Optimisation numérique

Basé sur le poly de Quentin Mérigot

Luca Nenna*

2022-2023

Table des matières

*Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France. e-mail : `luca.nenna@universite-paris-saclay.fr`

Introduction

Motivation Dans de nombreuses applications, la formulation naturelle du problème qu'on cherche à résoudre est un problème d'optimisation :

— Dans la méthode des moindres carrés, on remplace un système linéaire $Ax = b$ surdéterminé et/ou n'ayant pas de solution (par exemple car certaines des égalités se contredisent) par le problème d'optimisation suivant :

$$\min_{x\in\mathbb{R}^N} \|Ax - b\|^2 . \tag{1}$$

Le minimiseur x^* de ce problème vérifie "au mieux" la famille d'équation $Ax = b$. Au contraire, lorsqu'un système linéaire $Ax = b$ admet plusieurs solutions, on peut en sélectionner une en considérant le problème

$$\min_{x\in K} \|x\|^2 \qquad K = \{x \in \mathbb{R}^N \mid Ax = b\} \tag{2}$$

— Si $\bar{x} \in \mathbb{R}^N$ représente un signal 1D échantilloné avec du bruit ($\bar{x}_i$ représentant par exemple la mesure effectuée en un temps t_i), on peut débruiter le signal en considérant le problème d'optimisation suivant, où $\lambda > 0$ est un paramètre :

$$\min_{x\in\mathbb{R}^N} \|x - \bar{x}\|^2 + \lambda \sum_{1\leq i\leq N-1} |x_{i+1} - x_i|^2 \tag{3}$$

un compromis entre deux comportements : x^* doit être proche de $\bar{x}$ (c'est le rôle du premier terme $\|x - \bar{x}\|^2$ de la fonction optimisée) mais doit également être "régulier", au sens où deux valeurs successives x_i et x_{i+1} doivent être proches (second terme $\sum_i |x_{i+1} - x_i|^2$).

— En finance, on peut considérer le problème de l'optimisation de portefeuille. Étant donnés N actifs, il s'agit de déterminer le pourcentage $x_i \geq 0$ du portefeuille que l'on investit dans l'actif i. Comme on souhaite investir 100% du portefeuille, ce problème d'optimisation est accompagné d'une contrainte $\sum_{1\leq i\leq N} x_i = 1$. On pourra donc considérer des problèmes d'optimisation *avec contraintes* de la forme

$$\min_{x\in\Delta} f(x) \text{ où } \Delta = \{x \in \mathbb{R}^N \mid \forall i, x_i \geq 0 \text{ et } \sum_i x_i = 1\}. \tag{4}$$

La fonction f est typiquement de la forme $f(x) = \frac{1}{\varepsilon} |\langle c|x\rangle - r|^2 + \langle x|Qx\rangle$: $c \in \mathbb{R}^N$ représente le rendement des actifs et le premier terme cherche à fixer

le niveau de rendement $\langle c|x\rangle = \sum_i c_i x_i$ à r. Le second terme de la fonction $\langle x|Qx\rangle$ est une mesure de risque : Q est une matrice symétrique mesurant les corrélations entre actifs, et on cherche un investissement minimisant cette corrélation.
- En apprentissage automatique (*machine learning*), de nombreux problèmes peuvent être formulés comme des problèmes d'optimisation. Nous verrons par exemple des problèmes de classification, que l'on résoudra par régression logistique ou par machine à vecteurs support (*support vector machine*).

Pour plus d'exemple, on renvoie au livre de Boyd et Vanderberghe, qui est disponible gratuitement (en anglais) en ligne : `http://web.stanford.edu/~boyd/cvxbook/`.

Problèmes avec/sans contrainte On peut séparer les problèmes en deux grandes classes. Il y a d'une part les problèmes d'optimisation sans contraintes, où l'on cherche à minimiser une fonctionnelle sur $\mathbb{R}^d$ ou sur son domaine de définition *ouvert* (les problèmes (1), (3) et la régression logistique sont de ce type). D'autre part, les problèmes d'optimisation avec contraintes, où l'on cherche à minimiser sur l'ensemble des points de $\mathbb{R}^N$ vérifiant un certain nombre de contraintes d'égalité ou d'inégalité (les problèmes (2), (4) et les machines à vecteurs support sont de ce type).

Convexité Tout les algorithmes et exemples présentés dans ce cours relèvent de l'optimisation *convexe*, où aussi bien la fonction optimisée que le domaine d'optimisation sont supposés convexes. La raison fondamentale pour laquelle on se restreint à ce cas est que pour les problèmes d'optimisation convexe, un minimiseur local est *automatiquement* minimiseur global.

Chapitre 1

Existence, unicité, convexité

Contents

Dans cette première partie, on s'intéresse à un problème de minimisation d'une fonction $f \in \mathcal{C}^1(\mathbb{R}^d)$:

$$\inf_{x \in \mathbb{R}^d} f(x) \tag{P}$$

Définition 1. On appelle :

(i) *infimum* ou *valeur du problème* de (P) la valeur $\inf_{\mathbb{R}^d} f$.

(ii) *minimiseur global* (ou simplement minimiseur) de (P) tout élément $x^* \in \mathbb{R}^d$ vérifiant $f(x^*) = \inf_{\mathbb{R}^d} f$. On note $\arg\min_{\mathbb{R}^d} f$ l'ensemble des minimiseurs de f (qui peut être vide), i.e.

$$\arg\min_{\mathbb{R}^d} f = \{x \in \mathbb{R}^d \mid f(x) = \inf_{\mathbb{R}^d} f\}.$$

(iii) On appelle *suite minimisante* pour (P) toute suite $x^{(0)}, \dots, x^{(k)}, \dots$ d'éléments de $\mathbb{R}^d$ telle que $\lim_{k\to+\infty} f(x^{(k)}) = \inf_{x\in\mathbb{R}^d} f(x)$.

Remarque 1. Il est possible que le problème (P) n'admette pas de minimiseur : penser par exemple à $f(x) = \exp(x)$ sur $\mathbb{R}$.

Lemme 1. *Il existe une suite minimisante pour le problème* (P).

Démonstration. Par définition de l'infimum, pour tout $k > 0$, il existe un élément $x^{(k)} \in \mathbb{R}^d$ tel que $\inf_{\mathbb{R}^d} f \leq f(x^{(k)}) \leq \inf_{\mathbb{R}^d} f + \frac{1}{k}$, soit $\lim_{k\to+\infty} f(x^{(k)}) = \inf_{\mathbb{R}^d} f$. $\square$

1.1 Existence

Proposition 2. *Soit $f \in \mathcal{C}^0(\mathbb{R}^d)$. On suppose de plus qu'il existe $x_0 \in \mathbb{R}^d$ tel que le sous-niveau $S = \{x \in \mathbb{R}^d \mid f(x) \leq f(x_0)\}$ est compact. Alors le problème d'optimisation* (P) *admet un minimiseur global.*

Démonstration. Si $f(x_0) = \inf_{\mathbb{R}^d} f$ on a déjà l'existence d'un minimum, à savoir le point x_0 lui-même. On suppose donc maintenant que $f(x_0) > \inf_{\mathbb{R}^d} f$. Soit $(x^{(k)})_{k\geq 0}$ une suite minimisante, qui vérifie donc $\lim_{k\to+\infty} f(x^{(k)}) = \inf_{\mathbb{R}^d} f < f(x_0)$. Alors, pour k suffisamment grand, on a $f(x^{(k)}) \leq f(x_0)$, soit $x^{(k)} \in S$. Comme l'ensemble S est compact, on peut extraire une sous-suite $(x^{(\sigma(k))})_{k\geq 0}$ qui converge vers un point $x^\infty \in S$. Alors, par continuité de f et par définition d'une suite minimisante on a $f(x^\infty) = \lim_{k\to+\infty} f(x^{(\sigma(k))}) = \inf_{\mathbb{R}^d} f$, et x^∞ minimise donc f sur $\mathbb{R}^d$. □

Corollaire 3. *Soit $f \in \mathcal{C}^0(\mathbb{R}^d)$ vérifiant*

$$\lim_{\|x\|\to+\infty} f(x) = +\infty,$$

c'est-à-dire que

$$\forall L \in \mathbb{R}, \exists R \geq 0 \text{ t.q. } \|x\| \geq R \Longrightarrow f(x) \geq L.$$

Alors (P) *admet un minimiseur global.*

Démonstration. Soit $x_0 \in \mathbb{R}^d$ quelconque et $S = \{x \in \mathbb{R}^d \mid f(x) \leq f(x_0)\}$. Pour montrer l'existence d'un minimiseur, il suffit de démontrer que S est compact. L'ensemble S est fermé comme sous-niveau d'une fonction continue. Supposons S non borné : pour tout k, il existe alors $x^{(k)} \in S$ tel que $\left\|x^{(k)}\right\| \geq qk$. Ainsi, $\lim_{k\to+\infty} x^{(k)} = +\infty$ et $f(x^{(k)}) \leq f(x_0)$, contredisant l'hypothèse. □

Corollaire 4. *Soit $f \in \mathcal{C}^0(\mathbb{R}^d)$ et vérifiant la propriété suivante :*

$$\forall x \in \mathbb{R}^d, f(x) \geq C\|x\|^p + D,$$

où $C > 0, D \in \mathbb{R}$ et $p > 0$. Alors le problème (P) *admet un minimiseur.*

Démonstration. Soit $x_0 \in \mathbb{R}^d$ quelconque et $S = \{x \in \mathbb{R}^d \mid f(x) \leq f(x_0)\}$. Pour pouvoir appliquer le théorème précédent, il suffit de démontrer que S est compact. Comme on sait déjà que S est fermé comme sous-niveau d'une fonction continue, il suffit de démontrer que cet ensemble est borné. Or, pour tout $x \in S$, on a

$$C\|x\|^p + D \leq f(x) \leq f(x^{(0)})$$

soit $\|x\|^p \leq E := \left|f(x^{(0)}) - D\right|/C$. Ainsi, l'ensemble S est contenu dans la boule centrée en 0 et de rayon $\sqrt[p]{E}$ et est donc borné. $\square$

1.2 Notions de calcul différentiel

Définition 2 (Différentiabilité)**.** Soit $f : \Omega \subset V \to W$ avec Ω ouvert. On dit que f est différentiable en $x_0 \in \Omega$ si et seulement si il existe une application linéaire continue $L \in \mathcal{L}(V, W)$ telle que

$$f(x_0 + h) \underset{h \to 0}{=} f(x_0) + L(h) + o(\|h\|).$$

L'application linéaire L est notée $\mathrm{d}_{x_0} f \in \mathcal{L}(V, W)$ et elle est dite différentielle de f en x_0

L'application f est alors $\mathcal{C}^1(\Omega)$ si elle est diffèrientable dans tout point x dans Ω et l'application

$$\begin{array}{rccl} \mathrm{d}f : & \Omega & \to & ((V, W), \|\cdot\|_{\mathrm{op}}) \\ & x_0 & \mapsto & \mathrm{d}_{x_0} f \end{array}$$

est continue.

Définition 3 (Dérivée directionnelle)**.** Soit $f : \Omega \subset V \to W$ avec Ω ouvert, $x_0 \in \Omega$ et $h \in V$. Quand elle existe, la limite

$$\lim_{t \to 0} \frac{f(x_0 + th) - f(x_0)}{t}$$

est appelée dérivée directionnelle de f dans la direction

Remarque 2. Si f est différentiable en x_0 alors elle admet une dérivée directionnelle dans toute direction $h \in V$ et

$$\mathrm{d}_{x_0} f(h) = \lim_{t \to 0} \frac{f(x_0 + th) - f(x_0)}{t}.$$

La réciproque n'est pas vraie en général ! Par exemple

$$f(x, y) = \begin{cases} \frac{y^2}{x} \text{ if } x \neq 0 \\ y \text{ if } x = 0. \end{cases}$$

Définition 4. Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$. Par le théorème de Riesz il existe un unique vecteur de V, noté $\nabla f(x_0)$ et appelé *gradient* de f en x_0, tel que

$$\mathrm{d}_{x_0} f(h) = \langle \nabla f(x_0), h \rangle \ \forall h \in V.$$

Soit $\langle \cdot, \cdot \rangle$ le produit scalaire usuel et $(e_i)_{1 \leq i \leq d}$ la base canonique de $\mathbb{R}^d$, on a

$$\nabla f(x) = \left(\frac{\partial f}{\partial e_i}(x) \right)_{1 \leq i \leq d} \quad \text{où } \frac{\partial f}{\partial e_i}(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(f(x + \varepsilon e_i) - f(x)).$$

Dans la suite on prendra toujours $V = \mathbb{R}^d$ et $W = \mathbb{R}$.

Exemple 1. Considérons $f(x) = \|x\|^2$ sur $\mathbb{R}^d$. En développant le carré de la norme, on obtient $f(x + v) = \|x\|^2 + \langle 2x|v \rangle + \|v\|^2$. On en déduit que $\nabla f(x) = 2x$, ce qui est conforme avec le calcul.

On dit que $f : \Omega \subset V \to W$ est deux fois différentiable en x_0 si elle est différentiable dans un voisinage de x_0 et $\mathrm{d}f$ est différentiable en x_0. On note cette dérivée $\mathrm{d}^2_{x_0}$ qui est un élément de $\mathcal{L}(V, \mathcal{L}(V, W))$. En particulier on a que $\mathrm{d}^2_{x_0} f(h, k)$ est la dérivée directionnelle de $x \mapsto \mathrm{d}_x f(h)$ dans la direction k.

Définition 5. Soit $f \in \mathcal{C}^2(\Omega)$, où $\Omega \subseteq \mathbb{R}^d$ est un ouvert. On appelle *hessienne* de f en $x_0 \in \Omega$ la matrice associée a la forme bilinéaire $\mathrm{d}^2_{x_0}$ dans la base canonique. En particulier

$$\mathrm{D}^2 f(x) = \left(\frac{\partial^2 f}{\partial e_i \partial e_j}(x) \right)_{1 \leq i,j \leq d},$$

où l'on a noté $(e_i)_{1 \leq i \leq d}$ la base canonique de $\mathbb{R}^d$. et où

1.3 Convexité et condition nécessaire et suffisante d'optimalité

1.3.1 Convexité

Définition 6. Une fonction $f : \mathbb{R}^d \to \mathbb{R}$ est *convexe* si

$$\forall (x, y) \in \mathbb{R}^d, \forall t \in [0, 1], \ f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y).$$

Exercice 1. Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction convexe. Montrer que l'ensemble

$$\{x \in \mathbb{R}^d \mid f(x) \leq C\}$$

est convexe quel que soit C. En déduire que l'ensemble des minimiseurs de (P) est un ensemble convexe fermé (possiblement vide).

Proposition 5. *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$. Alors les propositions suivantes sont équivalentes :*
(i) f *est convexe sur* $\mathbb{R}^d$,
(ii) $\forall x, y \in \mathbb{R}^d$, *la fonction* $g : t \in [0,1] \mapsto f((1-t)x + ty)$ *est convexe.*
(iii) $\forall x, y \in \mathbb{R}^d, f(y) \geq f(x) + \langle y - x | \nabla f(x) \rangle$,
(iv) $\forall x, y \in \mathbb{R}^d, \langle \nabla f(x) - \nabla f(y) | x - y \rangle \geq 0$.

Lemme 6. *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$. Étant donnés $x \in \Omega$ et $v \in \mathbb{R}^d$, on définit $x_t = x + tv$ et $g(t) = f(x_t)$. Alors,*

$$g'(t) = \langle \nabla f(x_t) | v \rangle. \tag{1.1}$$

Démonstration de la proposition 5. *(i)$\Longleftrightarrow$(ii)* conséquence directe de la définition.
(ii)$\Longrightarrow$(iii). Soit x, y dans $\mathbb{R}^d$ et $g : t \mapsto f(x_t)$ où $x_t = (1-t)x + tv = x + t(y - x)$, qui est convexe par hypothèse. Par convexité, on a $g(t) \geq g(0) + tg'(0)$, soit par le lemme $f(x) + \langle \nabla f(x) | y - x \rangle + \leq f(y)$.
(iii)$\Longrightarrow$(iv) Il suffit de sommer l'inégalité (iii) et la même inégalité où l'on a inversé le rôle de x et y.
(iv)$\Longrightarrow$(ii) Soit encore $g : t \mapsto f(x_t)$ où $x_t = (1-t)x + tv = x + t(y - x)$. Comme $g'(t) = \langle \nabla f(x_t) | y - x \rangle$, (lemme 6) l'inégalité (iv) appliquée en x_s et x_t (où $t > s$) nous donne

$$g'(t) - g'(s) = \langle \nabla f(x_t) - \nabla f(x_s) | y - x \rangle = \frac{1}{t-s} \langle \nabla f(x_t) - \nabla f(x_s) | x_t - x_s \rangle \geq 0,$$

et g' est donc croissante sur $[0,1]$. Ainsi, g est convexe. □

Lemme 7 (Taylor-Lagrange)**.** *Soit $f \in \mathcal{C}^2(\mathbb{R}^d)$, $x, v \in \mathbb{R}^d$ et $x_t = x + tv$. Alors,*

$$\forall t \geq 0, \exists s \in [0,t], \quad f(x_t) = f(x) + t \langle \nabla f(x) | v \rangle + \frac{t^2}{2} \langle D^2 f(x_s) v | v \rangle.$$

Lemme 8. *Soit $\Omega \subseteq \mathbb{R}^d$ ouvert, $f \in \mathcal{C}^2(\Omega)$, $x \in \Omega$ et $v \in \mathbb{R}^d$ et $g : t \mapsto f(x_t)$ où $x_t = x + tv$. Alors,*

$$g''(t) = \langle \mathrm{D}^2 f(x_t) v | v \rangle. \tag{1.2}$$

Démonstration. On fait le calcul en coordonnées, en notant $(e_k)_k$ la base canonique :

$$g(t) = f(x + \sum_k t v_k e_k)$$

$$g'(t) = \sum_i v_i \frac{\partial f}{\partial e_i} \left(x + \sum_k t v_k e_k \right) = \langle \nabla f(x_t) | v \rangle$$

$$g''(t) = \sum_i \sum_j v_i v_j \frac{\partial^2 f}{\partial e_j \partial e_i}\left(x + \sum_k t v_k e_k\right) = \langle \mathrm{D}^2 f(x_t) v | v\rangle \qquad \square$$

Proposition 9. *Soit $f \in \mathcal{C}^2(\mathbb{R}^d)$. Si $\forall x \in \mathbb{R}^d, \mathrm{D}^2 f(x) \succeq 0$, alors f est convexe.*

Démonstration de la proposition 9. Considérons $x, y \in \Omega$ et $g(t) = f(x_t)$ où $x_t = (1-t)x + ty$. Alors, $g''(t) = \langle \mathrm{D}^2 f(x_t)(y-x)|y-x\rangle$ est positif par hypothèse, de sorte que par Taylor-Lagrange

$$g(1) = g(0) + g'(0) + \frac{s^2}{2} g''(s) \geq g(0) + g'(0),$$

soit $f(y) \geq f(x) + \langle \nabla f(x)|y-x\rangle$. La proposition 5 montre que f est convexe. $\square$

1.3.2 Condition nécessaire d'optimalité

Théorème 10 (Fermat)**.** *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$, et x^* un minimiseur de* (P)*. Alors*

$$\nabla f(x^*) = 0. \quad \textit{(équation d'Euler)}$$

Remarque 3. La contraposée est fausse : prendre $f(x) = x^3$ sur $\Omega = \mathbb{R}$: le point 0 vérifie $f'(0) = 0$ mais n'est pas un minimiseur (même local).

Démonstration. Si x^* est un minimiseur de f sur Ω, on a pout tout $\varepsilon \in \mathbb{R}$, $f(x^* + \varepsilon e_i) \geq f(x^*)$. Ainsi,

$$\forall 0 < \varepsilon \leq \varepsilon_0, \quad \frac{f(x^* + \varepsilon e_i) - f(x^*)}{\varepsilon} \geq 0.$$

En passant à la limite, on obtient $\frac{\partial f}{\partial e_i}(x^*) \geq 0$. De même, en considérant le cas $\varepsilon < 0$

$$\forall -\varepsilon_0 \leq \varepsilon < 0, \quad \frac{f(x^* + \varepsilon e_i) - f(x^*)}{\varepsilon} \leq 0,$$

d'où l'on tire en passant à la limite $\frac{\partial f}{\partial e_i}(x^*) \leq 0$, soit *in fine* $\frac{\partial f}{\partial e_i}(x^*) = 0$. $\square$

1.3.3 Condition suffisante d'optimalité

Théorème 11. *Soit $\Omega \subseteq \mathbb{R}^d$ un ouvert convexe et $f \in \mathcal{C}^1(\Omega)$ convexe. Alors $x^* \in \Omega$ est un minimiseur de* (P) *si et seulement si $\nabla f(x^*) = 0$.*

Démonstration. Le théorème de Fermat nous donne déjà le sens direct. Pour la réciproque, il suffit de remarquer que si $\nabla f(x^*) = 0$, la proposition précédente donne

$$\forall y \in \Omega, f(y) \geq f(x^*) + \langle y - x^* | \nabla f(x^*) \rangle = f(x^*),$$

de sorte que x^* est bien un minimiseur de (P). $\square$

1.4 Stricte convexité et unicité du minimiseur

Définition 7. Une fonction $f : \mathbb{R}^d \to \mathbb{R}$ est strictement convexe si

$$\forall x \neq y \in \mathbb{R}^d, \forall t \in]0, 1[, \ f((1-t)x + ty) < (1-t)f(x) + tf(y).$$

Proposition 12. *Soit $f : \mathbb{R}^d \to \mathbb{R}$ strictement convexe. Alors f admet au plus un minimiseur sur $\mathbb{R}^d$.*

Démonstration. Par l'absurde, supposons que f admette deux minimiseurs distincts $x^* \neq y^* \in \mathbb{R}^d$. Alors, par stricte convexité de la fonction f on a $f(z^*) < \frac{1}{2}(f(x^*) + f(y^*)) = f(x^*)$, où $z^* = \frac{1}{2}(x^* + y^*)$, contredisant l'hypothèse que x^* minimise f. $\square$

Remarque 4. Cette proposition ne dit rien de l'existence d'un minimiseur.

Chapitre 2

Descente de gradient à pas optimal

On souhaite résoudre numériquement le problème de minimisation d'une fonction $f \in \mathcal{C}^0(\mathbb{R}^d)$. Comme en général il n'est pas raisonnable d'espérer calculer de manière exacte un minimiseur ou même la valeur de l'infimum du problème (P), on cherchera à l'*approcher*. Il s'agira de construire une suite $(x^{(k)})_{k\geq 0}$ de points vérifiant une des deux propriétés suivantes :

(a) la suite $x^{(k)}$ est minimisante pour (P), i.e. $\lim_{k\to+\infty} f(x^{(k)}) = \inf_{\mathbb{R}^d} f$.

(b) la suite $x^{(k)}$ converge vers un minimiseur de f sur $\mathbb{R}^d$.

Dans ce chapitre, nous nous intéresserons à la construction de suites $x^{(k)}$ vérifiant la seconde propriété (qui est bien sûr plus forte que la première).

2.1 Méthode de descente

Vocabulaire On appelle *méthode de descente* un procédé algorithmique permettant de construire itérativement une suite vérifiant (a) ou (b). Typiquement, une méthode de descente prend la forme suivante

$$\begin{cases} d^{(k)} = \dots & \textit{direction de descente} \\ t^{(k)} = \dots & \textit{pas de descente} \\ x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)} \end{cases}$$

Un tel algorithme est appelé méthode de descente si $f(x^{(k+1)}) \leq f(x^{(k)})$. Dans ce cours, on considèrera les possibilités suivantes :

(a) La *direction de descente* peut être égale à l'opposé du gradient, $d^{(k)} = -\nabla f(x^{(k)})$, auquel cas on parle de méthode de descente de gradient. Lorsque f est $\mathcal{C}^2$, il peut être plus avantageux de choisir $d^{(k)} = -\mathrm{D}^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$, auquel cas on parle de méthode de Newton.

(b) Le *pas de descente* peut être choisi constant ($t^{(k)} = \tau$), optimal (cf (2.1)), ou obtenu par des constructions un peu plus complexes, permettant de garantir la convergence de la méthode.

Définition 8. Soit $f \in \mathcal{C}^0(\mathbb{R}^d)$. On appelle *direction de descente* en $x \in \mathbb{R}^d$ tout vecteur v tel que $\exists \tau > 0, \forall t \in [0, \tau], f(x + tv) < f(x)$.

Exercice 2. Si $f \in \mathcal{C}^1(\mathbb{R}^d)$ et $\langle v | \nabla f(x) \rangle < 0$, alors v est une direction de descente.

Si f est différentiable en x, on a $f(x+tv) = f(x) + t\langle \nabla f(x) | v \rangle + \mathrm{o}(t)$. On cherche naturellement une direction de descente rendant le produit scalaire $\langle \nabla f(x) | v \rangle$ le plus petit possible, menant au choix de $d^{(k)} = -\nabla f(x^{(k)})$:

Lemme 13. *Soit* $f \in \mathcal{C}^1(\mathbb{R}^d)$ *et* $x_0 \in \mathbb{R}^d$. *Alors,* $\min_{\|v\|=1} \mathrm{d}_{x_0} f(v) = -\|\nabla f(x_0)\|$ *et si* $\nabla f(x_0) \neq 0$, *l'unique minimiseur est* $v = -\nabla f(x_0) / \|\nabla f(x_0)\|$.

Démonstration. On a $\mathrm{d}_{x_0} f(v) = \langle \nabla f(x_0) | v \rangle \geq -\|\nabla f(x_0)\| \|v\|$ par Cauchy-Schwarz, avec égalité si et seulement si v est positivement homogène à $-\nabla f(x_0)$. Comme $\|v\| = 1$, on a $v = -\nabla f(x_0) / \|\nabla f(x_0)\|$. $\square$

2.2 Descente de gradient à pas optimal

Définition 9. Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$. L'algorithme de descente de gradient à pas optimal est donné par :

$$\begin{cases} d^{(k)} = -\nabla f(x^{(k)}) \\ t^{(k)} \in \arg\min_t f(x^{(k)} + td^{(k)}) \\ x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)} \end{cases} \tag{2.1}$$

Remarque 5. Par construction, les itérées vérifient

$$f(x^{(k+1)}) \leq f(x^{(k)})$$
$$\langle \nabla f(x^{(k+1)}) | \nabla f(x^{(k)}) \rangle = 0.$$

Pour la deuxième égalité, il suffit de remarquer que si l'on pose $g(t) = f(x^{(k)} + td^{(k)})$, alors par définition de $t^{(k)}$,

$$g'(t^{(k)}) = 0 = \langle \nabla f(x^{(k)} + t^{(k)} d^{(k)}) | d^{(k)} \rangle = \langle \nabla f(x^{(k+1)}) | \nabla f(x^{(k)}) \rangle.$$

Remarque 6. Pour pouvoir mettre en œuvre cet algorithme il faut pouvoir calculer le pas optimal $t^{(k)}$ à chaque itération, ce qui implique de résoudre un problème d'optimisation (sur $\mathbb{R}$). Ceci n'est faisable de manière exacte que dans un nombre très limité de cas. En général, on préfèrera d'autre méthode de calcul du pas.

Exemple 2. Soit $f : x \in \mathbb{R}^d \mapsto \frac{1}{2}\langle Qx|x\rangle + \langle b|x\rangle$ où Q est une matrice symétrique définie positive. Alors f est strictement convexe et $\nabla f(x) = Qx + b$. Soit $x^{(k)} \in \mathbb{R}^d$, $d^{(k)} = -\nabla f(x^{(k)})$. Pour calculer le pas $t^{(k)} \in \mathbb{R}$, on cherche le minimum de $g : t \in \mathbb{R} \mapsto f(x_t)$ où $x_t = x^{(k)} + td^{(k)}$. La fonction g est convexe et atteint donc son minimum en l'unique point $t^{(k)}$ vérifiant $g'(t^{(k)}) = 0$. Or, $g'(t) = \langle \nabla f(x_t)|d^{(k)}\rangle$, soit

$$\begin{aligned} g'(t^{(k)}) = 0 &\iff \langle Q(x^{(k)} + t^{(k)}d^{(k)}) + b|d^{(k)}\rangle = 0 \\ &\iff t^{(k)}\langle Qd^{(k)}|d^{(k)}\rangle - \langle d^{(k)}|d^{(k)}\rangle = 0. \end{aligned}$$

Ainsi, $t^{(k)} = \langle d^{(k)}|d^{(k)}\rangle/\langle Qd^{(k)}|d^{(k)}\rangle$. En résumé, dans le cas d'une fonction f de la forme considérée, l'algorithme de descente de gradient à pas optimal s'écrit

$$\begin{cases} d^{(k)} = -(Qx^{(k)} + b) \\ t^{(k)} = \frac{\langle d^{(k)}|d^{(k)}\rangle}{\langle Qd^{(k)}|d^{(k)}\rangle} \\ x^{(k+1)} = x^{(k)} + t^{(k)}d^{(k)}. \end{cases} \tag{2.2}$$

On suppose dans la suite que l'ensemble

$$S = \{x \in \mathbb{R}^d \mid f(x) \leq f(x^{(0)})\} \text{ est compact,} \tag{2.3}$$

ce qui garantit (proposition 2) l'existence d'un minimiseur de f sur Ω.

Lemme 14. *Sous l'hypothèse* (2.3), *le minimum dans la définition du pas est atteint.*

Démonstration. Il s'agit de démontrer que la fonction $g : t \mapsto f(x^{(k)} + td^{(k)})$ atteint son minimum. On suppose que $d^{(k)}$ est non nul : sinon g est constante et atteint évidemment son minimum. Grâce à la proposition 2, il suffit de montrer que le sous-niveau $S_g := \{t \in \mathbb{R} \mid g(t) \leq f(x^{(k)})\}$ est compact. Ce sous-niveau est fermé comme image inverse du fermé $]-\infty, f(x^{(k)})]$ par la fonction continue g. Montrons maintenant que S_g est borné. Pour cela, nous utilisons l'hypothèse que S est compact donc borné : $\exists R \geq 0, \forall x \in S, \|x\| \leq R$. Si $t \in S_g$, alors $x^{(k)} + td^{(k)} \in S$, de sorte que

$$\left\| x^{(k)} + td^{(k)} \right\| \leq R$$

soit

$$|t| \leq \frac{R + \|x^{(k)}\|}{d^{(k)}}.$$

Ainsi S_g est compact et par la proposition 2, g atteint son minimum sur $\mathbb{R}$. □

Théorème 15. *Soit $f \in \mathcal{C}^2(\mathbb{R}^d)$ vérifiant*
(i) le sous-niveau $S = \{x \in \mathbb{R}^d \mid f(x) \leq f(x^{(0)})\}$ est compact.
(ii) $\exists M > 0,\ \forall x \in S,\ \mathrm{D}^2 f(x) \preceq M\mathrm{Id}$,
(iii) f est strictement convexe,
Alors les itérées de l'algorithme (2.1) convergent vers l'unique minimiseur global de f sur Ω.

On utilisera la proposition suivante :

Proposition 16. *Soit $(x_k)_{k\geq 1}$ une suite bornée de $\mathbb{R}^d$ admettant une unique valeur d'adhérence $\bar{x}$.* [a] *Alors, $\lim_{k\to+\infty} x_k = \bar{x}$.*

a. On rappelle que $\bar{x}$ est valeur d'adhérence de la suite $(x_k)_{k\geq 1}$ si et seulement si il existe une suite extraite $(x_{\sigma(k)})_{k\geq 1}$ dont la limite est $\bar{x}$

Démonstration du théorème 15. Soit $k \geq 1$ et $g(t) = f(x^{(k)} + td^{(k)}) = f(x^{(k)} - t\nabla f(x^{(k)}))$. Par définition du pas optimal $t^{(k)}$ on a

$$f(x^{(k+1)}) = \min_t f(x^{(k)} + td^{(k)}) = \min_t g(t),$$

et nous allons utiliser un développement de Taylor pour estimer le minimum de g. Par le lemme 8, et en posant $x_t = x^{(k)} + td^{(k)}$ on a

$$g'(t) = \langle \nabla f(x_t) | d^{(k)} \rangle, \qquad g''(t) = \langle \mathrm{D}^2 f(x_t) d^{(k)} | d^{(k)} \rangle.$$

Soit $\sigma = \{t \in \mathbb{R} \mid x_t \in S\}$ et $t \in \sigma$. Par Taylor-Lagrange, pour tout $t \in \sigma$, il existe $s \in [0, t]$ tel que $g(t) = g(0) + tg'(0) + \frac{t^2}{2} g''(s)$. Ainsi,

$$\begin{aligned} g(t) &= g(0) + tg'(0) + \frac{t^2}{2} g''(s) \\ &= f(x^{(k)}) - t \left\| \nabla f(x^{(k)}) \right\|^2 + \frac{t^2}{2} \langle \mathrm{D}^2 f(x_s) \nabla f(x^{(k)}) | \nabla f(x^{(k)}) \rangle \end{aligned}$$

Comme $s \in [0, t]$, alors $s = \gamma t$ avec $\gamma \in [0, 1]$ de sorte que $f(x_s) = f((1-\gamma)x_0 + \gamma x_t) \leq f(x^{(0)})$. Ainsi, $x_s \in S$. Par hypothèse, on a donc $\mathrm{D}^2 f(x_s) \leq M\mathrm{id}$, ce qui donne en utilisant (ii)

$$g(t) \leq f(x^{(k)}) + \left(\frac{M}{2} t^2 - t \right) \left\| \nabla f(x^{(k)}) \right\|^2$$

Le minimum de ce second membre est atteint en $t = 1/M$ et on a donc

$$f(x^{(k+1)}) \leq \min_t g(t) \leq f(x^{(k)}) - \frac{1}{2M} \left\| \nabla f(x^{(k)}) \right\|^2,$$

de sorte que

$$\left\| \nabla f(x^{(k)}) \right\|^2 \leq 2M(f(x^{(k)}) - f(x^{(k+1)})). \tag{2.4}$$

Ainsi, pour tout $K \geq 0$ on a

$$\sum_{0 \leq k \leq K} \left\| \nabla f(x^{(k)}) \right\|^2 \leq 2M(f(x^{(0)}) - f(x^{(K)})) \leq 2M(f(x^{(0)}) - \inf_{\mathbb{R}^d} f).$$

La série de terme général $\left\| \nabla f(x^{(k)}) \right\|^2$ est donc convergente, d'où l'on déduit que $\lim_{k \to +\infty} \left\| \nabla f(x^{(k)}) \right\| = 0$.

Montrons enfin que $\lim_{k \to +\infty} x^{(k)} = x^*$, où x^* est l'unique minimum de f sur $\mathbb{R}^d$, l'unicité provenant de la stricte convexité de f et de la proposition 12. Comme $f(x^{(k)}) \leq f(x^{(0)})$, le point $x^{(k)}$ appartient à S, qui est par hypothèse compact donc borné. Pour montrer que la suite $x^{(k)}$ converge vers x^*, il suffit par la proposition 16 de démontrer qu'elle admet x^* pour seule valeur d'adhérence. Soit donc $(x^{(\sigma(k))})$ une sous-suite convergeant vers une valeur d'adhérence $\bar{x} \in S$. Alors, comme $f \in \mathcal{C}^1(\mathbb{R}^d)$, $\nabla f(\bar{x}) = \lim_{k \to +\infty} \nabla f(x^{(\sigma(k))}) = 0$. Par convexité (théorème 11), on sait que $\bar{x}$ est un minimiseur de f sur $\mathbb{R}^d$. Par unicité du minimiseur, on en déduit que $\bar{x} = x^*$. $\square$

Bibliographie

[1] Hedy Attouch, Giuseppe Buttazzo, and Gérard Michaille. *Variational analysis in Sobolev and BV spaces : applications to PDEs and optimization*, volume 17. SIAM, 2014.

[2] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[3] Guillaume Carlier. *Classical and Modern Optimization*. World Scientific, 2022.

[4] Jean Dieudonné. Elements d'analyse. tome ii : Chapitres xii a xv. 1969.

[5] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2001.

[6] Robert R. Phelps. *Convex functions, monotone operators and differentiability*. Springer-Verlag, 1989.