 1.$$

Ici $g(x) = 1 - x$, et $\nabla g(x) = -1 \neq 0$. On cherche les points \bar{x} tel qu'il existe $\gamma \geq 0$

$$\underbrace{f'(\bar{x}) + \gamma(-1)}_{(1)} = 0 \quad \text{et} \quad \underbrace{\gamma(1 - \bar{x})}_{(2)} = 0.$$

L'équation (2) donne soit $\bar{x} = 1$ soit $\gamma = 0$. Dans le cas $\bar{x} = 1$, l'équation (1) donne $\gamma = 2$. Dans le cas $\gamma = 0$, l'équation (1) donne $\bar{x} = 0$, qui n'est pas un point admissible. Donc il existe un seul minimum local qui est $\bar{x} = 1$ pour $\gamma = 2$.

Il est bon d'avoir en mémoire un exemple où les contraintes ne sont pas qualifiées (c'est-à-dire $\nabla g(x^*) \neq 0$) et où la conclusion du Théorème 3.1 est fautive :

Exemple 3.1.4 Soit le problème :

$$\text{Minimiser } f(x) = x, x \in \mathbb{R}^n, \text{ sous la contrainte : } x^3 \geq 0.$$

Ici $g(x) = -x^3$, la solution de ce problème est : $x = 0$, et pourtant il n'existe pas de $\gamma \geq 0$ tel que $f'(0) + \gamma g'(0) = 0$ car $g'(0) = 0$ et le Théorème 3.1 ne s'applique pas.

3.1.2 Cas de contraintes d'inégalité et d'égalité

Dans le cas où il y a plusieurs contraintes d'égalité ou d'inégalité, le théorème devient :

Théorème 3.2 Si X est donné par (3.2). Si f, g, h sont des fonctions différentiables et si x^* est un minimum local de f sur X et si les contraintes sont qualifiées, alors il existe $\lambda^* \in \mathbb{R}^p, \gamma^* \in \mathbb{R}^q$ tels que

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^q \gamma_j^* \nabla g_j(x^*) = 0 \\ h_i(x^*) = 0, & i = 1, \dots, p \\ \gamma_j^* g_j(x^*) = 0, & j = 1, \dots, q \\ \gamma_j^* \geq 0, & j = 1, \dots, q. \end{cases} \quad (3.4)$$

Les points qui vérifient ces conditions sont appelés points KKT (Karush, Kuhn, Tucker)

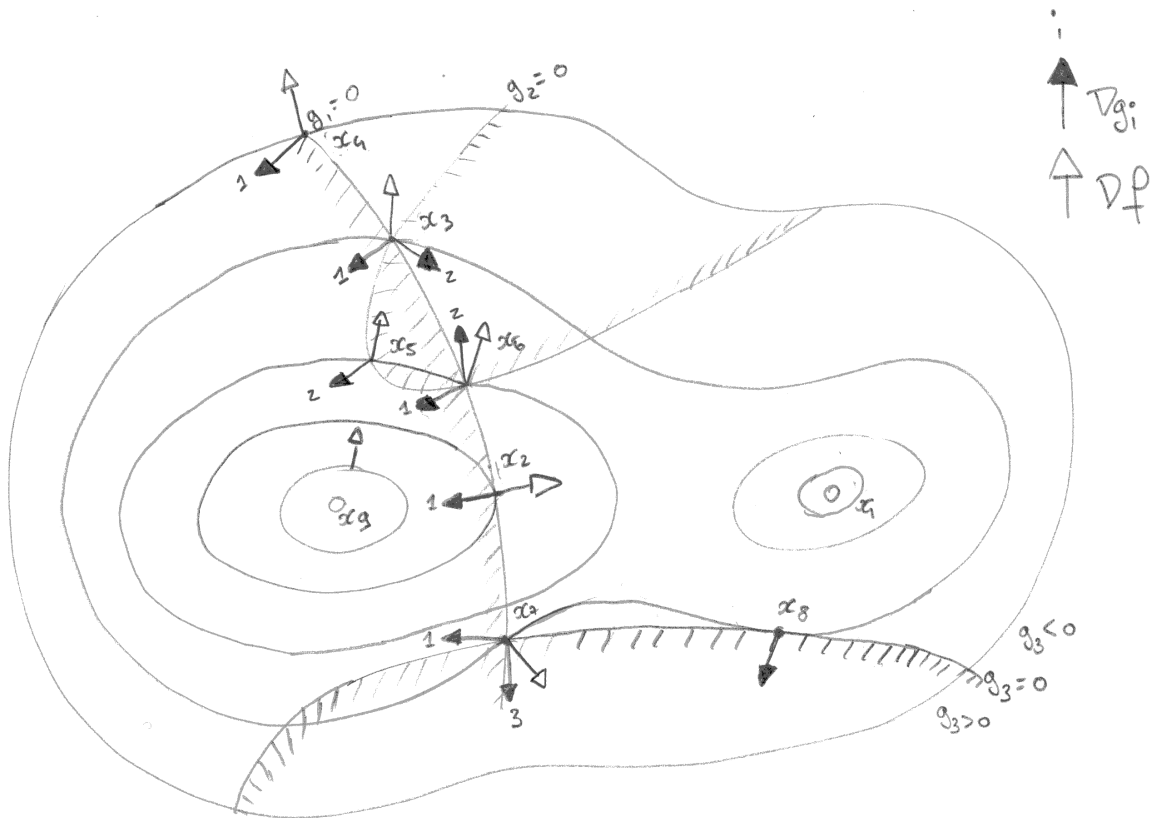
Concernant les contraintes d'inégalité, ce théorème est effectivement une extension raisonnable du théorème précédent. On peut aussi comprendre les contraintes d'égalité en scindant chaque contrainte d'égalité en deux contraintes d'inégalité : effectivement $h(x) = 0$ est équivalent à dire que $h(x) \leq 0$ et $-h(x) \leq 0$. Comme dans le cas d'une contrainte d'inégalité, il y a une hypothèse supplémentaire à ajouter, qui est la "qualification des contraintes". Il y a plusieurs types de qualification différente, mais une des plus générales est :

Définition 3.2 En un point $x \in X$, les contraintes sont dites **qualifiées** si et seulement si les vecteurs $(\nabla h(x)_i)_i$ sont linéairement indépendants et si il existe une direction d orthogonale à tous les vecteurs $\nabla h(x)_i$ telle que $\nabla g_j(x)^\top d < 0$ pour tout les contraintes actives.

S'il y a une seule contrainte d'inégalité, cette contrainte est qualifiée en un point si et seulement si son gradient est non nul quand elle est active. On retrouve les conditions du théorème précédent. Dans la définition du dessus, la direction d est appelée "direction ré-entrante". La démonstration de ce théorème se fera en 4ème année GMM, nous l'utiliserons cependant dans le cadre de la recherche de minimum locaux (ou globaux) qui sont soit des points de KKT soit des points où les contraintes ne sont pas qualifiées.

Il faut faire attention cependant, tous les points de KKT ne sont pas minima locaux de la même manière que tous les points critiques ne sont pas minima locaux.

Exemple 3.1.5 Dans l'exemple ci-dessous, seuls les points x_1 , x_2 et x_3 sont des points de KKT. Ils sont tous des minima globaux. Le point x_9 vérifie l'équation mais n'appartient pas à l'ensemble des points admissibles.



Exemple 3.1.6 Trouver les solutions de

$$\min 2x + 3y, \quad \text{s.c.} \quad x \geq 0 \text{ et } x^2 + y^2 \leq 1$$

Ici $f(x, y) = 2x + 3y$, $g_1(x, y) = -x$ et $g_2(x, y) = x^2 + y^2 - 1$ leurs gradients valent respectivement $(2, 3)$, $(-1, 0)$ et $(2x, 2y)$. La recherche des minima globaux se fait en trois étapes :

- Points où les contraintes ne sont pas qualifiées : Les points où ni g_1 ni g_2 ne sont actives sont des points où les contraintes sont qualifiées. Les points où une seule des deux contraintes est active et pas l'autre sont aussi des points où les contraintes sont qualifiées car leur gradient n'est jamais nul. Il ne reste plus que les points $(0, 1)$ et $(0, -1)$ où les

deux contraintes sont actives en même temps. Comme les gradients des contraintes en ces points sont non-colinéaires, il existe une direction rétrante. Ainsi, les contraintes sont qualifiées en tout point.

– Points de KKT : on doit résoudre le système

$$\begin{cases} 2 - \gamma_1 + 2\gamma_2 x = 0 \\ 3 + 2\gamma_2 y = 0 \end{cases} \quad \text{et} \quad \begin{cases} \gamma_1 x = 0 \\ \gamma_2 (x^2 + y^2 - 1) = 0 \end{cases} \quad \text{et} \quad \begin{cases} \gamma_1 \geq 0 \\ \gamma_2 \geq 0 \end{cases} \quad \text{et} \quad \begin{cases} x \geq 0 \\ x^2 + y^2 \leq 1 \end{cases}$$

L'unique solution est $(x, y, \gamma_1, \gamma_2) = (0, -1, 2, 3/2)$

– Comparaison des points trouvés : On doit calculer la valeur de f pour tous les points trouvés, ici on n'en a qu'un, c'est donc notre minimum global.

Il est important de connaître un exemple où les contraintes sont non qualifiées et où le minimum global n'est pas point de KKT :

Exemple 3.1.7 Trouver les solutions de

$$\min y, \quad \text{s.c.} \quad (x-1)^2 + y^2 \leq 1 \quad \text{et} \quad (x+1)^2 + y^2 \leq 1$$

Ici l'ensemble des contraintes est l'intersection de deux disques de rayon 1 et de centre $(-1, 0)$ et $(0, 1)$, ce qui donne l'unique point $(0, 0)$. Le minimum de f vaut donc $f(0, 0) = 0$ mais $\nabla f = (0, 1)$ alors qu'en ce point $\nabla g_1 = (-2, 0)$ et $\nabla g_2 = (2, 0)$, il est impossible d'écrire ∇f comme combinaison linéaire de ∇g_1 et ∇g_2 , donc le point n'est pas de KKT. Ceci est dû au fait que les contraintes ne sont pas qualifiées.

3.2 Lagrangien du problème

Le Lagrangien d'un problème d'optimisation sous contraintes d'égalité et d'inégalité est un objet dont nous ne pouvons qu'effleurer l'importance dans ce cours.

Définition 3.3 Soit $\lambda \in \mathbb{R}^p$ et $\gamma \in (\mathbb{R}_+)^q$. On définit :

$$\mathcal{L}(x, (\lambda, \gamma)) = f(x) + \sum_{i=1}^p \lambda_i h_i(x) + \sum_{j=1}^q \gamma_j g_j(x)$$

ou plus généralement

$$\mathcal{L}(x, (\lambda, \gamma)) = f(x) + \langle \lambda, h(x) \rangle_{\mathbb{R}^p} + \langle \gamma, g(x) \rangle_{\mathbb{R}^q}$$

le **Lagrangien** associé au problème (P).

Les paramètres (λ, γ) sont appelés **multiplicateurs de Lagrange**. A noter : le Lagrangien fait intervenir autant de multiplicateurs que de contraintes, et les multiplicateurs γ_j associés aux contraintes d'inégalité " $g_j(x) \leq 0$ " sont positifs ou nuls.

L'intérêt du Lagrangien est de ramener le problème (3.1)-(3.2) de minimisation sous contrainte de f sur X à un problème **sans contrainte** :

Définition-Proposition 3.1 *Le point x^* est un minimum global de f sur X (i.e. solution du problème $\min_{h(x)=0, g(x)\leq 0} f(x)$) si et seulement si il est aussi une solution de*

$$\min_{x \in \mathbb{R}^n} \max_{(\lambda, \gamma) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} \mathcal{L}(x, (\lambda, \gamma))$$

Preuve. Si $x \in X$ alors $\mathcal{L}(x, (\lambda, \gamma)) = f(x) + \langle \gamma, g(x) \rangle$ et le maximum de cette expression par rapport à γ est donné pour $\gamma = 0$ car $g(x) \leq 0$ et $\gamma \geq 0$. Si $x \notin X$ alors le maximum de \mathcal{L} par rapport au couple (λ, γ) vaut $+\infty$.

Ainsi

$$\max_{(\lambda, \gamma) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} \mathcal{L}(x, (\lambda, \gamma)) = \begin{cases} f(x) & \text{si } x \in X \\ +\infty & \text{si } x \notin X \end{cases}$$

d'où :

$$\min_{x \in \mathbb{R}^n} \max_{(\lambda, \gamma) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} \mathcal{L}(x, (\lambda, \gamma)) = \min_{x \in X} f(x),$$

et ainsi l'égalité des deux problèmes considérés. \square

Points-selles du Lagrangien

L'idée générale de l'introduction du Lagrangien est que l'on doit maximiser \mathcal{L} par rapport à sa deuxième variable et le minimiser par rapport à sa première variable, cela ressemble à la recherche d'un **point selle**.

Définition 3.4 *Soit $\mathcal{L} : U \times V \rightarrow \mathbb{R}$ une fonction à deux variables, le point $(x^*, p^*) \in U \times V$ est dit "point-selle" de \mathcal{L} sur $U \times V$ si et seulement si*

$$\mathcal{L}(x^*, p) \leq \mathcal{L}(x^*, p^*) \leq \mathcal{L}(x, p^*) \quad \forall (x, p) \in U \times V$$

Exemple 3.2.1 *Un point-selle d'une fonction à deux variables \mathcal{L} est donc un minimum en la première variable et un maximum en la deuxième variable. L'archétype de tous les points-selles est le point $(0, 0)$ de la fonction de \mathbb{R}^2 dans $\mathbb{R} : \mathcal{L}(u, v) = u^2 - v^2$.*

Théorème 3.3 *Considérons le problème sous contraintes fonctionnelles (3.1)-(3.2) :*

$$\begin{aligned} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{sous :} & h(x) = 0, g(x) \leq 0 \end{aligned}$$

où f, g et h sont supposées différentiables. On note X le domaine des contraintes et \mathcal{L} le Lagrangien associé à f et aux contraintes X .

Si (x^*, p^*) est un point-selle de \mathcal{L} sur $\mathbb{R}^n \times (\mathbb{R}^p \times (\mathbb{R}_+)^q)$ alors

- x^* est un minimum global de f sur X .
- On a l'équation suivante reliant x^* et $p^* = (\lambda^*, \gamma^*)$:

$$\nabla f(x^*) + \sum_i \lambda_i^* \nabla h_i(x^*) + \sum_j \gamma_j^* \nabla g_j(x^*) = 0 \quad \text{et } \forall j, \quad \gamma_j^* g_j(x^*) = 0 \quad (3.5)$$

Dans ce cas p^* est appelé le "multiplicateur de Lagrange" du point-selle (x^*, p^*) .

Avant de démontrer ce théorème, il faut bien en comprendre les implications. Ce théorème ne dit pas que tous les minima globaux de f sur X viennent de point-selle, (ou même que les points-selles du Lagrangien existent), il dit seulement que si on trouve un point-selle (x^*, p^*) , alors x^* est un minimum global de f sur X . De plus, on ne sait pas si tous les minima globaux obéissent à l'équation (3.5) (car il faut rajouter l'hypothèse de qualification des contraintes). En ce sens, ce théorème n'est pas un théorème de caractérisation des minima, au sens où on donne une condition suffisante mais non nécessaire. Il va de soit que ce théorème ne dit pas que tout (x^*, p^*) qui vérifie (3.5) est forcément point-selle (ce qui entraînerait que x^* est minimum global).

Preuve. On note $p = (\lambda, \gamma)$. On rappelle que les conditions de points-selles sont

$$\mathcal{L}(x^*, p) \leq \mathcal{L}(x^*, p^*) \leq \mathcal{L}(x, p^*) \quad \forall (x, p) \in \mathbb{R}^n \times P \quad (3.6)$$

En récrivant la première inégalité de (3.6), on obtient

$$f(x^*) + \langle \lambda, h(x^*) \rangle + \langle \gamma, g(x^*) \rangle \leq \mathcal{L}(x^*, p^*) \quad \forall \lambda \in \mathbb{R}^n, \gamma \in \mathbb{R}_+^p$$

On en conclut que $h(x^*) = 0$ et $g(x^*) \leq 0$, sinon en choisissant des λ et des γ très grand, on peut rendre le terme de gauche aussi grand que l'on veut. Ainsi x^* appartient bien à X . De plus, on obtient, en prenant $\lambda = \gamma = 0$:

$$f(x^*) \leq f(x^*) + \langle \gamma^*, g(x^*) \rangle$$

et comme γ^* est positif et $g(x^*)$ négatif, on en conclut que $\langle \gamma^*, g(x^*) \rangle = 0$. En récrivant la deuxième inégalité de (3.6), on obtient, en se limitant à des x dans X :

$$f(x^*) = \mathcal{L}(x^*, p^*) \leq f(x) + \langle \gamma^*, g(x) \rangle \quad \forall x \in X$$

et comme γ^* est positif et $g(x)$ est négatif pour tout x dans X , alors $f(x) + \langle \gamma^*, g(x) \rangle \leq f(x)$ et ainsi

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

Ainsi x^* est bien un minimum global de f sur X . Maintenant en reprenant la deuxième inégalité du Lagrangien :

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

On voit que x^* est un minimum sans contrainte de la fonctionnelle différentiable $g : x \mapsto \mathcal{L}(x, p^*)$ donc vérifie l'égalité d'Euler, $\nabla g(x^*) = 0$ qui est exactement la formule (3.5). On remarquera que pour l'établissement de cette formule, p^* est fixé et on ne doit dériver que par rapport à x .

□

Le théorème précédent assure que tout point-selle du Lagrangien est un minimum global de la fonction initiale et que tout point-selle vérifie l'équation (3.5). De plus on sait que si les contraintes sont qualifiées alors tout minimum local vérifie (3.5). La question qui se pose est de savoir quelles hypothèse ajouter pour qu'un point qui vérifie (3.5) soit un point-selle.

La réponse est donnée **en partie** par le théorème de Kuhn-Tucker (aussi quelquefois appelé Karush, Kuhn, Tucker).

Théorème 3.4 (Kuhn-Tucker) *Supposons X donné par des contraintes d'inégalité uniquement :*

$$X = \{x \in \mathbb{R}^n : g(x) \leq 0\},$$

On note :

$$\mathcal{L}(x, \gamma) = f(x) + \langle \gamma, g(x) \rangle$$

le Lagrangien associé à f avec les contraintes X . On suppose toutes les fonctions f et g_j convexes et différentiables. Si x^ est un point de X où les contraintes sont qualifiées, alors les propositions suivantes sont équivalentes :*

- x^* est un minimum global de f sur X .
- Il existe un $\gamma^* \in \mathbb{R}_+^q$ tel que (x^*, γ^*) est un point-selle du Lagrangien \mathcal{L} .
- Il existe un $\gamma^* \in \mathbb{R}_+^q$ tel que (x^*, γ^*) vérifie :

$$\nabla f(x^*) + \sum_{j=1}^q \gamma_j^* \nabla g_j(x^*) = 0 \quad \text{et} \quad \forall j = 1, \dots, q, \gamma_j^* g_j(x^*) = 0 \quad (3.5)$$

Preuve. Comme la deuxième assertion entraîne la première (Théorème 3.3) et que la première entraîne la troisième (Proposition 3.2), il suffit de montrer que la troisième assertion entraîne la première pour conclure. Soit un x^* tel que $\langle g, \gamma^* \rangle = 0$ et

$$\nabla f(x^*) + \sum_j \gamma_j^* \nabla g_j(x^*) = 0$$

Comme les g_j sont convexes et $\gamma^* \geq 0$, alors la fonction $L : x \mapsto \mathcal{L}(x, \gamma^*)$ est convexe. Et comme

$$\nabla L(x^*) = \nabla f(x^*) + \sum_j \gamma_j^* \nabla g_j(x^*) = 0,$$

alors x^* est le minimum global de L et on a bien

$$\mathcal{L}(x^*, \gamma^*) = L(x^*) \leq L(x) = \mathcal{L}(x, \gamma^*) \quad \forall x \in \mathbb{R}^n.$$

De plus, comme $\langle g(x^*), \gamma^* \rangle = 0$ et $g(x^*) \leq 0$, alors, pour tout $\gamma \geq 0$ $\mathcal{L}(x^*, \gamma) = f(x^*) + \langle g(x^*), \gamma \rangle \leq f(x^*) = \mathcal{L}(x^*, \gamma^*)$. Ainsi nous avons

$$\mathcal{L}(x^*, \gamma) \leq \mathcal{L}(x^*, \gamma^*) \leq \mathcal{L}(x, \gamma^*) \quad \forall x \in \mathbb{R}^n, \gamma \in P.$$

Ce qui prouve que (x^*, γ) est bien un point selle. □

En résumé

Les conditions nécessaires d'optimalité d'un problème d'optimisation avec contraintes d'inégalité et d'égalité, s'écrivent :



Proposition 3.2 (Conditions nécessaires d'optimalité) *Considérons le problème :*

$$(P) \quad \begin{array}{ll} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{sous :} & h(x) = 0, g(x) \leq 0 \end{array}$$

où f , g et h sont supposées différentiables. On note X le domaine des contraintes et \mathcal{L} le Lagrangien associé à f et aux contraintes X .

Si x^* est un minimum local de f sur X et si les contraintes sont qualifiées en x^* alors il existe des multiplicateurs $(\lambda^*, \gamma^*) \in \mathbb{R}^p \times \mathbb{R}_+^q$ tel que :

1. $\nabla_x \mathcal{L}(x, (\lambda^*, \gamma^*)) = 0$, soit :

$$\nabla f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^q \gamma_j^* \nabla g_j(x^*) = 0.$$

2. $\forall j = 1, \dots, q, \gamma_j^* g_j(x^*) = 0$ i.e. pour tout $j = 1, \dots, q$:

$$\gamma_j^* = 0 \quad \text{ou} \quad g_j(x^*) = 0.$$

Ces conditions sont appelées **relations de complémentarité**.

3. $\forall i = 1, \dots, p, h_i(x^*) = 0$.

4. $\forall j = 1, \dots, p, g_j(x^*) \leq 0$.

5. $\forall j = 1, \dots, q, \gamma_j^* \geq 0$.

TERMINOLOGIE :

- Le vecteur (λ^*, γ^*) est aussi appelé solution *duale* du problème (P) , x^* solution *primale* de (P) et $(x^*, \lambda^*, \gamma^*)$ solution *primate-duale* de (P) .
- On appelle *point stationnaire* du problème (P) tout point \bar{x} vérifiant les conditions nécessaires d'optimalité du premier ordre :

$$\left\{ \begin{array}{l} \nabla_x \mathcal{L}(\bar{x}, (\bar{\lambda}, \bar{\gamma})) = 0 \\ \bar{\gamma}_j g_j(\bar{x}) = 0, \quad j = 1, \dots, q \\ h_i(\bar{x}) = 0, \quad i = 1, \dots, p \\ \bar{\gamma}_j \geq 0, \quad j = 1, \dots, q. \end{array} \right.$$

pour un certain multiplicateur $(\bar{\lambda}, \bar{\gamma}) \in \mathbb{R}^p \times \mathbb{R}_+^q$.

- Les relations : $\gamma_j g_j(\bar{x}) = 0$ sont appelées *relations de complémentarité*. Elles sont trivialement satisfaites par toute contrainte j active en \bar{x} et indiquent que pour toute contrainte inactive, le multiplicateur γ_j correspondant est nul. Cela signifie que toute contrainte inactive à l'optimum aurait pu être relaxée.

On remarquera qu'il n'y a pas de condition sur les multiplicateurs λ_i^* associés aux contraintes d'égalité.

3.3 Algorithme du gradient projeté

Le but de cette section est d'élaborer un algorithme de résolution du problème :

$$\text{Minimiser } f(x), x \in \mathbb{R}^n, \text{ sous la contrainte : } x \in X, \quad (3.7)$$

dans le cas où X est un sous-ensemble convexe fermé non vide de \mathbb{R}^n , non nécessairement défini par des égalités ou des inégalités. La condition nécessaire d'optimalité locale est alors un peu plus générale que ce qui a été vu au paragraphe précédent :

Théorème 3.5 (Condition nécessaire d'optimalité locale) *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable et X un convexe fermé, non vide de \mathbb{R}^n . Soit $x^* \in \mathbb{R}^n$ un point de minimum local de f sur l'ensemble admissible X , alors*

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \quad \forall x \in X \quad (3.8)$$

Si de plus f est convexe alors la condition précédente devient suffisante, c'est-à-dire que tout $x^ \in X$ vérifiant l'équation (3.8) est un minimum global de f sur X .*

La méthode du gradient projeté s'inspire des méthodes de gradient décrites dans le chapitre précédent. L'idée de base consiste à suivre la direction de plus profonde descente, comme dans le cas sans contrainte :

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

où $s_k > 0$ est choisi de sorte que : $f(x_k + s_k d_k) < f(x_k)$. Toutefois, si $x_k \in X$, rien ne garantit que x_{k+1} appartienne également à X . Pour résoudre ce problème, si d'aventure on obtient un point x_{k+1} non admissible, on le projette celui-ci sur l'ensemble de contraintes X .

Définition-Proposition 3.2 (Projection sur un convexe) *Soit X convexe fermé, non vide de \mathbb{R}^n . La projection d'un point $x \in \mathbb{R}^n$ sur X , notée $p_X(x)$, est définie comme l'unique solution de :*

$$\min_{y \in X} \frac{1}{2} \|x - y\|_2^2 \quad (3.9)$$

De plus $p_X(x)$ est l'unique élément de X qui vérifie :

$$\langle p_X(x) - x, y - p_X(x) \rangle \geq 0, \quad \forall y \in X. \quad (3.10)$$

Preuve. On définit la fonctionnelle $f : y \in \mathbb{R}^n \mapsto \frac{1}{2} \|x - y\|_2^2$, dont le gradient vaut $\nabla f(y) = y - x$ et la Hessienne $H[f](x) = Id$. La fonctionnelle f est strictement convexe. Comme l'ensemble des contraintes est fermé borné et f est continue, le Théorème 1.1 nous assure l'existence d'un minimum global. Comme l'ensemble des contraintes est convexe et la fonction f est strictement convexe, le Théorème 1.3 nous assure qu'il est unique. Comme f est strictement convexe et que l'ensemble des contraintes est convexe, la condition (3.10) est une condition nécessaire et suffisante pour déterminer l'unique minimiseur de (3.9). \square

On remarque en particulier que si $x \in X$, alors nécessairement : $p_X(x) = x$.

ALGORITHME DU GRADIENT PROJETÉ.

Données: f , p_X l'opérateur de projection sur X , x_0 un point initial et $\varepsilon > 0$ la précision demandée

Sortie: une approximation x^* de la solution

1. $k := 0$;
2. Tant que critère d'arrêt non satisfait,
 - (a) $y_k = x_k - s\nabla f(x_k)$ où s est le pas calculé par la méthode de gradient choisie;
 - (b) *Projection sur X* : $x_{k+1} = p_X(y_k)$; $k := k + 1$;
3. Retourner x_k .

Remarque 3.1 *Il est important de remarquer que le calcul à l'étape 2(b) du projeté sur X , peut parfois être aussi difficile que le problème initial. En effet y_k est obtenu en résolvant le problème :*

$$\begin{aligned} \min_{y \in \mathbb{R}^n} \quad & \frac{1}{2} \|x_k - s\nabla f(x_k) - y\|_2^2 \\ \text{s.c.} \quad & x \in X. \end{aligned}$$

Il s'agit donc de résoudre un problème d'optimisation sur un convexe, avec une fonction objectif convexe. Lorsque le domaine X des contraintes est simple (contraintes de bornes en particulier), c'est faisable. Dès que les contraintes ne sont pas des contraintes de bornes, le calcul de la projection devient beaucoup plus délicat.

Vérifions que la direction $d_k = x_{k+1} - x_k$, si elle est non nulle, est bien une direction de descente de f en x_k .

Lemme 3.1 *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ suppose différentiable et $X \subset \mathbb{R}^n$ un convexe fermé, non vide. Notons x_k l'itéré courant et :*

$$d(s) = p_X(x_k - s\nabla f(x_k)) - x_k, \quad s > 0$$

Si $d(s)$ est non nulle, alors $d(s)$ est une direction de descente pour tout $s > 0$.

Preuve. Soit $s > 0$ fixé. Supposons : $d(s) = p_X(x_k - s\nabla f(x_k)) - x_k \neq 0$. Il s'agit de démontrer que $d(s)$ est une direction de descente de f en x_k , autrement dit que $\nabla f(x_k)^\top d(s) < 0$.

D'après la caractérisation (3.10) de la projection sur un convexe, on peut écrire :

$$\forall y \in X, \langle p_X(x_k - s\nabla f(x_k)) - (x_k - s\nabla f(x_k)), y - p_X(x_k - s\nabla f(x_k)) \rangle \geq 0,$$

D'où, pour tout $y \in X$: $\langle d(s) + s\nabla f(x_k), y - x_k - d(s) \rangle \geq 0$. Puisque $x_k \in X$, on choisit $y = x_k$, soit :

$$-\langle d(s) + s\nabla f(x_k), d(s) \rangle \geq 0, \quad \text{ou encore : } \langle \nabla f(x_k), d(s) \rangle \leq -\frac{1}{s} \langle d(s), d(s) \rangle \leq 0.$$

Par hypothèse, $d(s) \neq 0$ ce qui implique : $\langle \nabla f(x_k), d(s) \rangle < 0$. □

Remarque 3.2 La direction $d(s)$ possède les propriétés suivantes :

1. Si $d(s) = 0$, alors : $p_X(x_k - s\nabla f(x_k)) = x_k$. Cela signifie que la direction choisie par l'algorithme de gradient est orthogonale à l'ensemble X des contraintes en x_k . Le point x_k est alors un point stationnaire car la condition nécessaire d'optimalité (3.5) est satisfaite.
2. Supposons $d(s) \neq 0$. Alors x_k et $p_X(x_k - s\nabla f(x_k))$ sont des points admissibles du problème (3.7). La convexité de X nous garantit alors : $\forall \alpha \in [0, 1], x_k + \alpha d(s) \in X$.

Exemple 3.3.1 On veut résoudre par une méthode de gradient projeté le problème suivant :

$$\min_{(x,y) \in \mathbb{R}^2} f(x,y) = \frac{1}{2}x^2 + \frac{7}{2}y^2 \quad \text{sous} \quad -x + y = 1.$$

Le domaine des contraintes $X = \{(x,y) \in \mathbb{R}^2 / -x + y = 1\}$ est un convexe fermé. Pour l'avoir résolu au chapitre 2, on sait que le problème hors contrainte admet un minimum global en $(0,0)$. Cependant $(0,0)$ ne satisfait pas la contrainte, ce n'est donc pas un point admissible du problème avec contrainte.

Afin de mettre en oeuvre l'algorithme de gradient projeté, il faut choisir les méthodes de calcul des pas s_k et α_k aux étapes 2(a) et 2(c) de notre algorithme :

- étape 2(a) : on choisit une méthode de gradient pour le calcul du pas s_k .
- étape 2(c) : en première approche, on choisit un pas fixe $\alpha_k = 1$, ce qui implique : $x_{k+1} = y_k = p_X(x_k - s\nabla f(x_k))$.

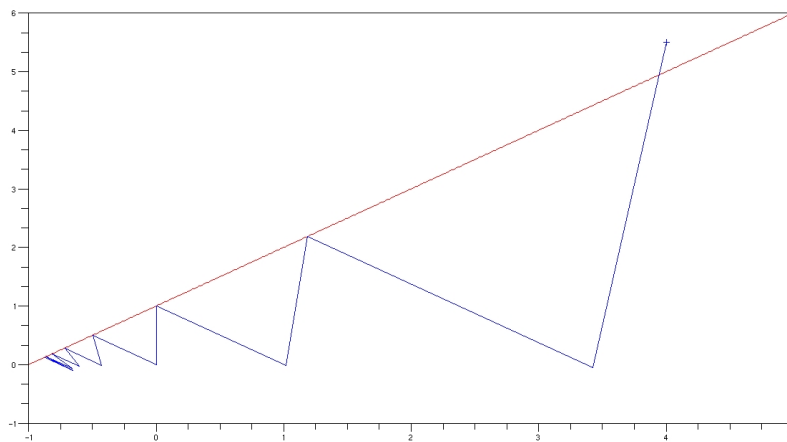


FIGURE 3.1 – Itérations successives (avant et après projection) de l'algorithme du gradient projeté - exemple 3.3.1 à partir du point $(4, 5.5)$

Le comportement numérique de la méthode de gradient projeté ainsi définie, est illustré par la figure 3.1 et le tableau 3.1 des itérations. On observe une convergence de la suite d'itérés générés à partir du point $x_0 = (4, 5.5)$ vers le point $(-0.875, 0.125)$.

k	x_k	$x_k - s_k \nabla f(x_k)$	$\ d_k\ = \ y_k - x_k\ $
0	[4 5.5]	[3.4232925 - 0.0508095]	4.3472097
1	[1.1862415 2.1862415]	[1.0159064 - 0.0112495]	1.6743058
2	[0.0023285 1.0023285]	[0.0019958 - 9.462e - 08]	0.7089885
3	[-0.4990021 0.5009979]	[-0.4264826 - 0.0086691]	0.3091099
4	[-0.7175758 0.2824242]	[-0.6037028 - 0.0313035]	0.1413186
5	[-0.8175032 0.1824968]	[-0.6619890 - 0.0605186]	0.0618727
6	[-0.8612538 0.1387462]	[-0.6636631 - 0.0840740]	0.0178399
7	[-0.8738685 0.1261315]	[-0.6570769 - 0.0929058]	0.0015879
8	[-0.8749913 0.1250087]	[-0.6562565 - 0.0937435]	0.0000123
9	[-0.875 0.125]	[-0.65625 - 0.09375]	7.301e - 10

TABLE 3.1 – Itérations de la méthode du gradient projeté : on note $y_k = p_X(x_k - s_k \nabla f(x_k))$. Le critère d'arrêt $\|d_k\| < \varepsilon$ est satisfait en 10 itérations pour une précision $\varepsilon = 10^{-5}$.

Vérifions analytiquement ce résultat. On cherche un point $(x, y) \in \mathbb{R}^2$ vérifiant la contrainte : $y = 1 + x$ et minimisant f , ce qui revient à minimiser sans contrainte l'application :

$$\tilde{f} : x \mapsto f(x, 1 + x) = 4x^2 + 7x + \frac{7}{2}.$$

Remarquons que \tilde{f} est strictement convexe ; d'après le théorème 2.2 appliqué au point $x^* = (-\frac{7}{8}, \frac{1}{8}) = (-0.875, 0.125)$, \tilde{f} admet un point de minimum global en x^* .

Annexe A

Compléments

A.1 Rappels de calcul différentiel

Différentiabilité et gradient

Nous commençons par rappeler la notion de différentiabilité :

Définition A.1 Soit $J : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle. J est différentiable au point x ssi il existe L_x une application linéaire de \mathbb{R}^n dans \mathbb{R} et une fonction ε de $\mathbb{R}^n \times \mathbb{R}^n$ dans \mathbb{R} telle que

$$J(x + \mathbf{h}) = J(x) + L_x(\mathbf{h}) + \|\mathbf{h}\| \varepsilon(\mathbf{h}, x), \quad \forall \mathbf{h} \in \mathbb{R}^n, \quad (\text{A.1})$$

avec

$$\varepsilon(\mathbf{h}, x) \xrightarrow{\mathbf{h} \rightarrow 0} 0. \quad (\text{A.2})$$

Comme L_x est une application linéaire de \mathbb{R}^n dans \mathbb{R} , il existe un vecteur appelé gradient de f au point x tel que

$$L_x(\mathbf{h}) = \langle \nabla J(x), \mathbf{h} \rangle, \quad \forall \mathbf{h} \in \mathbb{R}^n \quad (\text{A.3})$$

Lorsqu'il existe, le gradient peut être exprimé à l'aide des dérivées partielles

$$\nabla J(x) = \begin{bmatrix} \partial_{x_1} J(x) \\ \partial_{x_2} J(x) \\ \dots \\ \partial_{x_n} J(x) \end{bmatrix}. \quad (\text{A.4})$$

Aller plus loin : Nous avons choisi d'introduire la forme linéaire L_x (appelée la différentielle au point x) dans (A.1) au lieu de mettre la formule de Taylor :

$$J(x + \mathbf{h}) = J(x) + \langle \nabla J(x), \mathbf{h} \rangle + \|\mathbf{h}\| \varepsilon(\mathbf{h}, x), \quad \forall \mathbf{h} \in \mathbb{R}^n,$$

Il y a une raison fondamentale à cela. En effet, si on change le produit scalaire de \mathbb{R}^n , alors on va changer la formule (A.3) donc le gradient mais pas la différentielle. En effet, la formule (A.1) est indépendante du choix du produit scalaire, cependant la formule (A.3) n'est que l'application du théorème suivant au produit scalaire usuel :

Si L est une forme linéaire sur H et (\cdot, \cdot) est un produit scalaire, il existe un seul et unique vecteur $u \in H$ tel que $L(v) = (u, v)$ pour tout $v \in H$.

Ainsi, si on change le produit scalaire, on change la définition du gradient.

On donne un théorème qui permet de dire qu'une fonction est différentiable en calculant son gradient.

Définition-Proposition A.1 Soit $J : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle telle que chacune des composantes (les dérivées partielles) de ∇J existe et sont continues en un point x . Alors J est différentiable au point x et on dit que J est C^1 en ce point.

Dérivées d'ordre supérieur

Définition A.2 On dit que $J : \mathbb{R}^n \rightarrow \mathbb{R}$ est deux fois différentiable au point $x \in \mathbb{R}^n$ ssi J est différentiable et chacune des composantes de ∇J est différentiable. Dans ce cas on appelle "Hessienne de J " la matrice :

$$H[J](x) = [\partial_i \partial_j J(x)]_{1 \leq i, j \leq n}. \quad (\text{A.5})$$

De la même manière qu'il est plus simple de montrer qu'une fonction est C^1 que de montrer qu'une fonction est différentiable, il est plus souvent plus facile d'établir qu'une fonction est C^2 plutôt que deux fois différentiable.

Définition-Proposition A.2 Soit une fonctionnelle $J : \mathbb{R}^n \rightarrow \mathbb{R}$ dont chacune des composantes de $H[J](x)$ est continue (les dérivées partielles secondes sont continues au point x), alors J est deux fois différentiable et on dit que J est C^2 au point x .

Théorème A.1 (de Schwarz) Soit $J : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle de classe C^2 , alors, pour tout x , la matrice $H[J](x)$, est symétrique.

Formules de Taylor

On rappelle les notations de Landau :

$$\boxed{\begin{aligned} g(h) = \mathcal{O}(h^p) &\Leftrightarrow \begin{cases} \exists a > 0, \exists M > 0, |h| < a \Rightarrow \left| \frac{g(h)}{h^p} \right| \leq M : \\ \text{"}g(h) \text{ tend vers } 0 \text{ avec } h \text{ aussi vite que } h^p \text{"}. \end{cases} \\ g(h) = o(h^p) &\Leftrightarrow \begin{cases} \lim_{h \rightarrow 0} \frac{g(h)}{h^p} = 0, \\ \text{"}g(h) \text{ tend vers } 0 \text{ avec } h \text{ plus vite que } h^p \text{"} \end{cases} \end{aligned}}$$

On rappelle que la formules de Taylor s'écrivent pour une fonction f de \mathbb{R} dans \mathbb{R} :

$$f(x+h) = f(x) + \underbrace{hf'(x)}_{(1)} + \underbrace{\frac{h^2}{2}f''(x)}_{(2)} + \dots + \frac{h^n}{n!}f^{(n)}(x) + o(h^n). \quad (\text{A.6})$$

On remarque que, dans cette équation, le terme linéaire en h du développement de Taylor (le terme (1)) fait intervenir la dérivée de f alors que le terme quadratique (le terme (2)) fait intervenir la dérivée seconde.

Proposition A.1 Soit $J : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle de classe \mathcal{C}^2 , les formules de Taylor s'écrivent :

$$J(x+h) = J(x) + \langle \nabla J(x), h \rangle + \frac{1}{2} \langle h, H[J](x) h \rangle + o(\|h^2\|). \quad (\text{A.7})$$

Proposition A.2 Soit $J : \mathbb{R}^n \rightarrow \mathbb{R}^p$ une fonctionnelle. On dit que J est de classe \mathcal{C}^2 si et seulement si chacune de ses composantes $J_i : \mathbb{R}^n \rightarrow \mathbb{R}$ est de classe \mathcal{C}^2 . Dans ce cas, les formules de Taylor s'écrivent :

$$J(x+h) = J(x) + L_x(h) + Q_x(h, h) + o(\|h^2\|). \quad (\text{A.8})$$

où L_x est une application linéaire de \mathbb{R}^n dans \mathbb{R}^p et Q_x est une application bilinéaire de $\mathbb{R}^n \times \mathbb{R}^n$ dans \mathbb{R}^p . Le terme $Q_x(h, h)$ est donc un terme quadratique. On a les formules suivantes pour L_x et Q_x , en notant $J_i(x)$ la $i^{\text{ème}}$ composante de $J(x)$

$$\begin{aligned} (L_x(h))_i &= \sum_{k=1}^n \frac{\partial J_i(x)}{\partial x_k} h^k \quad \forall i = 0, \dots, p \\ (Q_x(a, b))_i &= \sum_{k=1}^n \sum_{l=1}^n \frac{\partial^2 J_i(x)}{\partial x_k \partial x_l} a^k b^l \quad \forall i = 0, \dots, p \end{aligned} \quad (\text{A.9})$$

La matrice $n \times p$ qui représente l'application L_x est appelée Jacobienne J au point x .

Tout ceci se montre en écrivant l'équation A.6 pour chaque composante de J .

A.2 Quelques démonstrations

A.2.1 Hessienne et convexité : démonstration du Théorème 1.5

Théorème 1.5. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 . On note $H[f](x)$ sa hessienne en x .

- Si $H[f](x)$ est semidéfinie positive pour tout $x \in \mathbb{R}^n$, alors f est convexe.
- Si $H[f](x)$ est définie positive pour tout $x \in \mathbb{R}^n$, alors f est strictement convexe.

Pour montrer ce résultat nous avons besoin du lemme technique suivant.

Lemme A.1 Soit $\phi \in \mathcal{C}^2([0, 1])$.

- Si $\phi^{(2)}(x) > 0$, alors $\phi(\lambda) < (1-\lambda)\phi(0) + \lambda\phi(1)$ pour $\lambda \in]0, 1[$.
- Si $\phi^{(2)}(x) \geq 0$, alors $\phi(\lambda) \leq (1-\lambda)\phi(0) + \lambda\phi(1)$.

Preuve. Comme $\phi^{(2)}(x) > 0$ (resp. \geq) pour $0 \leq x \leq 1$, ϕ' est strictement croissante (resp. croissante) et par conséquent

$$\begin{aligned} \phi(\lambda) - \phi(0) &= \int_0^\lambda \phi'(t) dt < \lambda \phi'(\lambda) \quad (\text{resp. } \leq), \\ \phi(1) - \phi(\lambda) &= \int_\lambda^1 \phi'(t) dt > (1-\lambda) \phi'(\lambda) \quad (\text{resp. } \geq). \end{aligned}$$

En groupant ces deux inégalités, nous tirons

$$\frac{\phi(\lambda) - \phi(0)}{\lambda} < \left(\phi'(\lambda) < \right) \frac{\phi(1) - \phi(\lambda)}{1 - \lambda} \quad (\text{resp. } \leq);$$

c'est-à-dire : $\phi(\lambda) < (1 - \lambda)\phi(0) + \lambda\phi(1)$ (resp. \leq). \square

Preuve du théorème 1.5. Supposons $H[f](z)$ symétrique définie positive. Soient x et y deux éléments distincts de \mathbb{R}^n . Introduisons la fonction $\phi : [0, 1] \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 définie par

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

Utilisons le développement à l'ordre deux de ϕ pour calculer la dérivée seconde de ϕ

$$\phi(\lambda + h) = f((\lambda + h)x + (1 - \lambda - h)y) = f(\lambda x + (1 - \lambda)y + h(x - y)).$$

Avec $t = \lambda x + (1 - \lambda)y$, nous avons

$$\begin{aligned} \phi(\lambda + h) &= f(t) + h[\nabla f(t)]^\top (x - y) + \frac{1}{2}h^2(x - y)^\top H[f](t)(x - y) \\ &\quad + h^2\|x - y\|^2\varepsilon(t, h(x - y)), \end{aligned}$$

et, par conséquent

$$\phi(\lambda + h) = \phi(\lambda) + h([\nabla f(t)]^\top (x - y)) + \frac{h^2}{2} \left((x - y)^\top H[f](t)(x - y) \right) + h^2\varepsilon(h)$$

Comme ϕ est de classe \mathcal{C}^2 , la dérivée première et la dérivée seconde de ϕ sont donc données par

$$\begin{aligned} \phi'(\lambda) &= [\nabla f(\lambda x + (1 - \lambda)y)]^\top (x - y), \\ \phi^{(2)}(\lambda) &= (x - y)^\top \underbrace{H[f](\lambda x + (1 - \lambda)y)}_{\in \mathbb{R}^{n \times n}} (x - y). \end{aligned}$$

Donc la fonction $\phi^{(2)}(\lambda)$ est strictement positive car $H[f](\lambda x + (1 - \lambda)y)$ est symétrique définie positive et $x \neq y$. Le lemme A.1 s'écrit :

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

puisque : $\phi(0) = f(y)$, $\phi(1) = f(x)$ et $\phi(\lambda) = f(\lambda x + (1 - \lambda)y)$.

Ceci prouve que f est strictement convexe. Si pour tout x , $H[f](x)$ est positive, une preuve similaire nous permet de montrer que f est convexe. \square

A.2.2 Méthode de Gauss-Newton pour la résolution des problèmes de moindres carrés

Problème de moindres carrés

- $F : \mathbb{R}^p \longrightarrow \mathbb{R}^n, n > p : F(x) = \begin{bmatrix} F_1(x_1, \dots, x_p) \\ \dots \dots \\ F_n(x_1, \dots, x_p) \end{bmatrix}.$
- $r(x) = \frac{1}{2} \|F(x)\|^2 = \frac{1}{2} \langle F(x), F(x) \rangle.$
- On cherche $x^* = \operatorname{argmin} \{r(x)\} \Leftrightarrow r(x^*) = \min \left\{ \frac{1}{2} \sum_{i=1}^n F_i^2(x), x \in \mathbb{R}^p \right\}.$

Utilisation de la méthode de Newton

- Condition nécessaire d'optimalité du premier ordre : $\nabla r(x^*) = 0.$
- Résolution par la méthode de Newton : pour $k \geq 0,$

$$x^{(k+1)} = x^{(k)} - (H_r(x^{(k)}))^{-1} \nabla r(x^{(k)}). \quad (\text{A.10})$$

Calcul de $\nabla r(x)$: on peut le faire en **identifiant de la partie linéaire** d'un développement au voisinage de x :

$$\begin{aligned} 2 r(x+h) &= \langle F(x+h), F(x+h) \rangle, \\ &= \langle F(x) + J_F(x)h + \mathcal{O}(\|h\|^2), F(x) + J_F(x)h + \mathcal{O}(\|h\|^2) \rangle, \\ &= \langle F(x), F(x) \rangle + \langle F(x), J_F(x)h \rangle + \langle J_F(x)h, F(x) \rangle + \mathcal{O}(\|h\|^2). \end{aligned}$$

- Comme en dimension 1, on assimile les termes qui seront au moins $\mathcal{O}(\|h\|^2)$: $\langle F(x), \mathcal{O}(\|h\|^2) \rangle + \langle J_F(x)h, J_F(x)h \rangle + \dots = \mathcal{O}(\|h\|^2).$
- Pour $A \in M_{n,p}(\mathbb{R}), x \in \mathbb{R}^n$ et $y \in \mathbb{R}^p,$ on a $\langle x, Ay \rangle = \langle A^T x, y \rangle :$

$$r(x+h) = r(x) + \langle J_F(x)^T F(x), h \rangle + \mathcal{O}(\|h\|^2).$$

- Par identification,

$$\nabla r(x) = J_F(x)^T F(x). \quad (\text{A.11})$$

Calcul de $\nabla r(x)$: on peut le faire en **calculant les dérivées partielles** :

- Pour tout $i, 1 \leq i \leq p,$

$$\partial_{x_i} r(x) = \partial_{x_i} \left(\frac{1}{2} \sum_{j=1}^n F_j^2(x) \right) = \sum_{j=1}^n (\partial_{x_i} F_j(x)) F_j(x).$$

– En prenant en compte toute les lignes,

$$\begin{bmatrix} \partial_{x_1} r(x) \\ \dots \\ \partial_{x_p} r(x) \end{bmatrix} = \begin{bmatrix} \partial_{x_1} F_1(x) & \dots & \partial_{x_1} F_n(x) \\ \dots & & \dots \\ \partial_{x_p} F_1(x) & \dots & \partial_{x_p} F_n(x) \end{bmatrix} \begin{bmatrix} F_1(x) \\ \dots \\ F_n(x) \end{bmatrix} = \sum_{j=1}^n (\nabla F_j) F_j = J_F^T F.$$

Calcul de $H_r(x)$: on peut le faire par **identification de la partie quadratique** d'un développement au voisinage de x :

$$\begin{aligned} r(x+h) &= \frac{1}{2} \left\langle F(x) + J_F(x) h + \frac{1}{2} Q_F(h) + \mathcal{O}(\|h\|^3), F(x) + J_F(x) h + \frac{1}{2} Q_F(h) + \mathcal{O}(\|h\|^3) \right\rangle, \\ &= r(x) + \langle \nabla r(x), h \rangle + \frac{1}{2} \langle J_F(x) h, J_F(x) h \rangle \dots \\ &\dots + \frac{1}{4} \langle F(x), Q_F(h) \rangle + \frac{1}{4} \langle Q_F(h), F(x) \rangle + \mathcal{O}(\|h\|^3), \end{aligned} \tag{A.12}$$

– La partie quadratique de (A.12) s'identifie selon

$$\langle H_r(x) h, h \rangle = \langle J_F(x)^T J_F(x) h, h \rangle + \frac{1}{2} \langle F(x), Q_F(h) \rangle$$

– Suivant (A.9), $\langle F(x), Q_F(h) \rangle = \left(\sum_{i=1}^n F_i(x) h^T H_{F_i}(x) \right) h$, d'où

$$H_r = J_F^T J_F + \sum_{i=1}^n F_i H_{F_i}. \tag{A.13}$$

Calcul de $H_r(x)$: on peut le faire **en dérivant** $\nabla r(x)$:

$$- \nabla r = \sum_{j=1}^n (\nabla F_j) F_j.$$

– On utilise la règle de dérivation d'un produit : $(\nabla r)' = \sum_{j=1}^n (H_{F_j} F_j + (\nabla F_j) J_{F_j})$.

$$- J_{F_j} = (\nabla F_j)^T, \text{ et } \sum_{j=1}^n (\nabla F_j) (\nabla F_j)^T = [\nabla F_1 \mid \dots \mid \nabla F_n] \begin{bmatrix} \nabla F_1^T \\ \dots \\ \nabla F_n^T \end{bmatrix} = J_F^T J_F.$$

Calcul de $H_r(x)$: on peut **laborieusement** le faire en calculant les dérivées partielles secondes ; on ne le fait pas.