Algorithmes d'optimisation

7 avril 2020

# Table des matières

# 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

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

— 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 $\lim_{k \to +\infty} f(x^{(k)}) = \inf_{\mathbb{R}^d} f$. □

## 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 $x^\infty$ 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. □

## 1.2 Condition nécessaire d'optimalité

**Définition 2.** Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$. On appelle *gradient* de $f$ en $x_0 \in \mathbb{R}^d$ le vecteur des dérivées partielles, où l'on a noté $(e_i)_{1\leq i\leq d}$ la base canonique de $\mathbb{R}^d$ :

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

*Remarque* 2 (Calcul du gradient). Soit $f \in \mathcal{C}^0(\mathbb{R}^d)$ et $x \in \mathbb{R}^d$. On rappelle que si l'on peut écrire un développement limité d'ordre 1 pour $f$ en $x$, de la forme

$$f(x+v) = f(x) + \langle g|v\rangle + \mathrm{o}(\|v\|), \tag{1.1}$$

alors $f$ est différentiable au point $x$ et $\nabla f(x) = g$.

*Exemple* 1. Considérons $f(x) = \|x\|^2$ sur $\mathbb{R}^d$. En développant le carré de la norme, on obtient $f(x+v) = \|x\|^2 + \langle 2x|v\rangle + \|v\|^2$, qui est de le forme (1.1) avec $g = 2x$. On en déduit que $\nabla f(x) = 2x$, ce qui est conforme avec le calcul.

**Théorème 5** (Fermat)**.** *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$, et $x^*$ un minimiseur de* (P)*. Alors* $\nabla f(x^*) = 0$.

*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 $\Omega$, 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* $\frac{\partial f}{\partial e_i}(x^*) = 0$. □

## 1.3 Convexité et condition suffisante d'optimalité

**Définition 3.** 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 6.** *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 7.** *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,*

$$g'(t) = \langle \nabla f(x_t) | v \rangle. \tag{1.2}$$

*Démonstration de la proposition 6.* *(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 7) 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. □

**Théorème 8.** *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). □

## 1.4 Stricte convexité et unicité du minimiseur

**Définition 4.** 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 9.** *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 $f$. □

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

## 1.5 Convexité et hessienne

**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 des dérivées partielles secondes :

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

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

**Définition 6.** À toute matrice symétrique $A \in \mathcal{M}_{d,d}(\mathbb{R})$ on peut associer une fonction (appelée forme quadratique) $q_A : x \in \mathbb{R}^d \mapsto \langle x | Ax \rangle$.

(i) Une matrice symétrique $A$ est dite *positive* (ce qu'on note $A \succeq 0$) si et seulement si la forme quadratique associée $q_A$ est positive, i.e.

$$\forall x \in \mathbb{R}^d, \langle x | Ax \rangle \geq 0.$$

(ii) Une matrice symétrique est dite *définie positive* si

$$\forall x \in \mathbb{R}^d \setminus \{0\}, \langle x | Ax \rangle > 0.$$

(iii) Soient $A, B$ deux matrices symétriques. On note $A \succeq B$ si et seulement si $A - B \succeq 0$, ou de manière équivalente si $q_A \geq q_B$.

*Exemple* 2. En particulier, on $\lambda \mathrm{Id} \preceq A \preceq \Lambda \mathrm{Id}$ si et seulement si

$$\forall x \in \mathbb{R}^d, \quad \lambda \|x\|^2 \leq \langle x|Ax\rangle \leq \Lambda \|x\|^2 .$$

**Proposition 10.** *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.*

**Lemme 11** (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 12.** *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,*

$$g''(t) = \langle \mathrm{D}^2 f(x_t)v|v\rangle. \tag{1.3}$$

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

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

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

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

*Démonstration de la proposition 10.* 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 6 montre que $f$ est convexe. $\square$

# 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 7.** 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 $v = -\nabla f(x_0)/\|\nabla f(x_0)\|$. □

## 2.2 Descente de gradient à pas optimal

**Définition 8.** 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* 3. 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 $\Omega$.

**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

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

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 $\Omega$.*

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 $\bar{x}$.* [1] *Alors, $\lim_{k\to+\infty} x_k = \bar{x}$.*

1. 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 12, 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 9. 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 8), on sait que $\bar{x}$ est un minimiseur de $f$ sur $\mathbb{R}^d$. Par unicité du minimiseur, on en déduit que $\bar{x} = x^*$. □

# Chapitre 3

# Descente de gradient à pas optimal II

## 3.1 Forte convexité et stabilité du minimum

**Motivation.** Soit $f \in \mathcal{C}^0(\mathbb{R}^d)$ atteignant son minimum $x^*$ sur $\mathbb{R}^d$. On a vu dans la proposition 9 que si $f$ est strictement convexe, alors $f$ admet au plus un minimiseur $x^*$. Une manière de reformuler cette propriété est la suivante :

$$f(x) \leq f(x^*) \Longrightarrow x = x^*.$$

En pratique, nos algorithmes sont incapables de calculer $x^*$ mais permettent au mieux de calculer une suite $x^{(k)}$ tel que $f(x^{(k)}) \leq f(x^*)+\varepsilon_k$ où $\lim_{k\to+\infty} \varepsilon_k = 0$. On aimerait pouvoir en déduire que $x^{(k)}$ est "proche" de la solution $x^*$, i.e. $\left\|x^{(k)} - x^*\right\| \leq C\varepsilon_k^\alpha$ pour un certain exposant $\alpha > 0$ et une constante $C > 0$. Nous verrons qu'une telle inégalité est vraie pour $\alpha = \frac{1}{2}$ si la fonction $f$ est *fortement convexe*.

**Définition 9.** Un sous-ensemble $X$ de $\mathbb{R}^d$ est dit *convexe* s'il contient tout segment reliant deux de ses points, c'est-à-dire :

$$\forall (x, y) \in X, \forall t \in [0, 1],\ (1 - t)x + ty \in X.$$

**Définition 10.** Une fonction $f : \mathbb{R}^d \to \mathbb{R}$ est dite *$m$-fortement convexe* sur un ensemble convexe $X \subseteq \mathbb{R}^d$, où $m > 0$, si la fonction $f - \frac{m}{2}\|\cdot\|^2$ est convexe.

**Exercice 3.** Si $f$ est $m$-fortement convexe, alors elle est aussi strictement convexe.

**Proposition 17.** *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$ et $X \subseteq \mathbb{R}^d$ convexe. Alors les implications suivantes sont vraies (i)$\Longrightarrow$(ii)$\Longleftrightarrow$(iii)$\Longleftrightarrow$(iv), où*

*(i)* $f \in \mathcal{C}^2(\mathbb{R}^d)$ *et* $\forall x \in X,\ \mathrm{D}^2 f(x) \succeq m\mathrm{Id}$ *;*

*(ii)* $f$ *est* $m$*-fortement convexe sur* $X$ *;*

*(iii)* $\forall x, y \in X, \quad f(y) \geq f(x) + \langle \nabla f(x) | y - x \rangle + \frac{m}{2} \|x - y\|^2$

*(iv)* $\forall x, y \in X, \quad \langle \nabla f(y) - \nabla f(x) | y - x \rangle \geq m \|x - y\|^2$

*Démonstration.* La fonction $f$ est $m$-fortement convexe si et seulement si la fonction $g = f - \frac{m}{2}\|\cdot\|^2$ est convexe. L'équivalence est s'obtient et utilisant les proposition 6 10 pour caractériser la convexité de $g$. □

*Exemple* 4. La fonction $f : t \mapsto \exp(t)$ n'est pas fortement convexe sur $\mathbb{R}$, mais elle est $m$-fortement convexe sur tout segment $[a, b]$ pour $m = \min_{t\in[a,b]} \exp(t)$. De même, fonction $t \mapsto t^4$ n'est pas fortement convexe sur $\mathbb{R}$ mais est fortement convexe sur tout segment ne contenant pas l'origine.

**Corollaire 18.** *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$ admettant un minimiseur $x^* \in \mathbb{R}^d$. On suppose que $f$ est $m$-fortement convexe sur un ensemble convexe $S$ contenant $x^*$. Alors, pour tout $x \in S$,*

*(i)* $\|x - x^*\|^2 \leq \frac{2}{m}(f(x) - f(x^*))$

*(ii)* $\|x - x^*\| \leq \frac{1}{m}\|\nabla f(x)\|$

*(iii)* $f(x) - f(x^*) \leq \frac{1}{2m}\|\nabla f(x)\|^2$

*Remarque* 7. Le point (i) montre immédiatement qu'il n'existe pas d'autre minimiseur de $f$ dans $S$ : si $x$ est un autre minimiseur, on a $f(x) = f(x^*)$, de sorte que par (i), $x = x^*$. Mais il faut surtout retenir que si $x$ est "presqu'un minimiseur", au sens où $f(x) \leq f(x^*) + \varepsilon$, alors $x$ est "proche" du minimiseur $x^*$, plus précisément $\|x - x^*\| \leq \sqrt{2\varepsilon/m}$. On tire des conclusions similaires du point (ii) : si la condition d'optimalité ($\nabla f(x^*) = 0$) est "presque vérifiée", i.e. si $\|\nabla f(x)\| \leq \varepsilon$ est petit, alors $\|x - x^*\| \leq \varepsilon/m$.

*Exemple* 5. Pour $f(t) = t^4$, on a $x^* = 0$ et l'inégalité (i), qui s'écrit $t^2 \leq \frac{2}{m}t^4$, n'est vraie pour aucun $m > 0$.

*Démonstration.* (i) est une conséquence immédiate de la formulation équivalente de la forte convexité donnée dans la proposition 17.(ii) :

$$f(x) \geq f(x^*) + \langle x - x^* | \nabla f(x^*) \rangle + \frac{m}{2}\|x - x^*\| = f(x^*) + \frac{m}{2}\|x - x^*\|^2,$$

où l'on a utilisé $\nabla f(x^*) = 0$ par optimalité de $x^*$. (ii) s'obtient de la même manière, en utilisant proposition 17.(iii).

(iii) Par l'inégalité (ii) de la proposition 17, on a pour tout $x, y \in S$

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

Ainsi, $f(y) \geq \inf_z g(z)$ où $g(z) = Q(x,z)$. Comme $g$ est convexe sur $\mathbb{R}^d$, son unique minimiseur $z^*$ est solution de l'équation

$$\nabla g(z^*) = 0 = \nabla f(x) + m(z^* - x),$$

c'est-à-dire que $z^* = x - \frac{1}{m}\nabla f(x)$. Ainsi, nous venons de démontrer que

$$\forall y \in S, f(y) \geq g(z^*) = f(x) - \frac{1}{m}\|\nabla f(x)\|^2 + \frac{m}{2}\left\|\frac{1}{m}\nabla f(x)\right\|^2 = f(x) - \frac{1}{2m}\|\nabla f(x)\|^2 .$$

En prenant $y = x^*$ dans l'inégalité précédente, on obtient le résultat voulu. □

## 3.2 Vitesse de convergence du gradient à pas optimal

**Définition 11** (Convergence linéaire)**.** Soit $(u_k)_{k\geq 0}$ une suite de limite $u^*$. La suite $(u_k)$ converge *linéairement* vers $u^*$ s'il existe $0 \leq \kappa < 1$ et $k_0 \in \mathbb{N}$ telle que

$$\forall k \geq k_0, \|u_{k+1} - u^*\| \leq \kappa \|u_k - u^*\| .$$

**Théorème 19.** *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) $\exists M,\ \forall x \in S,\ \mathrm{D}^2 f(x) \preceq M\mathrm{Id}$,*
*(iii) $\exists m > 0,\ \forall x \in S,\ \mathrm{D}^2 f(x) \succeq m\mathrm{Id}$,*
*Alors les itérées de l'algorithme (2.1) convergent vers l'unique minimiseur global de $f$ sur $\mathbb{R}^d$, et de plus, en posant $c = M/m$,*

$$f(x^{(k+1)}) - f(x^*) \leq \left(1 - \frac{1}{c}\right)\left(f(x^{(k)}) - f(x^*)\right). \tag{3.1}$$

*Remarque* 8. En d'autres termes, $(f(x^{(k)}))_{k\geq 0}$ converge linéairement vers $f(x^*)$.

*Remarque* 9. Si une fonction $f \in \mathcal{C}^2(\mathbb{R}^d)$, vérifie $m\mathrm{Id} \preceq \mathrm{D}^2 f(x) \leq M\mathrm{Id}$, on pourra dira que le *conditionnement* de $f$ est majoré par $c = \frac{M}{m}$. L'inégalité (3.1) montre que cette quantité est cruciale pour comprendre la vitesse de convergence de l'algorithme de descente de gradient. Par exemple, si l'on souhaite estimer $f(x^*)$ à $\varepsilon > 0$ près, d'après l'inégalité (3.1), il suffit d'interrompre l'algorithme de descente de gradient à pas optimal après $k$ itérations où

$$\left(1 - \frac{1}{c}\right)^k (f(x^{(0)}) - f(x^{(*)})) \leq \varepsilon,$$

soit

$$k \geq \frac{\log\left(\frac{f(x^{(0)}) - f(x^{(*)})}{\varepsilon}\right)}{\log\left(\frac{1}{1-1/c}\right)} \sim_{c\to+\infty} c\log\left(\frac{f(x^{(0)}) - f(x^{(*)})}{\varepsilon}\right)$$

Ainsi, dans le cas $c \gg 1$, le nombre d'itération est *proportionnel* au conditionnement !

*Démonstration du théorème 19.* Le corollaire (18) nous donne

$$f(x^k) - f(x^*) \leq \frac{1}{2m} \left\| \nabla f(x^{(k)}) \right\|^2.$$

En combinant avec l'inégalité (2.4), on obtient

$$2m(f(x^k) - f(x^*)) \leq \left\| \nabla f(x^{(k)}) \right\|^2 \leq 2M(f(x^{(k)}) - f(x^{(k+1)})),$$

de sorte qu'en posant $c = M/m$, on obtient bien l'inégalité voulue. □

## 3.3 Une condition suffisante pour la forte convexité

**Proposition 20.** *Soit $f \in \mathcal{C}^2(\mathbb{R}^d)$ et $S \subseteq \mathbb{R}^d$ compact tel que $\forall x \in S$, $\mathrm{D}^2 f(x)$ est définie positive. Alors $\exists M \geq m > 0$ tels que*

$$\forall x \in S, m\mathrm{Id} \preceq \mathrm{D}^2 f(x) \preceq M\mathrm{Id}.$$

*Démonstration.* L'ensemble $K = \{(x, v) \in S \times \mathbb{R}^d \mid \|v\| = 1\}$ est compact comme fermé borné. La fonction $(x, v) \in K \mapsto \langle \mathrm{D}^2 f(x)v|v\rangle$ est continue. Elle es donc bornée sur $K$ et atteint ses bornes : ainsi

$$\forall (x, v) \in K, m \leq \langle \mathrm{D}^2 f(x)v|v\rangle \leq M,$$

et il existe $(x_m, v_m) \in K$ tels que $m = \langle \mathrm{D}^2 f(x_m)v_m|v_m\rangle > 0$ par hypothèses et comme $v_0 \neq 0$ (car $\|v_0\| = 1$). On en déduit l'inégalité voulue en remarquant que

$$\forall w \in \mathbb{R}^d, \langle \mathrm{D}^2 f(x)w|w\rangle = \langle \mathrm{D}^2 f(x)\frac{w}{\|w\|}|\frac{w}{\|w\|}\rangle \|w\|^2,$$

de sorte que $m\|w\|^2 \leq \langle \mathrm{D}^2 f(x)w|w\rangle \leq M\|w\|^2$ comme souhaité. □

Cette proposition permet d'énoncer une variante non-quantitative du théorème 19

**Corollaire 21.** *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) $\forall x \in S$, $\mathrm{D}^2 f(x)$ est définie positive.*
*Alors les itérées de l'algorithme (2.1) convergent vers l'unique minimiseur global $x^*$ de $f$ sur $\mathbb{R}^d$, et de plus $(f(x^{(k)}))_{k\geq 0}$ converge linéairement vers $f(x^*)$.*

# Chapitre 4

# Descente de gradient préconditionné à rebroussement

Une des objections à l'algorithme de descente de gradient à pas optimal est le calcul du pas demande de résoudre un problème d'optimisation 1D à chaque itération. Nous verrons dans ce chapitre une manière simple de choisir le pas qui donne les même garanties de convergence. Nous allons également un peu relaxer l'hypothèse que $d^{(k)} = -\nabla f(x^{(k)})$ afin de préparer le terrain à l'analyse de l'algorithme de Newton dans le chapitre suivant. Nous supposerons donc que la direction de descente est définie de la manière suivante, où $B^{(k)}$ est une matrice symétrique définie positive

$$d^{(k)} = -B^{(k)}\nabla f(x^{(k)}) \tag{4.1}$$

Le fait que $B^{(k)}$ est définie positive garantit que $\langle d^{(k)} | \nabla f(x^{(k)}) \rangle < 0$, c'est-à-dire que $d^{(k)}$ est bien une direction de descente.

*Remarque* 10. Une méthode de descente où la direction $d^{(k)}$ est définie par la relation (4.1) est souvent appelé *descente de gradient préconditionné*, et la matrice $B^{(k)}$ est appelée *préconditionneur*.

## 4.1 Choix du pas par rebroussement

### 4.1.1 Rebroussement naïf

Comme nous souhaitons construire une méthode de descente, il serait assez naturel de considérer le pas suivant,

$$t^{(k)} = \max\{t \geq 0 \mid \exists \ell \in \mathbb{N}, t = 2^{-\ell} \text{ et } f(x^{(k)} + td^{(k)}) \leq f(x^{(k)})\}.$$

Autrement dit étant donné $x^{(k)}, d^{(k)}$ on teste les pas de $1, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^\ell}$ et on s'arrête dès que que la condition de descente $f(x^{(k)} + 2^{-\ell}d^{(k)}) \leq f(x^{(k)})$ est satisfaite. Cet algorithme, assez naturel, est en fait non-convergent en général. Considérons $f(x) = x^2$ sur $\mathbb{R}$ et $x^{(0)} = 1$, $d^{(k)} = -f'(x^{(k)}) = -2x^{(k)}$. Alors le pas $t^{(k)} = 1$ est admissible et on obtient donc $x^{(k)} = (-1)^k$ qui ne converge pas vers le minimum de $f$.

### 4.1.2 Rebroussement d'Armijo

L'idée est de renforcer la condition de descente $f(x^{(k)}+t^{(k)}d^{(k)}) \leq f(x^{(k)})$ par la condition plus forte (4.2) appellée "condition d'Armijo" :

**Lemme 22.** *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$, $v \in \mathbb{R}^d$ tel que $\langle \nabla f(x)|v\rangle < 0$, et $0 < \alpha < \frac{1}{2}$. Alors, il existe $t_0 > 0$ tel que $\forall t \in [0, t_0]$,*

$$f(x+tv) \leq f(x) + \alpha t \langle \nabla f(x)|v\rangle. \tag{4.2}$$

*Démonstration.* C'est une conséquence immédiate de l'égalité

$$f(x+tv) = f(x) + \alpha t\langle \nabla f(x)|v\rangle + (1-\alpha)t\langle \nabla f(x)|v\rangle + \mathrm{o}(t). \qquad \square$$

**Définition 12** (Pas d'Armijo)**.** Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$. Étant donné $x^{(k)}$, $d^{(k)}$ on appellera *pas d'Armijo* de paramètres $0 < \alpha < \frac{1}{2}$ et $0 < \beta < 1$ le pas $t^{(k)}$ défini par

$$\begin{aligned} t^{(k)} = \max\{t \mid \ell \in \mathbb{N}, t = \beta^\ell, \\ f(x^{(k)} + td^{(k)}) \leq f(x^{(k)}) + t\alpha\langle \nabla f(x^{(k)})|d^{(k)}\rangle\} \end{aligned} \tag{4.3}$$

*Remarque* 11. En d'autres termes, le pas d'Armijo peut être calculé par la procédure

```
pas_armijo(x^(k), d^(k)) :
    t ← 1
    tant que f(x^(k) + t d^(k)) > f(x^(k)) + t α ⟨∇f(x^(k)) | d^(k)⟩ :
        t ← β t
    retourner t
```

**Définition 13** (Méthode de descente avec pas d'Armijo)**.** On appelle algorithme de descente de gradient préconditionné la méthode itérative suivante, où pour tout $k \in \mathbb{N}$, $B^{(k)} = A^{(k)}A^{(k)}$ et $A^{(k)}$ est une matrices symétriques définies positives :

$$\begin{cases} d^{(k)} = -B^{(k)}\nabla f(x^{(k)}) \\ t^{(k)} = \text{ donné par (4.3)} \\ x^{(k)} = x^{(k)} + t^{(k)}d^{(k)}. \end{cases}$$

## 4.2 Convergence de l'algorithme de descente de gradient préconditionné à rebroussement

**Théorème 23.** *Soit $\Omega \subseteq \mathbb{R}^d$ un ouvert convexe et $f \in \mathcal{C}^2(\Omega)$ vérifiant*
*(i) le sous-niveau $S = \{x \in \Omega \mid f(x) \leq f(x^{(0)})\}$ est compact.*
*(ii) $\exists 0 < \lambda < \Lambda,\ \forall x \in S, \forall k \in \mathbb{N},\ \lambda \mathrm{Id} \preceq A^{(k)} \mathrm{D}^2 f(x) A^{(k)} \preceq \Lambda \mathrm{Id}$,*
*Alors les itérées $x^{(k)}$ de l'algorithme de descente avec pas d'Armijo (Déf. 13) convergent vers l'unique minimiseur global de $f$ sur $\Omega$, et de plus, en posant $c = 2\alpha\lambda \min(1, \beta/\Lambda)$,*

$$f(x^{(k+1)}) - f(x^*) \leq (1 - c)\left(f(x^{(k)}) - f(x^*)\right). \tag{4.4}$$

*Remarque* 12. Cette analyse suggère une manière simple d'améliorer l'algorithme de descente de gradient : il suffit de choisir $B^{(k)}$ tel que le conditionnement $\Lambda/\lambda$ soit aussi proche de 1 que possible. Dans le cas $f(x_1, x_2) = Kx_1^2 + x_2^2$, dont la hessienne est $H = \mathrm{diag}(K, 1)$ (constante), l'idéal est bien sûr de prendre $B = A^2$ où $A = \mathrm{diag}(K^{-1/2}, 1)$, soit $B = H^{-1}$. Plus généralement, le choix $B^{(k)} = \mathrm{D}^2 f(x^{(k)})^{-1}$ est souvent judicieux : il donne ce qu'on appelle la "méthode de Newton".

**Lemme 24.** *Soit $f \in \mathcal{C}^2(\Omega)$ et $B^{(k)} = (A^{(k)})^2$ telles que que*

$$\forall x \in S,\ A^{(k)} \mathrm{D}^2 f(x) A^{(k)} \preceq \Lambda \mathrm{Id}.$$

*Alors, le pas d'Armijo défini par (4.3) vérifie*

$$t^{(k)} \geq \min\left(1, \frac{\beta}{\Lambda}\right).$$

*Remarque* 13. Cette inégalité permet de dire que l'algorithme `pas_armijo` calculant le pas d'Armijo termine dès que $\beta^k \leq \beta/\Lambda$. En d'autres termes, le nombre d'itération de l'algorithme de recherche du pas est au plus $\log(\Lambda/\beta)/\log(1/\beta)$.

*Démonstration.* Étant donné $t \in \mathbb{R}$, on pose $x_t = x^{(k)} + td^{(k)}$. Si $t$ est tel que $x_t \in S$, alors le segment $[x_0, x_t]$ est inclus dans $S$ et on montre donc comme dans la preuve du théorème 15 que

$$f(x_t) = f(x^{(k)}) + t\langle \nabla f(x^{(k)}) | d^{(k)} \rangle + \frac{t^2}{2} \langle D^2 f(x_s) d^{(k)} | d^{(k)} \rangle$$

Par (4.1), on a $d^{(k)} = -B^{(k)} \nabla f(x^{(k)}) = -A^2 \nabla f(x^{(k)})$ où on a noté $A = A^{(k)}$. Ainsi,

$$\begin{aligned}
\langle D^2 f(x_s) d^{(k)} | d^{(k)} \rangle &= \langle D^2 f(x_s) A^2 \nabla f(x^{(k)}) | A^2 \nabla f(x^{(k)}) \rangle \\
&= \langle A D^2 f(x_s) A (A \nabla f(x^{(k)})) | A \nabla f(x^{(k)}) \rangle \\
&\leq \Lambda \left\| A \nabla f(x^{(k)}) \right\|^2 = \Lambda \langle A^2 \nabla f(x^{(k)}) | \nabla f(x^{(k)}) \rangle = -\langle \nabla f(x^{(k)}) | d^{(k)} \rangle
\end{aligned}$$

de sorte que

$$f(x_t) \leq f(x^{(k)}) + (1 - \frac{\Lambda}{2}t)t\langle \nabla f(x^{(k)}) | d^{(k)} \rangle.$$

Ainsi, la condition d'Armijo

$$f(x_t) \leq f(x^{(k)}) + \alpha t \langle \nabla f(x^{(k)}) | d^{(k)} \rangle$$

est vérifiée dès que $1 - \frac{\Lambda}{2}t \geq \alpha$. Comme $\alpha \leq \frac{1}{2}$, il suffit pour cela que $t \leq \Lambda$. □

**Lemme 25.** *Soit $X \subseteq \Omega \subseteq \mathbb{R}^d$ avec $X$ convexe et $\Omega$ ouvert. Soit $f \in \mathcal{C}^2(\Omega)$ et $A$ une matrice symétrique définie positive vérifiant $\forall x \in X, A\mathrm{D}^2 f(x)A \geq \lambda \mathrm{Id}$. Alors*

$$2\lambda(f(x) - f(x^*)) \leq \|A\nabla f(x)\|^2 .$$

*Démonstration.* On considère $g(y) = f(Ay)$. Alors $\nabla g(y) = A\nabla f(Ay)$ et $\mathrm{D}^2 g(y) = A\mathrm{D}^2 f(Ay)A$. On a supposé $A\mathrm{D}^2 f(x)A \geq \lambda \mathrm{Id}$, et par le corollaire (18) on a

$$2\lambda(g(y) - g(y^*)) \leq \|\nabla g(y)\|^2 ,$$

où $y^* = A^{-1}x^*$. Ainsi, en posant $y = A^{-1}x$, $2\lambda(f(x) - f(x^*)) \leq \|A\nabla f(x)\|^2$. □

*Démonstration du théorème 23.* Par définition du pas d'Armijo, on sait que

$$\begin{aligned} f(x^{(k+1)}) &\leq f(x^{(k)}) + \alpha t^{(k)} \langle \nabla f(x^{(k)}) | d^{(k)} \rangle \\ &= f(x^{(k)}) - \alpha t^{(k)} \langle \nabla f(x^{(k)}) | B^{(k)} \nabla f(x^{(k)}) \rangle \\ &= f(x^{(k)}) - \alpha t^{(k)} \langle A^{(k)} \nabla f(x^{(k)}) | A^{(k)} \nabla f(x^{(k)}) \rangle \\ &\leq f(x^{(k)}) - \varepsilon \left\| A^{(k)} \nabla f(x^{(k)}) \right\|^2 \text{ où } \varepsilon = \alpha \min(1, \beta/\Lambda) \end{aligned}$$

où on a utilisé le lemme 24 pour minorer $t^{(k)}$. En combinant avec le lemme précédent on obtient

$$\begin{aligned} &f(x^{(k+1)}) \leq f(x^{(k)}) - 2\lambda\varepsilon(f(x^k) - f(x^*)) \\ &f(x^{(k+1)}) - f(x^*) \leq (1 - c)(f(x^{(k)}) - f(x^*)) \text{ où } c = 2\lambda\varepsilon. \end{aligned}$$

□

# Chapitre 5

# Méthode de Newton

## 5.1 Méthode de Newton pure

### 5.1.1 Construction des itérées

Soit $f \in \mathcal{C}^2(\mathbb{R}^d)$, vérifiant $\forall x \in \mathbb{R}^d, \mathrm{D}^2 f(x) \succ 0$. La méthode de Newton est une méthode itérative, qui repose sur le développement de Taylor à l'ordre 2 de la fonction au point courant $x^{(k)}$ : $f(x^{(k)} + d) = g(d) + \mathrm{o}(\|d\|^2)$ où

$$g(d) = f(x^{(k)}) + \langle d | \nabla f(x^{(k)}) \rangle + \frac{1}{2} \langle \mathrm{D}^2 f(x^{(k)}) d | d \rangle.$$

Comme par hypothèse $\mathrm{D}^2 f(x^{(k)}) \succ 0$, la fonction $g$ est strictement convexe et admet un unique minimum, que l'on notera $d^{(k)}$. Ce minimum vérifie

$$\nabla g(d^{(k)}) = \mathrm{D}^2 f(x^{(k)}) d^{(k)} + \nabla f(x^{(k)}) = 0,$$

soit $d^{(k)} = -[\mathrm{D}^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)})$. On arrive à la définition suivante :

**Définition 14.** Les itérées de la méthode de Newton pure sont données par

$$\begin{cases} d^{(k)} = -[\mathrm{D}^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)}) \\ t^{(k)} = 1 \\ x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)} \end{cases}$$

Le terme "pur" fait référence au choix du pas $t^{(k)} = 1$, par opposition à la méthode de Newton amortie, introduite dans §5.2.

*Remarque* 14. La direction $d^{(k)}$ est appelée *direction de Newton*. Il s'agit d'une direction de descente si car $\langle \nabla f(x^{(k)}) | d^{(k)} \rangle = -\langle \nabla f(x^{(k)}) | \mathrm{D}^2 f(x^{(k)}) \nabla f(x^{(k)}) \rangle > 0$, mais en général il est possible que $f(x^{(k+1)}) \not\leq f(x^{(k)})$. La méthode de Newton pure n'est donc pas une méthode de descente.

*Remarque* 15. Dans le cas $d = 1$. Posons $h(x) = f'(x)$. La méthode de Newton s'écrit alors sous la forme

$$x^{(k+1)} = x^{(k)} - f'(x^{(k)}) / f''(x^{(k)}) = x^{(k)} - h(x^{(k)}) / h'(x^{(k)}).$$

On reconnaît alors la méthode de Newton "classique" permettant de trouver un zéro de la fonction $h = f'$, ce qui revient dans ce cas à trouver un minimum de la fonction convexe $f$.

### 5.1.2 Convergence quadratique locale

**Définition 15** (Convergence quadratique)**.** Soit $(u_k)_{k\geq 0}$ une suite de limite $u^*$. La suite $(u_k)$ converge *quadratiquement* vers $u^*$ s'il existe $\gamma > 0$ telle que

$$\|u_{k+1} - u^*\| \leq \gamma \|u_k - u^*\|^2 .$$

**Théorème 26.** *Soit $f \in \mathcal{C}^3(\mathbb{R}^d)$ vérifiant*
*(i) $f$ admet un minimiseur $x^*$ sur $\mathbb{R}$ ;*
*(ii) $\forall x \in \mathbb{R}^d, \mathrm{D}^2 f(x) \succ 0$.*
*Alors, il existe $r > 0$ tel que pour tout $x^{(0)} \in \mathrm{B}(x^*, r)$, la suite $(x^{(k)})$ construite par la méthode de Newton pure est définie pour tout $k \geq 0$ et converge quadratiquement vers $x^*$.*

*Remarque* 16. Soit $\varepsilon_k = \gamma \left\|x^{(k)} - x^*\right\|$, où $\gamma$ est la constante dans la définition de convergence quadratique. Alors, $\varepsilon_{k+1} \leq \varepsilon_k^2$, de sorte que $\varepsilon_k \leq \varepsilon_0^{2^k}$. Ainsi, si l'on souhaite une erreur $\varepsilon_k \leq \eta = 10^{-15}$, il suffit que $\varepsilon_0^{2^k} \leq \eta$ soit $2^k \log_2(\varepsilon_0) \leq \log_2(\eta)$. Ainsi, si $\varepsilon = 1/2$ et en prenant la minoration $\eta \geq 2^{-50}$, il suffit que $2^k \geq 50$, soit $k = 6$.

*Démonstration.* Soit $x^*$ l'unique minimiseur de $f$ sur $\mathbb{R}^d$ et $R > 0$. Comme $f$ est de classe $\mathcal{C}^3$, il existe une constante $L$ telle que $\forall x, x' \in K \cap \mathrm{B}(x^*, R), \left\|\mathrm{D}^2 f(x) - \mathrm{D}^2 f(x')\right\| \leq L \|x - x'\|$. Soit $x \in K$ et $x_t = (1-t)x^* + tx = x^* + t(x - x^*)$. Alors,

$$\begin{aligned}
\nabla f(x) &= \nabla f(x^*) + \int_0^1 \mathrm{D}^2 f(x_t)(x - x^*)\mathrm{d}t \\
&= \mathrm{D}^2 f(x)(x - x^*) + \int_0^1 (\mathrm{D}^2 f(x_t) - \mathrm{D}^2 f(x))(x - x^*)\mathrm{d}t \\
&= \mathrm{D}^2 f(x)(x - x^*) + R(x, x^*)
\end{aligned}$$

où

$$\|R(x, x^*)\| \leq L \|x - x^*\| \int_0^1 \|x_t - x\| \,\mathrm{d}t = \frac{L}{2} \|x - x^*\|^2$$

On considère maintenant l'application $N : x \in \mathbb{R}^d \mapsto x - \mathrm{D}^2 f(x)^{-1} \nabla f(x)$. Les itérées de l'algorithme de Newton vérifient $x^{(k+1)} = N(x^{(k)})$. On a de plus pour tout $x \in K$,

$$\|N(x) - x^*\| = \left\|x - \mathrm{D}^2 f(x)^{-1} \nabla f(x) - x^*\right\| \leq \left\|\mathrm{D}^2 f(x)^{-1}\right\| \|R(x, x^*)\|^2 \leq \frac{L}{2m} \|x - x^*\|^2 .$$

Pour que l'algorithme de Newton soit bien défini, on chercher $0 < r \leq R$ telle que $N(\mathrm{B}(x^*, r)) \subseteq \mathrm{B}(x^*, \min(r, R))$. Pour cela, il suffit que $\frac{L}{2m} r^2 \leq \min(r, R)$, et on peut donc prendre $r < \min(2m/L, \sqrt{2mR/L})$.

Si $x^{(0)} \in \mathrm{B}(x^*, r)$, on a alors $x^{(1)} = N(x^{(0)}) \in \mathrm{B}(x^*, r)$, et par récurrence la suite des itérées $x^{(k)}$ est bien définie pour tout $k \in \mathbb{N}$. De plus, on a bien

$$\left\|x^{(k+1)} - x^*\right\| = \left\|N(x^{(k)}) - x^*\right\| \leq \frac{L}{2m} \left\|x^{(k)} - x^*\right\|^2.$$

Soit $\varepsilon_k = \frac{L}{2m} \left\|x^{(k)} - x^*\right\|$, de sorte que $\varepsilon_{k+1} \leq \varepsilon_k^2$ soit par récurrence $\varepsilon_k \leq \varepsilon_0^{2^k}$. De plus, comme $r < 2m/L$, on a $\varepsilon_0 < 1$, de sorte que $\lim_{k \to +\infty} \varepsilon_k = 0$. Ceci prouve la convergence quadratique de la suite $x^{(k)}$ vers $x^*$. □

### 5.1.3 Invariance par reparamétrisation affine

Un autre avantage de la méthode de Newton que contrairement à la méthode de descente de gradient, elle est invariante par reparamétrisation affine (c'est-à-dire qu'elle ne dépend pas de la base choisie pour l'espace).

**Proposition 27.** *Soit $f : \mathbb{R}^d \to \mathbb{R}$, $A$ une matrice carrée inversible et $g : A^{-1}\Omega \to \mathbb{R}^d$ définie par $g(y) = f(Ay)$ . Soient maintenant*

$$x^{(k+1)} = x^{(k)} - \mathrm{D}^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

$$y^{(k+1)} = y^{(k)} - \mathrm{D}^2 g(y^{(k)})^{-1} \nabla g(y^{(k)})$$

$$\tilde{x}^{(k)} = A y^{(k)}.$$

*Si l'on suppose de plus que $x^{(0)} = \tilde{x}^{(0)}$, alors $\forall k \in \mathbb{N}$, $\tilde{x}^{(k)} = x^{(k)}$.*

*Démonstration.* Calculons d'abord le gradient et la hessienne de $g$, en les identifiants dans le développement de Taylor à l'ordre 2 :

$$\begin{aligned} g(y+v) = f(A(y+v)) &= f(Ay) + \langle \nabla f(Ay) | Av \rangle + \frac{1}{2} \langle \mathrm{D}^2 f(Ay) Av | Av \rangle + \mathrm{o}(\|Av\|^2) \\ &= g(y) + \langle A^T \nabla f(Ay) | v \rangle + \frac{1}{2} \langle A^T \mathrm{D}^2 f(Ay) Av | v \rangle + \mathrm{o}(\|v\|^2) \end{aligned}$$

On trouve donc $\nabla g(y) = A^T \nabla f(Ay)$ et $\mathrm{D}^2 g(y) = A^T \mathrm{D}^2 f(Ay) A$, ce qui donne

$$\begin{aligned} y^{(k+1)} &= y^{(k)} - \mathrm{D}^2 g(y^{(k)})^{-1} \nabla g(y^{(k)}) \\ &= y^{(k)} - (A^T \mathrm{D}^2 f(Ay^{(k)}) A)^{-1} A^T \nabla f(Ay^{(k)}) \\ &= y^{(k)} - A^{-1} \mathrm{D}^2 f(Ay^{(k)})^{-1} \nabla f(Ay^{(k)}) \end{aligned}$$

Ainsi, en multipliant cette égalité par $A$ on obtient

$$\tilde{x}^{(k+1)} = \tilde{x}^{(k)} - \mathrm{D}^2 f(\tilde{x}^{(k)})^{-1} \nabla f(\tilde{x}^{(k)}).$$

□

### 5.1.4 Non-convergence globale

Dans ce paragraphe, nous construisons un exemple explicite de fonction pour laquelle la méthode de Newton "pure" ne converge pas pour tout $x^{(0)}$. Considérons $f : x \in \mathbb{R} \mapsto \sqrt{x^2+1}$. Alors,

$$f'(x) = \frac{x}{\sqrt{1+x^2}} \quad f''(x) = \frac{1}{(\sqrt{x^2+1})^{3/2}},$$

de sorte que $f$ est convexe et même $\frac{1}{\sqrt{r^2+1}^{3/2}}$-fortement convexe sur l'intervalle $[-r, r]$. L'unique minimiseur de $f$ sur $\mathbb{R}$ est $x^* = 0$.

Calculons maintenant les itérées de la méthode de Newton : $d^{(k)}$ est définie par l'équation

$$f''(x^{(k)})d^{(k)} = -f'(x^{(k)}),$$
$$d^{(k)} = x^{(k)}((x^{(k)})^2 + 1)$$

Ainsi, l'itérée $x^{(k+1)}$ est définie par

$$x^{(k+1)} = x^{(k)} + d^{(k)} = (x^{(k)})^3$$

La suite $x^{(k)}$ définie par cette relation peut avoir trois comportements. Si $\left|x^{(0)}\right| < 1$, alors la suite $x^{(k)}$ converge vers $0 = x^*$ (avec une vitesse cubique !). Si $\left|x^{(0)}\right| > 0$, alors la suite $(\left|x^{(k)}\right|)$ tend vers $+\infty$, là encore très vite. Si $\left|x^{(k)}\right| = 1$, la suite est stationnaire en $1 \neq x^*$ (si $x^{(0)} = 1$) ou alterne entre les deux valeurs $\pm 1$.

## 5.2 Méthode de Newton amortie

Nous allons voir dans cette partie qu'une modification simple la méthode de Newton pure permet de la rendre globalement convergence.

**Définition 16.** On appelle algorithme Newton amortie la méthode itérative

$$\begin{cases} d^{(k)} = -\mathrm{D}^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}) \\ t^{(k)} = \text{pas\_armijo}(x^{(k)}, d^{(k)}) \\ x^{(k)} = x^{(k)} + t^{(k)} d^{(k)}. \end{cases},$$

où pas_armijo a été défini dans le chapitre précédent.

La méthode de Newton amortie est un cas particuler d'algorithme de descente de gradient préconditionné, où l'on a choisi comme préconditionneur $B^{(k)} = \mathrm{D}^2 f(x^{(k)})$.

**Théorème 28.** *Soit $f \in \mathcal{C}^3(\mathbb{R}^d)$ vérifiant*
*(i) le sous-niveau $S = \{x \in \mathbb{R}^d \mid f(x) \leq f(x^{(0)})\}$ est compact.*
*(ii) $\forall x \in S,\ \mathrm{D}^2 f(x) \succ 0$.*
*Alors les itérées $x^{(k)}$ de la méthode de Newton amortie convergent vers l'unique minimiseur global de $x^*$ de $f$ sur $\mathbb{R}^d$. En outre, il existe $k_0 \in \mathbb{N}$ et une constante $\gamma > 0$ tel que $\gamma \left\|x^{k_0} - x^*\right\| < 1$ et*

$$\forall k \geq k_0, \quad \left\|x^{k+1} - x^*\right\| \leq \gamma \left\|x^k - x^*\right\|^2.$$

*Remarque* 17. Un intérêt pratique de cet algorithme est que le choix de $t^{(k)}$ par rebroussement d'Armijo permet de "automatiquement" de passer d'un régime où la convergence est linéaire ($k \leq k_0$), et expliquée par l'analyse de la méthode de gradient préconditionnée et un second régime ($k \geq k_0$) où on observe une convergence quadratique.

La preuve du théorème se fait en deux étapes. Le lemme suivant permet de vérifier les hypothèses du théorème (23) sur la convergence des méthodes de gradient préconditionné. On obtient alors que $\lim_{k\to+\infty} x^{(k)} = x^*$.

**Lemme 29.** *Pour tout $k \in \mathbb{N}$, il existe une matrice symétrique $A^{(k)} \succ 0$ telle que $(A^{(k)})^2 = B^{(k)}$. De plus, il existe $0 < \lambda < \Lambda$ tels que*

$$\forall k \in \mathbb{N}, \forall x \in S, \quad \lambda \mathrm{Id} \leq A^{(k)} \mathrm{D}^2 f(x) A^{(k)} \preceq \Lambda \mathrm{Id}$$

*Démonstration.* Étant donné $k \in \mathbb{N}$, on définit $B^{(k)} = \mathrm{D}^2 f(x) \succ 0$. Il existe donc une matrice orthogonale $P$ et une matrice diagonale $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ telle que $B^{(k)} = P^T D P$. La matrice $A^{(k)} = P^T \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_d^{1/2}) P$ est symétrique définie positive et vérifie $(A^{(k)})^2 = B^{(k)}$. De plus,

$$\langle A^{(k)} \mathrm{D}^2 f(x) A^{(k)} v | v \rangle = \langle \mathrm{D}^2 f(x) A^{(k)} v | A^{(k)} v \rangle.$$

Or, d'après la proposition 20, nous savons qu'il existe $0 < m < M$ tels que $\forall x \in S, m\mathrm{Id} \preceq \mathrm{D}^2 f(x) \preceq M\mathrm{Id}$, de sorte que

$$m \left\|A^{(k)} v\right\|^2 \leq \langle A^{(k)} \mathrm{D}^2 f(x) A^{(k)} v | v \rangle \leq M \left\|A^{(k)} v\right\|^2.$$

Or,

$$\left\|A^{(k)} v\right\|^2 = \langle A^{(k)} v | A^{(k)} v \rangle = \langle (A^{(k)})^2 v | v \rangle = \langle \mathrm{D}^2 f(x^{(k)}) v | v \rangle,$$

d'où

$$m^2 \|v\|^2 \leq \langle A^{(k)} \mathrm{D}^2 f(x) A^{(k)} v | v \rangle \leq M \|v\|^2.$$

et on peut donc prendre $\lambda = m^2$, $\Lambda = M^2$. □

On montre ensuite que pour $k \gg 1$, on a nécessairement $t^{(k)} = 1$, de sorte que les itérées de la méthode de Newton amortie coïncident alors avec celles de Newton pure pour $k \gg 1$. Par le théorème 26, nous aurons donc démontré le résultat de convergence quadratique.

**Lemme 30.** *Il existe* $k_0 \in \mathbb{N}$ *tel que* $\forall k \geq k_0, t^{(k)} = 1$.

*Démonstration.* Admis. □

# Chapitre 6

# Projection et gradient projeté

On s'intéresse désormais à des problèmes d'optimisation sous contraintes :

$$\min_{x \in K} f(x) \tag{6.1}$$

où $K \subseteq \mathbb{R}^d$ est un ensemble convexe fermé et $f : \mathbb{R}^d \to \mathbb{R}$ est convexe.

Dans le cas de l'optimisation sous contrainte ($K$ fermé),

$$x^* \in \arg\min_K f \not\Longrightarrow \nabla f(x^*) = 0.$$

*Exemple* 6. $K = [1, 2] \subseteq \mathbb{R}$ et $f(x) = x^2$. Le minimum de $f$ sur $[1, 2]$ est atteint au point $x^* = 1$, mais évidemment $f'(x^*) = 2 \neq 0$.

Pour pouvoir écrire des algorithmes d'optimisation, il faut d'abord comprendre la conditions d'optimalité.

**Proposition 31** (Existence). *Le problème d'optimisation sous contraintes* (6.1) *admet une solution si une des conditions suivantes est vérifiée :*

*(i) $K$ est compact, non vide, et $f$ est continue ;*

*(ii) $K$ est fermé, non vide, $f$ est continue et $\lim_{\|x\| \to +\infty} f(x) = +\infty$ ;*

**Proposition 32** (Unicité). *Si la fonction $f$ est strictement convexe et si l'ensemble $K$ est convexe, alors le problème d'optimisation sous contraintes* (6.1) *admet au plus une solution.*

## 6.1 Projection sur un convexe fermé

**Théorème 33** (Caractérisation de la projection). *Soit $K \subseteq \mathbb{R}^d$ convexe fermé, non vide. Alors, pour tout point $x \in \mathbb{R}^d$, le problème de minimisation*

$$\min_{q \in K} \|q - x\|^2 \tag{6.2}$$

*admet un unique minimiseur $p$. De plus, $p$ est caractérisé par la condition*

$$p \in K \text{ et } \forall q \in K, \langle x - p | p - q \rangle \geq 0. \tag{6.3}$$

**Définition 17** (Projection). Soit $K \subseteq \mathbb{R}^d$ convexe fermé non vide et $x \in \mathbb{R}^d$. L'unique minimiseur de (6.2) est appelé *projection de $x$ sur $K$* et noté $\mathrm{p}_K(x)$.

*Démonstration du théorème 33.* Soit $f(q) = \|x - q\|^2$, qui est convexe. Montrons que si $p$ est un minimiseur de (6.2), alors $p$ vérifie (6.3). Soit $q \in K$ et $q_t = (1-t)p + tq$ pour $t \in [0,1]$. Par convexité, $q_t \in K$. De plus, comme $p$ est un minimiseur de $f$ sur $K$, on a $f(q_t) \geq f(p)$, ce qui implique

$$\forall 0 < t < 1, \quad \frac{f(q_t) - f(p)}{t} \geq 0 \Longrightarrow \langle \nabla f(p) | q - p \rangle = 2\langle p - x | q - p \rangle \geq 0.$$

Réciproquement, considérons un point $\bar{p}$ vérifiant (6.3). Par convexité de $f$, on a

$$\forall q \in K, f(q) \geq f(p) + \langle q - p | \nabla f(p) \rangle = f(p) + 2 \underbrace{\langle q - p | p - x \rangle}_{\geq 0} \geq f(p).$$

Ainsi, $p$ minimise $f$ sur $K$ et est donc un minimiseur de (6.2). □

**Corollaire 34.** *Soit $K \subseteq \mathbb{R}^d$ convexe fermé non vide et $x, y \in \mathbb{R}^d$. Alors,*
*(i)* $\langle x - y | \mathrm{p}_K(x) - \mathrm{p}_K(y) \rangle \geq \|\mathrm{p}_K(x) - \mathrm{p}_K(y)\|^2$
*(ii)* $\|\mathrm{p}_K(x) - \mathrm{p}_K(y)\| \leq \|x - y\|$

*Démonstration.* Par le théorème précédent, avec $p = \mathrm{p}_K(x)$ et $q = \mathrm{p}_K(y)$ on a

$$\langle x - \mathrm{p}_K(x) | \mathrm{p}_K(x) - \mathrm{p}_K(y) \rangle \geq 0.$$

En inversant les rôles, on obtient

$$\langle y - \mathrm{p}_K(y) | \mathrm{p}_K(y) - \mathrm{p}_K(x) \rangle \geq 0.$$

En sommant ces deux inégalités on obtient (i). Le point (ii) s'obtient à partir de (i) en utilisant l'inégalité de Cauchy-Schwarz. □

*Exemple* 7. Soit $e \in \mathbb{R}^d$ non nul et $K = \{\lambda e \mid \lambda \in \mathbb{R}\}$. Alors,

$$\mathrm{p}_K(x) = \frac{\langle x|e\rangle}{\|e\|^2} e.$$

*Exemple* 8. Soit $K = \{x \in \mathbb{R}^d \mid Ax = b\}$ et $x \in \mathbb{R}^d$. Par (6.3), la projection de $x$ sur $K$ est caractérisée par

$$p = p_K(x) \iff p \in K \text{ et } \forall q \in K, \langle x - p|p - q\rangle \geq 0 \iff p \in K \text{ et } \forall v \in \mathrm{Ker}A, \langle x - p|v\rangle = 0$$
$$p \in K \text{ et } \iff x - p \in (\mathrm{Ker}A)^\perp.$$

de plus, $(\mathrm{Ker}A)^\perp = \mathrm{Im}A^T$ car

$$\begin{aligned} v \in \mathrm{Ker}A &\iff \forall w, \langle Av|w\rangle = 0 \\ &\iff \forall w, \langle v|A^T w\rangle = 0 \\ &\iff w \in (\mathrm{Im}A^T)^\perp, \end{aligned}$$

d'où l'on déduit $\mathrm{Ker}A = (\mathrm{Im}A^T)^\perp$ soit $(\mathrm{Ker}A)^\perp = \mathrm{Im}A^T$.

*Proposition* 35. *Si* $\mathrm{Ker}A^T = \{0\}$ *et* $K = \{x \mid Ax = b\}$, *alors*

$$\mathrm{p}_K(x) = x - A^T(AA^T)^{-1}(Ax - b).$$

*Démonstration.* Par la caractérisation précédente, $p = \mathrm{p}_K(x)$ si et seulement si $p \in K$ et si $x - p \in \mathrm{Im}A^T$, c'est-à-dire s'il existe $w$ tel que $p = x - A^T w$ et $Ap = b$. Le vecteur $w$ est donc caractérisé par

$$A(x - A^T)w = b, \quad \text{i.e. } AA^T w = Ax - b,$$

soit $w = (AA^T)^{-1}(Ax - b)$. Ceci donne la formule souhaitée pour $p$. □

*Exemple* 9. Soit $K = \{q \in \mathbb{R}^d \mid \forall 1 \leq i \leq d, q_i \geq 0\}$.

Si $d = 1$, alors $\mathrm{p}_K(x) = \max(x, 0) =: x^+$. Pour montrer cela, nous utilisons comme précédemment la caractérisation (6.3). Le point $p = x^+ \in K$ est la projection de $x$ sur $K$ si

$$\forall q \in K, \langle x - p|p - q\rangle = (x - x^+)(x^+ - q) \geq 0.$$

Or, si $x \geq 0$, on a $x - x^+ = 0$ de sorte que l'inégalité est vraie. Si $x < 0$, $x - x^+ < 0$ et $x^+ - q < 0$, de sorte que l'inégalité est aussi vraie. De manière générale, nous avons la proposition suivante :

*Proposition* 36. *Soit* $K = \{q \in \mathbb{R}^d \mid \forall 1 \leq i \leq d, q_i \geq 0\}$ *et* $x \in \mathbb{R}^d$. *Alors,*

$$\mathrm{p}_K(x) = (x_1^+, \ldots, x_d^+).$$

*Exemple* 10. Soit $K = K_1 \times \ldots \times K_\ell \subseteq \mathbb{R}^d$ où les $K_i \subseteq \mathbb{R}^{d_i}$ sont des convexes fermés non vides et où $d = d_1 + \ldots + d_\ell$. Alors,

$$\mathrm{p}_K(\underbrace{x_1}_{\in \mathbb{R}^{d_1}}, \ldots, \underbrace{x_\ell}_{\in \mathbb{R}^{d_\ell}}) = (\mathrm{p}_{K_1}(x_1), \ldots, \mathrm{p}_{K_\ell}(x_\ell))$$

*Exemple* 11. Soit $K = \{x \in \mathbb{R}^d \mid \|x\| \leq 1\}$ où $\|\cdot\|$ est la norme euclidienne. Alors,

$$\mathrm{p}_K(x) \begin{cases} x & \text{si } \|x\| \leq 1 \\ \frac{x}{\|x\|} & \text{sinon} \end{cases}.$$

Le cas où $x \in K$ est évident ; supposons donc que $x \notin K$ et posons $p := x/\|x\|$. Alors,

$$\forall q \in K, \langle x - p | p - q \rangle = \left( \frac{1}{\|x\|} - 1 \right) \langle x | p - q \rangle$$

Or, par Cauchy-Scharwz et en utilisant $\langle x|p\rangle = \|x\|$, $\langle x|p-q\rangle \geq \|x\| - \|x\|\,\|q\| \geq 0$. Par la caractérisation (6.3) on obtient comme souhaité $\mathrm{p}_K(x) = p$.

## 6.2 Condition d'optimalité pour l'optimisation sous contraintes

### 6.2.1 Condition nécessaire et suffisante d'optimalité

Dans cette partie, nous démontrons plusieurs CNS d'optimalité pour l'optimisation sous contraintes.

À nouveau, $x^*$ peut minimiser une fonction $f$ sur un compact $K$ alors que $\nabla f(x^*) \neq 0$!. Pour écrire des algorithmes d'optimisation corrects, nous devons quelle est la bonne condition nécessaire et suffisante d'optimalité.

**Théorème 37.** *Soit $K \subseteq \mathbb{R}^d$ un convexe fermé non vide et $f \in \mathcal{C}^1(\mathbb{R}^d)$ une fonction convexe. Alors, les affirmations suivantes sont équivalentes :*

*(i)* $x^* \in \arg\min_K f$
*(ii)* $\forall y \in K, \langle x^* - y | - \nabla f(x^*) \rangle \geq 0$ (formulation variationnelle)
*(iii)* $\exists \tau > 0,\ \forall y \in K, \mathrm{p}_K(x^* - \tau \nabla f(x^*)) = x^*$ (formulation géométrique)
*(iv)* $\forall \tau > 0,\ \forall y \in K, \mathrm{p}_K(x^* - \tau \nabla f(x^*)) = x^*$

*Remarque* 18. Ce théorème ne dit rien de l'existence ni ce l'unicité, il permet seulement de caractériser l'optimalité d'un point pour un problème d'optimisation sous contraintes.

*Démonstration.* (ii) $\Longrightarrow$ (i) : Par convexité de $f$, on sait que

$$\forall y \in \mathbb{R}^d,\ f(y) \geq f(x^*) + \langle y - x^* | \nabla f(x^*) \rangle.$$

Or, si $y \in K$, alors $\langle x^* - y | - \nabla f(x^*) \rangle \geq 0$ par hypothèse, de sorte que

$$\forall y \in \mathbb{R}^d,\ f(y) \geq f(x^*),$$

i.e. $x^*$ est le minimum de $f$ sur $K$.

(i) $\Longrightarrow$ (ii) : Soit $y \in K$ et $y_t = (1-t)x^* + ty$ pour $t \in [0,1]$. Par convexité de $K$, $y_t \in K$ pour tout $t \in [0,1]$. De plus, comme $x^*$ est un minimiseur de $f$ sur $K$, on a $f(y_t) \geq f(x^*)$, ce qui implique

$$\forall 0 < t < 1, \quad \frac{f(x^* + t(y-x^*)) - f(x^*)}{t} \geq 0.$$

En passant à la limite $t \to 0, t > 0$, on obtient $\langle \nabla f(x^*) | y - x^* \rangle \geq 0$.

(iv) $\Longleftrightarrow$ (iii) $\Longleftrightarrow$ (ii). Soit $t > 0$. Par caractérisation de la projection du point $x = x^* - t\nabla f(x^*))$ sur le convexe fermé $K$, on a

$$\begin{aligned} \mathrm{p}_K(x^* - t\nabla f(x^*)) = x^* &\Longleftrightarrow \forall y \in K, \langle (x^* - t\nabla f(x^*)) - x^* | x^* - y \rangle \geq 0 \\ &\Longleftrightarrow \forall y \in K, \langle -\nabla f(x^*) | x^* - y \rangle \geq 0, \end{aligned}$$

où on a utilisé $t > 0$ pour obtenir la deuxième équivalence. □

*Exemple* 12. Soit $K = \mathbb{R}^d_+ \subseteq \mathbb{R}^d$. On a démontré précédemment que la projection d'un point $x$ sur $K$ est donnée par

$$\mathrm{p}_K(x) = (\max(x_1, 0), \ldots, \max(x_d, 0)).$$

Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$ une fonction convexe. Combiné au théorème précédent, on obtient

$$\begin{aligned} x^* \in \arg\min_K f &\Longleftrightarrow \forall t > 0,\ \mathrm{p}_K(x^* - t\nabla f(x^*)) = x^*, \\ &\Longleftrightarrow \forall t > 0, \forall i, \max\left(x_i^* - t\frac{\partial f}{\partial e_i}(x^*), 0\right) = x_i^* \\ &\Longleftrightarrow \forall i, \forall t > 0, \max\left(x_i^* - t\frac{\partial f}{\partial e_i}(x^*), 0\right) = x_i^* \end{aligned}$$

On vérifie alors que cette propriété est équivalente à

$$\begin{cases} \frac{\partial f}{\partial e_i} = 0 & \text{si } x_i^* > 0 \\ \frac{\partial f}{\partial e_i} \geq 0 & \text{si } x_i^* = 0. \end{cases}$$

### 6.2.2 Algorithme du gradient projeté

**Définition 18.** Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$ et $K \subseteq \mathbb{R}^d$ un ensemble convexe fermé non vide. L'algorithme du gradient projeté à pas constant $\tau > 0$ est donné par :

$$x^{(k+1)} = \mathrm{p}_K(x^{(k)} - \tau\nabla f(x^{(k)})) \tag{6.4}$$

*Remarque* 19. Cette méthode ne mérite le nom d'algorithme que lorsqu'on sait calculer explicitement la projection sur l'ensemble $K$ !

**Théorème 38.** *On suppose que $f \in \mathcal{C}^2(\mathbb{R}^d)$, et que*

$$\exists 0 < m \leq M \text{ t.q. } \forall x \in \mathbb{R}^d, \quad m\mathrm{Id} \preceq \mathrm{D}^2 f(x) \preceq M\mathrm{Id}. \tag{6.5}$$

*Alors la suite définie par (6.4) converge vers l'unique $x^* \in \arg\min_K f$ si le pas $\tau$ vérifie*

$$0 < \tau < 2m/M^2.$$

On peut écrire $x^{(k+1)} = F(x^{(k)})$ où $F(x) := \mathrm{p}_K(x - \tau\nabla f(x))$. Si la suite $x^{(k)}$ converge vers un point $\bar{x} \in \mathbb{R}^d$, alors en passant à la limite dans l'équation $x^{(k+1)} = F(x^{(k)})$ et en utilisant la continuité de $F$, on déduit que

$$\bar{x} = F(\bar{x}) = p_K(\bar{x} - \tau\bar{x}),$$

de sorte que par le théorème 37, $\bar{x}$ minimise $f$ sur $K$. Il reste donc à démontrer que la suite $x^{(k)}$ définie par la récurrence (6.4) converge. Pour cela, nous appliquerons le théorème du point fixe contractant.

**Théorème 39** (Point fixe)**.** *Soit $F : \mathbb{R} \to \mathbb{R}^d$ une application contractante, c'est-à-dire qu'il existe $\kappa \in ]0,1[$ tel que $\|F(x) - F(y)\| \leq \kappa \|x - y\|$. Alors :*

*(i) $F$ admet un unique point fixe $x^*$ ;*

*(ii) pour tout $x^{(0)} \in \mathbb{R}^d$, la suite définie par $x^{(k+1)} = F(x^{(k)})$ converge vers $x^*$.*

Nous utiliserons également le lemme suivant.

**Lemme 40.** *Soit $f \in \mathcal{C}^2(\mathbb{R}^d)$ vérifiant (6.5). Alors, l'application $G : x \mapsto x - \tau\nabla f(x)$ vérifie*

$$\|G(x) - G(y)\|^2 \leq \left(1 - 2\tau m + M^2\tau^2\right) \|x - y\|^2$$

*Démonstration.* On applique la formule de Taylor avec reste intégral à l'application $h(t) = \nabla f(x_t)$, avec $x_t = x + tv$ et $v = y - x$, ce qui donne

$$h(1) - h(0) = \nabla f(y) - \nabla f(x) = \int_0^1 D^2 f(x_t) v \mathrm{d}t.$$

On en déduit que

$$\begin{aligned}
\|\nabla f(y) - \nabla f(x)\|^2 &= \left\| \int_0^1 D^2 f(x_t) v \mathrm{d}t \right\|^2 \\
&\leq \int_0^1 \left\| D^2 f(x_t) v \right\|^2 \mathrm{d}t \\
&\leq \int_0^1 \left\| D^2 f(x_t) \right\|^2 \|v\|^2 \mathrm{d}t \\
&\leq M^2 \|v\|^2 ,
\end{aligned}$$

où l'on utilise que pour une matrice symétrique la norme matricielle induite par le produit scalaire est égale au maximum des valeurs absolues des valeurs propres,qui est ici majoré par $M$. De plus, comme $\mathrm{D}^2 f(x) \succeq m\mathrm{Id}$ pour tout $x \in \mathbb{R}^d$,

$$\begin{aligned}\|\nabla f(y) - \nabla f(x)\| \, x - y = \langle \int_0^1 D^2 f(x_t) v \mathrm{d}t | v \rangle \\ &= \int_0^1 \langle D^2 f(x_t) v | v \rangle \mathrm{d}t \\ &\geq m \|v\|^2 .\end{aligned}$$

Ainsi, pour tout $x, y \in \mathbb{R}^d$,

$$\begin{aligned}\|G(x) - G(y)\|^2 &= \|x - y - \tau(\nabla f(x) - \nabla f(y))\|^2 \\ &= \|x-y\|^2 - 2\tau\langle x - y | \nabla f(x) - \nabla f(y)\rangle + \tau^2 \|\nabla f(x) - \nabla f(y)\|^2 \\ &\leq \left(1 - 2\tau m + M^2\tau^2\right) \|x - y\|^2\end{aligned}$$

□

*Démonstration du théorème 38.* On pose $F(x) = p_K(G(x))$ où $G(x) = x - \tau\nabla f(x)$, de sorte que $x^{(k+1)} = F(x^{(k)})$. En utilisant que la projection $\mathrm{p}_K$ est 1-Lipschitzienne, on a

$$\|F(x) - F(y)\| = \|\mathrm{p}_K(G(x)) - \mathrm{p}_K(G(y))\| \leq \|G(x) - G(y)\| .$$

En combinant avec le lemme précédent, on trouve

$$\|F(x) - F(y)\|^2 \leq \|G(x) - G(y)\|^2 \leq \left(1 - 2\tau m + M\tau^2\right) \|x - y\|^2 .$$

Cette application est contractante si

$$1 - 2\tau m + M^2\tau^2 \leq 1 \Longleftrightarrow \tau(M^2\tau - 2m),$$

ce qui est vrai si $\tau > 0$ et $\tau < 2m/M^2$. Par le théorème du point fixe, on en déduit que dans ce cas l'application $F$ possède un unique point fixe. Par le théorème 37, on sait qu'un point $\bar{x}$ minimise $f$ sur $K$ si et seulement si $F(\bar{x}) = \bar{x}$. On déduit de ce qui précède que le problème de minimisation $\min_K f$ possède une seule solution $x^*$, qui est l'unique point fixe de $F$. Par le (ii) du théorème du point fixe, la suite $x^{(k)}$ définie par $x^{(k)} = F(x^{(k)})$ converge vers $x^*$. □

# Chapitre 7

# Optimisation avec contraintes d'inégalités

Dans ce chapitre on s'intéresse à des problèmes d'optimisation $\min_K f$ où l'ensemble de contraintes $K$ est défini par des inégalités : étant données des fonctions convexes $c_1, \dots, c_\ell \in \mathcal{C}^1(\mathbb{R}^d)$, on définit l'ensemble $K$

$$K = \{x \in \mathbb{R}^d \mid c_1(x) \leq 0, \dots, c_\ell(x) \leq 0\}.$$

On appelle chacune des fonctions $c_1, \dots, c_\ell$ une *contrainte* du problème et on note souvent le problème par

$$P := \min_{c_1(x)\leq 0,\dots,c_\ell(x)\leq 0} f(x)$$

**Lemme 41.** *Si les fonctions $c_1, \dots, c_\ell : \mathbb{R}^d \to \mathbb{R}$ sont continues et convexes, alors l'ensemble $K = \{x \in \mathbb{R}^d \mid \forall 1 \leq i \leq \ell,\ c_i(x) \leq 0\}$ est fermé et convexe.*

*Démonstration.* Montrons d'abord que $K$ est fermé : soit $(x_n)_n$ une suite d'éléments de $K$ convergent vers une limite $x$. Par hypothèse, comme $x_n \in K$, $c_i(x_n) \leq 0$. En passant à la limite $n \to +\infty$ et en utilisant la continuité de $c_i$ on en déduit que $c_i(x) \leq 0$. Ainsi $x$ appartient à $K$, et $K$ est donc fermé.

On montre maintenant la convexité de $K$. Soient $x, y \in K$ et $x_t = (1-t)x + ty$. Comme $x, y \in K$, on a pour tout $i$, $c_i(x) \leq 0$ et $c_i(y) \leq 0$. Pour $t \in [0,1]$ on a donc

$$c_i(x_t) \leq (1-t)c_i(x) + tc_i(y) \leq 0,$$

de sorte que $x_t \in K$. Ceci démontre que $K$ est convexe. □

*Exemple* 13 (Simplexe). Considérons l'ensemble $K$ définit par

$$K = \left\{ x \in \mathbb{R}^d \mid \forall i, x_i \geq 0 \text{ et } \sum_{1\leq i\leq d} x_i = 1 \right\},$$

qui décrit l'ensemble des mesures de probabilités supportés sur $\{1, \dots, d\}$. On peut écrire cet ensemble sous la forme ci-dessus avec

$$c_1(x) = -x_1, \dots, c_d(x) = -x_d, c_{d+1}(x) = \sum_{1\leq i\leq d} x_i - 1, c_{d+2}(x) = -(\sum_{1\leq i\leq d} x_i - 1).$$

## 7.1 Méthode de pénalisation

L'idée de cette méthode est d'approcher le problème d'optimisation avec contraintes $\min_K f$ où $K = \{x \mid \forall i,\ c_i(x) \leq 0\}$ par un problème d'optimisation sans contraintes $\min_{\mathbb{R}^d} f_\varepsilon$ où $f_\varepsilon$ est la somme de $f$ et de termes qui "explosent" lorsque $c_i(x) > 0$. Précisément, pour $\varepsilon > 0$ on pose

$$P_\varepsilon := \min_{x\in\mathbb{R}^d} f_\varepsilon(x) \text{ où } f_\varepsilon(x) := f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{\ell} \max(c_i(x),0)^2.$$

Dans cette formulation du problème, les points $x \in \mathbb{R}^d$ tels que $c_i(x) > 0$ sont *pénalisés* au sens où si $\varepsilon$ est très petit, $\frac{1}{\varepsilon}\max(c_i(x),0)^2$ peut être très grand, ce qui dissuade le choix de ce point dans le problème d'optimisation. Lorsque $\varepsilon \to 0$, les points vérifiants sont "infiniment pénalisés" et deviennent en fait interdits :

$$\forall x \in \mathbb{R}^d, \quad \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\sum_{i=1}^{\ell} \max(c_i(x),0)^2 = \begin{cases} 0 & \text{si } c_i(x) \leq 0 \\ +\infty & \text{sinon} \end{cases}$$

Cette intuiton est précisée par la proposition suivante.

**Proposition 42.** *Supposons que $f \in \mathcal{C}^0(\mathbb{R}^d)$ est strictement convexe et vérifie $\lim_{\|x\|\to+\infty} f(x) = +\infty$. Alors :*

*(i) Les problèmes $P$ et $P_\varepsilon$ (pour $\varepsilon > 0$) admettent chacun un unique minimiseur, noté $x^*$ et $x^*_\varepsilon$.*

*(ii) $\lim_{\varepsilon\to 0} x^*_\varepsilon = x^*$.*

*Démonstration.* (i) L'existence d'une solution au problème d'optimisation sans contrainte $P_\varepsilon$ se déduit du fait que $f_\varepsilon$ est continue et que

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

tandis que l'existence de solution au problème avec contraintes $P$ se déduit de ce que $\lim_{\|x\|\to+\infty} f(x) = +\infty$ et que $K$ est fermé (Lemme 41). Pour l'unicité de la solution au problème $P$, il suffit de remarquer que $f$ est strictement convexe (hypothèse) et que $K$ est convexe (Lemme 41). Enfin, pour l'unicité de la solution au problème $P$ il faut montrer que $f_\varepsilon$ est strictement convexe. Soient $x, y \in \mathbb{R}^d$, $t \in [0,1]$ et $x_t = (1-t)x + ty$. Alors,

$$c_i(x_t) \leq (1-t)c_i(x) + tc_i(y)$$

de sorte que

$$\max(c_i(x_t),0) \leq \max((1-t)c_i(x)+tc_i(y),0) \leq (1-t)\max(c_i(x),0)+t\max(c_i(y),0).$$

Ainsi, en utilisant la convexité de $r \in \mathbb{R} \mapsto r^2$, on a

$$\begin{aligned}\max(c_i(x_t),0)^2 &\leq [(1-t)\max(c_i(x),0) + t\max(c_i(y),0)]^2 \\ &\leq (1-t)\max(c_i(x),0)^2 + t\max(c_i(y),0)^2\end{aligned}$$

On en déduit que les fonctions $x \mapsto \max(c_i(x), 0)^2$ sont convexes, et $f_\varepsilon$ est donc strictement convexe comme combinaison linéaire à coefficients positifs de fonctions convexes et d'une fonction strictement convexe.

(ii) Soit $x^* \in K$ le minimiseur de $P$. Alors, comme $c_i(x^*) \leq 0$, on a

$$P_\varepsilon \leq f_\varepsilon(x^*) = f(x^*) + \frac{1}{\varepsilon}\sum_{i=1}^{\ell} \max(c_i(x^*), 0)^2 = f(x^*) = P.$$

Ainsi,

$$P_\varepsilon = f(x_\varepsilon^*) + \frac{1}{\varepsilon}\sum_{1\leq i\leq \ell} \max(c_i(x_\varepsilon^*), 0)^2 \leq P.$$

Comme $\lim_{\|x\|\to+\infty} f(x) = +\infty$, la fonction $f$ est minorée, par exemple $f \geq m \in \mathbb{R}$. On déduit donc que

$$\forall i \in \{1, \dots, \ell\}, \quad \frac{1}{\varepsilon}\max(c_i(x_\varepsilon^*), 0)^2 \leq \frac{1}{\varepsilon}\sum_{1\leq j\leq \ell} \max(c_j(x_\varepsilon^*), 0)^2 \leq P - f(x_\varepsilon^*) \leq P - m,$$

ou encore

$$\max(c_i(x_\varepsilon^*), 0)^2 \leq \varepsilon(P - m).$$

Soit $\bar{x}$ une valeur d'adhérence de la suite $x_\varepsilon^*$ lorsque $\varepsilon \to 0$. Alors, par continuité,

$$\max(c_i(\bar{x}), 0)^2 \leq 0,$$

ce qui montre que $\bar{x} \in K$. De plus, pour tout $\varepsilon > 0$,

$$f(x_\varepsilon) \leq P_\varepsilon \leq P,$$

de sorte que $f(\bar{x}) \leq P$. On en déduit que $\bar{x}$ minimise $f$ sur $K$, soit $\bar{x} = x^*$. Pour conclure, on remarque que la suite $(x_\varepsilon^*)$ est bornée et admet une unique valeur d'adhérence, de sorte que $\lim_{\varepsilon\to 0} x_\varepsilon^* = x^*$. □

**Lemme 43.** *Soit $p : t \in \mathbb{R} \mapsto \max(t, 0)^2$. Alors $p \in \mathcal{C}^1(\mathbb{R})$ et $p'(t) = 2\max(t, 0)$*

*Démonstration.* Comme $p(t) = O(t^2)$, $p$ est différentiable en zéro et $p'(0) = 0$. D'autre part, sur $\mathbb{R}_+^*$, $p(t) = t^2$ et $p'(t) = 2t$, et sur $\mathbb{R}_-^*$, $p(t) = 0$, soit $p'(t) = 0$. Ainsi,

$$p'(t) = \begin{cases} 0 & \text{si } t < 0 \\ 0 & \text{si } t = 0 \\ 2t & \text{si } t > 0 \end{cases}$$

Ainsi, $p'(t)$ est continue et $p(t) = 2\max(t, 0)$. □

**Proposition 44.** *Supposons que $f \in \mathcal{C}^1(\mathbb{R}^d)$ est strictement convexe. Alors,*

$$x_\varepsilon^* \in \arg\min_{\mathbb{R}^d} f_\varepsilon \iff \nabla f(x_\varepsilon^*) + \frac{2}{\varepsilon}\sum_{i=1}^{\ell} \max(c_i(x_\varepsilon^*), 0)\nabla c_i(x_\varepsilon^*) = 0. \tag{7.1}$$

*Démonstration.* Nous avons déjà démontré que $f_\varepsilon$ est strictement convexe, et $f_\varepsilon \in \mathcal{C}^1(\mathbb{R}^d)$ par le lemme précédente, de sorte que par caractérisation de l'optimalité dans le problème d'optimisation *sans contrainte* $P_\varepsilon$,

$$x_\varepsilon^* \in \arg\min_{\mathbb{R}^d} f_\varepsilon \iff \nabla f_\varepsilon(x_\varepsilon^*) = 0.$$

D'autre part, en utilisant la fonction $p$ introduite dans le lemme précédent on peut écrire

$$f_\varepsilon(x) = f(x) + \frac{1}{\varepsilon} \sum_{1 \leq i \leq \ell} p(c_i(x)),$$

soit

$$\begin{aligned} \nabla f_\varepsilon(x) &= \nabla f(x) + \frac{2}{\varepsilon} \sum_{1 \leq i \leq \ell} p'(c_i(x)) \nabla c_i(x). \\ &= \nabla f(x) + \frac{2}{\varepsilon} \sum_{i=1}^{\ell} \max(c_i(x), 0) \nabla c_i(x) \end{aligned}$$

On en déduit l'équivalence annoncée. □

## 7.2 Théorème de Karush-Kush-Tucker

L'objet du reste de ce chapitre est d'énoncer et de démontrer le théorème de Karush-Kush-Tucker. Ce théorème permet de donner une condition nécessaire et suffisante d'optimalité pour les problèmes d'optimisation avec contraintes d'inégalité.

**Théorème 45** (Karush-Kush-Tucker)**.** *Soient $f, c_i : \mathbb{R}^d \to \mathbb{R}$ vérifiant :*
- *$f \in \mathcal{C}^1(\mathbb{R}^d)$ convexe ;*
- *$c_i(x) = \langle a_i | x \rangle - b_i$ pour $\forall i \in \{1, \dots, \ell\}$ ;*

*et soit*

$$K = \{x \in \mathbb{R}^d \mid \forall i \in \{1, \dots, \ell\},\ c_i(x) \leq 0\}.$$

*Alors,*

$$x^* \in \arg\min_K f \iff \exists \lambda \in \mathbb{R}^\ell \ t.q. \begin{cases} -\nabla f(x^*) = \sum_{i=1}^{\ell} \lambda_i \nabla c_i(x^*) & \\ x^* \in K & \\ \lambda_i \geq 0 & \forall i \in \{1, \dots, \ell\} \\ \lambda_i c_i(x^*) = 0 & \forall i \in \{1, \dots, \ell\} \end{cases}$$

*Remarque* 20. Par analogie avec le théorème des extrémas liés, le vecteur $\lambda$ apparaissant dans ce théorème est souvent appelé *multiplicateur de Lagrange*.

*Remarque* 21. Les quatres conditions apparaissant dans ce théorème ont un nom :
- la première condition ($-\nabla f(x^*) = \dots$) est appelée *condition d'équilibre* ;
- la deuxième ($x^* \in K$) est l'*admissibilité du point $x$* ;
- la troisième ($\lambda_i \geq 0$) est l'*admissibilité du multiplicateur de Lagrange $\lambda$* ;
- et la quatrième ($\lambda_i c_i(x)$) est la *condition de complémentarité.*

Il est important de n'oublier aucune de ces quatre conditions lorsqu'on applique le théorème de Karush-Kuhn-Tucker (souvent abrégé KKT).

Nous commençons pas démontrer le sens direct de ce théorème, en utilisant la méthode de pénalisation, et plus précisément en passant à la limite dans l'équation d'optimalité du problème pénalisé (7.1).

*Démonstration du sens direct ($\Longrightarrow$) du théorème 45.* Soit $x^*$ un minimiseur de $f$ sur $K$. On commence la démonstration en supposant que $f$ est strictement convexe, et on montrera comment en déduire le cas général. Soit $\varepsilon_N = \frac{1}{N}$ et $x_N^* := x_{\varepsilon_N}^*$ le minimiseur du problème pénalisé $P_{\varepsilon_N}$ :

$$x_N^* \in \arg\min_{x\in\mathbb{R}^d} f(x) + \frac{1}{\varepsilon}\sum_{i=1}^{\ell} \max(c_i(x),0)^2.$$

Par la condition d'optimalité (7.1) pour le problème $P_{\varepsilon_N}$, $x_N^*$ vérifie

$$\nabla f(x_N^*) + \frac{2}{\varepsilon_N}\sum_{i=1}^{\ell} \max(c_i(x_N^*),0)\nabla c_i(x_N^*) = 0.$$

Soit $I$ l'ensemble des $i \in \{1,\dots,\ell\}$ tels que $c_i(x^*) < 0$. Par la proposition 42, $x_N^*$ converge vers $x^*$, de sorte qu'il existe $N_0$ tel que

$$\forall N \geq N_0, \forall i \in I, c_i(x_N^*) < 0.$$

Ainsi, pour $N \geq N_0$, et en utilisant $\nabla c_i(x_N^*) = a_i$ (car $c_i(x) = \langle x|a_i\rangle - b_i$),

$$\begin{aligned} -\nabla f(x_N^*) &= \frac{2}{\varepsilon_N}\sum_{i=1}^{\ell} \max(c_i(x_N^*),0)\nabla c_i(x_N^*) \\ &= \frac{2}{\varepsilon_N}\sum_{i\notin I} \max(c_i(x_N^*),0)a_i \\ &\in C \end{aligned}$$

où $C$ est l'ensemble des combinaisons linéaires à coefficients positifs des $(a_i)_{i\notin I}$, que l'on peut écrire

$$C := \left\{ \sum_{1\leq i\leq N} \lambda_i a_i \mid \lambda \in \mathbb{R}_+^{\ell} \text{ tq } \forall i \in I, \lambda_i = 0 \right\}.$$

Comme l'ensemble $C$ est fermé (lemme 46), en utilisant la continuité de $\nabla f$ ($f \in \mathcal{C}^1(\mathbb{R}^d)$) et $\lim_{N\to+\infty} x_N^* = x_N$, on en déduit que

$$-\nabla f(x^*) \in C.$$

Ainsi, il existe $\lambda \in \mathbb{R}_+^N$ tel que

$$-\nabla f(x^*) = \sum_i \lambda_i a_i = \sum_i \lambda_i \nabla c_i(x^*),$$

qui par définition de l'ensemble $C$ vérifie en outre $\forall i \in I,\ \lambda_i = 0$. Ainsi, si $i \in I$, $c_i(x^*)\lambda_i = 0$, et si $i \notin I$, $c_i(x^*) = 0$ et on a aussi $c_i(x^*)\lambda_i = 0$.

Pour finir, nous avions supposé que la fonction $f$ était strictement convexe. Si ça n'est pas le cas, on remplace $f$ par $\tilde{f}(x) = f(x) + \|x - x^*\|^2$. La fonction $\tilde{f}$ est strictement convexe (somme de convexe et strictement convexe) et a aussi $x^*$ pour minimiseur. On peut donc lui appliquer la démonstration précédente : il existe $\lambda \in \mathbb{R}^\ell$ vérifiant

$$\begin{cases} -\nabla \tilde{f}(x^*) = \sum_{1 \le i \le \ell} \lambda_i \nabla c_i(x^*) & \\ x^* \in K & \\ \lambda_i \ge 0 & \forall i \in \{1, \dots, \ell\} \\ \lambda_i c_i(x) = 0 & \forall i \in \{1, \dots, \ell\}. \end{cases}$$

Pour conclure, il suffit de remarquer que $\nabla \tilde{f}(x^*) = \nabla f(x^*)$. □

**Lemme 46.** *Soit $a_1, \dots, a_\ell \in \mathbb{R}^d$ et $I \subseteq \{1, \dots, \ell\}$. Alors l'ensemble $C$ défini ci-dessous est fermé :*

$$C := \left\{ \sum_{1 \le i \le N} \lambda_i a_i \mid \lambda \in \mathbb{R}^\ell_+ \ tq\ \forall i \in I, \lambda_i = 0 \right\}.$$

*Démonstration.* Admis. □

*Démonstration du sens indirect ($\Longleftarrow$) du théorème 45.* Soit $x^* \in \mathbb{R}^d$ et $\lambda \in \mathbb{R}^\ell$ vérifiant les quatre conditions du théorèmes. Comme la fonction $f$ est convexe,

$$\begin{aligned} \forall x \in \mathbb{R}^d,\ f(x) &\ge f(x^*) + \langle x - x^* | \nabla f(x^*) \rangle \\ &= f(x^*) - \sum_{i=1}^{\ell} \lambda_i \langle x - x^* | \nabla c_i(x^*) \rangle, \end{aligned} \tag{7.2}$$

où l'on a utilisé la condition d'équilibre ($-\nabla f(x^*) = ...$) pour obtenir l'égalité de la deuxième ligne. Comme la contrainte $c_i$ est convexe (car $c_i(x) = \langle x | a_i \rangle - b_i$), on a

$$\forall x \in \mathbb{R}^d,\ c_i(x) \ge c_i(x^*) + \langle x - x^* | \nabla c_i(x^*) \rangle,$$

soit

$$\forall x \in \mathbb{R}^d,\ -\langle x - x^* | \nabla c_i(x^*) \rangle \ge c_i(x^*) - c_i(x), \tag{7.3}$$

En combinant les inégalités (7.2) et (7.3), et en utilisant à nouveau la condition d'admissibilité de $\lambda$ ($\lambda_i \ge 0$), on en déduit

$$\forall x \in \mathbb{R}^d,\ f(x) \ge f(x^*) + \sum_{i=1}^{\ell} \lambda_i (c_i(x^*) - c_i(x)).$$

Par la condition de complémentarité ($\lambda_i c_i(x^*) = 0$), on a

$$\forall x \in \mathbb{R}^d,\ f(x) \ge f(x^*) - \sum_{i=1}^{\ell} \lambda_i c_i(x).$$

Si $x \in K$, alors $c_i(x) \leq 0$. En combinant avec la condition d'admissibilité de $\lambda$ ($\lambda_i \geq 0$), on obtient $\lambda_i c_i(x) \leq 0$, ce qui donne finalement

$$\forall x \in K,\ f(x) \geq f(x^*).$$

Le point $x^*$ est dans $K$ (par la condition d'admissibilité) et $f(x^*) \leq f(x)$ pour tout autre $x \in K$ : ceci montre bien que $x^* \in \arg\min_K f$. □

# Chapitre 8

# Dualité lagrangienne

### Motivation : calcul des multiplicateurs de Lagrange

À nouveau on s'intéresse à un problème d'optimisation sous contraintes :

$$P := \min_{x \in K} f(x) \text{ avec } K = \{x \in \mathbb{R}^d \mid \forall i \in \{1, \dots, \ell\}, c_i(x) \leq 0\}, \tag{8.1}$$

où

(i) $f \in \mathcal{C}^1(\mathbb{R}^d)$ est une fonction convexe,

(ii) $c_i(x) = \langle x | a_i \rangle - b_i$

La proposition suivante montre que si l'on connait les multiplicateurs de Lagrange (le vecteur $\lambda \in \mathbb{R}^\ell$ dont l'existence est garantie par le théorème de Karush-Kuhn-Tucker, théorème 45) alors il est possible de transformer le problème d'optimisation avec contraintes d'inégalités (8.1) en un problème d'optimisation sans contraintes !

**Proposition 47.** *Soit $x^*$ un minimiseur de (8.1) et $\lambda \in \mathbb{R}^\ell$ les multiplicateurs de Lagrange donnés par le théorème KKT (théorème 45). Alors,*

$$x^* \in \arg\min_{\mathbb{R}^d} f_\lambda \text{ où } f_\lambda = f + \sum_{1 \leq i \leq \ell} \lambda_i c_i. \tag{8.2}$$

*Démonstration.* Les hypothèses du théorème KKT sont vérifiées, et il existe donc $\lambda \in \mathbb{R}^\ell$ vérifiant, parmi d'autres conditions,

$$\begin{cases} -\nabla f(x^*) = \sum_{1 \leq i \leq \ell} \lambda_i \nabla c_i(x^*), \\ \forall i, \lambda_i \geq 0 \end{cases}$$

Posons $f_\lambda = f + \sum_{1 \leq i \leq \ell} \lambda_i c_i$. Comme $\lambda_i \geq 0$, $f_\lambda$ est une combinaison linéaire à coefficients positifs de fonctions convexes et est donc convexe. La condition d'équilibre s'écrit :

$$\nabla f_\lambda(x^*) = 0.$$

Comme $f_\lambda$ est convexe, cette condition garantit que $x^*$ la minimise sur $\mathbb{R}^d$. □

Une question très naturelle est alors celle de trouver (en pratique !) ces multiplicateurs de Lagrange $\lambda$. Dans ce chapitre, nous verrons qu'ils sont eux aussi solutions d'un problème d'optimisation "dual". En conséquence, on verra l'algorithme d'Uzawa, un algorithme simple à mettre en œuvre et assez général pour l'optimisation avec contraintes d'inégalités.

## 8.1 Problème dual et dualité faible

On considère la fonction suivante

$$L : (x, \lambda) \in \mathbb{R}^d \times \mathbb{R}^\ell \mapsto f(x) + \sum_{1 \leq i \leq \ell} \lambda_i c_i(x). \tag{8.3}$$

La proposition suivante montre que le problème d'optimisation sous contraintes (8.1) peut être réécrit comme la recherche d'un "point selle" de la fonction $L$, c'est-à-dire l'infimum en $x \in \mathbb{R}^d$ du suprémum en $\lambda \in \mathbb{R}^\ell_+$ de $L$.

**Proposition 48** (Formulation "point selle")**.** *Soit $f, c_1, \ldots, c_\ell$ des fonctions quelconques sur $\mathbb{R}^d$, $K = \{x \mid \forall i \in \{1, \ldots, \ell\},\ c_i(x) \leq 0\}$ et $L$ défini par* (8.3)*. Alors*

$$P = \inf_{x \in K} f(x) = \inf_{x \in \mathbb{R}^d} \sup_{\lambda \in \mathbb{R}^\ell_+} L(x, \lambda). \tag{8.4}$$

*Démonstration.* Montrons d'abord que

$$\sup_{\lambda_i \in \mathbb{R}_+} \lambda_i c_i(x) = \begin{cases} 0 & \text{si } c_i(x) \leq 0, \\ +\infty & \text{sinon.} \end{cases} \tag{8.5}$$

En effet, si $c_i(x) \leq 0$, alors pour tout $\lambda_i \geq 0$, $\lambda_i c_i(x) \leq 0$ avec égalité lorsque $\lambda_i = 0$, ainsi,

$$\sup_{\lambda_i \in \mathbb{R}_+} \lambda_i c_i(x) = 0.$$

Si en revanche, $c_i(x) > 0$, alors

$$\sup_{\lambda_i \in \mathbb{R}_+} \lambda_i c_i(x) \geq \lim_{n \in \mathbb{N}} n c_i(x) = +\infty.$$

On déduit de l'équation (8.5) que

$$\begin{aligned} \sup_{\lambda \in \mathbb{R}^\ell_+} \sum_{1 \leq i \leq \ell} \lambda_i c_i(x) &= \sum_{1 \leq i \leq \ell} \sup_{\lambda_i \in \mathbb{R}_+} \lambda_i c_i(x) \\ &= \begin{cases} +\infty & \text{si } \exists i \in \{1, \ldots, \ell\},\ \text{tq } c_i(x) > 0 \\ 0 & \text{si } \forall i, c_i(x) \leq 0 \end{cases} \\ &= \begin{cases} +\infty & \text{si } x \notin K \\ 0 & \text{si } x \in K. \end{cases} \end{aligned}$$

On s'en suit que

$$\sup_{\lambda\in\mathbb{R}^\ell_+} L(x,\lambda) = \begin{cases} +\infty & \text{si } x \notin K \\ f(x) & \text{si } x \in K, \end{cases}$$

soit

$$\inf_{x\in\mathbb{R}^d} \sup_{\lambda\in\mathbb{R}^\ell_+} L(x,\lambda) = \inf_{x\in K} f(x) \qquad \square$$

**Définition 19** (Problème dual)**.** On appelle problème dual associé au problème d'optimisation avec contraintes d'inégalités (8.4) le problème de point selle

$$D := \sup_{\lambda\in\mathbb{R}^d_+} \inf_{x\in\mathbb{R}^d} L(x,\lambda). \tag{8.6}$$

*Remarque* 22. Pour passer de la formulation "point-selle" du problème primal (8.4) au problème dual (8.6), on a simplement inversé l'infimum et le suprémum.

*Exemple* 14 (Exemple de calcul du problème dual). On considère $f(x) = \frac{1}{2}\|x\|^2$ sur $\mathbb{R}^d$ et $c_i(x) = \langle a_i|x\rangle - b_i$ pour $i \in \{1,\dots,\ell\}$ et on pose

$$L(x,\lambda) = \|x\|^2 + \sum_{1\le i\le \ell} \lambda_i(\langle a_i|x\rangle - v_i).$$

Notons $A \in \mathcal{M}_{\ell,d}(\mathbb{R})$ la matrice dont la $i$ème ligne est le vecteur $a_i$ et $b = (b_1,\dots,b_\ell)$. Alors, $(Ax-v)_i = \langle a_i|x\rangle - b_i$ de sorte que

$$\sum_{1\le i\le \ell} \lambda_i(\langle a_i|x\rangle - b_i) = \langle \lambda|Ax-b\rangle,$$

soit

$$L(x,\lambda) = \|x\|^2 + \langle\lambda|Ax-b\rangle.$$

Le problème dual s'écrit

$$D := \sup_{\lambda\in\mathbb{R}^d_+} \inf_{x\in\mathbb{R}^d} L(x,\lambda).$$

La fonction $f_\lambda : x \mapsto L(x,\lambda)$ est convexe, et $\nabla f_\lambda(x) = x + A^T\lambda$, de sorte que le minimiseur de $\inf_{x\in\mathbb{R}^d} L(x,\lambda) = \inf_{x\in\mathbb{R}^d} f_\lambda(x)$ est atteint en un point $x_\lambda$ vérifiant

$$\nabla f_\lambda(x_\lambda) = 0 \Longleftrightarrow x_\lambda + A^T\lambda = 0,$$

soit $x_\lambda = -A^T\lambda$. Ainsi,

$$\begin{aligned} \inf_{x\in\mathbb{R}^d} L(x,\lambda) &= \inf_{x\in\mathbb{R}^d} f_\lambda(x) = f_\lambda(x_\lambda) \\ &= f(x_\lambda) + \langle\lambda|Ax_\lambda - b\rangle \\ &= \frac{1}{2}\left\|-A^T\lambda\right\|^2 + \langle\lambda| - AA^T\lambda - b\rangle \\ &= -\frac{1}{2}\left\|-A^T\lambda\right\|^2 - \langle\lambda|b\rangle \end{aligned}$$

Ainsi,

$$D = \sup_{\lambda\in\mathbb{R}^\ell_+} \inf_{x\in\mathbb{R}^d} L(x,\lambda) = \sup_{\lambda\in\mathbb{R}^\ell_+} -\frac{1}{2}\left\|-A^T\lambda\right\|^2 - \langle\lambda|b\rangle.$$

*Remarque* 23. Dans l'exemple précédent, le problème dual peut avoir beaucoup moins de variables que le problème primal (si $\ell \ll d$). De plus, les contraintes sont beaucoup plus simples, de la forme $\lambda_i \geq 0$.

**Proposition 49** ("Dualité faible"). *Soit $f, c_1, \ldots, c_\ell$ des fonctions quelconques sur $\mathbb{R}^d$, $K = \{x \mid \forall i, c_i(x) \leq 0\}$ et $L$ défini par* (8.3). *Alors*

$$P \geq D, \tag{8.7}$$

*où $P$ est défini par* (8.1) *et $D$ par* (8.6)

*Remarque* 24. On parle de dualité faible quand on sait montrer que $P \geq D$ et de dualité forte lorsqu'on montrer que $P = D$. La dualité faible est toujours vraie tandis que la dualité forte demande un peu plus d'hypothèses.

*Démonstration.* Soit $x \in K$ et $\lambda \in \mathbb{R}^\ell_+$. Alors,

$$\inf_{\tilde{x} \in K} L(\tilde{x}, \lambda) \leq L(x, \lambda) \leq \sup_{\tilde{\lambda} \in \mathbb{R}^\ell_+} L(x, \tilde{\lambda}),$$

soit

$$\inf_{\tilde{x} \in K} L(\tilde{x}, \lambda) \leq \sup_{\tilde{\lambda} \in \mathbb{R}^\ell_+} L(x, \tilde{\lambda})$$

Ainsi, en prenant le suprémum en $\lambda \in \mathbb{R}^\ell_+$ dans le membre de gauche, on a

$$\sup_{\lambda \in \mathbb{R}^\ell_+} \inf_{\tilde{x} \in K} L(\tilde{x}, \lambda) \leq \sup_{\tilde{\lambda} \in \mathbb{R}^\ell_+} L(x, \tilde{\lambda}).$$

On prend maintenant l'infimum en $x \in K$ dans le membre de droite :

$$\sup_{\lambda \in \mathbb{R}^\ell_+} \inf_{\tilde{x} \in K} L(\tilde{x}, \lambda) \leq \inf_{x \in K} \sup_{\tilde{\lambda} \in \mathbb{R}^\ell_+} L(x, \tilde{\lambda}).$$

En utilisant l'équation (8.4) et la définition de $D$ (eq. 8.6) cette inégalité dit exactement que $D \leq P$. □

## 8.2 Dualité forte

**Théorème 50** (Dualité forte). *Soit $f \in \mathcal{C}^1(\mathbb{R}^d)$ une fonction convexe, $c_i(x) = \langle a_i | x \rangle - b_i$ pour $1 \leq i \leq \ell$ et $K = \{x \in \mathbb{R}^d \mid \forall i, c_i(x) \leq 0\}$. Alors :*

*(i) Si le problème* (8.1) *admet un minimiseur $x^*$, alors $P = D$ (où $D$ est définit dans* (8.6)*) et le maximum dans $D$ est atteint.*

*(ii) Si le problème* (8.1) *admet un minimiseur $x^*$ et que $\lambda^*$ est n'importe quel maximiseur du problème dual (c'est-à-dire $\lambda^* \in \arg\max_{\mathbb{R}^+} g$ avec $g(\lambda) = \inf_{x \in \mathbb{R}^d} L(x, \lambda)$), alors $x^*$ est minimiseur du problème d'optimisation* sans contrainte

$$x^* \in \arg\min_{x \in \mathbb{R}^d} L(x, \lambda).$$

*Démonstration.* (i) Si $x^*$ est minimiseur de $f$ sur $K$ (i.e. $P = f(x^*)$), on est dans les conditions d'application du théorème KKT. Ainsi, il existe $\lambda \in \mathbb{R}^\ell$ vérifiant les quatre conditions.

$$\begin{cases} -\nabla f(x^*) = \sum_{1\leq i\leq \ell} \lambda_i \nabla c_i(x^*) & \\ x^* \in K & \\ \lambda_i \geq 0 & \forall i \in \{1,\dots,\ell\} \\ \lambda_i c_i(x) = 0 & \forall i \in \{1,\dots,\ell\} \end{cases}$$

La condition d'équilibre peut se mettre sous la forme

$$\nabla f_\lambda(x^*) = \nabla f(x^*) + \sum_{1\leq i\leq \ell} \lambda_i \nabla c_i(x^*) = 0,$$

où $f_\lambda = L(\cdot,\lambda) = f + \sum_{1\leq i\leq \ell} \lambda_i c_i$. La fonction $f_\lambda$ est convexe comme combinaison linéaire à coefficients positifs de fonctions convexes, et comme $\nabla f_\lambda(x^*) = 0$, on en déduit donc que $x^* \in \arg\min_{\mathbb{R}^d} f_\lambda$. Ainsi,

$$\begin{aligned} D &= \sup_{\tilde\lambda \in \mathbb{R}^\ell_+} \inf_{x\in\mathbb{R}^d} L(x,\tilde\lambda) \\ &\geq \inf_{x\in\mathbb{R}^d} L(x,\lambda) \\ &= f_\lambda(x^*) \\ &= f(x^*) + \sum_{1\leq i\leq \ell} \lambda_i c_i(x^*) \\ &= f(x^*), \end{aligned}$$

où l'on a utilisé la condition de complémentarité $\lambda_i c_i(x^*) = 0$ pour obtenir la dernière égalité. Ainsi $D \geq f(x^*) = P$. Comme de plus on sait que $D \leq P$ (Proposition 49) on en déduit la dualité forte $D = P$.

(ii) Soit $x^* \in K$ un minimiseur du problème primal et $\lambda^* \in \mathbb{R}^\ell_+$ un maximiseur du problème dual. En utilisant (i), on sait que $P = D$, de sorte que

$$\begin{aligned} f(x^*) = P = D &= \sup_{\lambda\in\mathbb{R}^\ell_+} \inf_{x\in\mathbb{R}^d} L(x,\lambda) \\ &= \inf_{x\in\mathbb{R}^d} L(x,\lambda^*) \\ &\leq L(x^*,\lambda^*) = f(x^*) + \sum_{1\leq i\leq \ell} \lambda_i^* c_i(x^*). \end{aligned} \tag{8.8}$$

On en déduit que

$$\sum_{1\leq i\leq \ell} \lambda_i^* c_i(x^*) \leq 0.$$

Comme $\lambda_i^* \geq 0$ et $c_i(x^*) \leq 0$, on peut en déduire que

$$\sum_{1\leq i\leq \ell} \lambda_i^* c_i(x^*) = 0.$$

Ainsi, la seule inégalité dans (8.8) doit en fait être une égalité :

$$\inf_{x\in\mathbb{R}^d} L(x,\lambda^*) = L(x^*,\lambda^*),$$

ce qui montre que $x^*$ minimise $L(\cdot,\lambda^*) = f_{\lambda^*}$ sur $\mathbb{R}^d$. □

## 8.3 Algorithme d'Uzawa

**Notation et hypothèses** Dans la suite, pour tout $x, y \in \mathbb{R}^d$, et on note $x \leq y$ si et seulement si pour tout $i \in \{1,\dots,d\}$, $x_i \leq y_i$. Soit $A \in \mathcal{M}_{\ell,d}(\mathbb{R})$ et $b \in \mathbb{R}^\ell$. Pour $i \in \{1,\dots,\ell\}$, on note $c_i(x) = \langle a_i|x\rangle - b_i$ où $a_i$ est la $i$ème ligne de $A$, et on pose

$$K = \{x \in \mathbb{R}^d \mid \forall i \in \{1,\dots,d\}, c_i(x) \leq 0\} = \{x \in \mathbb{R}^d \mid Ax \leq b\}.$$

Sous les hypothèses du théorème de dualité forte (Théorème 50), on sait alors que

$$P = \min_K f(x) = \min_{Ax\leq b} f(x) = \min_{x\in\mathbb{R}^d} f(x) + \langle \lambda^*|Ax-b\rangle,$$

où $\lambda^* \in \mathbb{R}^\ell_+$ est un maximiseur du problème dual

$$D = \max_{\lambda\in\mathbb{R}^\ell_+} g(\lambda), \quad g(\lambda) = \inf_{x\in\mathbb{R}^d} L(x,\lambda), \text{ où } L(x,\lambda) = f(x) + \sum_{1\leq i\leq d} \lambda_i c_i(x).$$

L'algorithme d'Uzawa est alors simplement l'algorithme du gradient projeté pour résoudre le problème dual D, c'est-à-dire maximiser la fonction $g$ sur $\mathbb{R}^\ell_+$. Cet algorithme peut être mis en œuvre en pratique car on sait projeter sur l'ensemble de contrainte $\mathbb{R}^\ell_+$, et qu'il est parfois possible de calculer $g$ et $\nabla g$ explicitement (par exemple lorsque $f$ est par exemple quadratique). Le choix de $\lambda^{(0)} \in \mathbb{R}^\ell_+$ peut être arbitraire (typiquement, on prendra $\lambda^{(0)} = 0_{\mathbb{R}^\ell}$) :

$$\begin{cases} x^{(k)} \in \arg\min_{x\in\mathbb{R}^d} L(x,\lambda^{(k)}) \\ \gamma^{(k)} = \nabla g(\lambda^{(k)}) \\ \lambda^{(k+1)} = \mathrm{p}_{\mathbb{R}^\ell_+}(\lambda^{(k)} + \tau\gamma^{(k)}) \end{cases}$$

*Remarque* 25. Comme on cherche à maximiser la fonction $g$ sur $\mathbb{R}^\ell_+$, on fait de la montée plutôt que de la descente de gradient.

### 8.3.1 Algorithme d'Uzawa dans le cas quadratique

À partir de maintenant, on suppose que $f$ est une fonction quadratique. Plus précisément on suppose qu'il existe une matrice symétrique définie positive $Q$ et un vecteur $e$ tel que $f(x) = \frac{1}{2}\langle x|Qx\rangle + \langle e|x\rangle$, et on considère donc le problème de minimisation sous contraintes

$$P = \min_K f(x) \text{ où } f(x) = \frac{1}{2}\langle x|Qx\rangle + \langle e|x\rangle \text{ et } K = \{x \in \mathbb{R}^d \mid Ax \leq b\}. \tag{8.9}$$

**Lemme 51.** *La fonction* $g(\lambda) = \inf_{x\in\mathbb{R}^d} L(x,\lambda)$ *est concave, et*

$$\begin{cases} g(\lambda) = -\left(\frac{1}{2}\langle x_\lambda | Q x_\lambda\rangle + \langle \lambda | b\rangle\right), \\ \nabla g(\lambda) = A x_\lambda - b \end{cases}$$

*où*

$$x_\lambda = -Q^{-1}(e + A^T\lambda) \in \arg\min_{x\in\mathbb{R}^d} L(x,\lambda).$$

*Démonstration.* Commençons par calculer la fonction $L$ plus explicitement :

$$\begin{aligned} L(x,\lambda) &= f(x) + \sum_{1\le i\le d} \lambda_i c_i(x) \\ &= f(x) + \sum_{1\le i\le d} \lambda_i(\langle a_i | x\rangle - b_i) \\ &= f(x) + \sum_{1\le i\le d} \lambda_i (Ax-b)_i \\ &= \frac{1}{2}\langle x | Qx\rangle + \langle e | x\rangle + \langle \lambda | Ax - b\rangle \end{aligned}$$

En particulier, on voit que la fonction $f_\lambda = L(\cdot,\lambda)$ est strictement convexe (somme de la fonction strictement convexe $x \mapsto \frac{1}{2}\langle x|Qx\rangle$ et de fonctions convexes). Ainsi, elle admet au plus un minimiseur. De plus,

$$\nabla f_\lambda(x) = Qx + e + A^T\lambda,$$

de sorte que le minimiseur de $f_\lambda$ sur $\mathbb{R}^d$ est atteint en l'unique point $x_\lambda$ vérifiant

$$Qx_\lambda + e + A^T\lambda = 0, \text{ i.e. } x_\lambda = -Q^{-1}(e + A^T\lambda).$$

Ainsi,

$$\begin{aligned} g(\lambda) &= \inf_{x\in\mathbb{R}^d} f_\lambda(x) = f_\lambda(x_\lambda) \\ &= \frac{1}{2}\langle x_\lambda | Q x_\lambda\rangle + \langle e | x_\lambda\rangle + \langle \lambda | A x_\lambda - b\rangle \\ &= \frac{1}{2}\langle x_\lambda | Q x_\lambda\rangle + \langle e + A^t\lambda | x_\lambda\rangle - \langle \lambda | b\rangle \end{aligned}$$

En utilisant $Qx_\lambda + e + A^T\lambda = 0$, on trouve

$$\begin{aligned} g(\lambda) &= -\left(\frac{1}{2}\langle x_\lambda | Q x_\lambda\rangle + \langle \lambda | b\rangle\right) \\ &= -\left(\frac{1}{2}\langle Q^{-1}(e + A^T\lambda) | QQ^{-1}(e + A^T\lambda)\rangle + \langle \lambda | b\rangle\right) \\ &= -\left(\frac{1}{2}\langle Q^{-1}(e + A^T\lambda) | e + A^T\lambda\rangle + \langle \lambda | b\rangle\right) \\ &= -\left(\frac{1}{2}\langle AQ^{-1}A^T\lambda | \lambda\rangle + \langle AQ^{-1}e | \lambda\rangle + \frac{1}{2}\langle Q^{-1}e | e\rangle + \langle \lambda | b\rangle\right). \end{aligned}$$

On en déduit d'abord que $g$ est concave : il suffit de remarquer que $AQ^{-1}A^T$ est symétrique positive (exercice !). On peut aussi se servir de cette expression pour calculer le gradient de $g$ :

$$\nabla g(\lambda) = -AQ^{-1}A^T\lambda - AQ^{-1}e - b = Ax_\lambda - b. \qquad \square$$

**Corollaire 52.** *L'algorithme d'Uzawa pour le problème de minimisation sous contraintes* (8.9) *s'écrit de la manière suivante,*

$$\begin{cases} x^{(k)} = -Q^{-1}(e + A^T\lambda^{(k)}) \\ \gamma^{(k)} = Ax^{(k)} - b \\ \lambda^{(k+1)} = \max(\lambda^{(k)} + \tau\gamma^{(k)}, 0) \end{cases} \tag{8.10}$$

### 8.3.2 Convergence de l'algorithme d'Uzawa

On conclut le cours par la démonstration de convergence de l'algorithme d'Uzawa.

**Théorème 53.** *Soient* $(x^{(k)}, \lambda^{(k)}) \in \mathbb{R}^d \times \mathbb{R}^\ell$ *les itérées de l'algorithme d'Uzawa* (8.10) *pour le problème d'optimisation sous contraintes* (8.9). *On suppose que*

$$\tau \in \left]0, \frac{2m}{\|A\|^2}\right[,$$

*où* $m$ *est la plus petite valeur propre de* $Q$, *et* $\|A\|$ *est la norme d'opérateur de* $A$ *(i.e.* $\|A\| = \max_{x\neq 0} \|Ax\| / \|x\|$*). Alors,*

$$\lim_{k\to+\infty} x^{(k)} = x^*,$$

*où* $x^*$ *est l'unique minimiseur de* (8.10).

*Démonstration.* Le problème primal admet un minimiseur $x^*$ car la fonction $f$ est quadratique (donc continue) avec $Q$ définie positive, ce qui garantit que $\lim_{\|x\|\to+\infty} f(x) = +\infty$. Par le théorème de dualité forte (Théorème 50), on sait que le problème dual admet aussi un maximiseur $\lambda^*$, qui n'est pas nécessairement unique. De plus, comme la fonction $g \in \mathcal{C}^1(\mathbb{R}^d)$ est concave (Lemme 51), $\lambda^*$ maximise $g$ sur $\mathbb{R}^\ell_+$ si et seulement si (Théorème 37)

$$\lambda^* \in \arg\max_{\mathbb{R}^\ell_+} g \iff \forall\tau > 0, \lambda^* = \mathrm{p}_{\mathbb{R}^\ell_+}(\lambda^* + \tau\nabla g(\lambda^*)).$$

Le théorème 50 nous dit aussi que $x^*$, l'unique minimiseur du problème primal, $x^* \in \arg\min_K f$ est caractérisé par $x^* \in \arg\min_{x\in\mathbb{R}^d} L(x, \lambda^*)$. Ainsi, en utilisant les notations du Lemme 51, $x^* = x_{\lambda^*}$, de sorte qu'en utilisant l'expression de $\nabla g$ donnée dans ce même lemme, on obtient

$$\nabla g(\lambda^*) = Ax_{\lambda^*} - b = Ax^* - b.$$

Ainsi, $\lambda^*$ maximise $g$ sur $\mathbb{R}_+^\ell$ si et seulement si $\lambda^* = \mathrm{p}_{\mathbb{R}_+^\ell}(\lambda^* + \tau(Ax^* - b))$. À partir de maintenant, on note $\lambda^*$ un minimiseur (peu importe lequel) du problème dual. On a alors, par définition de l'algorithme et en utilisant que $\mathrm{p}_{\mathbb{R}_+^\ell}$ est 1-Lipschitzienne,

$$\begin{aligned}
\left\|\lambda^{(k+1)} - \lambda^*\right\|^2 &= \left\|\mathrm{p}_{\mathbb{R}_+^\ell}(\lambda^{(k)} - \tau(Ax^{(k)} - b)) - \mathrm{p}_{\mathbb{R}_+^\ell}(\lambda^* - \tau(Ax^* - b))\right\|^2 \\
&\leq \left\|\lambda^{(k)} - \tau(Ax^{(k)} - b)) - (\lambda^* - \tau(Ax^* - b))\right\|^2 \\
&= \left\|\lambda^{(k)} - \lambda^*\right\|^2 + \tau^2 \left\|A(x^{(k)} - x^*)\right\|^2 + 2\tau\langle\lambda^{(k)} - \lambda^* | A(x^{(k)} - x^*)\rangle
\end{aligned}$$

On rappelle (Lemme 51) que $x_\lambda = -Q^{-1}(e + A^T\lambda)$, et on s'en sert pour majorer le produit scalaire :

$$\begin{aligned}
\langle\lambda^{(k)} - \lambda^* | A(x^{(k)} - x^*)\rangle &= \langle A^T(\lambda^{(k)} - \lambda^*) | x^{(k)} - x^*\rangle \\
&= -\langle Q(x^{(k)} - x^*) | x^{(k)} - x^*\rangle \\
&\leq -m\left\|x^{(k)} - x^*\right\|^2,
\end{aligned}$$

où $m$ est la plus petite valeur propre de $Q$. Ainsi,

$$\begin{aligned}
\left\|\lambda^{(k+1)} - \lambda^*\right\|^2 &\leq \left\|\lambda^{(k)} - \lambda^*\right\|^2 + \tau^2 \left\|A(x^{(k)} - x^*)\right\|^2 + 2\tau\langle\lambda^{(k)} - \lambda^* | A(x^{(k)} - x^*)\rangle \\
&\leq \left\|\lambda^{(k)} - \lambda^*\right\|^2 + \left(\tau^2 \|A\|^2 - 2\tau m\right)\left\|x^{(k)} - x^*\right\|^2 \\
&\leq \left\|\lambda^{(k)} - \lambda^*\right\|^2
\end{aligned}$$

La dernière inégalité vient de l'hypothèse sur $\tau$. Ainsi, la suite $r_k := \left\|\lambda^{(k+1)} - \lambda^*\right\|^2$ est décroissante et minorée (par 0) donc convergente vers une certaine limite $r^*$. De plus, l'inégalité du dessus nous donne

$$\left(2\tau m - \tau^2 \|A\|^2\right)\left\|x^{(k)} - x^*\right\|^2 \leq r_k - r_{k-1},$$

où le second membre tend $r^* - r^*$ lorsque $k \to +\infty$. Comme $2\tau m - \tau^2 \|A\|^2 > 0$ par hypothèse, on en déduit que

$$\lim_{k \to +\infty} \left\|x^{(k)} - x^*\right\|^2 = 0 \qquad \square$$