

---

OPTIMISATION : NOTES DE COURS

---

## Table des matières

<b>1</b>	<b>Calcul différentiel</b>	<b>2</b>
1.1	Dérivabilité . . . . .	2
1.1.1	Cas $X = \mathbb{R}^p$ et $Y = \mathbb{R}^q$ . . . . .	2
1.2	Dérivées secondes . . . . .	3
1.3	Formules de Taylor . . . . .	4
1.4	Gradient et Hessienne . . . . .	4
<b>2</b>	<b>Existence de minima et fonctions convexes</b>	<b>4</b>
2.1	Existence d'un minimum . . . . .	4
2.2	Fonctions convexes . . . . .	6
2.3	Existence de minimiseur pour des problèmes avec contraintes . . . . .	8
2.3.1	Contraintes d'égalités . . . . .	8
2.3.2	Contraintes d'inégalité . . . . .	9
<b>3</b>	<b>Algorithmes</b>	<b>10</b>
3.1	Optimisation sans contrainte . . . . .	10
3.1.1	Méthodes de descente . . . . .	10
3.1.2	Méthodes de gradient à pas fixe, variable et optimal . . . . .	11
3.2	Optimisation avec contraintes . . . . .	11
3.2.1	Contraintes d'égalité . . . . .	11
3.2.2	Contraintes convexes . . . . .	12

## Introduction à l'optimisation continue

Soient  $X$  un espace de Banach,  $\Omega$  une partie de  $X$  et  $A$  un sous-ensemble de  $\Omega$ . On considère une fonction  $J : \Omega \rightarrow \mathbb{R}$ .

On s'intéresse dans ce cours aux deux types de problèmes suivants

(i)

Trouver  $x_* \in \Omega$  vérifiant  $J(x_*) = \inf_{x \in \Omega} J(x)$ .

(ii)

Trouver  $x_* \in A \subsetneq \Omega$  vérifiant  $J(x_*) = \inf_{x \in A} J(x)$ .

Le premier problème est appelé *problème d'optimisation (continue) sans contrainte*. Le second problème est un *problème d'optimisation (continue) avec contraintes*. Ces contraintes peuvent être de type

★ convexes :  $A$  est un ensemble convexe fermé

★ égalité

$$A = \{x \in \Omega, g_i(x) = 0, 1 \leq i \leq p\} \quad (1)$$

★ inégalité

$$A = \{x \in \Omega, h_i(x) \leq 0, 1 \leq i \leq q\} \quad (2)$$

Dans (1) et (2), les fonctions  $h_i$  et  $g_i$  sont au moins continues.

La fonction  $J$  est appelée la fonction **coût** (ou encore fonction objectif, ou critère) et le sous-ensemble  $A$  définissant les contraintes, l'ensemble des *éléments admissibles*.

## 1 Calcul différentiel

On pourra trouver une introduction au calcul différentiel dans le livre de Ciarlet [Cia82, Chapitre 7].

### 1.1 Dérivabilité

**Définition 1.1.** On dit qu'une fonction  $f : \Omega \rightarrow Y$  où  $\Omega$  est un ouvert de  $X$  est dérivable en un point  $x$  de  $\Omega$  s'il existe une application linéaire continue  $df(x) : X \rightarrow Y$  telle que pour tout  $h$  dans un voisinage de 0 on a

$$f(x+h) = f(x) + df(x)h + o(\|h\|).$$

Si  $f$  est dérivable en tout point de  $\Omega$ , on dit que  $f$  est dérivable sur  $\Omega$ .

Si de plus l'application  $x \mapsto df(x)$  est continue, on dit que  $f$  est continûment dérivable dans  $\Omega$ .

On rappelle qu'une fonction dérivable en un point est continue en ce point.

#### 1.1.1 Cas $X = \mathbb{R}^p$ et $Y = \mathbb{R}^q$

Dans le cas où  $Y = \mathbb{R}^q$ , en utilisant le fait que  $\mathbb{R}^q = \bigtimes_{i=1}^q \mathbb{R}$ , on peut écrire  $f$  selon ses

composantes  $f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_q(x) \end{bmatrix}.$

Dans ce cas,  $f$  est dérivable en  $x \in \Omega$  si et seulement si chacune de ses composantes  $f_j$  est dérivable en  $x$ . On a alors

$$df(x) = \begin{bmatrix} df_1(x) \\ \vdots \\ df_q(x) \end{bmatrix}. \quad (3)$$

Si  $\Omega \subset \mathbb{R}^p$ , alors on peut écrire  $x \in \mathbb{R}^p$  en fonction de ses composantes  $x = [x_1, \dots, x_p]^T$  et considérer, par exemple pour  $(x_1, \dots, x_{k-1}, x^{(k+1)}, \dots, x_p)$  fixé la fonction  $y_k \mapsto f(x_1, \dots, x_{k-1}, y_k, x^{(k+1)}, \dots, x_p)$ . Si cette fonction est dérivable en  $x_k$ , on note  $\partial_k f(x)$  cette dérivée appelée *dérivée partielle de la fonction  $f$  au point  $x$  par rapport à la  $k$ -ième variable*.

**Proposition 1.2.** Si  $f : \Omega \mapsto \mathbb{R}^q$ , où  $\Omega$  est un ouvert de  $\mathbb{R}^p$  est dérivable en  $x \in \Omega$ , alors  $f$  admet des dérivées partielles par rapport à toutes ses variables. De plus pour  $h \in \mathbb{R}^p$ , on a

$$\begin{aligned} df(x)h &= \sum_{k=1}^p \partial_k f(x) h_k, \quad \text{où } h = \begin{bmatrix} h_1 \\ \vdots \\ h_p \end{bmatrix} \\ &= \begin{bmatrix} \partial_1 f_1(x) & \partial_2 f_1(x) & \dots & \partial_p f_1(x) \\ \partial_1 f_2(x) & \partial_2 f_2(x) & \dots & \partial_p f_2(x) \\ \vdots & \vdots & & \vdots \\ \partial_1 f_q(x) & \partial_2 f_q(x) & \dots & \partial_p f_q(x) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}. \end{aligned}$$

La matrice  $(\partial_i f_j(x))_{1 \leq i \leq q, 1 \leq j \leq p} \in \mathbb{R}^{q \times p}$  est appelée matrice jacobienne de  $f$  en  $x$ .

*Démonstration.* La preuve est immédiate en utilisant la linéarité de  $df(x)$  et en regardant la définition de la dérivée pour  $h = te_k$  où  $e_k$  est le  $k$ -ième vecteur de la base canonique.  $\square$

**Remarque 1.3.** La réciproque est fausse. La fonction  $f(x, y) = \begin{cases} 0 & \text{si } xy = 0 \\ 1 & \text{sinon} \end{cases}$  n'est pas continue en 0 et admet des dérivées partielles selon les deux variables.

Un autre contre-exemple classique est la fonction  $f(x, y) = \frac{2xy}{\sqrt{x^2+y^2}}$ . Cette fonction peut être prolongée continûment en 0 et admet des dérivées partielles selon toutes les directions mais n'est pas dérivable en 0.

**Théorème 1.4** (Composition des dérivées). Soient  $f : \Omega \subset \mathbb{R}^p \mapsto \mathbb{R}^q$  une fonction dérivable en  $x \in \Omega$ ,  $g : \tilde{\Omega} \subset \mathbb{R}^q \mapsto \mathbb{R}^r$  une fonction dérivable en  $f(x)$ . On suppose que  $f(\Omega) \subset \tilde{\Omega}$  alors  $g \circ f$  est dérivable en  $x$  et

$$d(g \circ f)(x) = dg(f(x))df(x),$$

où de façon matricielle

$$\begin{aligned} \begin{bmatrix} \partial_1(g \circ f)_1(x) & \dots & \partial_p(g \circ f)_1(x) \\ \vdots & & \vdots \\ \partial_1(g \circ f)_r(x) & \dots & \partial_p(g \circ f)_r(x) \end{bmatrix} &= \\ & \begin{bmatrix} \partial_1 g_1(f(x)) & \dots & \partial_q g_1(f(x)) \\ \vdots & & \vdots \\ \partial_1 g_r(f(x)) & \dots & \partial_q g_r(f(x)) \end{bmatrix} \begin{bmatrix} \partial_1 f_1(x) & \dots & \partial_p f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_q(x) & \dots & \partial_p f_q(x) \end{bmatrix} \end{aligned} \quad (4)$$

## 1.2 Dérivées secondes

On considère une fonction  $f : \Omega \subset \mathbb{R}^p \mapsto \mathbb{R}^q$  une fonction dérivable sur  $\Omega$ .

Sa dérivée est alors une fonction de  $\mathbb{R}^p$  dans les applications linéaires de  $\mathbb{R}^p$  dans  $\mathbb{R}^q$ , que l'on peut identifier à l'espace des matrices  $\mathbb{R}^{q \times p}$ . Si la dérivée  $df$ , vue en tant que fonction de  $\mathbb{R}^p$  dans  $\mathbb{R}^{q \times p}$  est elle-même dérivable en un point  $x \in \Omega$ , on note  $d^2 f(x)$  sa dérivée seconde, qui est une application linéaire de  $\mathbb{R}^p$  dans  $\mathbb{R}^{q \times p}$ .

On peut alors associer  $d^2 f(x)$  à une application bilinéaire  $\mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^q$ , que l'on notera aussi  $d^2 f(x)$  par abus de notation.

Si  $d^2 f(x)$  existe pour tout  $x \in \Omega$ , alors on dit que  $f$  est dérivable deux fois dans  $\Omega$ . Si de plus,  $x \mapsto d^2 f(x)$  est continue, on dit que  $f$  est deux fois continûment dérivable dans  $\Omega$ .

**Remarque 1.5.** Par le théorème de Schwarz, si  $f$  est deux fois continûment dérivable dans  $\Omega$ , alors  $d^2f(x)$  est une application bilinéaire symétrique.

On se place maintenant dans le cas où  $q = 1$  et où  $f : \Omega \subset \mathbb{R}^p \rightarrow \mathbb{R}$  est deux fois continûment dérivable dans  $\Omega$ . On peut alors écrire pour tout  $x, h, k \in \mathbb{R}^p$

$$d^2f(x)[h, k] = \sum_{i,j=1}^p h_i k_j d^2f(x)[e_i, e_j],$$

où par définition  $d^2f(x)[e_i, e_j] = \partial_i(\partial_j f)(x) = \partial_j(\partial_i f)(x)$ . On note alors ces dérivées  $\partial_{ij}^2 f(x)$ , que l'on appelle *dérivées partielles secondes de  $f$  en  $x$* .

### 1.3 Formules de Taylor

**Théorème 1.6.** Soit  $f : \Omega \subset \mathbb{R}^p \rightarrow \mathbb{R}$  est deux fois continûment dérivable dans  $\Omega$ . Pour tout  $x \in \Omega$  et  $h \in \mathbb{R}^p$ , tel que  $x + h \in \Omega$ , on a

$$f(x + h) = f(x) + df(x)h + \frac{1}{2}d^2f(x)[h, h] + o(\|h\|^2). \quad (5)$$

### 1.4 Gradient et Hessienne

Si  $f : \Omega \subset \mathbb{R}^p \rightarrow \mathbb{R}$  est deux fois continûment dérivable et si  $\mathbb{R}^p$  est muni d'un produit scalaire  $\langle \cdot, \cdot \rangle$ , l'application linéaire  $df(x) : \mathbb{R}^p \rightarrow \mathbb{R}$  peut être identifiée par le théorème de représentation de Ritz avec un unique vecteur de  $\mathbb{R}^p$ , noté  $\nabla f(x)$  vérifiant

$$\forall h \in \mathbb{R}^p, df(x)h = \langle \nabla f(x), h \rangle.$$

**Remarque 1.7.** Si le produit scalaire change, le gradient est différent !

La matrice hessienne de  $f$  est la matrice symétrique notée  $Hf(x) \in \mathcal{M}_p(\mathbb{R})$  des dérivées partielles secondes

$$Hf(x) = \begin{bmatrix} \partial_{11}^2 f(x) & \dots & \partial_{1n}^2 f(x) \\ \vdots & \ddots & \vdots \\ \partial_{n1}^2 f(x) & \dots & \partial_{nn}^2 f(x) \end{bmatrix}.$$

On peut alors réécrire les formules de Taylor en utilisant le gradient et la matrice hessienne : pour tout  $x \in \Omega$  et  $h \in \mathbb{R}^p$  tel que  $x + h \in \Omega$ , on a

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2}\langle h, Hf(x)h \rangle + o(\|h\|^2).$$

## 2 Existence de minima et fonctions convexes

### 2.1 Existence d'un minimum

On considère une fonction  $J : \Omega \subset X \rightarrow \mathbb{R}$  où  $\Omega$  est ouvert et  $X$  est un espace de Banach.

**Définition 2.1.** Soit  $J : \Omega \rightarrow \mathbb{R}$ ,  $A \subset \Omega$ .

★  $J$  admet un minimum global sur  $A$  si

$$\exists x_* \in A, \quad \forall x \in A, J(x) \geq J(x_*).$$

$x_*$  est alors appelé minimiseur de  $J$  sur  $A$ .

★  $J$  admet un minimum local en  $x_* \in A$  si

$$\exists \Omega \in \mathcal{V}(x_*), \quad \forall x \in \Omega \cap A, J(x) \geq J(x_*).$$

$x_*$  est alors appelé minimiseur local de  $J$  sur  $A$ .

**Proposition 2.2.** Soit  $x_* \in \Omega$  un minimiseur local de  $J$ . Si  $J$  est dérivable continûment sur  $\Omega$ , alors  $dJ(x_*) = 0$ .

Il est essentiel que  $\Omega$  soit ouvert pour avoir  $dJ(x_*) = 0$ . Cette condition est nécessaire mais pas suffisante pour garantir que  $x_*$  est un minimum local de  $J$ . On peut affiner ce résultat si on a accès à la seconde dérivée de  $J$ .

**Proposition 2.3.** Soit  $J$  une fonction deux fois dérivable sur  $\Omega$  un ouvert de  $X$  un espace de Banach.

- (i) si  $x_*$  est un minimiseur local de  $J$ , alors  $d^2J(x_*)$  est semi-définie positive, c'est-à-dire que pour tout  $h \in X$ , on a  $d^2J(x_*)[h, h] \geq 0$ .
- (ii) si  $x_* \in \Omega$  vérifie  $dJ(x_*) = 0$  et il existe  $\alpha > 0$ , telle que pour tout  $h \in X$ ,  $d^2J(x_*)[h, h] \geq \alpha \|h\|^2$ , alors  $x_*$  est un minimiseur local de  $J$ .

*Démonstration.* La preuve repose sur la formule de Taylor (5). □

**Définition 2.4** (Coercivité). Soit  $J : X \rightarrow \mathbb{R}$  une fonction où  $X$  est un espace de Banach. On dit que  $J$  est coercive si  $\lim_{\|x\| \rightarrow \infty} J(x) = \infty$ .

En dimension finie, si  $J$  est continue et coercive, on peut alors montrer l'existence d'un minimiseur.

**Proposition 2.5.** Soit  $J : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction continue et coercive. Alors  $J$  admet au moins un minimiseur  $x_*$ , i.e. pour  $x \in \mathbb{R}^d$ ,  $J(x_*) \leq J(x)$ .

Avant de passer à la preuve de ce résultat, on énonce la définition d'une suite minimisante.

**Définition 2.6.** On appelle suite minimisante de  $J$  dans  $A$  une suite  $(u_n)_{n \in \mathbb{N}}$  d'éléments de  $A$  telle que

$$\lim_{n \rightarrow +\infty} J(u_n) = \inf_{x \in A} J(x).$$

*Démonstration.* On considère une suite minimisante  $(x^{(n)})_{n \in \mathbb{N}}$ , par coercivité, cette suite est bornée. On en extrait une sous-suite  $(J(x^{(\phi(n))}))$  convergente car  $(x^{(n)})$  est une suite en dimension finie. Par continuité de  $J$ ,  $(J(x^{(\phi(n))}))$  tend vers le minimum de  $J$  sur  $\mathbb{R}^d$ . □

En dimension infinie, comme un ouvert borné n'est pas nécessairement compact, il n'est pas possible de conclure sur l'existence d'un minimiseur. Habituellement, il faut une hypothèse de compacité ou une hypothèse de continuité par passage à la limite faible (voir Remarque 2.15).

## 2.2 Fonctions convexes

On rappelle qu'un ensemble convexe  $C$  d'un espace de Banach  $X$  est un ensemble tel que pour tout  $(x, y) \in C^2$  et  $t \in [0, 1]$ , on a  $tx + (1 - t)y \in C$ .

**Définition 2.7.** Soit  $C$  un convexe d'un ;  $J : C \rightarrow \mathbb{R}$  est convexe si

$$\forall (x, y) \in C \times C, \quad \forall t \in [0, 1], \quad J(tx + (1 - t)y) \leq tJ(x) + (1 - t)J(y).$$

**Proposition 2.8.** Soit  $U$  un **ouvert convexe** de  $\mathbb{R}^n$  et  $J$  une fonction différentiable de  $U$  dans  $\mathbb{R}$ . On a équivalence entre

- (i)  $J$  est convexe
- (ii) Le graphe de  $J$  est "au-dessus de ses tangentes"

$$\forall (x, y) \in U \times U, \quad J(y) \geq J(x) + \langle \nabla J(x), y - x \rangle$$

- (iii) L'application  $\nabla J : U \rightarrow \mathbb{R}^n$  est "monotone"

$$\forall (x, y) \in U \times U, \quad \langle \nabla J(y) - \nabla J(x), y - x \rangle \geq 0$$

De plus il y a équivalence entre la convexité stricte de  $J$  et les inégalités (ii) et (iii) rendues strictes, pour  $x \neq y$ .

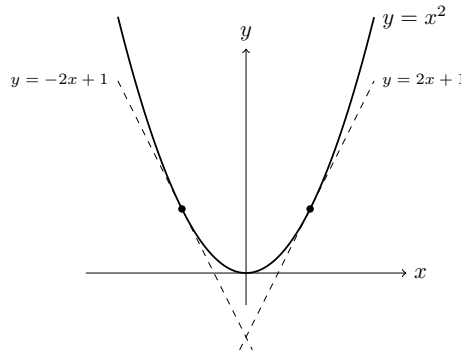


FIGURE 1 – Illustration du graphe d'une fonction convexe au dessus de ses tangentes

**Proposition 2.9.** Soit  $U$  un **ouvert convexe** de  $\mathbb{R}^n$  et  $J$  une fonction deux fois différentiable de  $U$  dans  $\mathbb{R}$ . On a équivalence entre

- (i)  $J$  est convexe
- (ii)  $H_J(x)$  est semi-définie positive pour tout  $x \in \mathbb{R}^n$  :

$$\forall (x, h) \in \mathbb{R}^n \times \mathbb{R}^n, \quad \langle H_J(x)h, h \rangle \geq 0.$$

Une fonction convexe n'a pas nécessairement de minimum (par exemple la fonction  $\exp$  sur  $\mathbb{R}$ ). Une sous-classe de fonctions convexes qui admettent un minimum est la classe des fonctions dites  $\alpha$ -convexes.

## Fonctions $\alpha$ -convexes

**Définition 2.10.** Soit  $C$  un convexe d'un espace vectoriel normé,  $J : C \rightarrow \mathbb{R}$  est  $\alpha$ -convexe si

$$\forall (x, y) \in C \times C, \quad \forall t \in [0, 1], \quad J(tx + (1-t)y) \leq tJ(x) + (1-t)J(y) - \frac{\alpha}{2}t(1-t)\|x - y\|^2.$$

**Proposition 2.11.** Soit  $C$  un convexe d'un espace vectoriel normé,  $J : C \rightarrow \mathbb{R}$ . On a les équivalences suivantes

$$\begin{aligned} J \text{ est } \alpha\text{-convexe} &\Leftrightarrow J - \frac{\alpha}{2}\|\cdot\|^2 \text{ est convexe} \\ &\Leftrightarrow \forall (x, y) \in C \times C, J\left(\frac{x+y}{2}\right) \leq \frac{J(x) + J(y)}{2} - \frac{\alpha}{8}\|x - y\|^2 \end{aligned}$$

**Proposition 2.12.** Soit  $U$  un ouvert convexe de  $\mathbb{R}^n$  et  $J$  une fonction différentiable de  $U$  dans  $\mathbb{R}$ . On a les équivalences suivantes

$$\begin{aligned} J \text{ est } \alpha\text{-convexe} &\Leftrightarrow \forall (x, y) \in U \times U, \quad J(y) \geq J(x) + \langle \nabla J(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2 \\ &\Leftrightarrow \forall (x, y) \in U \times U, \quad \langle \nabla J(y) - \nabla J(x), y - x \rangle \geq \alpha\|x - y\|^2 \end{aligned}$$

**Proposition 2.13.** Soit  $U$  un ouvert convexe de  $\mathbb{R}^n$  et  $J$  une fonction deux fois différentiable de  $U$  dans  $\mathbb{R}$ . Alors

$$J \text{ est } \alpha\text{-convexe} \Leftrightarrow \forall (x, h) \in U \times \mathbb{R}^n, \langle HJ(x)h, h \rangle \geq \alpha\|h\|^2.$$

**Théorème 2.14** (Existence d'un minimum, cas fortement convexe.). Soit  $J : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction  $\alpha$ -convexe continue. Alors il existe un unique minimum  $x_*$  de  $J$  sur  $\mathbb{R}^d$  et on a

$$\forall x \in \mathbb{R}^d, \quad \|x - x_*\|^2 \leq \frac{2}{\alpha}[J(x) - J(x_*)]. \quad (6)$$

*Démonstration.* Une fonction  $\alpha$ -convexe est nécessairement coercive, donc  $J$  admet un minimiseur  $x_*$ . Par  $\alpha$ -convexité, on a alors pour tout  $x \in \mathbb{R}^d$

$$J(x) \geq J(x_*) + \langle \nabla J(x_*), x - x_* \rangle + \frac{\alpha}{2}\|x - x_*\|^2 \geq J(x_*) + \frac{\alpha}{2}\|x - x_*\|^2.$$

Ainsi, si  $x \neq x_*$ ,  $J(x) > J(x_*)$ , d'où l'unicité du minimum. □

**Remarque 2.15.** Si  $J : X \rightarrow \mathbb{R}$  est une fonction  $\alpha$ -convexe dérivable sur  $X$  où  $X$  est un espace de Banach, on peut montrer l'existence et l'unicité du minimiseur. Pour l'existence, la suite minimisante est nécessairement bornée dans  $X$ . On peut en extraire une sous-suite, notée  $(x^{(n)})$ , qui converge faiblement vers  $x_*$ . On a alors par définition de la convexité

$$\forall n \in \mathbb{N}, J(x^{(n)}) \geq J(x_*) + dJ(x_*)(x^{(n)} - x_*).$$

L'application  $dJ(x_*)$  est aussi continue pour la topologie faible, donc  $\lim_{n \rightarrow \infty} dJ(x_*)(x^{(n)} - x_*) = 0$  et  $\lim_{n \rightarrow \infty} J(x^{(n)}) \geq J(x_*)$ . Or  $(x^{(n)})$  est une suite minimisante donc  $x_*$  est un minimiseur.

La dérivabilité de  $J$  n'est pas nécessaire pour montrer l'existence du minimum, on peut la remplacer en utilisant la notion de sous-différentielle pour les fonctions convexes.

## 2.3 Existence de minimiseur pour des problèmes avec contraintes

On considère une fonction continue  $J : \Omega \rightarrow \mathbb{R}$  où  $\Omega$  est un ouvert de  $X$  un espace de Banach. Soit  $A \subset \Omega$  un ensemble fermé, appelé dans la suite *ensemble admissible*. Dans cette partie, le problème qui nous intéresse est le problème d'optimisation sous contraintes

$$\text{Trouver } x_* \in A \subset \Omega \text{ vérifiant } J(x_*) = \inf_{x \in A} J(x). \quad (7)$$

On peut alors énoncer un résultat d'existence analogue au cas sans contrainte.

**Proposition 2.16** (Existence d'un minimiseur). *Soit  $J : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction continue coercive et  $A$  un ensemble fermé de  $\mathbb{R}^d$ . Alors le problème (7) admet une solution  $x_* \in A$ .*

*Démonstration.* On considère une suite minimisante. Par coercivité, cette suite est bornée. Il existe une sous-suite convergente. Par fermeture de  $A$ , la limite de cette sous-suite est bien dans  $A$ .  $\square$

### 2.3.1 Contraintes d'égalités

Dans le cas de contraintes d'égalité,  $A$  est décrit par un ensemble de fonctions  $(g_i)_{1 \leq i \leq p}$  de  $\Omega$  dans  $\mathbb{R}$  supposées continues. L'ensemble admissible  $A$  est alors donné par

$$A = \{x \in \Omega \mid \forall 1 \leq i \leq p, g_i(x) = 0\}.$$

**Théorème 2.17** (Extrema liés). *Soient  $J, g_1, \dots, g_p : \mathbb{R}^n \rightarrow \mathbb{R}$  des fonctions continûment dérivables et  $A = \{x \in \mathbb{R}^n, \forall 1 \leq i \leq p, g_i(x) = 0\}$ . On suppose*

- ★ *qu'il existe  $x_* \in A$  tel que  $J(x_*) = \inf_{x \in A} J(x)$*
  - ★ *que la famille  $(\nabla g_1(x_*), \dots, \nabla g_p(x_*))$  est libre (condition de qualification des contraintes).*
- Alors il existe des réels  $\lambda_1, \dots, \lambda_p$  (appelés multiplicateurs de Lagrange) tels que*

$$\nabla J(x_*) + \sum_{i=1}^p \lambda_i \nabla g_i(x_*) = 0. \quad (8)$$

*Cette équation est appelée équation d'Euler-Lagrange.*

Nous allons montrer ce théorème dans le cas plus simple où  $\begin{bmatrix} g_1(x) \\ \vdots \\ g_p(x) \end{bmatrix} = Cx + b$ , pour une matrice  $C \in \mathcal{M}_{p \times n}(\mathbb{R})$  et  $b \in \mathbb{R}^p$ .

*Démonstration du Théorème 2.17 dans un cas simplifié.* Soit  $x_* \in A$  un minimiseur du problème sous contraintes d'égalité. Pour tout  $x \in A$  et  $t \in \mathbb{R}$ , on a  $(1-t)x_* + tx \in A$ . Par définition de  $x_*$ , on a alors

$$J((1-t)x_* + tx) \geq J(x_*).$$

Donc on a  $\langle \nabla J(x_*), x - x_* \rangle = 0$  pour tout  $x \in A$ . Or  $x - x_*$  vérifie  $C(x - x_*) = 0$  donc  $x - x_* \in \text{Ker}(C)$ . Il suffit de montrer que  $\text{Ker}(C) = \text{Im}(C^T)^\perp$  pour montrer le résultat. En effet,  $\nabla(Cx + b) = C^T$  et la condition  $\langle \nabla J(x_*), x - x_* \rangle = 0$  pour tout  $x - x_* \in \text{Im}(C^T)^\perp$  s'écrit  $\nabla J(x_*) \in \text{Im}(C^T)$  donc il existe  $\Lambda \in \mathbb{R}^p$  tel que  $\nabla J(x_*) = -C^T \Lambda$ .

Montrons donc que  $\text{Ker}(C) = \text{Im}(C^T)^\perp$ . Pour  $x \in \text{Ker}(C)$ , on a pour tout  $y \in \mathbb{R}^p$ ,  $\langle y, Cx \rangle = 0 \Leftrightarrow \langle C^T y, x \rangle = 0$ , donc par équivalence  $\text{Ker}(C) = \text{Im}(C^T)^\perp$ .  $\square$



**Remarque 2.18.** Dans le cas général, la preuve repose sur des arguments de géométrie différentielle. Soit  $G(x) = [g_1(x), \dots, g_p(x)]^T$ . Si  $\nabla G(x_*)$  est de rang  $p$ , localement dans un voisinage  $V$  de  $x_*$ ,  $A \cap V$  définit une variété différentielle de dimension  $p$ , dont l'espace tangent en  $x_*$  est défini par  $T_{x_*}(A \cap V) = \{\nabla g_1(x_*), \dots, \nabla g_p(x_*)\}^\perp$ . En reprenant le même schéma que la preuve précédente, on fixe  $y \in T_{x_*}(A \cap V)$  et  $t \mapsto x(t)$  une trajectoire telle que pour tout  $t \in [-1, 1]$ ,  $x(t) \in A \cap V$ ,  $x(0) = x_*$  et  $\dot{x}(0) = y$ . Dans ce cas, par les mêmes calculs, on montre que  $\langle \nabla J(x_*), y \rangle = 0$ , donc  $\nabla J(x_*) \perp T_{x_*}(A \cap V) = \{\nabla g_1(x_*), \dots, \nabla g_p(x_*)\}^\perp$ . On en déduit que  $\nabla J(x_*) \in \text{Vect}(\nabla g_1(x_*), \dots, \nabla g_p(x_*))$ .

### 2.3.2 Contraintes d'inégalité

Dans le cas de contraintes d'inégalité,  $A$  est décrit par un ensemble de fonctions  $(h_i)_{1 \leq i \leq q}$  de  $\Omega$  dans  $\mathbb{R}$  supposées continues. L'ensemble admissible  $A$  est alors donné par

$$A = \{x \in \Omega \mid \forall 1 \leq i \leq q, h_i(x) \leq 0\}. \quad (9)$$

Le problème d'optimisation avec contraintes d'inégalité est donné par

$$\min_{x \in A} J(x), \quad (10)$$

où  $J$  est une fonction de  $\Omega \subset X$ , où  $X$  est un espace de Banach, dans  $\mathbb{R}$ .

Ces problèmes sont en fait équivalents aux problèmes d'optimisation avec contraintes d'égalité mais avec des variables additionnelles.

Pour simplifier la discussion, supposons que  $q = 1$  et considérons le problème d'optimisation

$$\min_{x \in \mathbb{R}^n, h(x) \leq 0} J(x),$$

où  $J : \mathbb{R}^n \rightarrow \mathbb{R}$  est continue. Supposons que  $x_*$  est une solution du problème d'optimisation au dessus et posons  $\varepsilon_* \geq 0$  tel que  $h(x_*) = -\varepsilon_*^2$ .

Soit  $\tilde{h}(x, \varepsilon) = h(x) + \varepsilon^2$  et considérons le problème d'optimisation avec contraintes d'égalité

$$\min_{(x, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}, \tilde{h}(x, \varepsilon) = 0} J(x).$$

Pour tout  $\varepsilon \in \mathbb{R}$  et tout  $x \in \mathbb{R}^n$  tels que  $\tilde{h}(x, \varepsilon) = 0$ , on a  $h(x) = -\varepsilon^2 \leq 0$ . Donc l'ensemble des  $x \in \mathbb{R}^n$  vérifiant  $\tilde{h}(x, \varepsilon) = 0$  pour un certain  $\varepsilon \in \mathbb{R}$  est inclus dans  $\{x \mid h(x) \leq 0\}$ . On a donc

$$\min_{(x, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}, \tilde{h}(x, \varepsilon) = 0} J(x) \geq \min_{x \in \mathbb{R}^n, h(x) \leq 0} J(x).$$

Or  $(x_*, \varepsilon_*)$  vérifie  $\tilde{h}(x_*, \varepsilon_*) = 0$  donc

$$\min_{(x, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}, \tilde{h}(x, \varepsilon) = 0} J(x) = \min_{x \in \mathbb{R}^n, h(x) \leq 0} J(x).$$

On peut écrire alors l'équation d'Euler-Lagrange pour le problème augmenté d'une variable et sous l'hypothèse que  $\begin{bmatrix} \nabla h(x_*) \\ 2\varepsilon_* \end{bmatrix} \neq 0$ , il existe  $\lambda \in \mathbb{R}$  tel que

$$\begin{bmatrix} \nabla J(x_*) \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} \nabla h(x_*) \\ 2\varepsilon_* \end{bmatrix} = 0.$$

On a alors la dichotomie suivante

- (i) si  $h(x_*) = 0$ , alors  $\varepsilon_* = 0$  donc on retrouve les équations d'Euler-Lagrange du cas d'égalité  $\nabla J(x_*) + \lambda \nabla h(x_*) = 0$
- (ii) si  $h(x_*) < 0$ , nécessairement  $\lambda = 0$ , donc  $\nabla J(x_*) = 0$  : le minimum étant atteint dans l'intérieur de l'espace admissible  $A$ , la contrainte est alors invisible (notons que ceci est une particularité du cas avec une unique contrainte d'inégalité).

Ce résultat peut alors être généralisé pour des problèmes d'optimisation avec contraintes d'inégalité ou de contraintes mixtes. Soit  $I_0(x_*)$  l'ensemble des contraintes saturées au point  $x_*$  :

$$I_0(x_*) = \{i \in \{1, \dots, q\} / h_i(x_*) = 0\}.$$

**Définition 2.19.** On dit que les contraintes sont qualifiées au point  $x_*$  si les vecteurs  $\nabla h_i(x_*)$ , pour  $i \in I_0(x_*)$ , sont linéairement indépendants.

**Théorème 2.20** (Conditions de Karush-Kuhn-Tucker (KKT)). Soient  $J$  et  $(h_i)_{1 \leq i \leq q}$  des fonctions continûment dérivables de  $\mathbb{R}^n$  dans  $\mathbb{R}$ . On suppose les contraintes qualifiées au point  $x_*$ . Alors une condition nécessaire pour que  $x_*$  soit un minimum de  $J$  sur l'ensemble  $A = \{x \in \mathbb{R}^n, h_i(x) \leq 0, \forall 1 \leq i \leq q\}$  est qu'il existe des réels positifs  $\lambda_1, \lambda_2, \dots, \lambda_q$  (appelés multiplicateurs de Kuhn-Tucker ou de Lagrange généralisés) tels que

$$\nabla J(x_*) + \sum_{i=1}^q \lambda_i \nabla h_i(x_*) = 0, \text{ où } \lambda_i h_i(x_*) = 0 \text{ pour } i \in \{1, \dots, q\}. \quad (11)$$

Cette équation est appelée conditions de Karush-Kuhn-Tucker (KKT).

## 3 Algorithmes

### 3.1 Optimisation sans contrainte

#### 3.1.1 Méthodes de descente

**Définition 3.1.** Soit  $J : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction dérivable,  $x \in \mathbb{R}^n$ ,  $d \in \mathbb{R}^n$ . On dit que  $d$  est une direction de descente si

$$\langle \nabla J(x), d \rangle < 0.$$

**Proposition 3.2.** Soit  $J : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction dérivable,  $x \in \mathbb{R}^n$  tel que  $\nabla J(x) \neq 0$ ,  $d \in \mathbb{R}^n$ . Si  $d$  est une direction de descente en  $x$ , alors il existe  $\eta > 0$  tel que

$$\forall \alpha \in [0, \eta[, \quad J(x + \alpha d) < J(x).$$

Les algorithmes de descente sont de la forme

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

où  $d^{(k)}$  est la direction de descente et  $t_k$  est appelé le pas de descente.

### 3.1.2 Méthodes de gradient à pas fixe, variable et optimal

Les méthodes de gradient consistent à choisir comme direction de descente  $d^{(k)} = -\nabla J(x^{(k)})$ . Ce choix est motivé par le fait que par la formule de Taylor

$$J(x + td) = J(x) + t\langle \nabla J(x), d \rangle + o(t\|d\|),$$

donc pour  $t\|d\|$  petit, le terme dominant est le produit scalaire  $\langle \nabla J(x), d \rangle$  qui est maximal pour  $d$  dans la direction de  $\nabla J(x)$ . Elles s'écrivent donc

$$x^{(k+1)} = x^{(k)} - t_k \nabla J(x^{(k)}).$$

Pour  $t_k = t$  constant, on obtient la méthode de *gradient à pas fixe*. Sinon, on parle de méthode de *gradient à pas variable*. En particulier, si on prend  $t_k$  (s'il existe) réalisant

$$J(x^{(k)} - t_k \nabla J(x^{(k)})) = \inf_{t \in \mathbb{R}} J(x^{(k)} - t \nabla J(x^{(k)}))$$

on obtient la méthode du *gradient à pas optimal*.

**Théorème 3.3.** Soit  $J : \mathbb{R}^n \rightarrow \mathbb{R}$  continûment dérivable, vérifiant

- ★  $J$  est  $\alpha$ -convexe
- ★  $\nabla J$  est  $M$ -lipschitzienne
- ★  $\exists c_1 \in \mathbb{R}, c_2 \in \mathbb{R}$  tel que pour tout  $k \in \mathbb{N}$ ,  $0 < c_1 \leq t_k \leq c_2 < \frac{2\alpha}{M^2}$ .

Soit  $x_* \in \mathbb{R}^n$  l'unique minimiseur de  $J : J(x_*) = \min_{x \in \mathbb{R}^n} J(x)$ . Alors la méthode de gradient à pas variable

$$x^{(k+1)} = x^{(k)} - t_k \nabla J(x^{(k)})$$

converge vers  $x_*$  pour tout choix de  $x_0$ , de façon géométrique :

$$\exists \beta \in ]0, 1[, \quad \|x^{(k)} - x_*\| \leq \beta^n \|x_0 - x_*\|.$$

## 3.2 Optimisation avec contraintes

### 3.2.1 Contraintes d'égalité

**Méthode de Newton** On peut adapter la méthode de Newton à la minimisation d'une fonction  $J$  avec des contraintes égalité. Si l'ensemble des contraintes s'écrit

$$A = \{x \in \mathbb{R}^n, \quad g_i(x) = 0, \quad 1 \leq i \leq p\} \quad (12)$$

où  $g_1, \dots, g_p : \mathbb{R}^n \rightarrow \mathbb{R}$  sont continûment dérivables, tels que  $\nabla g_1(x_*), \dots, \nabla g_p(x_*)$  sont linéairement indépendants, la condition nécessaire d'optimalité au point  $x_*$  s'écrit d'après le Théorème 2.17

$$\nabla J(x_*) + \sum_{i=1}^p \lambda_i \nabla g_i(x_*) = 0, \quad (13)$$

où  $\lambda_1, \dots, \lambda_p$  sont les multiplicateurs de Lagrange. La méthode consiste alors, en posant

$$F(x, \lambda) = \begin{pmatrix} \nabla J(x) + \langle \lambda, G(x) \rangle \\ G(x) \end{pmatrix}, \quad \text{avec} \quad G(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_p(x) \end{pmatrix},$$

à déterminer  $x_*$  en résolvant l'équation  $F(x, \lambda) = 0$  par une méthode de Newton.

**Méthode de pénalisation** On suppose ici que l'ensemble admissible  $A$  est tel qu'il existe une fonction continue  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^+$  tel que

$$x \in A \iff \phi(x) = 0. \quad (14)$$

L'idée de la méthode consiste à remplacer  $J$  par une fonction  $J_\varepsilon$  définie par

$$J_\varepsilon(x) = J(x) + \frac{1}{\varepsilon}\phi(x) \quad (15)$$

et à minimiser  $J_\varepsilon$  pour des valeurs de  $\varepsilon$  de plus en plus petites. Le résultat suivant vient justifier cette méthode dans le cas où  $\phi$  et  $J$  sont convexe et strictement convexe respectivement.

**Proposition 3.4.** *Soit  $J : \mathbb{R}^n \rightarrow \mathbb{R}^+$  strictement convexe, coercive, définie sur  $\mathbb{R}^n$ . On suppose qu'il existe une fonction  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^+$  convexe continue vérifiant (14), et on définit  $J_\varepsilon$  par (15). Alors*

- (i)  $\forall \varepsilon > 0, \exists! x_\varepsilon \in \mathbb{R}^n, J_\varepsilon(x_\varepsilon) = \min_{x \in \mathbb{R}^n} J_\varepsilon(x).$
- (ii)  $\exists! x_* \in A, \lim_{\varepsilon \rightarrow 0} x_\varepsilon = x_*$  et  $J(x_*) = \min_{x \in A} J(x).$

*Démonstration.* A  $\varepsilon > 0$  fixé, l'existence et l'unicité de  $x_\varepsilon \in \mathbb{R}^n$  provient de la stricte convexité de  $J_\varepsilon$ . Idem pour  $x_* \in A$ . On a

$$J(x_\varepsilon) \leq J(x_\varepsilon) + \frac{1}{\varepsilon}\phi(x_\varepsilon) \leq J(x_*) + \underbrace{\frac{1}{\varepsilon}\phi(x_*)}_{=0 \text{ car } x_* \in A} = J(x_*)$$

Comme  $J$  est coercive, l'ensemble  $\{x_\varepsilon, \varepsilon > 0\}$  est borné. On peut donc en extraire une sous-suite (encore notée  $x_\varepsilon$ ) qui converge vers un élément  $y \in \mathbb{R}^n$ . Par passage à la limite dans l'inégalité précédente, on a  $J(y) \leq J(x_*)$ . De plus,

$$0 \leq \phi(x_\varepsilon) \leq \varepsilon \underbrace{J(x_*) - J(x_\varepsilon)}_{\text{terme borné}}$$

ce qui prouve par passage à la limite quand  $\varepsilon \rightarrow 0$  et continuité de  $\phi$  (en dimension finie toute fonction convexe est continue) que  $\phi(y) = 0$ , et donc  $y \in A$ . On en déduit que  $y = x_*$ . De plus, toute suite d'un compact possédant une seule valeur d'adhérence converge vers cette valeur d'adhérence.  $\square$

### 3.2.2 Contraintes convexes

**Méthode de projection** On suppose ici que l'ensemble  $A$  des contraintes est un convexe fermé, que l'on note ici  $C$ . On note  $P_C$  la projection sur  $C$  :

$$P_C(x) = \arg \min_{c \in C} \|x - c\|. \quad (16)$$

Etant donné un algorithme  $\mathcal{A}$  de minimisation sans contraintes, l'idée de la méthode de projection est de forcer la suite des itérés successifs à rester dans le convexe  $C$ , avec l'opérateur de projection  $C$ .

Nous allons d'abord établir que l'opérateur de projection sur un convexe fermé est bien défini, puis nous énoncerons la convergence de l'algorithme de projection dans le cas de l'algorithme de gradient à pas fixe.

---

**Algorithm 3.1** Algorithme de projection

---

```
1: Initialisation : Choix de  $x^{(0)} \in C$ .  
2: for  $k \geq 0$  jusqu'à critère d'arrêt do  
3:    $\hat{x}^{(k+1)} = \mathcal{A}(x^{(k)})$   
4:    $x^{(k+1)} = P_C(\hat{x}^{(k+1)})$   
5: end for
```

---

**Proposition 3.5** (Opérateur de projection sur un convexe fermé). *Si  $C$  est un ensemble convexe fermé non vide de  $\mathbb{R}^n$ , alors l'application  $P_C$  donnée par l'Equation (16) est bien définie. De plus on a*

- (i) *pour tout  $c \in C$ ,  $\langle x - P_C(x), c - P_C(x) \rangle \leq 0$*
- (ii) *pour tout  $x, y \in \mathbb{R}^n$ ,  $\|P_C(x) - P_C(y)\| \leq \|x - y\|$  ( $P_C$  est une application contractante).*

Si  $\mathcal{A}$  est l'algorithme du gradient à pas fixe, on peut montrer que cet algorithme converge pour des paramètres assez petits.

---

**Algorithm 3.2** Algorithme du gradient projeté

---

```
1: Initialisation : Choix de  $x^{(0)} \in C$ ,  $t > 0$ .  
2: for  $k \geq 0$  jusqu'à critère d'arrêt do  
3:    $\hat{x}^{(k+1)} = x^{(k)} - t \nabla J(x^{(k)})$   
4:    $x^{(k+1)} = P_C(\hat{x}^{(k+1)})$   
5: end for
```

---

**Théorème 3.6** (Convergence de l'algorithme du gradient projeté). *Soit  $J : \mathbb{R}^n \rightarrow \mathbb{R}$  continûment dérivable, vérifiant*

- ★  *$J$  est  $\alpha$ -convexe*
- ★  *$\nabla J$  est  $M$ -lipschitzienne.*

*Soit  $t > 0$  tel que  $t < \frac{2\alpha}{M^2}$ . Soit  $C$  un ensemble convexe fermé, non vide de  $\mathbb{R}^n$ . Soit  $x_* \in C$  l'unique minimiseur de  $J$  sur  $C$  :  $J(x_*) = \min_{x \in C} J(x)$ .*

*Alors l'algorithme du gradient projeté converge vers  $x_*$ .*

## Références

[Cia82] Philippe G Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. 1982.