UFR
DE MATHÉMATIQUES
ET INFORMATIQUE

Université
Paris Cité

Optimisation

Notes de cours

Master 1 Mathématiques, Modélisation, Apprentissage (MMA)

2023-2024

Quentin DENOYELLE
Bureau 812-D
quentin.denoyelle@parisdescartes.fr

Table des matières

1 Rappels et compléments de calculs différentiels **4**

1.1 Cadre et notations 4

1.2 Différentielle et gradient 5

1.2.1 Différentielle 5

1.2.2 Gradient 6

1.2.3 Dérivation des fonctions composées 7

1.3 Différentielle seconde et matrice hessienne 7

1.4 Formules de Taylor 9

2 Problèmes d'optimisation : Existence et unicité des solutions **10**

2.1 Cadre et vocabulaire 10

2.2 Généralités sur l'existence de solutions 11

2.2.1 Existence d'une suite minimisante 11

2.2.2 Coercivité et existence d'une solution 12

2.3 Extrema locaux et différentiabilité 13

2.3.1 Définitions 13

2.3.2 Extrema locaux et condition d'ordre un 13

2.3.3 Extrema locaux et conditions d'ordre deux 14

2.4 Ensembles convexes 15

2.5 Fonctions convexes 18

2.5.1 Définition et exemples 18

2.5.2 Caractérisation des fonctions convexes différentiables 19

2.5.3 Caractérisation des fonctions convexes deux fois différentiables 20

2.5.4 Problèmes d'optimisation convexes 21

2.6 Etude des fonctionnelles quadratiques 22

3 Algorithmes de descente de gradient pour les problèmes sans contraintes **25**

3.1 Méthode de descente 25

3.2 Algorithme de descente de gradient à pas optimal 26

3.2.1 Définition de l'algorithme et premières propriétés 26

3.2.2 Forte convexité 27

3.2.3 Convergence de l'algorithme du gradient à pas optimal 30

3.3 Descente de gradient préconditionné à rebroussement d'Armijo 33

3.3.1 Choix du pas par rebroussement d'Armijo 33

3.3.2 Algorithme de descente de gradient préconditionné 33

3.3.3 Convergence de l'algorithme de descente de gradient préconditionné à rebroussement d'Armijo 34

Introduction

Ce cours est une introduction aux problèmes d'optimisation. Le cours se focalise sur des problèmes d'optimisation sans contrainte pour les fonctions suffisamment différentiables en dimension finie. Après une introduction des différentes notions mathématiques nécessaires (rappels de calcul différentiel, conditions d'optimalité, convexité, etc.), une part importante est donnée à l'exposition des différents algorithmes classiques d'optimisation, l'étude théorique de leur convergence, ainsi que leur mise en œuvre pratique. Le langage Python sera utilisé en séance de Travaux Pratiques (TP).

L'auteur remercie Bruno Galerne qui est à l'origine de ce poly, Joan Glaunès pour ses nombreux conseils, et enfin Quentin Mérigot car le Chapitre 3 de ce poly est fortement inspiré de ses notes de cours `http://quentin.mrgt.fr/cours/m315/`.

Les principaux ouvrages de référence pour ce cours sont :

[ROUVIÈRE] François ROUVIÈRE, *Petit guide de calcul différentiel à l'usage de la license et de l'agrégation*, troisième édition, Cassini, 2009

[CIARLET] Philippe G. CIARLET, *Introduction à l'analyse numérique matricielle et à l'optimisation*, cinquième édition, Dunod, 1998

[BOYD & VANDENBERGHE] Stephen BOYD and Lieven VANDENBERGHE *Convex Optimization*, Cambridge University Press, 2004.
Ouvrage téléchargeable gratuitement ici :
`http://stanford.edu/~boyd/cvxbook/`

[ALLAIRE & KABER] Grégoire ALLAIRE et Sidi Mahmoud KABER, *Algèbre linéaire numérique*, Ellipses, 2002

La page web dédiée à ce cours est ici :
`https://qdenoyelle.github.io/M1_Optim/`

Chapitre 1

Rappels et compléments de calculs différentiels

La référence principale pour ce chapitre est le chapitre 2 de [ROUVIÈRE]. Voir aussi le chapitre A.4 de [BOYD & VANDENBERGHE] et le chapitre 7 de [CIARLET].

1.1 Cadre et notations

Espaces. Dans ce cours on se placera toujours sur des espaces vectoriels normés de dimensions finis que l'on identifie à $\mathbb{R}^n$, munis de leur base canonique notée $(e_1, \dots, e_n)$, $n \geqslant 1$. Par convention, on identifiera très souvent les vecteurs de $\mathbb{R}^n$ à des vecteurs colonnes. On note $\langle \cdot, \cdot \rangle$ le produit scalaire canonique et $\|\cdot\|$ la norme euclidienne associée.

Matrices. On note $\mathcal{M}_{m,n}(\mathbb{R})$ l'ensemble des matrices de taille $m \times n$ à coefficients réelles et $\mathcal{M}_n(\mathbb{R}) = \mathcal{M}_{n,n}(\mathbb{R})$ l'ensemble des matrices carrées de taille $n \times n$. La transposée d'une matrice A est notée A^T. On a donc pour tous $x, y \in \mathbb{R}^n$, $\langle x, y \rangle = x^T y$ et par conséquent, pour tout $A \in \mathcal{M}_{m,n}(\mathbb{R})$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$

$$\langle y, Ax \rangle = \left\langle A^T y, x \right\rangle .$$

Remarque 1.1 (Notation de la transposée)**.** La notation A^T correspond plutôt à une convention anglo-saxonne. Elle a été choisie pour ce polycopié car elle est plus simple à taper en LaTeX. Toutefois les étudiantes et étudiants sont libres d'utiliser la notation classique ${}^tA$ pour leurs prises de notes et leurs copies d'examen.

On rappelle que l'ensemble des matrices symétriques réelles peut être muni d'une relation d'ordre (partielle) définie par

$$A \preceq B \Leftrightarrow \forall x \in \mathbb{R}^n, x^T A x \leqslant x^T B x.$$

Pour vérifier que c'est bien une relation d'ordre, la seule difficulté concerne l'antisymétrie qui utilise le résultat de diagonalisation des matrices symétriques. Nous utiliserons cette relation d'ordre dans la suite du cours pour énoncer de manière concise les résultats de convergence des algorithmes étudiées.

Applications linéaires et matrices associées. On désigne par $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ l'ensemble des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^m$. Vu la convention sur les vecteurs considérés comme vecteurs colonnes, on identifiera souvent un élément de $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ à une matrice rectangulaire de taille $m \times n$ correspondant à la matrice de l'application dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$: si $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ alors pour tout $x \in \mathbb{R}^n$, $\varphi(x) = Ax$ avec A la matrice dont les colonnes sont les images par φ des vecteurs de la base canonique $(e_1, \dots, e_n)$ de la base canonique de $\mathbb{R}^n$

$$\varphi(x) = \varphi\left(\sum_{k=1}^n x_k e_k\right) = \sum_{k=1}^n x_k \varphi(e_k) = \begin{pmatrix} \varphi(e_1) & \cdots & \varphi(e_n) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = Ax.$$

1.2 Différentielle et gradient

Dans la suite de ce chapitre, Ω désigne un *ensemble ouvert* de $\mathbb{R}^n$.

1.2.1 Différentielle

Définition 1.2. Soit $f : \Omega \to \mathbb{R}^m$. La fonction f est *différentiable au point* $x \in \Omega$ si il existe une application linéaire (continue) $\ell \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ telle que pour tout $h \in \mathbb{R}^n$ au voisinage de 0, on ait

$$f(x+h) = f(x) + \ell(h) + \|h\| \, \varepsilon(h)$$

où $\varepsilon : \mathbb{R}^n \to \mathbb{R}$ est une fonction satisfaisant $\lim\limits_{h \to 0} \varepsilon(h) = 0$ (d'où $\|h\| \, \varepsilon(\|h\|) = o(\|h\|)$). Dans ce cas ℓ est unique et est appelée la différentielle de f au point x et notée $\mathrm{d}f(x)$.

La matrice de $\mathrm{d}f(x)$ (dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$) est appelée la matrice jacobienne de f au point x, notée $J_f(x)$.

On dit que f est *différentiable* sur Ω si f est différentiable en tout point de Ω.

On dit que f est *continûment différentiable*, notée $f \in \mathcal{C}^1(\Omega, \mathbb{R}^m)$, si f est différentiable sur Ω et l'application différentielle $\mathrm{d}f : x \in \Omega \mapsto \mathrm{d}f(x)$ est continue sur Ω.

La fonction affine $y \in \mathbb{R}^n \mapsto f(x) + \mathrm{d}f(x)(y - x)$ est l'approximation à l'ordre 1 de f au voisinage du point x.

Proposition 1.3 (Dérivées partielles)**.** Si $f = (f_1, \dots, f_m)$, avec $f_i : \mathbb{R}^n \mapsto \mathbb{R}$, est différentiable en $x \in \Omega$ alors f admet des dérivées partielles en x. Ainsi pour tout $j \in \{1, \dots, n\}$ on a

$$\partial_{x_j} f(x) = \lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = \mathrm{d}f(x)(e_j).$$

On a alors

$$\mathrm{d}f(x)(h) = \sum_{j=1}^n \partial_{x_j} f(x) h_j = J_f(x) h,$$

où $J_f(x) = \big(\partial_{x_j} f(x)\big)_{1 \leqslant j \leqslant n} = \big(\partial_{x_j} f_i(x)\big)_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n} \in \mathcal{M}_{m,n}(\mathbb{R})$.

Exercice 1.4. Soit A une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$ et $b \in \mathbb{R}^m$. Montrer que la fonction $f : \mathbb{R}^n \to \mathbb{R}^m$ définie par $f(x) = Ax + b$ est différentiable sur $\mathbb{R}^n$ et calculer sa différentielle en tout point $x \in \mathbb{R}^n$. En déduire que f est $\mathcal{C}^1$ (et même $\mathcal{C}^\infty$) sur $\mathbb{R}^n$.

Solution de l'exercice 1.4. Pour tous $x, h \in \mathbb{R}^n$,

$$f(x+h) = A(x+h) + b = f(x) + Ah$$

donc f est différentiable en x et $\mathrm{d}f(x) = A$. C'est vrai pour tout $x \in \mathbb{R}^n$ donc f est différentiable sur $\mathbb{R}^n$. Puis comme l'application différentielle $\mathrm{d}f : x \in \mathbb{R}^n \mapsto \mathrm{d}f(x)$ est constante, c'est en particulier une application continue, d'où f est bien continûment différentiable sur $\mathbb{R}^n$.

1.2.2 Gradient

Dans ce cours on s'intéressera plus particulièrement à des fonctions à valeurs réelles, ce qui correspond au cas $m = 1$. Ce cas particulier important, nous amène à définir la notion fondamentale de gradient.

Définition 1.5 (Gradient d'une fonction). Si $f : \Omega \to \mathbb{R}$ est différentiable en $x \in \Omega \subset \mathbb{R}^n$ alors $\mathrm{d}f(x)$ est une forme linéaire sur $\mathbb{R}^n$ et donc il existe un unique vecteur, noté $\nabla f(x) \in \mathbb{R}^n$, tel que

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

et appelé *gradient de f* au point x.

Vu autrement de manière équivalente, la matrice $J_f(x)$ est une matrice ligne de taille $1 \times n$ et la transposée de cette matrice est un vecteur de $\mathbb{R}^n$ appelé *gradient de f* au point x et noté $\nabla f(x)$. Pour tout $h \in \mathbb{R}^n$,

$$\mathrm{d}f(x)(h) = \nabla f(x)^T h = \langle \nabla f(x),\, h \rangle.$$

Ainsi, pour $f : \Omega \to \mathbb{R}$, f est différentiable si et seulement si il existe un vecteur $v \in \mathbb{R}^n$ tel que

$$f(x+h) = f(x) + \langle v,\, h \rangle + o(h).$$

et alors $v = \nabla f(x)$.

$\nabla f(x)$ s'interprète comme le vecteur de plus forte augmentation de f au voisinage de x. En particulier, $\nabla f(x)$ est orthogonal au ligne de niveaux de la fonction f.

Exercice 1.6. Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$. Soit $f : \mathbb{R}^n \to \mathbb{R}$ l'application $f : x \mapsto \frac{1}{2} \langle Ax,\, x \rangle$.

1. Montrer que f est différentiable sur $\mathbb{R}^n$ et déterminer $\nabla f(x)$ pour tout x.
2. Quelle est l'expression de ∇f si A est symétrique ?
3. Quel est le gradient de l'application $x \mapsto \frac{1}{2}\|x\|^2$?

Solution de l'exercice 1.6.

1. Pour tous $x, h \in \mathbb{R}^n$,

$$\begin{aligned} f(x+h) &= \frac{1}{2} \langle A(x+h),\, x+h \rangle \\ &= \frac{1}{2} \langle Ax,\, x \rangle + \frac{1}{2} \langle Ax,\, h \rangle + \frac{1}{2} \langle Ah,\, x \rangle + \frac{1}{2} \langle Ah,\, h \rangle \\ &= f(x) + \frac{1}{2} \left\langle (A + A^T)x,\, h \right\rangle + \frac{1}{2} \langle Ah,\, h \rangle. \end{aligned}$$

Or par l'inégalité de Cauchy-Schwarz, $|\frac{1}{2} \langle Ah,\, h \rangle| \leqslant \frac{1}{2}\|Ah\|\|h\| \leqslant \frac{1}{2}\|A\|_{\mathcal{M}_n(\mathbb{R})}\|h\|^2$, donc c'est bien en o de $\|h\|$. D'où f est différentiable en tout point $x \in \mathbb{R}^n$ et

$$\nabla f(x) = \frac{1}{2}(A + A^T)x.$$

2. Si A est symétrique, alors $\nabla f(x) = \frac{1}{2}(A + A^T)x = Ax$.
3. C'est le cas particulier où $A = I_n$ qui est symétrique, donc $\nabla f(x) = x$.

1.2.3 Dérivation des fonctions composées

Théorème 1.7 (Dérivation des fonctions composées)**.** Soient $f : \mathbb{R}^n \to \mathbb{R}^m$ et $g : \mathbb{R}^m \to \mathbb{R}^p$ deux fonctions différentiables. Soit $h = g \circ f : \mathbb{R}^n \to \mathbb{R}^p$ la fonction composée définie par $h(x) = g(f(x))$. Alors h est différentiable sur $\mathbb{R}^n$ et pour tout $x \in \mathbb{R}^n$,

$$\mathrm{d}h(x) = \mathrm{d}g(f(x)) \circ \mathrm{d}f(x).$$

où $\mathrm{d}g(f(x)) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$ et $\mathrm{d}f(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. En terme de matrice jacobienne, cela s'écrit

$$J_h(x) = J_g(f(x))J_f(x).$$

Remarque 1.8. On peut énoncer une version locale du résultat précédent car, comme le suggère la formule $\mathrm{d}h(x) = \mathrm{d}g(f(x))\mathrm{d}f(x)$, pour que h soit différentiable en x, il suffit que f soit différentiable en x et que g soit différentiable en $f(x)$.

Exemple 1.9. Déterminons le gradient de l'application $h : \mathbb{R}^n \to \mathbb{R}$ définie par

$$h(x) = g(Ax + b)$$

où A est une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$ et $g : \mathbb{R}^m \to \mathbb{R}$ est une application différentiable. On a $h(x) = g \circ f(x)$ avec $f(x) = Ax + b$. Comme f est affine on a $\mathrm{d}f(x) = A$ en tout point x. On a donc d'après la règle de dérivation des fonctions composées

$$J_h(x) = J_g(f(x))J_f(x) = J_g(Ax + b)A.$$

Donc $\nabla h(x) = J_h(x)^T = A^T J_g(Ax + b)^T = A^T \nabla g(Ax + b)$.

1.3 Différentielle seconde et matrice hessienne

Définition 1.10 (Différentielle seconde)**.** Soit $f : \Omega \to \mathbb{R}^m$ une application différentiable sur Ω. On dira que f est deux fois différentiable en $a \in \Omega$ si l'application

$$\mathrm{d}f : x \mapsto \mathrm{d}f(x),$$

est différentiable en a. On note cette différentielle $\mathrm{d}^2 f(a) = \mathrm{d}(\mathrm{d}f)(a)$ et on l'appelle la différentielle seconde de f en a.

On dira que f est deux fois continûment différentiable sur Ω, noté $f \in \mathcal{C}^2(\Omega, \mathbb{R}^m)$, si f est deux fois différentiable en tout point de Ω et $x \mapsto \mathrm{d}^2 f(x)$ est continue sur Ω.

Comme $\mathrm{d}f : \Omega \subset \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, on a ainsi $\mathrm{d}^2 f(a) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ que l'on peut donc identifier à une application bilinéaire de $\mathbb{R}^n \times \mathbb{R}^n$ à valeurs dans $\mathbb{R}^m$.

Théorème 1.11 (Théorème de Schwarz)**.** Soit $f : \Omega \to \mathbb{R}^m$ une application deux fois différentiable en $a \in \Omega$. Alors $\mathrm{d}^2 f(a) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ est une application bilinéaire symétrique, c'est-à-dire

$$\forall (h, k) \in \mathbb{R}^n \times \mathbb{R}^n, \quad \mathrm{d}^2 f(a)(h, k) = \mathrm{d}^2 f(a)(k, h).$$

En pratique dans ce cours, nous allons très souvent considérer le cas particulier $m = 1$. La différentielle seconde en un point est alors une *forme* bilinéaire symétrique que l'on peut caractérisé par sa matrice dans la base canonique de $\mathbb{R}^n$.

Définition 1.12 (Matrice hessienne)**.** Soit $f : \Omega \to \mathbb{R}$ une application deux fois différentiable en $a \in \Omega$. Alors la forme bilinéaire symétrique $\mathrm{d}^2 f(a)$ peut s'identifier à sa matrice dans la base canonique de $\mathbb{R}^n$, appelée *matrice hessienne de f en a* et notée

$$\nabla^2 f(a) = \left(\mathrm{d}^2 f(e_i, e_j) = \partial^2_{x_i,x_j} f(a)\right)_{1\leqslant i,j\leqslant n}.$$

On a alors par définition

$$\forall (h,k) \in \mathbb{R}^n \times \mathbb{R}^n, \quad \mathrm{d}^2 f(x)(h,k) = \left\langle \nabla^2 f(x)h,\, k\right\rangle = k^T \nabla^2 f(x) h = h^T \nabla^2 f(x) k.$$

Pour ce cours on aura constamment besoin de calculer le gradient et la matrice hessienne de fonctionnelles $f : \Omega \to \mathbb{R}$ deux fois différentiable. En pratique on utilise la proposition suivante.

Proposition 1.13 (La matrice hessienne est la jacobienne du gradient)**.** Soit $f : \Omega \to \mathbb{R}$ une fonction deux fois différentiable au point $a \in \Omega$. Alors, la matrice hessienne $\nabla^2 f(a)$ de f au point a est la matrice jacobienne $J_{\nabla f}(a)$ de l'application gradient $x \mapsto \nabla f(x)$ au point a. Dit autrement

$$\forall h \in \mathbb{R}^n, \quad \mathrm{d}(\nabla f)(a)(h) = J_{\nabla f}(a) h = \nabla^2 f(a) h.$$

Notons que la première égalité est simplement la définition de la Jacobienne, c'est la deuxième qui correspond à cette proposition.

Démonstration. Posons $g(x) = \nabla f(x)$ pour tout $x \in \Omega$. Alors, par définition, la différentielle de $g : \mathbb{R}^n \to \mathbb{R}^n$ est caractérisée par sa matrice jacobienne

$$J_g(a) = \left(\partial_{x_j} g_i(a)\right)_{1\leqslant i,j\leqslant n} = \left(\partial_{x_j} \left(\partial_{x_i} f\right)(a)\right)_{1\leqslant i,j\leqslant n} = \left(\partial^2_{x_i,x_j} f(x)\right)_{1\leqslant i,j\leqslant n} = \nabla^2 f(x).$$

□

Exemple 1.14 (Composition avec une fonction affine)**.** Soit $g : \mathbb{R}^n \to \mathbb{R}$ définie par

$$g(x) = f(Ax + b)$$

où A est une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$ et $f : \mathbb{R}^m \to \mathbb{R}$ est une application deux fois différentiable sur $\mathbb{R}^m$. Alors, par composée, g est deux fois différentiable et

$$\nabla^2 g(x) = A^T \nabla^2 f(Ax + b) A,$$

formule. Ceci s'obtient facilement, en utilisant la proposition précédente, en prenant la différentielle de $x \in \mathbb{R}^n \mapsto \nabla g(x)$ dont l'expression $\nabla g(x) = A^T \nabla f(Ax + b)$ a été montrée précédemment.

1.4 Formules de Taylor

Les formules de Taylor se généralisent aux fonctions de plusieurs variables. On se limite aux fonctions à valeurs réelles.

Théorème 1.15 (Formules de Taylor pour les fonctions une fois dérivable)**.** Soit $f : \Omega \to \mathbb{R}$ une fonction.

(a) Définition de la différentielle = Formule de Taylor-Young à l'ordre 1 : Si f est différentiable en $x \in \Omega$, alors

$$f(x+h) = f(x) + \langle \nabla f(x),\, h\rangle + \|h\|\varepsilon(h) \quad \text{avec } \lim_{h\to 0} \varepsilon(h) = 0.$$

On considère maintenant un point h fixé tel que le segment $[x, x+h]$ soit inclus dans Ω.

(b) Inégalités des accroissements finis : Si f est continue sur Ω et différentiable sur $]x, x+h[$, alors

$$|f(x+h) - f(x)| \leqslant \sup_{y\in]x,x+h[} \|\nabla f(y)\|\|h\|.$$

(c) Formule de Taylor-Maclaurin : Si f est continue sur Ω et différentiable sur $]x, x+h[$, alors il existe $\theta \in]0, 1[$ tel que

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

(d) Formule de Taylor avec reste intégral : Si $f \in \mathcal{C}^1(\Omega)$ alors

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

Preuve. On applique les formules de Taylor à la fonction $\varphi(t) = f(x+th)$, $t \in [0, 1]$. $\square$

Théorème 1.16 (Formules de Taylor pour les fonctions deux fois dérivable)**.** Soit $f : \Omega \to \mathbb{R}$ une fonction.

(a) Formule de Taylor-Young à l'ordre 2 : Si f est différentiable dans Ω et deux fois différentiable en $x \in \Omega$, alors

$$f(x+h) = f(x) + \langle \nabla f(x),\, h\rangle + \frac{1}{2}\left\langle \nabla^2 f(x)h,\, h\right\rangle + \|h\|^2\varepsilon(h) \quad \text{avec } \lim_{h\to 0} \varepsilon(h) = 0.$$

On considère maintenant un point h fixé tel que le segment $[x, x+h]$ soit inclus dans Ω.

(b) Formule des accroissements finis généralisée : Si $f \in \mathcal{C}^1(\Omega)$ et f est deux fois différentiable sur $]x, x+h[$, alors

$$|f(x+h) - f(x) - \langle \nabla f(x),\, h\rangle\,| \leqslant \frac{1}{2} \sup_{y\in]x,x+h[} \|\nabla^2 f(y)\|_{\mathcal{M}_n(\mathbb{R})}\|h\|^2.$$

où $\|\cdot\|_{\mathcal{M}_n(\mathbb{R})}$ désigne la norme subordonnée des matrices pour la norme euclidienne.

(c) Formule de Taylor-Maclaurin : Si $f \in \mathcal{C}^1(\Omega)$ et f est deux fois différentiable sur $]x, x+h[$, alors il existe $\theta \in]0, 1[$ tel que

$$f(x+h) = f(x) + \langle \nabla f(x),\, h\rangle + \frac{1}{2}\left\langle \nabla^2 f(x+\theta h)h,\, h\right\rangle.$$

(d) Formule de Taylor avec reste intégral : Si $f \in \mathcal{C}^2(\Omega)$ alors

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

Chapitre 2

Problèmes d'optimisation : Existence et unicité des solutions

Une référence pour ce chapitre est le chapitre 8 de [CIARLET].

2.1 Cadre et vocabulaire

Définition 2.1. On appelle *problème d'optimisation* tout problème de la forme

$$\text{Trouver } x^\star \in U \text{ satisfaisant, } f(x^\star) = \inf_{x \in U} f(x), \tag{2.2}$$

où $U \subset \mathbb{R}^n$ et $f : V \to \mathbb{R}$ une fonction avec $U \subset V \subset \mathbb{R}^n$.

Remarque 2.3.

1. Formuler un problème d'optimisation sous la forme d'une minimisation plutôt qu'une maximisation est un choix arbitraire qui dépend d'ailleurs souvent du contexte applicatif. Tous les résultats que nous verrons dans la suite du cours peuvent être facilement adaptés à ce point de vue équivalent car $\sup_{x \in U} f(x) = -\inf_{x \in U} -f(x)$.
2. $\inf_{x \in U} f(x)$ est une notation pour $\inf \{f(x) : x \in U\}$. On note aussi parfois $\inf_U f$ (on omet l'argument de f). Ainsi s'intéresser au problème d'optimisation (2.2), c'est se poser la question si l'ensemble $\{f(x) : x \in U\} \subset \mathbb{R}$ admet une borne inférieure (on dit aussi dans ce cas que f admet une borne inférieure) et si la borne inférieure est atteinte, c'est-à-dire si l'ensemble admet un minimum.

   Lorsque l'on sait que l'inf est atteint, en général on écrit plutôt le problème d'optimisation sous la forme

   $$\text{Trouver } x^\star \in U \text{ satisfaisant, } f(x^\star) = \min_{x \in U} f(x).$$

   Dans le cas où $\{f(x) : x \in U\}$ n'admet pas de borne inférieure (f n'admet pas d'infimum), on écrira parfois que $\inf_U f = -\infty$.
3. Notez qu'en pratique, lorsque l'on parle d'un problème d'optimisation, on omet souvent le *"Trouver $x^\star \in U$"* et on écrit directement

   $$\inf_{x \in U} f(x). \tag{2.4}$$

   On sait alors implicitement que l'on va s'intéresser à l'existence de cette borne inférieure, si elle est atteinte (l'inf est un min), si il y a un unique $x^\star \in U$ tel que $f(x^\star)$ réalise le minimum de f sur U etc...

4. Par abus de langage, on écrit parfois à la place de (2.4) $\min_{x\in U} f(x)$, sans que l'on sache d'avance si la borne inférieure de f existe et est atteinte.
5. Enfin, on réécrit également (2.2) sous la forme

$$\text{Trouver } x^\star \in \underset{x\in U}{\operatorname{argmin}} f(x),$$

où $\operatorname{argmin}_{x\in U} f(x) \overset{\text{def.}}{=} \{\tilde{x} \in U : f(\tilde{x}) = \inf_U f\}$. Remarquez que cet ensemble peut-être vide si l'inf n'est pas atteint ou pire si l'inf n'existe pas.

Définition 2.5. Soit $U \subset V \subset \mathbb{R}^n$ et $f : V \to \mathbb{R}$. Soit le problème d'optimisation

$$\inf_U f.$$

On appelle alors
— f la *fonction objectif* du problème d'optimisation,
— un élement $x^\star \in U$ tel que $f(x^\star) = \inf_{x\in U} f(x)$, *une solution du problème d'optimisation*.

On dira que
— le problème d'optimisation est *sans contraintes* si $U = \mathbb{R}^n$ et *sous contraintes* sinon,
— le problème est *convexe* si f et U sont convexes.

Il est souvent impossible en pratique de trouver explicitement une solution $x^\star$. Un des buts de l'optimisation est alors de proposer des algorithmes dits *itératifs* permettant d'approcher une solution $x^\star$ au sens où, partant d'un vecteur initial $x^{(0)}$ quelconque, on construit explicitement une suite $(x^{(k)})_{k\geqslant 0}$ convergeant vers une solution $x^\star$.

Dans ce cours on s'intéressera à la résolution numérique de problèmes d'optimisation convexes, sans contraintes (et de dimension finie : on considèrera toujours que l'on travaille dans $\mathbb{R}^n$). On étudiera également les propriétés de convergence des algorithmes rencontrés.

Dans ce chapitre, on va énoncer des conditions d'existence et d'unicité de solutions de problèmes d'optimisation (pas forcément sans contraintes ni convexe).

2.2 Généralités sur l'existence de solutions

2.2.1 Existence d'une suite minimisante

Le résultat suivant nous apprend que lorsque qu'une fonction est minorée (on dit aussi *bornée inférieurement* en optimisation) sur un ensemble, alors il est possible d'approcher via une suite d'évaluations de la fonction, son infimum sur l'ensemble.

Proposition 2.6. Soit $f : U \to \mathbb{R}$ une fonction minorée sur $U \subset \mathbb{R}^n$ un ensemble non vide. Alors f admet une borne inférieure sur U et il existe une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de U telle que

$$f(x_n) \underset{n\to+\infty}{\longrightarrow} \inf_U f. \tag{2.7}$$

On dit que $(x_n)_{n\in\mathbb{N}}$ est une suite minimisante.

Démonstration. La preuve est basée sur la propriété de la borne inférieure et la caractérisation de la borne inférieure.

L'ensemble $\{f(x) : x \in U\}$ est un sous ensemble de $\mathbb{R}$ non vide (car U non vide) et minorée (car f minorée). Or $\mathbb{R}$ satisfait la propriété de la borne inférieure, donc l'ensemble $\{f(x) : x \in U\}$ admet une borne inférieure. Cela revient exactement à dire que $\inf_U f \in \mathbb{R}$.

Soit $n \in \mathbb{N}$. Soit $\varepsilon = \frac{1}{n+1} > 0$, alors d'après la caractérisation de la borne inférieure, il existe un élément de l'ensemble $\{f(x) : x \in U\}$, que l'on note $f(x_n)$ avec $x_n \in U$, tel que $\inf_U f \leqslant f(x_n) < \inf_U f + \varepsilon$. On a donc construit une suite $(x_n)_{n\in\mathbb{N}}$ de U, qui par passage à la limite dans l'inégalité précédente vérifie

$$f(x_n) \underset{n\to+\infty}{\longrightarrow} \inf_U f. \tag{2.8}$$

□

Il est important de noter que, a priori, on ne sait rien de la convergence de la suite $(x_n)_{n\in\mathbb{N}}$. Par exemple elle peut très bien être divergente. Le résultat nous dit que c'est *la suite des évaluations de* f, $(f(x_n))_{n\in\mathbb{N}}$, qui converge vers l'infimum de f. De même, on se sait pas si l'infimum est atteint (i.e. l'inf est un min), ni encore moins si $(x_n)_{n\in\mathbb{N}}$ se rapproche d'un minimum. Pour répondre à ces questions, il faut des hypothèses supplémentaires sur la fonction f et sur U.

La section suivante donne des conditions sur f et U qui assurent l'existence d'un minimum.

2.2.2 Coercivité et existence d'une solution

La première question que l'on peut se poser quand on étudie un problème d'optimisation est celle de l'existence d'une solution. Si on cherche à minimiser sur $U \subset \mathbb{R}^n$ une fonction continue $f : \mathbb{R}^n \to \mathbb{R}$, alors il est bien connu que si U est compact (*i.e.* fermé et borné [1]) la fonction f est bornée et atteint ses bornes sur U. Elle admet donc au moins un minimum $x^\star \in U$.

La notion de coercivité permet d'étendre ce résultat au cas des fonctions définies sur un domaine qui est uniquement fermé (donc possiblement non borné).

Définition 2.9 (Fonction coercive)**.** Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dite *coercive* si

$$\lim_{\|x\|\to+\infty} f(x) = +\infty. \tag{2.10}$$

Théorème 2.11. Soit $U \subset \mathbb{R}^n$ non vide et fermée de $\mathbb{R}^n$, et $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue et coercive. Alors le problème d'optimisation

$$\inf_{x\in U} f(x),$$

admet au moins une solution.

Preuve. Soit x_0 un point quelconque de U. Comme f est coercive, il existe $r > 0$ tel que pour tout $x \in U$ vérifiant $\|x\| > r$, on a $f(x) > f(x_0)$. Ainsi

$$\{x \in U : f(x) \leqslant f(x_0)\} \subset U \cap \overline{B}(0, r),$$

1. $U \subset \mathbb{R}^n$ est compact si et seulement si U est fermé et borné dans $\mathbb{R}^n$. La condition est nécessaire si on remplace $\mathbb{R}^n$ par un espace vectoriel normé quelconque, mais la réciproque n'est plus vraie en général. On peut démontrer qu'elle est vraie pour tout U si et seulement si l'espace vectoriel normé est de dimension finie (théorème de Riesz).

où $\overline{B}(0,r)$ est la boule fermée de centre 0 et de rayon r.

Or l'ensemble $U \cap \overline{B}(0,r)$ est fermé (intersection de fermés) et borné donc compact. Puisque f est continue sur $\mathbb{R}^n$, f est bornée et atteint ses bornes sur $U \cap \overline{B}(0,r)$. Il existe donc $x^\star \in U \cap \overline{B}(0,r)$ tel que $f(x^\star) = \min_{U \cap \overline{B}(0,r)} f$, et en particulier $f(x^\star) \leqslant f(x_0)$.

Soit $x \in U$. Si $\|x\| \leqslant r$, alors $f(x) \geqslant f(x^\star)$. Sinon si $\|x\| > r$, alors $f(x) > f(x_0) \geqslant f(x^\star)$. On a donc montré que dans tous les cas, $f(x^\star) \leqslant f(x)$, i.e. $f(x^\star) = \min_U f$. □

Exemple 2.12 (Projection sur un convexe fermé)**.** Soit C un ensemble convexe fermé non vide de $\mathbb{R}^n$. Soit $x_0 \in \mathbb{R}^n$. Considérons le problème qui revient à trouver (si possible) l'élément de C qui est le plus proche (pour la distance euclidienne) de x_0. On le formalise sous la forme du problème d'optimisation suivant

$$\inf_{c \in C} \|x_0 - c\| .$$

Alors ce problème admet au moins une solution $c^\star \in C$ grâce au théorème précédent. En effet la fonction $f : x \in \mathbb{R}^n \mapsto \|x_0 - x\|$ est continue sur $\mathbb{R}^n$ et coercive par l'inégalité triangulaire inverse : $\|x_0 - x\| \geqslant \|x\| - \|x_0\|$.

On peut montrer (voir le cours d'analyse fonctionnelle) que la solution est unique et que c'est en fait la projection orthogonale de x_0 sur C (ne pas hésiter à faire un dessin pour s'en convaincre). En fait, ce résultat se démontre dans le cas plus général où $\mathbb{R}^n$ est remplacé par un espace de Hilbert. C'est un résultat fondamental de l'analyse fonctionnelle. On le retrouve ici manière détournée et élémentaire. Nous n'avons d'ailleurs pas eu besoin de la convexité pour obtenir l'existence. La convexité permet de démontrer l'unicité.

2.3 Extrema locaux et différentiabilité

2.3.1 Définitions

Définition 2.13. Soit $f : U \to \mathbb{R}$, où $U \subset \mathbb{R}^n$.

1. On dit que la fonction f admet en un point $x \in U$ un *minimum local* (respectivement un *maximum local*), s'il existe un $r > 0$ tel que pour tout $y \in U \cap B(x,r)$, $f(y) \geqslant f(x)$ (resp. $f(y) \leqslant f(x)$).
2. On dit que la fonction admet un *extremum local* en x, si elle admet soit un minimum soit un maximum local en x.
3. Enfin, on dit que f admet un minimum (resp. maximum) *global* en x sur U, si pour tout $y \in U$, $f(y) \geqslant f(x)$.

Par abus de langage, on dira que x est un minimum local pour dire que la fonction f admet un minimum local en x (idem pour maximum local, extremum local, minimum / maximum global).

On va maintenant chercher à caractériser les minimums (ou maximums) locaux des fonctions différentiables. Dans toute la suite du chapitre Ω désigne un sous-ensemble ouvert de $\mathbb{R}^n$.

2.3.2 Extrema locaux et condition d'ordre un

On commence par s'intéresser à une condition nécessaire d'extrémalité sur le gradient. Pour cela, on a besoin de la notion de point critique d'une fonction.

Définition 2.14 (Point critique)**.** Soit $f : \Omega \to \mathbb{R}$ une fonction différentiable en $x \in \Omega$. On dira que x est un *point critique* de la fonction f si $\nabla f(x) = 0$.

Alors on a le théorème suivant.

Proposition 2.15 (Condition nécessaire d'extremum local)**.** Soit $f : \Omega \to \mathbb{R}$. Si la fonction f admet un extremum local en $x \in \Omega$ et si elle est différentiable en ce point, alors x est un point critique de f.

Démonstration. Soit $h \in \mathbb{R}^n$. Comme Ω est un ouvert, il existe $\varepsilon > 0$, tel que pour tout $t \in]-\varepsilon, \varepsilon[$, $x+th \in \Omega$. La fonction $\varphi : t \mapsto f(x+th)$ est donc définie sur $]-\varepsilon, \varepsilon[$. Elle est dérivable en $t = 0$, et d'après la formule de dérivation des fonctions composées, $\varphi'(0) = \mathrm{d}f(x)(h) = \langle \nabla f(x),\, h\rangle$. φ ayant un extremum en $t = 0$, on sait que $\varphi'(0) = 0$ (en utilisant le cas bien connu des fonctions réelles de la variable réelle). D'où $\langle \nabla f(x),\, h\rangle = 0$. Ceci est vrai pour tout $h \in \mathbb{R}^n$, donc on a bien $\nabla f(x) = 0$. $\square$

Remarque 2.16.

1. La conclusion du théorème est fausse si Ω n'est pas un ouvert. En effet il suffit de prendre la fonction $x \in [0,1] \mapsto x$, qui admet un minimum global en 0 sans que sa dérivée soit nulle.
2. La réciproque du théorème est fausse en général. Par exemple la fonction $x \in \mathbb{R} \mapsto x^3$ admet un point critique en 0 qui n'est pas un extremum local.

On remarque donc que s'intéresser uniquement aux points d'annulation du gradient n'est pas suffisant pour déterminer la nature des extrema locaux (ni même pour les identifier). Pour cela, il nous faut une information sur la courbure locale de la fonction autour des points critiques. Cette information est contenue dans la différentielle seconde.

2.3.3 Extrema locaux et conditions d'ordre deux

On s'intéresse maintenant aux conditions nécessaires et suffisantes faisant intervenir la matrice hessienne.

Proposition 2.17 (Condition nécessaire de minimum local sur la différentielle seconde)**.** Soit $f : \Omega \to \mathbb{R}$. Si la fonction f admet un minimum local en un point $x \in \Omega$ et si f est deux fois différentiable en x, alors pour tout $h \in \mathbb{R}^n$,

$$\langle \nabla^2 f(x)h, h\rangle \geqslant 0 \quad \text{(ou encore } \mathrm{d}^2 f(x)(h,h) \geqslant 0\text{)},$$

autrement dit la matrice hessienne $\nabla^2 f(x)$ est positive.

Preuve. Soit $h \in \mathbb{R}^n$. Comme x est un minimum local de f, il existe $r > 0$ tel que pour tout $t \in]-r, r[$, $x + th \in \Omega$ et $f(x+th) \geqslant f(x)$. La formule de Taylor-Young donne

$$f(x+th) = f(x) + t\langle \nabla f(x), h\rangle + \frac{t^2}{2}\langle \nabla^2 f(x)h, h\rangle + t^2\|h\|^2\varepsilon(th)$$

avec $\lim\limits_{y\to 0} \varepsilon(y) = 0$. D'après le Proposition 2.15, on a également $\nabla f(x) = 0$. Ainsi,

$$0 \leqslant f(x+th) - f(x) = \frac{t^2}{2}\langle \nabla^2 f(x)h, h\rangle + t^2\|h\|^2\varepsilon(th)$$

En divisant par $\frac{t^2}{2}$, lorsque $t \neq 0$, on déduit que pour tout $t \in]-r, r[$, $t \neq 0$

$$\left\langle \nabla^2 f(x)h,\, h\right\rangle + 2\,\|h\|^2\,\varepsilon(th) \geqslant 0.$$

En faisant tendre t vers 0 on obtient bien que $\langle \nabla^2 f(x)h, h\rangle \geqslant 0$. $\square$

Cependant il est possible que la matrice hessienne soit positive en un point critique sans que ce dernier soit un minimum local. On pensera à l'exemple de la selle de singe $f : (x, y) \in \mathbb{R}^2 \mapsto x^3 - 3xy^2$ en $(0, 0)$ (traité en TD/TP). Il faut donc des conditions plus fortes sur la matrice hessienne.

Théorème 2.18 (Condition suffisante de minimum local sur la différentielle seconde)**.** Soit $f : \Omega \to \mathbb{R}$ une fonction deux fois différentiable en $x \in \Omega$, un point critique de f. Si

$$\forall h \in \mathbb{R}^n \setminus \{0\}, \quad \langle \nabla^2 f(x)h, h\rangle > 0$$

c'est-à-dire si les valeurs propres de $\nabla^2 f(x)$ sont toutes strictement positives, ou dit autrement si $\nabla^2 f(x)$ est une matrice symétrique définie positive, alors la fonction f admet un minimum local en x.

Preuve. Comme $\nabla^2 f(x)$ est définie positive, il existe $\alpha > 0$ (en prenant par exemple la plus petite valeur propre de $\nabla^2 f(x)$) tel que

$$\forall h \in \mathbb{R}^n, \quad \langle \nabla^2 f(x)h, h\rangle \geqslant \alpha \|h\|^2,$$

Il existe $r_0 > 0$ tel que $B(x, r_0) \subset \Omega$. Alors pour tout $h \in B(0, r_0)$, d'après la formule de Taylor-Young on a

$$f(x+h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle \nabla^2 f(x)h, h\rangle + \|h\|^2 \varepsilon(h) \geqslant f(x) + (\frac{1}{2}\alpha - |\varepsilon(h)|)\|h\|^2$$

avec $\lim_{h \to 0} \varepsilon(h) = 0$. Il existe donc $r_1 > 0$, que l'on peut choisir vérifiant $r_1 \leqslant r_0$, et pour tout $h \in B(0, r_1)$, $|\varepsilon(h)| \leqslant \frac{1}{2}\alpha$. Ainsi, pour tout $h \in B(0, r_1)$, on a $f(x+h) \geqslant f(x)$. Donc x est bien un minimum local de f. □

Remarque 2.19. On notera que si $\nabla^2 f(x)$ est définie négative en un point critique x alors x est un maximum local.

Exemple 2.20. La fonction $f : (x, y) \in \mathbb{R}^2 \mapsto x^2 - y^2$ admet un point critique en $(0, 0)$, mais c'est ni un minimum local ni un maximum local. En effet, les valeurs propres de $\nabla^2 f(0, 0)$ sont -2 et 2. En fait c'est un *point col*, aussi appelé *point selle*, car $x \mapsto f(x, 0) = x^2$ admet un minimum local en 0 et $y \mapsto f(0, y) = -y^2$ admet un maximum local en 0. Au voisinage de $(0, 0)$ la fonction ressemble à une selle de cheval.

2.4 Ensembles convexes

On rappelle qu'étant donnés deux vecteurs x et $y \in \mathbb{R}^n$, $[x, y]$ désigne le segment entre x et y, à savoir

$$[x, y] = \{tx + (1-t)y, \ t \in [0, 1]\}.$$

Définition 2.21 (Ensembles convexes)**.** On dira qu'un ensemble $U \subset \mathbb{R}^n$ est *convexe* si

$$\forall x, y \in U, \quad [x, y] \subset U,$$

soit encore si

$$\forall x, y \in U, \forall t \in [0, 1], \quad tx + (1-t)y \in U.$$

(autrement dit U contient tout segment rejoignant n'importe quel couple de ses points).

Exemple 2.22.

- Un sous-espace vectoriel est convexe
- Un hyperplan est convexe.
- La boule unité d'une norme est convexe.
- Toute intersection d'ensembles convexes est convexe.
- Un hyper-rectangle $[a_1, b_1] \times \cdots \times [a_n, b_n]$ est convexe. Plus généralement le produit cartésien $C = C_1 \times \cdots \times C_k$ d'ensembles convexes $C_1 \subset \mathbb{R}^{n_1}, \ldots, C_n \subset \mathbb{R}^{n_k}$ est un ensemble convexe de l'espace produit $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$.
- L'image d'un ensemble convexe par une application linéaire est convexe (voir exercice ci-dessous). En particulier, les translations, rotations, dilatations, projections d'ensembles convexes sont convexes.

Exercice 2.23. Soit $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ une application linéaire.

1. Montrer que si $U \subset \mathbb{R}^n$ est convexe alors l'image directe $W = \varphi(U) = \{\varphi(x),\ x \in U\}$ de U par φ est un ensemble convexe de $\mathbb{R}^m$.
2. Montrer que si $W \subset \mathbb{R}^m$ est un ensemble convexe alors l'image réciproque $U = \varphi^{-1}(W) = \{x \in \mathbb{R}^n,\ \varphi(x) \in W\}$ est un ensemble convexe de $\mathbb{R}^n$.

Solution de l'exercice 2.23.

1. Soit y_1 et $y_1 \in W$. Alors il existe x_1 et x_2 dans U tel que $y_1 = \varphi(x_1)$ et $y_2 = \varphi(x_2)$. Soit $t \in [0, 1]$. Alors, par linéarité,
$$tw_1 + (1-t)w_2 = t\varphi(x_1) + (1-t)\varphi(x_2) = \varphi(tx_1 + (1-t)x_2).$$
Or comme U est convexe, $tx_1 + (1-t)x_2 \in U$, et donc, $ty_1 + (1-t)y_2 \in W = \varphi(U)$. $W = \varphi(U)$ est bien un ensemble convexe.
2. Soit x_1 et $x_2 \in U = \varphi^{-1}(W)$ et $t \in [0, 1]$. Comme W est convexe
$$\varphi(tx_1 + (1-t)x_2) = t\varphi(x_1) + (1-t)\varphi(x_2) \in W.$$
Donc $tx_1 + (1-t)x_2 \in U = \varphi^{-1}(W)$, c'est bien un ensemble convexe.

Nous avons vu que sur $\mathbb{R}^n$ (ou sur un ouvert non vide de $\mathbb{R}^n$), que si un point est un minimul local alors c'est un point d'annulation du gradient (un point critique). Le résultat suivant s'intéresse au même problème, mais où cette fois le point est un minimum local *par rapport* à un sous ensemble convexe.

Théorème 2.24 (Condition nécessaire de minimum local sur un ensemble convexe)**.** Soit $f : \Omega \to \mathbb{R}$ et U une partie convexe de Ω. Si la fonction f est différentiable en un point $x \in U$ et si elle admet en x un minimum local par rapport à l'ensemble U, alors

$$\forall y \in U, \quad \langle \nabla f(x), y - x \rangle \geqslant 0 \quad \text{(ou encore } \mathrm{d}f(x)(y-x) \geqslant 0\text{)}.$$

Preuve. Soit y un point quelconque de l'ensemble U. U étant convexe, les points $x + t(y - x)$, pour $t \in [0, 1]$, sont tous dans U. La différentiabilité de f en x permet d'écrire

$$f(x + t(y-x)) - f(x) = t\langle \nabla f(x), y - x \rangle + t\|y - x\|\varepsilon(t), \tag{2.25}$$

avec $\lim_{t\to 0} \varepsilon(t) = 0$. Pour tout $t > 0$ suffisamment petit, le membre de gauche est positif. Après simplification par t, par passage à la limite on a donc nécessairement $\langle \nabla f(x), y - x \rangle \geqslant 0$. $\square$

On voit donc que dans ce cas de la restriction de la recherche d'un minimum local sur un sous ensemble convexe de $\mathbb{R}^n$, on perd la nullité du gradient. On a seulement une inégalité. Voir la remarque 2.27 ci dessous pour une interprétation. Si l'ensemble convexe U est un sous espace affine, alors cette inégalité prend une forme particulièrement simple, comme le montre le corollaire ci-dessous.

Corollaire 2.26. Dans le théorème précédent, si U est un sous-espace affine de $\mathbb{R}^n$, c'est-à-dire $U = x + F$ avec F un sous-espace vectoriel de V, alors

$$\nabla f(x) \in F^{\perp}.$$

Preuve. Soit $h \in F$. Alors pour $y = x + h \in U$, on a d'après le théorème précédent

$$\langle \nabla f(x),\, h\rangle = \langle \nabla f(x),\, y - x\rangle \geqslant 0.$$

On a aussi, $z = x - h \in U$, comme $-h \in F$, donc $\langle \nabla f(x), -h\rangle \geqslant 0$. Ainsi $\langle \nabla f(x),\, h\rangle = 0$. Comme c'est vrai pour tout $h \in F$, on obtient $\nabla f(x) \in F^{\perp}$. $\square$

Remarque 2.27. L'interprétation géométrique du résultat précédent est très importante. Soit $x \in U$. Définissons H_x le demi espace fermé suivant

$$H_x = \{x + d \in \mathbb{R}^n : d \in \mathbb{R}^n, \langle \nabla f(x),\, d\rangle \geqslant 0\}.$$

Si $x \in U$ est un minimum local de f sur le convexe U tel que $\nabla f(x) \neq 0$, alors $U \subset H_x$.

En effet, la condition $\langle \nabla f(x), y - x\rangle \geqslant 0$, pour tout $y \in U$, signifie que l'angle formé par les vecteurs $\nabla f(x)$ et $y - x$ (qui est un vecteur orienté vers l'intérieur de U) est un angle aigu i.e. $\langle \nabla f(x),\, d\rangle \geqslant 0$. Ceci est tout à fait attendu car le gradient indique une direction d'augmentation (locale) des valeurs prises par la fonction. Ainsi, si l'angle était obtus par rapport à une certaine direction $y - x$ avec $y \in U$, on aurait $\langle \nabla f(x), y - x\rangle < 0$ et donc l'application $t \mapsto f(\underbrace{x + t(y - x)}_{\in U \text{ car } U \text{ convexe}})$ serait strictement décroissante dans un voisinage de 0. On trouverait donc $t_0 \in]0, 1[$ tel que pour tout $t \in]0, t_0[$, $f(x + t(y - x)) < f(x)$, ce qui contredirait la minimalité locale de f en $x \in U$ sur U.

Dans le cas d'un espace affine $U = u + F$, il faut nécessairement que l'angle soit droit car sinon on peut toujours réussir à se déplacer dans une direction de décroissance de f, comme F est stable par passage à l'opposée.

Dans ce cours on ne considérera pas de problème de minimisation sous contrainte convexe générale. En revanche, on sera souvent amené à minimiser des fonctionnelles sur des sous-espaces affines, et en premier lieu des droites. On retrouvera donc souvent cette condition d'orthogonalité.

Remarque 2.28. Supposons que x est un minimum local de f sur un ouvert Ω de $\mathbb{R}^n$. On déjà vu que x est un point critique de f, i.e. $\nabla f(x) = 0$. Redémontrons ce résultat grâce au corollaire précédent.

Comme Ω ouvert, il existe $\varepsilon_0 > 0$, tel que $B(x, \varepsilon_0) \subset \Omega$. Soit $h \in \mathbb{R}^n$, alors comme x est un minimum local de f sur Ω, c'est aussi un minimum local de f par rapport à l'ensemble convexe $(x + \mathrm{Vect}(h)) \cap B(x, \varepsilon) \subset \Omega$. Ainsi par le corollaire, on a $\nabla f(x) \in \mathrm{Vect}(h)^{\perp}$ (le fait que l'on ait intersecté le sous espace affine de $\mathbb{R}^n$: $x + \mathrm{Vect}(h)$ avec un voisinage ouvert de x ne change pas le résultat du corollaire (adapter la preuve en exercice)). Comme h est quelconque dans $\mathbb{R}^n$, on a donc $\nabla f(x) \in (\mathbb{R}^n)^{\perp} = \{0\}$.

On remarque donc que le fait d'être minimum local par rapport à un ouvert force le gradient à valoir 0. En effet il doit pointer vers l'extérieur de toute droite passant par le minimum local : il n'y a pas d'autre échappatoire que d'être le vecteur nul.

2.5 Fonctions convexes

2.5.1 Définition et exemples

Définition 2.29 (Fonctions convexes)**.** Soit $U \subset \mathbb{R}^n$ un ensemble convexe. Soit $f : U \to \mathbb{R}$ une fonction à valeurs réelles.

— f est *convexe* si

$$\forall x, y \in U,\ \forall t \in [0,1], \quad f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y).$$

— f est *strictement convexe* si

$$\forall x, y \in U,\ x \neq y,\ \forall t \in]0,1[, \quad f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

Une fonction f est (strictement) *concave* si son opposée $x \mapsto -f(x)$ est (strictement) convexe.

Remarque 2.30. On peut également restreindre t à $]0,1[$ pour la définition de la convexité.

Exemple 2.31. Voici quelques exemples de fonctions convexes :

— Sur $\mathbb{R}$, la fonction $x \mapsto x^2$ est strictement convexe.
— Sur $\mathbb{R}$, la fonction $x \mapsto |x|$ est convexe mais pas strictement convexe.
— De même, sur $\mathbb{R}^n$, la fonction $x \mapsto \|x\|^2$ est strictement convexe, la fonction $x \mapsto \|x\|_1$ est convexe mais pas strictement convexe.
— Le sup d'une famille quelconque de fonctions convexes est convexe.
— La composée d'une fonction affine et d'une fonction convexe est convexe (voir ci-dessous) : pour $A \in M_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$, et $f : \mathbb{R}^m \to \mathbb{R}$ convexe alors la fonction $g : \mathbb{R}^n \to \mathbb{R}$ définie par $g(x) = f(Ax + b)$ est convexe.

Définition 2.32. Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$. On appelle ensemble de sous-niveau α de f l'ensemble

$$C_\alpha = \{x \in U,\ f(x) \leqslant \alpha\}.$$

Proposition 2.33. Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \to \mathbb{R}$ une fonction convexe sur U. Alors pour tout $\alpha \in \mathbb{R}$, l'ensemble C_α est convexe. En particulier, l'ensemble des minimums globaux de f est un ensemble convexe (qui peut-être vide).

Preuve. Soit $\alpha \in \mathbb{R}$. Soit x_1 et x_2 dans C_α et $t \in [0,1]$. Alors, comme f est convexe,

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2) \leqslant t\alpha + (1-t)\alpha = \alpha$$

donc $tx_1 + (1-t)x_2 \in C_\alpha$ ce qui prouve bien que C_α est convexe. L'ensemble des minimums globaux de f n'est autre que l'ensemble de sous-niveau $\alpha^* = \min_{x \in U} f(x)$ de f, et il est donc bien convexe. □

Remarque 2.34. La réciproque de la proposition précédente est fausse. Il existe des fonctions non convexes dont tous les ensembles de sous-niveaux sont convexes, comme par exemple $x \mapsto -e^{-x}$ sur $\mathbb{R}$.

Un autre exemple : sur $\mathbb{R}$, tous les ensembles de niveaux de la fonction $\mathbb{1}_{[0,+\infty[}$ sont convexes mais cette fonction n'est pas convexe. En effet,

$$C_\alpha = \{x \in \mathbb{R},\ \mathbb{1}_{[0,+\infty[}(x) \leqslant \alpha\} = \begin{cases} \emptyset & \text{si } \alpha < 0, \\ ]-\infty, 0[& \text{si } \alpha \in [0,1[, \\ \mathbb{R} & \text{si } \alpha > 1, \end{cases}$$

qui sont bien tous des ensembles convexes. En revanche la fonction n'est pas convexe car par exemple

$$1 = f(0) > \frac{1}{2}f(1) + \frac{1}{2}f(-1) = \frac{1}{2}.$$

2.5.2 Caractérisation des fonctions convexes différentiables

Le théorème ci-dessous exprime le fait qu'une fonction différentiable est convexe si et seulement si son graphe est au-dessus de chacun des ses plans tangents. Ce résultat est fondamental !

Théorème 2.35 (Convexité et différentiabilité)**.** Soit $f : \Omega \to \mathbb{R}$ une fonction différentiable sur l'ouvert Ω et soit $U \subset \Omega$ un sous-ensemble convexe.

a) La fonction f est convexe sur U si et seulement si pour tout $x, y \in U$,

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle \quad \text{(ou encore } f(y) \geqslant f(x) + \mathrm{d}f(x)(y - x)\text{)}.$$

b) La fonction f est strictement convexe sur U si et seulement si pour tout $x, y \in U$, $x \neq y$,

$$f(y) > f(x) + \langle \nabla f(x), y - x\rangle \quad \text{(ou encore } f(y) > f(x) + \mathrm{d}f(x)(y - x)\text{)}.$$

Preuve. a) Faisons le sens direct. Soit x, y deux points distincts de U et $t \in]0, 1[$. Comme f est convexe,

$$f((1 - t)x + ty) \leqslant (1 - t)f(x) + tf(y)$$

et on a donc

$$\frac{f(x + t(y - x)) - f(x)}{t} \leqslant f(y) - f(x).$$

En passant à la limite $t \to 0$ on a

$$\langle \nabla f(x), y - x\rangle = \lim_{t\to 0} \frac{f(x + t(y - x)) - f(x)}{t} \leqslant f(y) - f(x).$$

Réciproquement supposons que pour tout $x, y \in U$,

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle.$$

Soit x, y deux points distincts de U et $t \in]0, 1[$. En appliquant l'inégalité aux deux couples $(tx + (1 - t)y, y)$ et $(tx + (1 - t)y, x)$ on a

$$f(y) \geqslant f(tx + (1 - t)y) + \langle \nabla f(tx + (1 - t)y), t(y - x)\rangle,$$

et

$$f(x) \geqslant f(tx + (1 - t)y) + \langle \nabla f(tx + (1 - t)y), (1 - t)(x - y)\rangle.$$

En multipliant par $(1 - t)$ et t ces deux inégalités, on obtient en les sommant

$$tf(x) + (1 - t)f(y) \geqslant f(tx + (1 - t)y),$$

donc f est bien convexe.

b) La preuve de l'implication indirecte est identique en remplaçant les inégalités larges par des inégalités strictes. En revanche pour l'implication directe, le passage à la limite change les inégalités strictes en inégalités larges, donc on ne peut pas conclure aussi rapidement.

Soit t, t' tels que $0 < t < t' < 1$. Alors, comme f est strictement convexe et $(1-t)x + ty \in [x, (1-t')x + t'y]$ avec

$$(1-t)x + ty = (1 - \frac{t}{t'})x + \frac{t}{t'}((1-t')x + t'y),$$

on a

$$f((1-t)x + ty) < (1 - \frac{t}{t'})f(x) + \frac{t}{t'}f((1-t')x + t'y).$$

D'où,

$$\frac{f(x + t(y-x)) - f(x)}{t} < \frac{f(x + t'(y-x)) - f(x)}{t'} < f(y) - f(x).$$

La dernière inégalité étant obtenu en appliquant cette fois la strict convexité de f sur $[x, y]$ pour le point $(1-t')x + t'y$. En passant à la limite $t \to 0$, on a alors

$$\langle \nabla f(x), y - x \rangle \leqslant \frac{f(x + t'(y-x)) - f(x)}{t'} < f(y) - f(x).$$

□

2.5.3 Caractérisation des fonctions convexes deux fois différentiables

Théorème 2.36 (Convexité et différentiabilité seconde)**.** Soit $f : \Omega \to \mathbb{R}$ une fonction deux fois différentiable sur $\Omega \subset \mathbb{R}^n$ un ouvert, et soit $U \subset \Omega$ un sous-ensemble convexe.

a) La fonction f est convexe sur U si et seulement si pour tout $x, y \in U$,

$$\langle \nabla^2 f(x)(y-x), y-x \rangle \geqslant 0$$

b) Si pour tout $x, y \in U$, $x \neq y$,

$$\langle \nabla^2 f(x)(y-x), y-x \rangle > 0$$

alors f est strictement convexe sur U.

En particulier, si $\Omega = U$ est un ouvert convexe, alors

a) f est convexe sur Ω si et seulement si pour tout $x \in \Omega$ la matrice hessienne $\nabla^2 f(x)$ est positive.

b) Si pour tout $x \in \Omega$ la matrice hessienne $\nabla^2 f(x)$ est définie positive, alors f est strictement convexe sur Ω.

Preuve. Soit x, y deux points distincts de U. Alors, comme f est deux fois différentiable, d'après la formule de Taylor-Maclaurin il existe $z \in]x, y[$ tel que

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\langle \nabla^2 f(z)(y-x), y-x \rangle.$$

Mais $z \in]x, y[$, donc il existe $t \in]0, 1[$ tel que $z = (1-t)x + ty$, soit $z - x = t(y - x)$. Ainsi

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2t^2} \underbrace{\langle \nabla^2 f(z)(z-x), (z-x)\rangle}_{=\langle \nabla^2 f(z)(x-z),(x-z)\rangle}.$$

Si par hypothèse $\langle \nabla^2 f(z)(x-z), (x-z)\rangle$ est positif (resp. strictement positif) on déduit du théorème de caractérisation des fonctions convexes différentiables que f est convexe (resp. strictement convexe).

Il reste à montrer que si f est convexe alors pour tout $x, y \in U$, $\langle \nabla^2 f(x)(y-x), y-x\rangle \geqslant 0$. Soit $x, y \in U$. En appliquant la formule de Taylor-Young en x pour l'accroissement $t(y-x)$ (avec $t \in]0, 1]$),

$$f(x + t(y-x)) = f(x) + t\langle \nabla f(x), y - x\rangle + \frac{t^2}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle + t^2\|y-x\|^2\varepsilon(t)$$

avec $\lim_{t\to 0} \varepsilon(t) = 0$. Donc

$$0 \leqslant f(x + t(y-x)) - f(x) - t\langle \nabla f(x), y-x\rangle = \frac{t^2}{2}\left(\langle \nabla^2 f(x)(y-x), y-x\rangle + 2\|y-x\|^2\varepsilon(t)\right)$$

et on en déduit que $\langle \nabla^2 f(x)(y-x), y-x\rangle \geqslant 0$ par passage à la limite, après avoir simplifié par t^2.

Le cas où $U = \Omega$ est une conséquence directe de la réécriture des inégalités où cette fois $y - x$ peut décrire toutes les directions de $\mathbb{R}^n$. On peut aussi le démontrer rapidement en étudiant la différence entre f et ses approximations au premier ordre. En effet, pour $x \in \Omega$,

$$g(y) = f(y) - f(x) - \langle \nabla f(x), y - x\rangle$$

est une fonction convexe (en tant que somme de fonctions convexes), deux fois différentiable et telle que $\nabla^2 f(y) = \nabla^2 g(y)$. Comme $g(y) \geqslant 0$ et que $g(x) = 0$, x est un minimum global de f et donc nécessairement pour tout $h \in \mathbb{R}^n$, $\langle \nabla^2 f(x)h, h\rangle \geqslant 0$. $\square$

2.5.4 Problèmes d'optimisation convexes

On rappelle qu'un problème d'optimisation

$$\inf_{x\in U} f(x),$$

est dit *convexe* si U et f sont convexes. Vis-à-vis de l'optimisation, la convexité joue un rôle crucial puisqu'elle permet d'assurer qu'un minimum local est en fait un minimum global, comme précisé par le résultat suivant.

Théorème 2.37 (Minimum de fonctions convexes)**.**

a) Si une fonction convexe $f : U \to \mathbb{R}$ admet un minimum local en un point $x \in U$, elle y admet en fait un minimum global sur U.

b) Si une fonction $f : U \to \mathbb{R}$ strictement convexe admet un minimum local en un point $x \in U$, alors x est l'unique minimum global de f sur U.

c) Soit $f : \Omega \to \mathbb{R}$ une fonction différentiable sur un ouvert Ω et convexe sur un ensemble convexe $U \subset \Omega$. Soit $x \in U$.
Alors x est un minimum global de f sur U si et seulement si,
$$\forall y \in U, \quad \langle \nabla f(x), y - x \rangle \geqslant 0.$$
d) Soit $f : \Omega \to \mathbb{R}$ une fonction convexe différentiable sur $\Omega \subset \mathbb{R}^n$ ouvert convexe. Soit $x \in \Omega$. Alors un x est un minimum global de f si et seulement si x est un point critique de f.

Preuve.

a) Soit $y \in U$. Comme précédemment, la convexité entraîne que pour tout $t \in]0, 1[$,
$$x + t(y - x) \in U \text{ et } f(y) - f(x) \geqslant \frac{f(x + t(y - x)) - f(x)}{t}.$$
Comme x est un minimum local, il existe un t_0 assez petit tel que $f(x+t_0(y-x))-f(x) \geqslant 0$. Mais alors, $f(y) - f(x) \geqslant 0$, donc x est bien un minimum global.

b) Si f est strictement convexe et que x est un minimum local de f, alors d'après la précédente question, x est un minimum global de f. On montre qu'il est nécessairement unique. Soit $x' \in U$ tel que $f(x') = f(x)$ (i.e. x' est un autre minimum global de f sur U). Supposons par l'absurde que $x \neq x'$. Alors par strict convexité de f, pour $t \in]0, 1[$, $f((1-t)x + tx') < (1-t)f(x) + tf(x') = \min_U f$. Mais on doit forcément avoir $f((1-t)x + tx') \geqslant \min_U f$. C'est donc absurde. D'où $x' = x$.

c) La condition est nécessaire d'après le Théorème 2.24. Réciproquement, d'après le Théorème 2.35, pour tout $y \in U$, $f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle \geqslant f(x)$ donc x est bien un minimum global sur U.

d) On sait que $\nabla f(x) = 0$ est une condition nécessaire pour être un minimum global. Montrons que c'est une condition suffisante si f est convexe. D'après le Théorème 2.35, si $x \in \Omega$ est tel que $\nabla f(x) = 0$, alors pour tout $y \in \Omega$, $f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle = f(x)$, donc x est bien un minimum global.

□

Remarque 2.38. — Une fonction non strictement convexe peut admettre plusieurs minimums locaux. Cependant, comme on l'a vu, l'ensemble des minimums globaux forme un ensemble convexe.
— On a aussi démontré dans ce résultat qu'une fonction strictement convexe admet au plus un minimum global : si elle admet un minimum local alors il est unique et est global, mais la fonction peut aussi ne pas en admettre. Il existe des fonctions strictement convexe qui n'ont pas de point critique (par exemple $x \in \mathbb{R}_+^* \mapsto -\ln(x)$).
— Le théorème précédent est fondamental pour la suite de ce cours. Sauf exception, en pratique on ne s'intéressera qu'à des problèmes d'optimisation convexes.

2.6 Etude des fonctionnelles quadratiques

Définition 2.39. On appelle *fonctionnelle quadratique* toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ de la forme
$$f(x) = \frac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle + c,$$
où $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice carré symétrique, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

La théorème suivante résume les propriétés des fonctionnelles quadratiques.

Théorème 2.40 (Propriétés des fonctionnelles quadratiques)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonctionnelle quadratique, donc de la forme $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c$, avec A symétrique. Alors,

a) f est $\mathcal{C}^2$ sur $\mathbb{R}^n$ (et même $\mathcal{C}^\infty$).

b) Pour tout $x \in \mathbb{R}^n$,

$$\nabla f(x) = Ax - b \quad \text{et} \quad \nabla^2 f(x) = A.$$

c) f est convexe si et seulement si A est positive.

d) f est strictement convexe si et seulement si A est définie positive.

e) $\inf\limits_{x \in \mathbb{R}^n} f(x)$ est fini (f admet un infimum sur $\mathbb{R}^n$) si et seulement si A est positive et le système linéaire $Ax = b$ admet (au moins) une solution.

   Dans le cas où l'une des deux assertions équivalentes est satisfaite, alors l'ensemble des solutions de $Ax = b$ est l'ensemble des minimums globaux de f.

Démonstration. a) f est polynomiale en les coefficients de $x \in \mathbb{R}^n$ donc f est $\mathcal{C}^2$ sur $\mathbb{R}^n$ (et même $\mathcal{C}^\infty$).

b) On écrit d'abord la définition de la différentiabilité pour f en $x \in \mathbb{R}^n$, pour tout $h \in \mathbb{R}^n$ on a

$$\begin{aligned} f(x+h) &= f(x) + \frac{1}{2}\langle Ax,\, h\rangle + \frac{1}{2}\langle Ah,\, x\rangle - \langle b,\, h\rangle + \frac{1}{2}\langle Ah,\, h\rangle\,, \\ &= f(x) + \langle Ax - b,\, h\rangle + o(\|h\|), \end{aligned}$$

   où on a utilisé que A est symétrique. Donc $\nabla f(x) = Ax - b$. Enfin, la différentielle de ∇f en x est la matrice A donc $\nabla^2 f(x) = A$.

c) Comme f est $\mathcal{C}^2$ sur $\mathbb{R}^n$, d'après le théorème 2.36, f est convexe sur $\mathbb{R}^n$ si et seulement si $\nabla^2 f(x) = A$ est positive pour tout $x \in \mathbb{R}^n$.

d) De nouveau grâce au théorème 2.36, si A est définie positive alors f est strictement convexe. Par contre nous avons vu qu'en général la réciproque est fausse (exemple $x \mapsto x^4$). Cependant pour les fonctionnelles quadratiques, elle est vraie. En effet d'après le théorème 2.35, f est strictement convexe est équivalent à pour tout $x, y \in \mathbb{R}^n$, $x \neq y$, $f(y) > f(x) + \langle \nabla f(x),\, y - x\rangle$. En prenant $x = 0$, on obtient directement $\langle Ay,\, y\rangle > 0$. D'où A définie positive.

e) Commençons par le sens indirect. Le fait que A soit positive signifie par c) que f est convexe et le fait que $Ax = b$ admet une solution $x^\star \in \mathbb{R}^n$ implique $\nabla f(x^\star) = 0$ par b), c'est-à-dire $x^\star$ est un point critique de f. Par le théorème 2.37, $x^\star$ est un minimum global de f d'où $\inf_{\mathbb{R}^n} f = f(x^\star) > -\infty$.

   Pour le sens direct, montrons tout d'abord que A est nécessairement positive. Supposons par l'absurde qu'il existe une valeur propre $\lambda < 0$. Soit $v \in \mathbb{R}^n \setminus \{0\}$ tel que $Av = \lambda v$. Pour tout $t \in \mathbb{R}$, on a alors

$$f(tv) = \frac{t^2}{2}\lambda \|v\|^2 - t\,\langle b,\, v\rangle + c,$$

   et comme $\|v\|^2 \neq 0$, on en déduit $\lim\limits_{t \to +\infty} f(tv) = -\infty$, ce qui contredit que $\inf_{\mathbb{R}^n} f > -\infty$. Donc toutes les valeurs propres de A sont positives. Comme A est symétrique, cela implique

que A est une matrice positive. Montrons maintenant que $Ax = b$ admet au moins une solution. Raisonnons de nouveau par l'absurde. Cela signifie que $b \notin \operatorname{Im} A$. Comme A est symétrique, $\mathbb{R}^n$ se décompose suivant la somme directe orthogonale suivante

$$\mathbb{R}^n = \operatorname{Ker} A \overset{\perp}{\oplus} \operatorname{Im} A.$$

On a donc $b = b_1 + b_2$ avec $b_1 \in \operatorname{Ker} A$, $b_2 \in \operatorname{Im} A$ et $\langle b_1, b_2 \rangle = 0$. On a forcément $b_1 \neq 0$ car sinon $b = b_2 \in \operatorname{Im} A$, d'où pour tout $t \in \mathbb{R}$

$$f(tb_1) = \frac{1}{2} t^2 \left\langle \underbrace{Ab_1}_{=0}, b_1 \right\rangle - t \langle b, b_1 \rangle + c = -t \|b_1\|^2 + c,$$

ce qui implique $\lim\limits_{t \to +\infty} f(tb_1) = -\infty$ car $\|b_1\|^2 \neq 0$. Ceci contredit $\inf_{\mathbb{R}^n} f > -\infty$. D'où $Ax = b$ admet au moins une solution.

On a donc démontré que $\inf_{\mathbb{R}^n} f > -\infty$ est équivalent à A positive et $Ax = b$ admet au moins une solution.

On a aussi vu que l'ensemble des points critiques de f est l'ensemble des solutions du système linéaire $Ax = b$. Ainsi, si l'une des deux conditions équivalentes est satisfaite, on déduit que f est convexe puisque A est positive. Or on a vu (Théorème 2.37) que l'ensemble des points critiques d'une fonction convexe est l'ensemble de ses minimiseurs globaux. Par conséquent on déduit bien finalement que l'ensemble des solutions du système linéaire $Ax = b$ est non vide et correspond aux minimiseurs globaux de f.

$\square$

Remarque 2.41. Le théorème précédent montre en particulier que résoudre le problème d'optimisation associé à f revient à résoudre le système linéaire $Ax = b$.

Chapitre 3

Algorithmes de descente de gradient pour les problèmes sans contraintes

La référence principale pour ce chapitre est le chapitre 9 de [BOYD & VANDENBERGHE]. Mais il est surtout largement inspiré des notes du cours d'optimisation de Quentin Mérigot `http://quentin.mrgt.fr/cours/m315/`, que je remercie !

On souhaite résoudre de manière approchée le problème d'optimisation $\min_{\mathbb{R}^n} f$, où $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction $\mathcal{C}^2$, convexe et admettant un unique minimum global. Pour cela, nous allons construire une suite $(x^{(k)})_{k\in\mathbb{N}}$ de $\mathbb{R}^n$, via des algorithmes dit de descente, qui va converger, modulo des hypothèses additionnelles, vers le minimum global de f.

Dans ce chapitre nous énoncerons tout d'abord quelques généralités sur les algorithmes dit de descente. Puis nous étudierons en détail l'algorithme de descente de gradient à pas optimal. Et enfin nous verrons l'algorithme de descente de gradient préconditionné à rebroussement d'Armijo, qui fournit une réponse aux limitations du précédent.

3.1 Méthode de descente

Définition 3.1 (Direction de descente)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$. On dira que $d \in \mathbb{R}^n$ est une direction de descente pour f en $x \in \mathbb{R}^n$ si il existe $t_0 > 0$ tel que pour tout $t \in]0, t_0]$, $f(x + td) < f(x)$.

Remarque 3.2. Une direction de descente pour f en x est donc une direction dans laquelle on diminue strictement la fonction localement.

Définition 3.3 (Méthode de descente)**.** On appelle *méthode de descente* pour f, tout algorithme qui définit des suites $(d^{(k)})_{k\in\mathbb{N}}$, $(t^{(k)})_{k\in\mathbb{N}}$, $(x^{(k)})_{k\in\mathbb{N}}$ où pour tout $k \in \mathbb{N}$,

$$\begin{aligned}
&d^{(k)} \in \mathbb{R}^n \text{ est une direction de descente pour } f \text{ en } x^{(k)},\\
&t^{(k)} \in \mathbb{R} \text{ est appelé pas de descente},\\
&x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)},
\end{aligned}$$

et tel que $f(x^{(k+1)}) \leqslant f(x^{(k)})$ (avec égalité seulement si $x^{(k)}$ est un minimum de f auquel cas la suite $(x^{(k)})_{k\in\mathbb{N}}$ devient stationnaire).

Une méthode de descente est donc entièrement caractérisée lorsque des procédés pour déterminer les directions de descente successives et les pas de descente associés ont été choisis.

Lorsque f est différentiable, les directions de descente sont caractérisées simplement à partir du gradient de f, d'après la proposition suivante.

Proposition 3.4. Soit $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$. Supposons f est différentiable en x. Si $\langle \nabla f(x),\, d\rangle < 0$, alors d est une direction de descente pour f en x.

Démonstration. Comme f est différentiable en x, on a pour tout $t \in \mathbb{R}$, $f(x+td) = f(x) + t\,\langle \nabla f(x),\, d\rangle + o(t)$. Par conséquent si $\langle \nabla f(x),\, d\rangle < 0$, alors nécessairement pour tout t suffisamment petit, on a $f(x+td) < f(x)$, i.e. d est une direction de descente (pour f en x). $\square$

Une direction de descente classique, et que l'on considèrera dans tout ce chapitre, est la direction donné par l'opposé du gradient.

Proposition 3.5. Soit $f : \mathbb{R}^n \to \mathbb{R}$ et $x \in \mathbb{R}^n$. Supposons f est différentiable en x, et $\nabla f(x) \neq 0$. Alors $-\nabla f(x)$ est une direction de descente pour f en x.

Démonstration. Pour $d = -\nabla f(x)$, on a $\langle \nabla f(x),\, d\rangle = -\|\nabla f(x)\|^2 < 0$, d'où d'après Proposition 3.5, $-\nabla f(x)$ est une direction de descente pour f en x. $\square$

Remarque 3.6. Si x n'est pas un point critique pour f, alors la direction $-\frac{\nabla f(x)}{\|\nabla f(x)\|}$ est la direction qui minimise sur la sphère unité de $\mathbb{R}^n$ l'application $v \mapsto \langle \nabla f(x),\, v\rangle$. C'est donc la direction de plus forte pente pour la norme euclidienne. Si l'on considère pour ce critère une autre norme sur $\mathbb{R}^n$ que la norme euclidienne, alors la direction de plus forte pente sur la sphère associée sera différente.

Comme nous allons le voir dans la suite, un algorithme de descente de gradient est tout simplement une méthode de descente pour laquelle la directe de descente est donnée par l'opposé du gradient.

3.2 Algorithme de descente de gradient à pas optimal

3.2.1 Définition de l'algorithme et premières propriétés

Définition 3.7 (Gradient à pas optimal)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur $\mathbb{R}^n$. L'algorithme de descente de gradient à pas optimal est donné pour tout $k \in \mathbb{N}$ par

$$\begin{aligned}
&d^{(k)} = -\nabla f(x^{(k)}),\\
&t^{(k)} \in \operatorname*{argmin}_{t\in\mathbb{R}} f(x^{(k)} + td^{(k)}),\\
&x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}.
\end{aligned}$$

pourvu que pour tout $k \in \mathbb{N}$, l'argmin soit non vide. On dit alors que $t^{(k)}$ est un pas de descente optimal.

Remarque 3.8. On est bien en présence d'une méthode de descente comme donnée à la Définition 3.3.

Soit $(x^{(k)})_{k\in\mathbb{N}}$ une suite des itérés produits par l'algorithme de descente de gradient à pas optimal. Une des particularités de cet algorithme réside dans le fait que les gradients successifs en les $x^{(k)}$ sont orthogonaux

$$\forall k \in \mathbb{N}, \quad \nabla f(x^{(k)}) \perp \nabla f(x^{(k+1)}),$$

ou de manière équivalente

$$\forall k \in \mathbb{N}, \quad (x^{(k+1)} - x^{(k)}) \perp (x^{(k+2)} - x^{(k+1)}),$$

puisque $x^{(k+1)} - x^{(k)} = -t^{(k)}\nabla f(x^{(k)})$, pour tout k. En effet, si on note $g : t \mapsto f(x^{(k)} + td^{(k)})$, alors par définition de $t^{(k)}$, on déduit $g'(t^{(k)}) = 0$. Or $g'(t) = \left\langle \nabla f(x^{(k)} + td^{(k)}),\, d^{(k)} \right\rangle = -\left\langle \nabla f(x^{(k)} + td^{(k)}),\, \nabla f(x^{(k)}) \right\rangle$.

La première question qui se pose naturellement est celle de l'existence d'un pas optimal. Dans le cas où f est une fonctionnelle quadratique, la réponse est positive. Ce pas est même unique et a une expression simple.

Proposition 3.9. Soit f une fonctionnelle quadratique strictement convexe, dont l'expression est donnée pour tout $x \in \mathbb{R}^n$ par : $f(x) = \frac{1}{2}\langle Ax,\, x\rangle - \langle b,\, x\rangle + c$. Alors l'algorithme de descente de gradient à pas optimal est bien définie pour f et de plus il s'écrit pour tout $k \in \mathbb{N}$

$$\begin{aligned}
d^{(k)} &= -(Ax^{(k)} - b),\\
t^{(k)} &= \frac{\left\langle d^{(k)},\, d^{(k)} \right\rangle}{\left\langle Ad^{(k)},\, d^{(k)} \right\rangle}, \text{ si } d^{(k)} \neq 0,\\
x^{(k+1)} &= x^{(k)} + t^{(k)}d^{(k)}.
\end{aligned}$$

Démonstration. Soit $k \in \mathbb{N}$. Supposons construits jusqu'au rang k les élements $d^{(0)}, \ldots, d^{(k)}$, $t^{(0)}, \ldots, t^{(k-1)}$ et $x^{(0)}, \ldots, x^{(k)}$.

Comme f est une fonctionnelle quadratique, on sait que $d^{(k)} = -\nabla f(x^{(k)}) = -(Ax^{(k)} - b)$. Montrons qu'il est possible de trouver un pas de descente optimal $t^{(k)} \in \operatorname{argmin}_{t\in\mathbb{R}} f(x^{(k)} + td^{(k)})$. Dans le situation où $d^{(k)} = 0$, alors $x^{(k)}$ est le minimum global de f et donc on peut prendre n'importe quelle valeur pour $t^{(k)}$ et la suite des itérées $x^{(0)}, \ldots, x^{(k)}, x^{(k+1)}, \ldots$ est stationnaire à partir (au moins) du rang k. Supposons donc $d^{(k)} \neq 0$, montrons alors qu'il y a un unique pas optimal et qu'il satisfait l'expression de l'énoncé. Soit $g : t \in \mathbb{R} \mapsto f(x^{(k)} + td^{(k)})$, alors g est $\mathcal{C}^2$ sur $\mathbb{R}$ et est strictement convexe puisque f l'est. Pour tout $t \in \mathbb{R}$, $g'(t) = \left\langle \nabla f(x^{(k)} + td^{(k)}),\, d^{(k)} \right\rangle = \left\langle Ax^{(k)} + tAd^{(k)} - b,\, d^{(k)} \right\rangle = -\left\langle d^{(k)},\, d^{(k)} \right\rangle + t\left\langle Ad^{(k)},\, d^{(k)} \right\rangle$. D'où g admet un unique point critique $\frac{\left\langle d^{(k)},\, d^{(k)} \right\rangle}{\left\langle Ad^{(k)},\, d^{(k)} \right\rangle}$ et comme g est strictement convexe, c'est son unique minimum global. On a donc bien $t^{(k)} = \frac{\left\langle d^{(k)},\, d^{(k)} \right\rangle}{\left\langle Ad^{(k)},\, d^{(k)} \right\rangle}$. $\square$

Remarque 3.10. Dans la précédente proposition 3.9, comme f est strictement convexe, A est définie positive et donc pour tout $k \in \mathbb{N}$, $t^{(k)} > 0$.

En général l'existence et l'unicité de ce pas optimal ne sont pas assurés.

3.2.2 Forte convexité

Pour étudier la convergence des algorithmes de descente de gradient, une notion importante dont on a besoin est celle de la forte convexité.

Définition 3.11 (Fonction fortement convexe)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$, $C \subset \mathbb{R}^n$ un convexe et $m > 0$. On dira que f est m-fortement convexe sur C si la fonction $f - \frac{m}{2}\|\cdot\|^2$ est convexe sur C.

Cette notion de convexité est plus forte que celle de convexité ou de strict convexité, puisque une fonction m-fortement convexe est nécessairement strictement convexe.

Proposition 3.12. Soit $f : \mathbb{R}^n \to \mathbb{R}$, $C \subset \mathbb{R}^n$ un convexe et $m > 0$. On suppose que f est m-fortement convexe sur C. Alors f est strictement convexe sur C.

Démonstration. Soit $x, y \in C$ et $t \in]0,1[$. Il s'agit de démontrer que $f((1-t)x+ty) < (1-t)f(x)+tf(y)$. Or on sait que $g = f - \frac{m}{2}\|\cdot\|^2$ est convexe sur C, donc

$$f((1-t)x+ty) - \frac{m}{2}\|(1-t)x+ty\|^2 \leqslant (1-t)f(x)+tf(y) - \frac{m}{2}(1-t)\|x\|^2 - \frac{m}{2}\|y\|^2,$$

et comme $\frac{m}{2}\|\cdot\|^2$ est strictement convexe, on a $\frac{m}{2}\|(1-t)x+ty\|^2 - \frac{m}{2}(1-t)\|x\|^2 - \frac{m}{2}\|y\|^2 < 0$, d'où

$$f((1-t)x+ty) < (1-t)f(x)+tf(y).$$

□

Il existe différentes manières de caractériser les fonctions fortement convexes, comme le montre la proposition suivante.

Proposition 3.13. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^1$, $C \subset \mathbb{R}^n$ un convexe et $m > 0$. Alors on a les équivalences suivantes :

1. f est m-fortement convexe sur C,
2. pour tous $x, y \in C$, $f(y) \geqslant f(x) + \langle \nabla f(x),\, y-x\rangle + \frac{m}{2}\|y-x\|^2$,
3. pour tous $x, y \in C$, $\langle \nabla f(y) - \nabla f(x),\, y-x\rangle \geqslant m\|y-x\|^2$.

De plus si f est $\mathcal{C}^2$ alors f est m-fortement convexe sur C si et seulement si pour tous $x, y \in C$, $\langle \nabla^2 f(x)(y-x),\, y-x\rangle \geqslant m\|y-x\|^2$.

Et si en plus $C = \mathbb{R}^n$, alors f est m-fortement convexe sur C si et seulement si pour tous $x \in \mathbb{R}^n$, $\nabla^2 f(x) \succeq mI_n$.

Démonstration.

1 ⇔ 2. La fonction $g = f - \frac{m}{2}\|\cdot\|^2$ est convexe sur C si et seulement si pour tous $x, y \in C$, $g(y) \geqslant g(x) + \langle \nabla g(x),\, y-x\rangle$. Or $\nabla g(x) = \nabla f(x) - mx$. Donc g est convexe si et seulement si pour tous $x, y \in C$, $f(y) \geqslant f(x) + \langle \nabla f(x),\, y-x\rangle - m\langle x,\, y-x\rangle + \frac{m}{2}\|y\|^2 - \frac{m}{2}\|x\|^2 = f(x) + \langle \nabla f(x),\, y-x\rangle + \frac{m}{2}\|y-x\|^2$.

2 ⇒ 3. Soit $x, y \in C$. Il suffit d'appliquer l'inégalité en (x, y) puis (y, x) et de sommer.

3 ⇒ 1. Soit $x, y \in C$ et $h : t \in [0,1] \mapsto g(x + t(y-x))$. La fonction h est $\mathcal{C}^1$ sur $[0,1]$ et pour tout $t \in [0,1]$, $h'(t) = \langle \nabla g(x+t(y-x)),\, y-x\rangle$. Par 3, on déduit que si $t \leqslant s$ alors $h'(t) \leqslant h'(s)$. D'où h' est une fonction croissante, donc h est convexe. On obtient ainsi que pour tout $t \in [0,1]$, $g((1-t)x+ty) = h(t) = h((1-t)\cdot 0 + t\cdot 1) \leqslant (1-t)h(0) + th(1) = (1-t)g(x) + tg(y)$. Donc g est convexe sur C.

Si f est $\mathcal{C}^2$, alors pour tout $x \in \mathbb{R}^n$, $\nabla^2 g(x) = \nabla^2 f(x) - mI_n$. Donc f est m-fortement convexe sur C si et seulement si g est convexe sur C donc si et seulement si

$$\begin{aligned} &\Leftrightarrow \forall x, y \in C, \quad \left\langle \nabla^2 g(x)(y-x),\, y-x\right\rangle \geqslant 0, \\ &\Leftrightarrow \forall x, y \in C, \quad \left\langle \nabla^2 f(x)(y-x),\, y-x\right\rangle \geqslant m\|y-x\|^2. \end{aligned}$$

Enfin si $C = \mathbb{R}^n$, alors g est convexe sur $\mathbb{R}^n$ si et seulement si $\nabla^2 g(x) \succeq 0$ i.e. $\nabla^2 f(x) \succeq mI_n$. □

La forte convexité permet d'assurer automatiquement l'existence et unicité d'un minimum global.

Proposition 3.14. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^1$ sur $\mathbb{R}^n$ et m-fortement convexe sur $\mathbb{R}^n$. Alors f est coercive et admet un minimum global sur $\mathbb{R}^n$. Ce dernier est unique car puisque f est m-fortement convexe, elle est strictement convexe

Démonstration. D'après la Proposition 3.13, pour tout $x, y \in \mathbb{R}^n$, on a

$$f(y) \geqslant f(x) + \langle \nabla f(x), \, y - x \rangle + \frac{m}{2} \|y - x\|^2 .$$

En choisissant $x = 0$, on remarque alors que $f(y) \underset{\|y\| \to +\infty}{\longrightarrow} +\infty$. D'où f est coercive. Elle est de plus continue car $\mathcal{C}^1$, donc admet un minimum global d'après le Théorème 2.11. Comme f est strictement convexe sur $\mathbb{R}^n$ grâce à la Proposition 3.12, ce minimum global est bien unique. □

De la Proposition 3.13, on déduit les inégalités suivantes qui vont jouer un rôle important dans la suite pour l'étude de la convergence des algorithmes.

Corollaire 3.15. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^1$ m-fortement convexe sur C, un sous ensemble convexe de $\mathbb{R}^n$. On suppose que f admet un minimum global x^* sur $\mathbb{R}^n$ et tel que $x^* \in C$. Alors pour tout $x \in C$, on a :

1. $\|x - x^*\|^2 \leqslant \frac{2}{m}(f(x) - f(x^*))$,
2. $f(x) - f(x^*) \leqslant \frac{1}{2m} \|\nabla f(x)\|^2$,
3. $\|x - x^*\| \leqslant \frac{1}{m} \|\nabla f(x)\|$.

Démonstration.

1. Comme f est m-fortement convexe, on déduit de la Proposition 3.13 que pour tout $x \in \mathbb{R}^n$ on a $f(x) \geqslant f(x^*) + \left\langle \underbrace{\nabla f(x^*)}_{=0}, \, x - x^* \right\rangle + \frac{m}{2} \|x - x^*\|^2$. D'où l'inégalité souhaitée.
2. De même on a pour tous $x, y \in C$, $f(y) \geqslant f(x) + \langle \nabla f(x), \, y - x \rangle + \frac{m}{2} \|y - x\|^2$. Donc pour tous $x, y \in C$, $f(y) \geqslant \min_{z \in \mathbb{R}^n} h(z)$ où $h : z \mapsto f(x) + \langle \nabla f(x), \, z - x \rangle + \frac{m}{2} \|z - x\|^2$. Or h est convexe et admet $z = x - \frac{1}{m} \nabla f(x)$ comme unique point critique (calculer le gradient et l'annuler). On a donc pour tous $x, y \in \mathbb{R}^n$, $f(y) \geqslant h(x - \frac{1}{m} \nabla f(x)) = f(x) - \frac{1}{2m} \|\nabla f(x)\|^2$. Ainsi on obtient l'inégalité souhaité en prenant $y = x^*$.
3. Il suffit de mettre les deux inégalités précédentes bout à bout et de passer à la racine carré. □

Remarque 3.16. Si $x \in C$ est "presque" un minimiseur de f, dans le sens où il existe $\varepsilon > 0$ "petit" tel que $f(x) \leqslant f(x^*) + \varepsilon$, alors la première inégalité du Collaire 3.15 donne $\|x - x^*\| \leqslant \sqrt{\frac{2\varepsilon}{m}}$. Ceci signifie alors que x est nécessairement "proche" du minimum global.

De même si la condition d'optimilité ($\nabla f(x) = 0$) est "presque" satisfaite, i.e. $\|\nabla f(x)\| \leqslant \varepsilon$ pour $\varepsilon > 0$ "petit", alors $\|x - x^*\| \leqslant \frac{\varepsilon}{m}$ et de nouveau x est nécessaire "proche" du minimum global.

3.2.3 Convergence de l'algorithme du gradient à pas optimal

Dans cette section, nous allons montrer que sous certaines hypothèses, l'algorithme de descente de gradient à pas optimal produit des itérés $(x^{(k)})_{k\in\mathbb{N}}$ qui convergent vers le minimum global de la fonction impliquée. Nous allons également étudier à quelle vitesse se fait la convergence. Pour cela nous allons avoir besoin de la notion suivante.

Définition 3.17 (Convergence linéaire)**.** On dit que la suite $(u^{(k)})_{k\in\mathbb{N}}$ converge linéairement vers u^* si il existe $0 \leqslant \alpha < 1$ et $k_0 \in \mathbb{N}$ tel que pour tout $k \geqslant k_0$

$$\|u_{k+1} - u^*\| \leqslant \alpha \|u_k - u^*\| .$$

On dira que la convergence est super-linéaire si $\limsup_{k\to+\infty} \frac{\|u_{k+1}-u^*\|}{\|u_k-u^*\|} = 0$: le α de la définition précédente peut-être choisi arbitrairement petit.

Remarque 3.18. L'utilisation du mot *linéaire* pour nommer la convergence précédente vient du fait qu'à chaque itération de la suite le nombre de décimales de précision gagner dans l'approximation de la limite est $-\log_{10}(\alpha)$. Dit autrement le nombre de décimale exacte dans l'approximation de u^* par u_k dépend linéairement de k.

Le résultat suivant donne la convergence de l'algorithme de descente de gradient à pas optimal.

Théorème 3.19. Soit $x^{(0)} \in \mathbb{R}^n$. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^2$, convexe et vérifiant
— l'ensemble $C = \{x \in \mathbb{R}^n : f(x) \leqslant f(x^{(0)})\}$ est compact,
— il existe $0 < m \leqslant M$ tels que pour tout $x \in C$, on a $mI_n \leqslant \nabla^2 f(x) \leqslant MI_n$.
Alors les itérées $(x^{(k)})_{k\in\mathbb{N}}$ de l'algorithme de descente de gradient à pas optimal appliqué à f convergent vers l'unique minimum global, noté x^*, de f sur $\mathbb{R}^n$. De plus en posant $c = \frac{M}{m} \geqslant 1$, on a

$$\forall k \in \mathbb{N}, \quad f(x^{(k+1)}) - f(x^*) \leqslant (1 - \frac{1}{c})(f(x^{(k)}) - f(x^*)),$$

c'est-à-dire $(f(x^{(k)}))_{k\in\mathbb{N}}$ converge linéairement vers $f(x^*)$.

Démonstration.
Existence et unicité du minimum global. Comme C est compact, f admet un minimum global sur C que l'on note $x^* \in C$. Ce minimum global est unique sur C car f est m-fortement convexe sur C donc en particulier strictement convexe. Puisque pour tout $x \notin C$, $f(x) > f(x^{(0)})$, les minimums globaux de f sur $\mathbb{R}^n$ sont nécessairement dans C. D'où x^* est l'unique minimum global de f sur C.

Bonne définition de l'algorithme. Ce n'est pas évident que l'algorithme soit bien définie, c'est-à-dire produit une suite d'itérés $(x^{(k)})_{k\in\mathbb{N}}$, car le pas de descente est construit à travers une étape non triviale de minimisation (on choisit le meilleur pas par définition). Soit $k \in \mathbb{N}$ et supposons construits les éléments $d^{(0)}, \dots, d^{(k)}$, $t^{(0)}, \dots, t^{(k-1)}$ et $x^{(0)}, \dots, x^{(k)}$. Montrons que l'on peut définir $t^{(k)}$.

Notons $\mathcal{S}^{(k)} = \{t \in \mathbb{R} : f(x^{(k)} + td^{(k)}) \leqslant f(x^{(0)})\}$. Comme C est compact il existe $r > 0$ tel que pour tout $x \in C$, $\|x\| \leqslant r$. L'ensemble $\mathcal{S}^{(k)}$ est non vide car $0 \in \mathcal{S}^{(k)}$. Soit $t \in \mathcal{S}^{(k)}$, on a donc $\left\|x^{(k)} + td^{(k)}\right\| \leqslant r$, d'où par l'inégalité triangulaire $|t| \left\|d^{(k)}\right\| \leqslant r + \left\|x^{(k)}\right\|$. Si $d^{(k)} = 0$

alors $x^{(k)} = x^*$ et n'importe quel $t^{(k)} \in \mathbb{R}$ est un pas optimal. Les itérés forment alors une suite stationnaire valant x^* (l'algorithme a convergé en temps fini). Sinon on obtient

$$|t| \leqslant \frac{r + \|x^{(k)}\|}{\|d^{(k)}\|}.$$

Donc l'ensemble $\mathcal{S}^{(k)}$ est non vide, borné et fermé par continuité de f, d'où $\mathcal{S}^{(k)}$ est compact. Par continuité de $g : t \in \mathbb{R} \mapsto f(x^{(k)} + td^{(k)})$, on déduit alors qu'il existe bien $t^{(k)} \in \operatorname{argmin}_{\mathcal{S}^{(k)}} g$ un pas optimal.

Majoration de $\left\|\nabla f(x^{(k)})\right\|^2$. Soit $k \in \mathbb{N}$ et supposons que $\nabla f(x^{(k)}) \neq 0$ (car sinon encore une fois l'algorithme a convergé : $x^{(k)} = x^*$). On considère $g : t \in \mathbb{R} \mapsto f(x^{(k)} + td^{(k)})$ et on pose pour tout $t \in \mathbb{R}$, $x_t = x^{(k)} + td^{(k)}$. On a par définition $f(x^{(k+1)}) = \min_{\mathbb{R}} g$. Soit $t \in \mathcal{S}^{(k)}$, alors comme g est $\mathcal{C}^2$ sur $\mathbb{R}$, d'après la formule de Taylor-Lagrange il existe $s \in [0, t]$ tel que $g(t) = g(0) + tg'(0) + \frac{t^2}{2}g''(s) = f(x^{(k)}) + t\left\langle \nabla f(x^{(k)}),\, d^{(k)} \right\rangle + \frac{t^2}{2}\left\langle \nabla^2 f(x_s)d^{(k)},\, d^{(k)} \right\rangle$.

Or $s \in \mathcal{S}^{(k)}$, car comme C est convexe (puisque f est convexe) et $x_s \in [x^{(k)}, x_t]$ avec $x^{(k)}, x_t \in C$, on a $x_s \in C$. Donc par hypothèse, $\nabla^2 f(x_s) \leqslant MI_n$. D'où

$$g(t) \leqslant f(x^{(k)}) - t\left\|\nabla f(x^{(k)})\right\|^2 + \frac{t^2}{2}M\left\|\nabla f(x^{(k)})\right\|^2 = f(x^{(k)}) + (\frac{t^2}{2}M - t)\left\|\nabla f(x^{(k)})\right\|^2. \tag{3.20}$$

Le second membre de cette dernière égalité atteint son minimum en $t = \frac{1}{M}$. De plus $\frac{1}{M} \in \mathcal{S}^{(k)}$ [1]. En effet si on considère l'ensemble $\mathcal{T} = \{t \in [1, \frac{2}{M}] : \forall t' \leqslant t, t' \in \mathcal{S}^{(k)}\}$, ce dernier est non vide, majorée donc admet une borne supérieure τ et $\tau \in \mathcal{T}$ (car $\mathcal{T}$ est définie par une condition fermée). Comme $d^{(k)}$ est une direction de descente, il existe $t_0 > 0$ tel que $[0, t_0] \subset \mathcal{T}$. Donc $\tau > 0$. Si par l'absurde $\tau < \frac{1}{M}$, alors comme $\tau \in \mathcal{S}^{(k)}$ on a par l'inégalité (3.20) : $g(\tau) \leqslant f(x^{(k)}) + (\frac{\tau^2}{2}M - \tau)\left\|\nabla f(x^{(k)})\right\|^2 = f(x^{(k)}) - \varepsilon\left\|\nabla f(x^{(k)})\right\|^2$, avec $\varepsilon = -(\frac{\tau^2}{2}M - \tau) > 0$ car $\tau \in]0, \frac{1}{M}[$. Par continuité de g en τ, il existe donc $\eta > 0$ tel que pour tout $t \in [\tau, \tau + \eta]$, $g(t) \leqslant f(x^{(k)}) - \frac{\varepsilon}{2}\left\|\nabla f(x^{(k)})\right\|^2 < f(x^{(0)})$. Donc $\tau + \eta \in \mathcal{T}$, ce qui contredit le fait que $\tau = \sup \mathcal{T}$. D'où $\tau \geqslant \frac{1}{M}$ et $\frac{1}{M} \in \mathcal{S}^{(k)}$.

Ainsi, en appliquant (3.20) en $t = \frac{1}{M}$, on obtient

$$g(\frac{1}{M}) \leqslant f(x^{(k)}) - \frac{1}{2M}\left\|\nabla f(x^{(k)})\right\|^2.$$

Et comme par définition de $t^{(k)}$ on a en particulier $f(x^{(k+1)}) \leqslant g(\frac{1}{M})$, on déduit finalement que $f(x^{(k+1)}) \leqslant f(x^{(k)}) - \frac{1}{2M}\left\|\nabla f(x^{(k)})\right\|^2$, d'où

$$\left\|\nabla f(x^{(k)})\right\|^2 \leqslant 2M(f(x^{(k)}) - f(x^{(k+1)})).$$

Minoration de $\left\|\nabla f(x^{(k)})\right\|^2$. Comme f est m-fortement convexe sur C, on déduit d'après le Corollaire 3.15

$$2m(f(x^{(k)}) - f(x^*)) \leqslant \left\|\nabla f(x^{(k)})\right\|^2.$$

1. ce point est important, car si ce n'est pas vrai alors on ne peut appliquer l'inégalité (3.20) en $t = \frac{1}{M}$. On voit assez facilement que cela semble vrai car en appliquant l'inégalité (3.20), on obtient $g(\frac{1}{M}) \leqslant f(x^{(k)}) - \frac{1}{2M}\left\|\nabla f(x^{(k)})\right\|^2 < f(x^{(0)})$. Mais encore une fois, on ne peut appliquer cette inégalité que si l'on sait en amont que $\frac{1}{M} \in \mathcal{S}^{(k)}$... Il faut donc passer par le petit raisonnement légèrement lourd et technique qui suit pour s'en assurer. Il peut être sauté en première lecture.

Convergence linéaire de $(f(x^{(k)}))_{k\in\mathbb{N}}$**.** On a donc montré que pour tout $k \in \mathbb{N}$,

$$2m(f(x^{(k)}) - f(x^*)) \leqslant 2M(f(x^{(k)}) - f(x^{(k+1)})).$$

En posant $c = \frac{M}{m} \geqslant 1$, on a pour tout $k \in \mathbb{N}$,

$$f(x^{(k+1)}) - f(x^*) \leqslant (1 - \frac{1}{c})(f(x^{(k)}) - f(x^*)),$$

d'où $(f(x^{(k)}))_{k\in\mathbb{N}}$ converge linéairement car $1 - \frac{1}{c} \in [0, 1[$.

Convergence de $(x^{(k)})_{k\in\mathbb{N}}$ **vers** x^***.** De nouveau par le Corollaire 3.15, on obtient que

$$\left\|x^{(k)} - x^*\right\|^2 \leqslant \frac{2}{m}(f(x^{(k)}) - f(x^*)),$$

et comme le membre de droite tend vers 0, on déduit que $(x^{(k)})_{k\in\mathbb{N}}$ converge vers x^*. □

Remarque 3.21. La majoration de $f(x^{(k)}) - f(x^*)$ dans le théorème permet d'estimer combien il faut au minimum d'itérations pour que $f(x^{(k)}) - f(x^*) \leqslant \varepsilon$, avec $\varepsilon > 0$. En effet on a pour tout $k \in \mathbb{N}$, $f(x^{(k)}) - f(x^*) \leqslant \left(1 - \frac{1}{c}\right)^k (f(x^{(0)}) - f(x^*))$. Donc si $\left(1 - \frac{1}{c}\right)^k (f(x^{(0)}) - f(x^*)) \leqslant \varepsilon$, alors la condition $f(x^{(k)}) - f(x^*) \leqslant \varepsilon$ est satisfaite. En passant au $\log$, on obtient

$$k \geqslant \frac{-\log(\varepsilon) + \log(f(x^{(0)}) - f(x^*))}{-\log(1 - \frac{1}{c})}.$$

On remarque que plus $c \geqslant 1$ est proche de 1 et plus le terme de droite est petit, donc la convergence est rapide. De plus si c est très grand par rapport à 1, alors $\frac{-\log(\varepsilon)+\log(f(x^{(0)})-f(x^*))}{-\log(1-\frac{1}{c})} \simeq c(-\log(\varepsilon) + \log(f(x^{(0)}) - f(x^*)))$. Ainsi le nombre d'itérations est proportionnel à c. Plus c est grand et plus le nombre d'itérations nécessaire pour avoir $f(x^{(k)}) - f(x^*) \leqslant \varepsilon$ est grand.

Les deux résultats suivantes énoncent des conditions alternatives (plus fortes) pour lesquelles le théorème de convergence de l'algorithme de gradient à pas optimal s'applique également.

Corollaire 3.22. Soit $x^{(0)} \in \mathbb{R}^n$. Si $f : \mathbb{R}^n \to \mathbb{R}$ est $\mathcal{C}^2$ et il existe $0 < m \leqslant M$ tels que pour tout $x \in \mathbb{R}^n$, $mI_n \leqslant \nabla^2 f(x) \leqslant MI_n$, alors les hypothèses du théorème 3.19 sont satisfaites.

Démonstration. Comme f est m-fortement convexe sur $\mathbb{R}^n$, f est convexe et coercive. Ainsi l'ensemble $C = \{x \in \mathbb{R}^n : f(x) \leqslant f(x^{(0)})\}$ est compact. □

Corollaire 3.23. Soit $x^{(0)} \in \mathbb{R}^n$. Si $f : \mathbb{R}^n \to \mathbb{R}$ est $\mathcal{C}^2$, convexe et satisfaisant de plus
- $C = \{x \in \mathbb{R}^n : f(x) \leqslant f(x^{(0)})\}$ est compact,
- $\nabla^2 f(x)$ est définie positive pour tout $x \in C$,

alors les hypothèses du théorème 3.19 sont satisfaites.

Démonstration. Comme f est $\mathcal{C}^2$, les valeurs propres de $\nabla^2 f(x)$ dépendent continûment de $x \in C$. Comme C est compact, l'infimum et le supremum des valeurs propres de $\nabla^2 f(x)$ pour $x \in C$ sont donc finis et atteints. Notons respectivement m et M ces valeurs. Soit $x_0 \in C$ tel que m est la plus petite valeur propre de $\nabla^2 f(x_0)$. On a nécessairement $m > 0$ car $\nabla^2 f(x_0)$ est définie positive par hypothèse. De plus, par définition de m et M, on a $mI_n \leqslant \nabla^2 f(x) \leqslant MI_n$. □

En pratique, dans le cas d'une fonction f générale (en particulier non quadratique), on ne sait pas trouver un pas de descente optimal. Dans la section suivante, nous allons voir une méthode (rebroussement d'Armijo) qui permet de déterminer un pas de descente satisfaisant de sorte que l'algorithme de descente de gradient associé ait tout de même sensiblement les mêmes garantis de convergence que le théorème 3.19.

3.3 Descente de gradient préconditionné à rebroussement d'Armijo

3.3.1 Choix du pas par rebroussement d'Armijo

Définition 3.24 (Condition d'Armijo)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^1$, $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$. Soit $\alpha \in]0, \frac{1}{2}[$. On dira que $t \geqslant 0$ satisfait la condition d'Armijo si

$$f(x + td) \leqslant f(x) + \alpha t \langle \nabla f(x), d \rangle .$$

Remarque 3.25. Lorsque d est une direction de descente (ce qui sera toujours le cas, vu le contexte d'utilisation de cette définition), alors cette définition s'interprète géométriquement de la manière suivante : un pas $\tau > 0$ qui satisfait la condition d'Armijo assure que $f(x + \tau d)$ est en dessous, en $t = \tau$, de la droite de pente strictement négative $t \mapsto f(x) + \alpha t \langle \nabla f(x), d \rangle$. Notez que cette pente est négative mais plus grande (moins pentu) que celle de la tangente en 0 à $t \mapsto f(x + td)$ (valant $\langle \nabla f(x), d \rangle$) puisque $\alpha \in]0, \frac{1}{2}[$.

Définition 3.26 (Pas d'Armijo)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^1$, $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$. Soit $\alpha \in]0, \frac{1}{2}[$, $\beta \in]0, 1[$. On appelle *pas d'Armijo*, le réel $\tau > 0$ défini par

$$\tau = \max\{\beta^\ell : \ell \in \mathbb{N} \text{ et } f(x + \beta^\ell d) \leqslant f(x) + \alpha\beta^\ell \langle \nabla f(x), d \rangle\}. \tag{3.27}$$

Le pas d'Armijo est donc le plus grand réel positif qui s'écrit comme une puissance entière de β et qui satisfait la condition d'Armijo.

Remarque 3.28. Cette méthode de détermination d'un pas satisfaisant la condition d'Armijo est appelée méthode d'Armijo à rebroussement. Elle permet de s'assurer que le pas satisfaisant la condition d'Armijo ne sera pas trop petit (comme c'est un $\max$ parmi tous les β^ℓ, $\ell \in \mathbb{N}$) et qu'ainsi la diminution de f au nouvel itéré $x^{(k+1)}$ sera suffisante pour garantir de converger vers un minimum de f et au lieu de finir par stationner en un $x \in \mathbb{R}^2$ où $\nabla f(x) \neq 0$.

3.3.2 Algorithme de descente de gradient préconditionné

L'algorithme de descente de gradient préconditionné consiste à relâcher l'égalité $d^{(k)} = -\nabla f(x^{(k)})$, définissant la direction de descente, en $d^{(k)} = -B^{(k)} \nabla f(x^{(k)})$, où $B^{(k)}$ est une matrice symétrique définie positive.

Définition 3.29 (Descente de gradient préconditionné)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^1$. On appelle méthode de descente de gradient préconditionnée, l'algorithme suivant donné pour tout $k \in \mathbb{N}$ par

$$\begin{aligned} & d^{(k)} = -B^{(k)} \nabla f(x^{(k)}), \\ & t^{(k)} \text{ un pas de descente}, \\ & x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}, \end{aligned}$$

avec $B^{(k)}$ une matrice symétrique définie positive, appelée *préconditionneur*.

Remarque 3.30. Comme $B^{(k)}$ est une matrice symétrique définie positive, on a bien que $d^{(k)}$ est une direction de descente dès que $\nabla f(x^{(k)}) \neq 0$ puisque $\left\langle \nabla f(x^{(k)}),\, d^{(k)} \right\rangle < 0$.

De plus, puisque $B^{(k)}$ est une matrice symétrique définie positive, on a $B^{(k)} = (A^{(k)})^2$ avec $A^{(k)}$ une matrice symétrique définie positive uniquement définie à partir de $B^{(k)}$. Cette écriture de la matrice de préconditionnement $B^{(k)}$ comme le carré d'une matrice sera utile dans la section suivante pour étudier la convergence de l'algorithme.

3.3.3 Convergence de l'algorithme de descente de gradient préconditionné à rebroussement d'Armijo

On considère l'algorithme de la Définition 3.29 avec le pas de descente fixé comme étant le pas d'Armijo (voir Définition 3.26). Le résultat suivant précise sa convergence.

Théorème 3.31. Soit $x^{(0)} \in \mathbb{R}^n$ et $(B^{(k)})_{k\in\mathbb{N}}$ une suite de matrices carrés de taille n définies positives. On note $(A^{(k)})_{k\in\mathbb{N}}$ la suite des racines carrées des matrices $B^{(k)}$, $k \in \mathbb{N}$.

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction $\mathcal{C}^2$, convexe et vérifiant

— l'ensemble $C = \{x \in \mathbb{R}^n : f(x) \leqslant f(x^{(0)})\}$ est compact,

— il existe $0 < m \leqslant M$ tels que pour tout $x \in C$ et tout $k \in \mathbb{N}$, on a

$$mI_n \preceq A^{(k)} \nabla^2 f(x) A^{(k)} \preceq MI_n.$$

Alors les itérées $(x^{(k)})_{k\in\mathbb{N}}$ de l'algorithme de descente de gradient préconditionné (par la suite de matrices $(B^{(k)})_{k\in\mathbb{N}}$) à rebroussement d'Armijo appliqué à f convergent vers l'unique minimum global, noté x^*, de f sur $\mathbb{R}^n$. De plus en posant $\kappa = 2m\alpha \min(1, \frac{\beta}{M}) \in]0, 1[$, on a

$$\forall k \in \mathbb{N}, \quad f(x^{(k+1)}) - f(x^*) \leqslant (1 - \kappa)(f(x^{(k)}) - f(x^*)),$$

c'est-à-dire $(f(x^{(k)}))_{k\in\mathbb{N}}$ converge linéairement vers $f(x^*)$.

Démonstration.

Existence et unicité du minimum global x^* de f sur $\mathbb{R}^n$. L'existence de x^* se prouve exactement de la même manière que dans le Théorème 3.19 (c'est basé sur la compacité de C). On a nécessairement $x^* \in C$. Pour l'unicité, on montre que f est strictement convexe sur C : soit $x \in C$, $h \in \mathbb{R}^n \setminus \{0\}$ et posons $\tilde{h} = (A^{(k)})^{-1}h \neq 0$ alors $\langle \nabla^2 f(x)h,\, h\rangle = \left\langle \nabla^2 f(x) A^{(k)} (A^{(k)})^{-1} h,\, A^{(k)} (A^{(k)})^{-1} h \right\rangle = \left\langle A^{(k)} \nabla^2 f(x) A^{(k)} \tilde{h},\, \tilde{h} \right\rangle \geqslant m \left\| \tilde{h} \right\|^2 > 0$. D'où le minimum global x^* de f (sur $\mathbb{R}^n$) est unique sur C donc sur $\mathbb{R}^n$ (puisque n'importe quel minimum global de f sur $\mathbb{R}^n$ est forcément localisé dans C).

Bonne définition de l'algorithme. Soit $k \in \mathbb{N}$. Supposons que $\nabla f(x^{(k)}) \neq 0$, car sinon $x^{(k)} = x^*$ et l'algorithme devient stationnaire (et de plus on peut choisir n'importe quel pas $t^{(k)} \in \mathbb{R}$ puisque de toute façon $d^{(k)} = 0$). Alors $\left\langle \nabla f(x^{(k)}),\, d^{(k)} \right\rangle < 0$ et d'après la Proposition **??**, le pas d'Armijo $t^{(k)} > 0$ est uniquement défini. L'algorithme est donc bien posé, puisque l'on peut itérer en définissant $x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}$.

Montrons que $t^{(k)} \geqslant \min(1, \frac{\beta}{M})$**.** Soit $k \in \mathbb{N}$ fixé tel que $\nabla f(x^{(k)}) \neq 0$, i.e. en particulier $\left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle < 0$. Soit $\mathcal{S}^{(k)} = \{t \in \mathbb{R} : x_t = x^{(k)} + td^{(k)} \in C\}$. On sait que $\mathcal{S}^{(k)}$ est non vide car 0 et $t^{(k)}$ sont dans $\mathcal{S}^{(k)}$ (l'ensemble contient aussi un segment $[0, t_0]$ avec $t_0 > 0$ d'après la Proposition **??**). Soit $t \in \mathcal{S}^{(k)}$, alors par convexité de C (car f est convexe et C est un ensemble de sous-niveau de f), on a $[x_0, x_t] \subset C$. Comme f est $\mathcal{C}^2$, d'après la formule de Taylor-Lagrange, il existe $s \in [0, t]$ tel que

$$f(x_t) = f(x^{(k)}) + t \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle + \frac{t^2}{2} \underbrace{\left\langle \nabla^2 f(x_s) d^{(k)}, d^{(k)} \right\rangle}_{= \left\langle A^{(k)} \nabla^2 f(x_s) A^{(k)} A^{(k)} \nabla f(x^{(k)}), A^{(k)} \nabla f(x^{(k)}) \right\rangle} .$$

On a $x_s \in [x_0, x_t]$ donc $x_s \in C$ d'où par hypothèse

$$f(x_t) \leqslant f(x^{(k)}) + t \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle + \frac{t^2}{2} M \left\| A^{(k)} \nabla f(x^{(k)}) \right\|^2 .$$

Or $\left\| A^{(k)} \nabla f(x^{(k)}) \right\|^2 = - \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle$, d'où finalement

$$f(x_t) \leqslant f(x^{(k)}) + (t - \frac{t^2}{2} M) \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle . \tag{3.32}$$

Supposons que $t \in \mathcal{S}^{(k)}$, $t \neq 0$, soit tel que l'inégalité suivante soit vérifiée

$$f(x^{(k)}) + (t - \frac{t^2}{2} M) \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle \leqslant f(x^{(k)}) + \alpha t \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle . \tag{3.33}$$

Alors par (3.32), on déduit nécessairement que t satisfait la condition d'Armijo (voir Définition 3.24). Mais (3.33) est équivalente à $t - \frac{t^2}{2} M \geqslant \alpha t$, ou encore, comme $t \neq 0$, à $t \leqslant \frac{2(1-\alpha)}{M}$. Puisque $\alpha < \frac{1}{2}$, on déduit que $\frac{2(1-\alpha)}{M} > \frac{1}{M}$. D'où si $t \leqslant \frac{1}{M}$, nécessairement $t \leqslant \frac{2(1-\alpha)}{M}$ est satisfaite et donc également la condition d'Armijo.

Ainsi, si jamais $t^{(k)} = \beta^l$ avec $l \in \mathbb{N}^*$, on a $t^{(k)}$ satisfait la condition d'Armijo, mais pas β^{l-1} (car sinon contradiction avec le fait que $t^{(k)}$ est le plus grand des β^m, avec $m \in \mathbb{N}$, satisfaisant la condition d'Armijo). Donc nécessairement $\beta^{l-1} > \frac{1}{M}$, d'où $t^{(k)} = \beta^l > \frac{\beta}{M}$. Et on a bien $t^{(k)} \geqslant \min(1, \frac{\beta}{M})$. Sinon c'est que $t^{(k)} = 1$, mais alors de nouveau $t^{(k)} \geqslant \min(1, \frac{\beta}{M})$.

Dans tous les cas, on a bien montré que

$$t^{(k)} \geqslant \min(1, \frac{\beta}{M}). \tag{3.34}$$

Obtention de la convergence linéaire de $(f(x^{(k)}))_{k \in \mathbb{N}}$**.** Comme $t^{(k)}$ satisfait la condition d'Armijo, on a $f(x^{(k+1)}) \leqslant f(x^{(k)}) + \alpha t^{(k)} \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle$. D'où d'après (3.34), on obtient en posant $\varepsilon = \alpha \min(1, \frac{\beta}{M})$ que $f(x^{(k+1)}) \leqslant f(x^{(k)}) + \varepsilon \left\langle \nabla f(x^{(k)}), d^{(k)} \right\rangle = f(x^{(k)}) - \varepsilon \left\| A^{(k)} \nabla f(x^{(k)}) \right\|^2$, i.e.

$$\left\| A^{(k)} \nabla f(x^{(k)}) \right\|^2 \leqslant \frac{1}{\varepsilon} (f(x^{(k)}) - f(x^{(k+1)})). \tag{3.35}$$

Posons $g : y \in \mathbb{R}^n \mapsto f(A^{(k)} y)$. Alors g est $\mathcal{C}^2$ sur $\mathbb{R}^n$ (et convexe) et pour tout $y \in \mathbb{R}^n$, $\nabla g(y) = A^{(k)} \nabla f(A^{(k)} y)$, $\nabla^2 g(y) = A^{(k)} \nabla f(A^{(k)} y) A^{(k)}$. Par conséquent g est m-fortement convexe sur C. On déduit par le Corollaire 3.15 que $2m(g(y^{(k)}) - g(y^*)) \leqslant \left\|\nabla g(y^{(k)})\right\|^2$, avec $y^{(k)} = (A^{(k)})^{-1}x^{(k)}$ et $y^* = (A^{(k)})^{-1}x^*$ l'unique minimum global de g. Ainsi il découle que

$$2m(f(x^{(k)}) - f(x^*)) \leqslant \left\|A^{(k)}\nabla f(x^{(k)})\right\|^2.$$

En utilisant (3.35), on obtient alors

$$2m(f(x^{(k)}) - f(x^*)) \leqslant \frac{1}{\varepsilon}(f(x^{(k)}) - f(x^{(k+1)})).$$

Comme $\varepsilon > 0$, en multipliant l'inégalité par ε, en ajoutant (à gauche et à droite) $f(x^{(k+1)}) - f(x^*)$ et enfin en soustrayant (à gauche et à droite) $\kappa(f(x^{(k)}) - f(x^*))$ (où pour rappel $\kappa = 2m\varepsilon$), on obtient comme attendu

$$f(x^{(k+1)}) - f(x^*) \leqslant (1-\kappa)(f(x^{(k)}) - f(x^*)).$$

Notons que $\kappa < 1$ car si $\min(1, \frac{\beta}{M}) = \frac{\beta}{M}$ alors $\kappa \leqslant 2\alpha\beta < 1$, et sinon si $\min(1, \frac{\beta}{M}) = 1$ alors $\frac{\beta}{M} \geqslant 1$, i.e. $M \leqslant \beta < 1$ et donc $0 < m \leqslant M < 1$ d'où $\kappa = 2m\alpha < 2\alpha < 1$ [2]. On a donc bien la convergence linéaire de la suite $(f(x^{(k)}))_{k\in\mathbb{N}}$.

Convergence des itérés $(x^{(k)})_{k\in\mathbb{N}}$. La suite $(x^{(k)})_{k\in\mathbb{N}}$ admet au plus une valeur d'adhérence car si $\tilde{x}$ en est une alors par continuité de f et le fait que $(f(x^{(k)}))_{k\in\mathbb{N}}$ converge vers $f(x^*)$, on obtient par unicité de la limite $f(\tilde{x}) = f(x^*)$. Donc $\tilde{x}$ est un minimum global de f, d'où $\tilde{x} = x^*$ (par unicité du minimum global de f). Ainsi $(x^{(k)})_{k\in\mathbb{N}}$ est une suite d'éléments de C, un ensemble compact, qui admet au plus une valeur d'adhérence, nécessairement x^*, donc $(x^{(k)})_{k\in\mathbb{N}}$ converge vers x^*. □

Remarque 3.36. Comme $\kappa \sim 2\alpha\beta\frac{m}{M}$, on observe comme pour l'algorithme de descente de gradient à pas optimal que plus m est proche de M meilleure sera la convergence de l'algorithme. Mais cette fois, m et M ne sont plus des constantes figées du problème (par le choix de f). En effet, en choisissant les $A^{(k)}$ de manières appropriées, on peut espérer obtenir un meilleur conditionnement du problème par rapport au cas de l'algorithme sans préconditionneur ($B^{(k)} = I_n$).

Pour illustrer ceci, considérons l'exemple de $f : (x_1, x_2) \in \mathbb{R}^2 \mapsto \frac{1}{2}(Kx_1^2 + x_2^2)$ avec $K > 1$. Alors f est $\mathcal{C}^2$ et sa hessienne, en tout point, est diagonale avec $(K\ 1)$ sur la diagonale. Ainsi en choisissant pour tout $k \in \mathbb{N}$, $A^{(k)}$ la matrice diagonale de diagonale $(\frac{1}{\sqrt{K}}\ 1)$, on peut prendre $m = M = 1$ dans le théorème. On voit donc qu'en préconditionnant correctement, on a amélioré la convergence de l'algorithme de descente de gradient préconditionné (ici il converge en une itération...). De plus, le choix que l'on a fait pour la matrice $A^{(k)}$, et qui nous donne dans ce cas particulier le meilleur résultat possible, défini la matrice $B^{(k)} = (\nabla^2 f(x^{(k)}))^{-1}$. On peut se poser la question si, de manière générale, ce choix ne pourrait pas *toujours* s'avérer payant. Nous allons étudier cette question dans le chapitre suivant qui traite de la *méthode de Newton*.

2. On note ici l'importance d'avoir choisi $\alpha < \frac{1}{2}$ et $\beta < 1$, car sinon on n'aurait pas $\kappa < 1$.

Bibliographie

[ROUVIÈRE] François ROUVIÈRE, *Petit guide du calcul différentiel à l'usage de la license et de l'agrégation*, Cassini, 2009

[ALLAIRE & KABER] Grégoire ALLAIRE et Sidi Mahmoud KABER, *Algèbre linéaire numérique*, Ellipses, 2002

[CIARLET] Philippe G. CIARLET, *Introduction à l'analyse numérique matricielle et à l'optimisation*, cinquième édition, Dunod, 1998

[BOYD & VANDENBERGHE] Stephen BOYD and Lieven VANDENBERGHE *Convex Optimization*, Cambridge University Press, 2004