

Chapitre 1.

## Introduction à l'optimisation convexe

### 1.1 Conditions d'optimalité pour la minimisation sans contraintes

Dans ce chapitre, nous introduisons quelques définitions pour étudier des problèmes d'optimisation de la forme

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

Commençons par introduire les notions de minimum/maximum global d'une fonction  $f$  définie sur une partie  $S$  de  $\mathbb{R}^n$ .

#### 1.1.1 Minimum/Maximum global

**Définition 1** (Minimum/Maximum global). Soit  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  et  $x^* \in S$ . Alors

1.  $x^*$  est dit un minimum global de  $f$  sur  $S$  si  $f(x) \geq f(x^*)$  pour tout  $x \in S$ .
2.  $x^*$  est dit un minimum global strict de  $f$  sur  $S$  si  $f(x) > f(x^*)$  pour tout  $x^* \neq x \in S$ .
3.  $x^*$  est dit un maximum global de  $f$  sur  $S$  si  $f(x) \leq f(x^*)$  pour tout  $x \in S$ .
4.  $x^*$  est dit un maximum global strict de  $f$  sur  $S$  si  $f(x) < f(x^*)$  pour tout  $x^* \neq x \in S$ .

**Définition 2.** On note par  $\operatorname{argmin}_S f$  l'ensemble des minima globaux de  $f$  sur  $S$ . De même, on note par  $\operatorname{argmax}_S f$  l'ensemble des maxima globaux de  $f$  sur  $S$ .

**Exemple 1.** On considère  $f : (x, y) \in \mathbb{R}^2 \mapsto x + y$  et  $S = B_f(0, 1) = \{(x, y) : x^2 + y^2 \leq 1\}$ . On vérifie que  $\max_S f = \sqrt{2}$  et que  $\operatorname{argmax}_S f = \{(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})\}$ . En effet, pour  $(x, y) \in S$ , l'inégalité de Cauchy-Schwarz donne

$$f(x, y) = \langle (x, y), (1, 1) \rangle \leq \|(x, y)\|_2 \|(1, 1)\|_2 \leq \sqrt{2}.$$

De même, on a  $\min_S f = -\sqrt{2}$  et que  $\operatorname{argmin}_S f = \{(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})\}$ . On verra plus loin (ou l'année prochaine :) ) comment trouver résoudre ce type de problèmes grâce aux conditions KKT.

**Définition 3.** 1.  $x^*$  est dit un minimum locale de  $f$  sur  $S$  s'il existe  $\varepsilon > 0$  tel que  $f(x) \geq f(x^*)$  pour tout  $x \in S \cap B(x^*, \varepsilon)$ .

2.  $x^*$  est dit un minimum locale strict de  $f$  sur  $S$  s'il existe  $\varepsilon > 0$  tel que  $f(x) > f(x^*)$  pour tout  $x^* \neq x \in S \cap B(x^*, \varepsilon)$ .
3.  $x^*$  est dit un maximum global de  $f$  sur  $S$  s'il existe  $\varepsilon > 0$  tel que  $f(x) \leq f(x^*)$  pour tout  $x \in S \cap B(x^*, \varepsilon)$ .
4.  $x^*$  est dit un maximum global strict de  $f$  sur  $S$  s'il existe  $\varepsilon > 0$  tel que  $f(x) < f(x^*)$  pour tout  $x^* \neq x \in S \cap B(x^*, \varepsilon)$

### 1.1.2 Condition d'optimalité du premier ordre

On rappelle le résultat suivant, souvent connu sous la **règle de Fermat** (c.f Fig. 1.1).

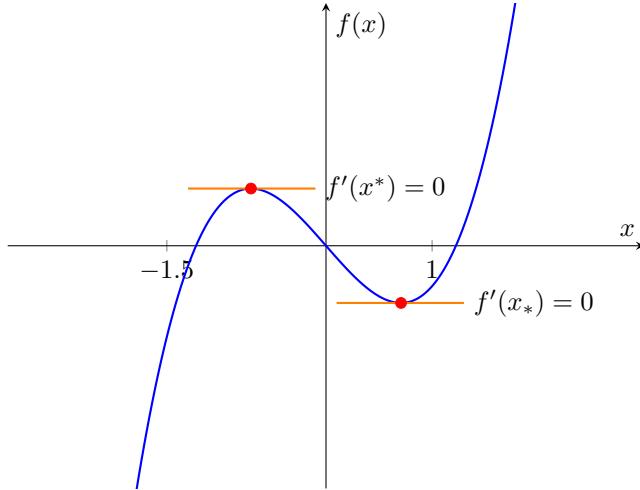


Figure 1.1: Illustration de la règle de Fermat: la fonction  $f$  a un minimum (respectivement maximum) local en  $x_*$  (respectivement  $x^*$ ). La dérivée de  $f$  s'annule en ces deux points. Géométriquement, la tangente de  $f$  en ces deux points est horizontale.

**Proposition 1 (Règle de Fermat).** Soit  $f : (a, b) \subset \mathbb{R} \rightarrow \mathbb{R}$  une fonction différentiable. Si  $x^* \in (a, b)$  est un minimum/maximum local de  $f$  alors  $f'(x^*) = 0$ .

**Preuve.** Supposons que  $x^*$  est un minimum local de  $f$  (le cas du maximum se démontre de façon analogue). Il existe  $\delta > 0$  tel que pour tout  $h$  avec  $0 < |h| < \delta$  et  $x^* + h \in (a, b)$  on a  $f(x^* + h) - f(x^*) \geq 0$ .

Comme  $f$  est différentiable en  $x^*$ , on peut écrire pour  $h \neq 0$  petit :

$$f(x^* + h) - f(x^*) = (f'(x^*) + \varepsilon(h))h,$$

avec  $\varepsilon(h) \rightarrow 0$  quand  $h \rightarrow 0$ . Pour  $h > 0$  suffisamment petit, l'inégalité  $f(x^* + h) - f(x^*) \geq 0$  donne

$$f'(x^*) + \varepsilon(h) \geq 0,$$

et pour  $h < 0$  suffisamment petit (donc  $-h > 0$  petit) l'inégalité  $f(x^* + h) - f(x^*) \geq 0$  devient

$$f'(x^*) + \varepsilon(h) \leq 0.$$

En faisant tendre  $h \rightarrow 0^+$  puis  $h \rightarrow 0^-$ , on obtient  $f'(x^*) \geq 0$  et  $f'(x^*) \leq 0$ , d'où  $f'(x^*) = 0$ .  $\square$

Le résultat suivant est une généralisation de la règle de Fermat.

**Théorème 1.** Soit  $f : S \rightarrow \mathbb{R}$  une fonction définie sur un ensemble  $S \subset \mathbb{R}^n$ . Supposons que  $x^* \in \text{int}(S)$  est un extrémum local et que toutes les dérivées partielles de  $f$  en  $x^*$  existent. Alors  $\nabla f(x^*) = 0$ .

**Preuve.** On définit, pour  $i \in \llbracket 1, n \rrbracket$ , la fonction de la variable réelle  $\phi(t) = f(x^* + te_i)$  ( $e_i$  étant le  $i$ ème vecteur de la base canonique de  $\mathbb{R}^n$ ). On remarquera que  $\phi$  est différentiable en 0 et  $\phi'(0) = \partial_{x_i} f(x^*)$ . Comme  $x^*$  est un extrémum local de  $f$ , il vient que  $t = 0$  est un extrémum local de  $\phi$ , et par la suite  $\phi'(0) = 0$ . Il s'en suit que  $\partial_{x_i} f(x^*) = 0$ , pour tout  $i \in \llbracket 1, n \rrbracket$ , i.e.,  $\nabla f(x^*) = 0$ .  $\square$

Le [Théorème 1](#) est donné une condition nécessaire d'optimalité, à savoir que le gradient d'une fonction en les extrema locaux intérieurs au domaine de la fonction est nul. La réciproque est fausse. Il existe des points (lesquelles ?) qui ne sont pas des extrema locaux et pourtant le gradient de la fonction en ces points est nul. Cela motive la définition suivante.

**Définition 4 (Point critique).** Soit  $f : S \rightarrow \mathbb{R}$  une fonction définie sur un ensemble  $S \subset \mathbb{R}^n$ . Supposons que  $x^* \in \text{int}(S)$  et que  $f$  est différentiable au voisinage de  $x^*$ . Alors  $x^*$  est un point stationnaire (ou critique) de  $f$  si  $\nabla f(x^*) = 0$ .

**Remarque 1.** Le [Théorème 1](#) affirme que les extrema locaux sont des points critiques.

**Exemple 2.** Considérons  $f(x) = x^3$ . On a que  $f'(x) = 3x^2$  et donc  $f'(0) = 0$  pourtant 0 n'est ni un minimum ni un maximum de  $f$ .

### 1.1.3 Conditions d'optimalité du second ordre

**Théorème 2 (Conditions nécessaires).** Soit  $f : S \rightarrow \mathbb{R}$  une fonction définie sur un ouvert  $S \subset \mathbb{R}^n$ . Supposons que  $f$  est deux fois continûment différentiable sur  $S$  et soit  $x^*$  un point stationnaire de  $f$ . Alors

1. Si  $x^*$  est un minimum local de  $f$  sur  $S$  alors  $\nabla^2 f(x^*) \succeq 0$ .
2. Si  $x^*$  est un maximum local de  $f$  sur  $S$  alors  $\nabla^2 f(x^*) \preceq 0$ .

**Preuve.** Comme  $x^*$  est un minimum local, on a  $\nabla f(x^*) = 0$ , et pour  $\varepsilon > 0$  suffisamment petit

$$f(x) = f(x^*) + \frac{1}{2} \langle \nabla^2 f(x^*)(x - x^*), x - x^* \rangle + o(\|x - x^*\|^2) \geq f(x^*), \text{ pour tout } x \in B(x^*, \varepsilon).$$

Posons  $v = \frac{x - x^*}{\|x - x^*\|}$ , on obtient en divisant par  $\|x - x^*\|^2$  dans l'inégalité précédente

$$\langle \nabla^2 f(x^*)v, v \rangle + o(1) \geq 0.$$

En prenant la limite quand  $x$  tend vers  $x^*$  on obtient que  $\langle \nabla^2 f(x^*)v, v \rangle \geq 0$  pour tout  $v$  tel que  $\|v\| = 1$ . Ainsi,  $\nabla^2 f(x^*)$  est semi-définie positive. On conclut en appliquant le même raisonnement à  $-f$ .  $\square$

Le résultat suivant donne une condition suffisante d'optimalité locale stricte.

**Théorème 3 (Condition suffisante).** Soit  $f : S \rightarrow \mathbb{R}$  une fonction définie sur un ouvert  $S \subset \mathbb{R}^n$ . Supposons que  $f$  est deux fois différentiable sur  $S$  et soit  $x^* \in \text{int}(S)$  un point stationnaire de  $f$ . Alors

1. Si  $\nabla^2 f(x^*) \succ 0$ , alors  $x^*$  est un minimum local strict de  $f$  sur  $S$ .
2. Si  $\nabla^2 f(x^*) \prec 0$ , alors  $x^*$  est un maximum local strict de  $f$  sur  $S$ .

**Preuve.** D'une part, comme  $\nabla^2 f(x^*) \succ 0$ , sa plus petite valeur propre  $\lambda_{\min} > 0$  et on a  $\nabla^2 f(x^*) \succeq \lambda_{\min} \text{Id}$ . D'autre part, comme  $f$  est deux fois différentiable et  $\nabla f(x^*) = 0$ , on a par ??

$$\begin{aligned} f(x) &= f(x^*) + \frac{1}{2} \langle \nabla^2 f(x^*)(x - x^*), x - x^* \rangle + o(\|x - x^*\|^2) \\ &\geq f(x^*) + \frac{1}{2} \lambda_{\min} \|x - x^*\|^2 + o(\|x - x^*\|^2) \end{aligned} \tag{1.1}$$

Comme  $o(r)/r \rightarrow 0$ , il existe  $r^* > 0$  tel que pour tout  $r \in (0, r^*]$  on ait  $o(r) \leq \frac{1}{4}r\lambda_{\min}$ . Il s'en

suit que pour tout  $x \in B(x^*, r)$  on a  $|o(\|x - x^*\|^2)| \leq \frac{1}{4}\lambda_{\min}\|x - x^*\|^2$  et par la suite

$$\begin{aligned} f(x) &= f(x^*) + \frac{1}{2}\langle \nabla^2 f(x^*)(x - x^*), x - x^* \rangle + o(\|x - x^*\|^2) \\ &\geq f(x^*) + \frac{1}{2}\lambda_{\min}\|x - x^*\|^2 + o(\|x - x^*\|^2) \\ &\geq f(x^*) + \frac{1}{4}\lambda_{\min}\|x - x^*\|^2 > f(x^*). \end{aligned} \tag{1.2}$$

□

Une preuve alternative est la suivante.

**Preuve.** Soit  $x^*$  un point stationnaire tel que  $\nabla^2 f(x^*) \succ 0$ , comme  $f$  est de classe  $C^2$ , sa Hesseinne est continue, et il existe donc un boule  $B(x^*, r) \subset S$  telle que  $\nabla^2 f(x) \succ 0$  pour tout  $x \in B(x^*, r)$ . Par ??, il existe  $z_x \in [x, x^*]$  tel que

$$f(x) = f(x^*) + \underbrace{\frac{1}{2}\langle \nabla^2 f(z_x)(x - x^*), x - x^* \rangle}_{>0},$$

donc  $f(x) > f(x^*)$  pour tout  $x \in B(x^*, r), x \neq x^*$ . Donc  $x^*$  est un minimum local stricte de  $f$  sur  $S$ . □

Les conditions d'optimalité que nous venons de présenter garantissent seulement l'optimalité **locale** des points stationnaires, puisqu'elles reposent sur une information locale — à savoir les valeurs du gradient et de la hessienne en un point donné. Le résultat suivant montre que, lorsque la hessienne est semi-définie positive partout (ça vous rappelle quelque chose ? :)), tout point stationnaire est en réalité un **minimum global**.

**Théorème 4 (Condition d'optimalité globale).** Soit  $f \in C^2(\mathbb{R}^n)$  telle que  $\nabla^2 f(x) \succeq 0$  pour tout  $x \in \mathbb{R}^n$ . Si  $x^*$  est un point stationnaire de  $f$ , alors  $x^*$  est un minimum global de  $f$ .

**Preuve.** Soit  $x \in \mathbb{R}^n$ . D'après ?? il existe  $y_x \in [x^*, x]$  tel que

$$f(x) - f(x^*) = \frac{1}{2}\langle \nabla^2 f(y_x)(x - x^*), x - x^* \rangle.$$

Puisque  $\nabla^2 f(y_x) \succeq 0$ , il vient  $f(x) - f(x^*) \geq 0$ , ce qui montre que  $f(x) \geq f(x^*)$  pour tout  $x \in \mathbb{R}^n$ . Ainsi,  $x^*$  est un **minimum global** de  $f$ . □

Avant de finir cette section, on introduit un autre type de points critiques évoqués par exemple dans l'[Exemple 2](#) et [Fig. 1.2](#).

**Définition 5 (Point selle).** Soit  $f : S \rightarrow \mathbb{R}$  une fonction définie sur un ouvert  $S \subseteq \mathbb{R}^n$ , supposée continûment différentiable sur  $S$ . Un point stationnaire  $x^* \in S$  est appelé *point selle* de  $f$  sur  $S$  s'il n'est ni un minimum local ni un maximum local de  $f$  sur  $S$ .

**Théorème 5 (Condition suffisante pour un point selle).** Soit  $f : S \rightarrow \mathbb{R}$  une fonction définie sur un ouvert  $S \subseteq \mathbb{R}^n$ , supposée deux fois continûment différentiable sur  $S$ . Soit  $x^* \in S$  un point stationnaire de  $f$ . Si la matrice Hessienne  $\nabla^2 f(x^*)$  est indéfinie, alors  $x^*$  est un point selle de  $f$  sur  $S$ .

**Preuve.** □

#### 1.1.4 Résultats d'existence

Jusqu'ici, nous avons défini les notions de minimum et de maximum comme s'ils existaient toujours. Il est pourtant fondamental de savoir si un problème du type

$$\min_{x \in C} f(x),$$

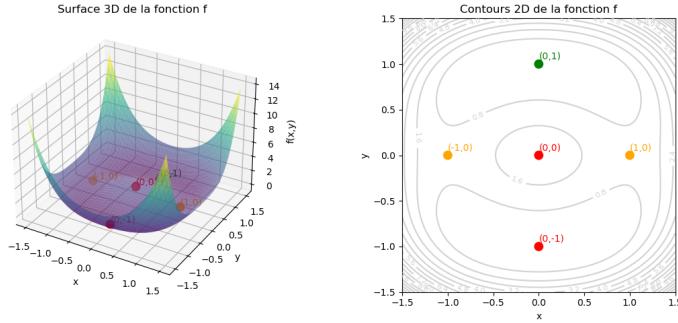


Figure 1.2: Points critiques de la fonction  $f(x, y) = (x^2 + y^2 - 1)^2 + (y^2 - 1)^2$ . Le point  $(0, 0)$  est un maximum local, les points  $(0, 1)$  et  $(0, -1)$  sont des minimas locaux stricts et aussi des minimas globaux tandis que  $(1, 0)$  et  $(-1, 0)$  sont des points-selles.

avec  $C = \mathbb{R}^n$  ou une partie de  $\mathbb{R}^n$ , admet effectivement une solution. Commençons par rappeler le résultat classique suivant.

**Théorème 6 (Weierstrass).** Soit  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue avec  $S$  un compact. Alors  $f$  est bornée et atteint ses bornes, i.e.,  $\exists x_* x^* \in S$  tels que

$$f(x^*) = \sup_{x \in S} f(x) \text{ et } f(x_*) = \inf_{x \in S} f(x).$$

Quand l'ensemble  $S$  n'est pas compact, on requiert la coercivité de  $f$ .

**Définition 6 (Coercivité).** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue. On dit que  $f$  est coercive si

$$\lim_{\|x\| \rightarrow \infty} f(x) = \infty.$$

**Théorème 7.** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue et coercive et  $S \subset \mathbb{R}^n$  un fermé non vide. Alors  $f$  admet un minimum global sur  $S$ .

**Preuve.** Soit  $x_0 \in S$ . Par coercivité, il existe  $K > 0$  tel qu'on ait  $f(x) > f(x_0)$  pour tout  $\|x\| > K$ . Si  $x^*$  est un minimum global de  $f$  sur  $S$ , alors  $f(x^*) \leq f(x_0)$  pour tout  $x \in S$  et en particulier pour tout  $x \in S \cap B(0, K)$ . Comme  $S$  est un fermé non vide,  $S \cap B(0, K)$  est un compact non vide et [Théorème 6](#) assure l'existence d'un minimum global de  $f$  sur  $S \cap B(0, K)$ , et donc sur  $S$ .  $\square$

En combinant la coercivité ([Définition 6](#)) et la semi-continuité inférieure ([??](#)) on peut étendre le résultat de Weierstrass:

**Théorème 8 (Existence d'un minimum).** Soit  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  une fonction **propre**, **semi-continuite inférieurement** et **coercive**. Alors  $f$  admet un minimiseur, c'est-à-dire qu'il existe  $x^* \in \mathbb{R}^n$  tel que

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

## 1.2 Optimisation convexe

### 1.2.1 Définitions et premières propriétés

Avant de finir ce chapitre, on va dégager quelques propriétés des problèmes d'optimisation convexe dans le sens suivant.

**Définition 7** (Problème convexe). Le problème d'optimisation

$$\min_{x \in C} f(x) \quad (1.3)$$

est dit convexe si  $C \subset \mathbb{R}^n$  un convexe et  $f$  une fonction convexe sur  $C$ .

Généralement, on rencontre des problèmes où l'ensemble des contraintes  $C$  combine à la fois des **contraintes d'égalité** et des **contraintes d'inégalité**. En effet, dans de nombreux problèmes pratiques, on cherche à déterminer un minimiseur  $x^* \in \mathbb{R}^n$  parmi tous les  $x \in \mathbb{R}^n$  vérifiant

$$g_i(x) = 0 \quad \text{et} \quad h_j(x) \leq 0,$$

où  $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$  sont des fonctions **affines**, et  $h_j : \mathbb{R}^n \rightarrow \mathbb{R}$  des fonctions **convexes**.

L'ensemble  $C$ , appelé **ensemble des contraintes** (ou **ensemble des solutions admissibles**), s'écrit alors

$$C = \bigcap_{j=1}^m \mathbf{Lev}(h_j, 0) \cap \bigcap_{i=1}^p g_i^{-1}(0),$$

c'est-à-dire comme l'intersection de sous-ensembles convexes, et est donc lui-même **convexe**.

Le résultat suivant affirme que pour les problèmes convexes, un minimum local est en réalité un minimum global.

**Théorème 9.** Soit  $f : C \rightarrow \mathbb{R}$  une fonction convexe définie sur un convexe  $C$ . Soit  $x^* \in C$  est un minimum local de  $f$  sur  $C$ . Alors  $x^*$  est un minimum global de  $f$  sur  $C$

**Preuve.** Soit  $x^*$  un minimum local de  $f$  sur  $C$ , il existe alors  $r > 0$  tel que  $f(x^*) \leq f(x)$  pour tout  $y \in C$  avec  $\|x^* - y\| \leq r$ . Pour montrer que  $f(x^*) \leq f(y)$  pour tout  $y \in C \setminus \{x^*\}$ , on considère un  $\bar{x} = (1-t)x^* + ty$  pour un certain  $t \in (0, 1]$  de telle façon que  $\|\bar{x} - x^*\| \leq r$ . On peut prendre par exemple  $t = r/\|x^* - y\|$ . Comme  $x^*$  est minimum local, il suit, par convexité de  $f$

$$f(x^*) \leq f(\bar{x}) = f(ty + (1-t)x^*) \leq tf(y) + (1-t)f(x^*),$$

soit  $tf(x^*) \leq tf(y)$  et donc  $f(x^*) \leq f(y)$ . □

De même on démontre que

**Théorème 10.** Soit  $f : C \rightarrow \mathbb{R}$  une fonction strictement convexe définie sur un convexe  $C$ . Soit  $x^* \in C$  est un minimum local de  $f$  sur  $C$ . Alors  $x^*$  est un minimum global strict de  $f$  sur  $C$ .

**Notation.** On note par  $\operatorname{argmin}_{x \in C} f(x)$  (des fois  $\mathcal{S}$ ) l'ensemble des minimiseurs de  $f$ , i.e., les solutions de (1.3).

**Théorème 11.** Soit  $f : C \rightarrow \mathbb{R}$  une fonction convexe, alors  $\mathcal{S}$  est convexe. Si de  $f$  est strictement convexe, alors  $\mathcal{S}$  contient au plus un élément.

**Preuve.** Si  $\mathcal{S} = \emptyset$ , rien à démontrer. Sinon, soient  $x, y \in \mathcal{S}$  et  $\lambda \in [0, 1]$ . Alors

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) = \lambda f^* + (1-\lambda)f^* = f^*.$$

Ce qui prouve la convexité. Maintenant, si  $f$  est strictement convexe et supposons qu'il existe  $x, y \in \mathcal{S}$  avec  $x \neq y$ . Par convexité de  $C$  on a  $\frac{1}{2}x + \frac{1}{2}y \in C$  et par convexité de  $f$ :

$$f\left(\frac{1}{2}x + \frac{1}{2}y\right) < \frac{1}{2}(x) + \frac{1}{2}f(y) = f^*,$$

cela contredit le fait que  $f^*$  est la valeur optimale. □

**Remarque 2.** Pour la convexité de  $\mathcal{S}$ , on peut remarquer que  $\mathcal{S} = \{x : f(x) \leq f^*\} \cap C = \text{Lev}(f, f^*) \cap C$  qui est convexe comme intersection de convexes.

### 1.2.2 Stationnarité

Nous avons introduit, dans la [Définition 4](#), la notion de **point critique** (ou **stationnaire**) dans le cas des problèmes **sans contraintes**, c'est-à-dire les zéros du gradient. Nous avons vu que cette condition constituait une **condition nécessaire d'optimalité**.

Nous allons maintenant étendre cette notion aux problèmes **avec contraintes** du type

$$\min_{x \in C} f(x), \quad (1.4)$$

où  $f \in C^1(C)$  et  $C \subset \mathbb{R}^n$  est un ensemble convexe fermé.

**Définition 8 (Point stationnaire pour (1.4)).** Soit  $f \in C^1(C)$  avec  $C \subset \mathbb{R}^n$  convexe fermé. On dit qu'un point  $x^* \in C$  est un **point stationnaire** du problème (1.4) si

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

**Remarque 3.** Lorsque  $C = \mathbb{R}^n$ , la condition de stationnarité s'écrit simplement

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \mathbb{R}^n.$$

En choisissant en particulier  $x = x^* - \nabla f(x^*)$ , on obtient

$$-\|\nabla f(x^*)\|^2 \geq 0,$$

ce qui impose  $\nabla f(x^*) = 0$ . Autrement dit, on retrouve la condition de stationnarité classique du cas non contraint, introduite en [Définition 4](#).

La notion de stationnarité introduite en [Définition 8](#) constitue en fait une **condition nécessaire d'optimalité** pour le problème (1.4).

**Théorème 12 (Condition de stationnarité).** Soit  $f \in C^1(C)$  avec  $C \subset \mathbb{R}^n$  convexe fermé, et soit  $x^* \in C$  un minimum local du problème (1.4). Alors  $x^*$  est un point stationnaire de (1.4).

De plus, si  $f$  est **convexe**, alors  $x^*$  est un point stationnaire de (1.4) si et seulement si  $x^*$  est une solution optimale du problème.

**Preuve.** Exercice. □

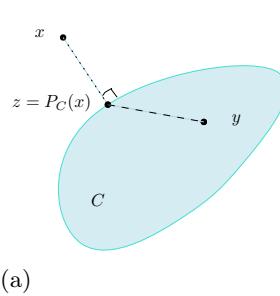


Figure 1.3: Projection orthogonale.

### 1.3 Projection orthogonale

Soit  $C \subset \mathbb{R}^n$  un convexe fermé non vide.

**Définition 9.** L'opérateur de projection orthogonale sur  $C$  est l'application  $P_C : \mathbb{R}^n \rightarrow C$  qui à  $x \in \mathbb{R}^n$  associe

$$P_C(x) = \operatorname{argmin}_{y \in C} \|y - x\|^2 \quad (1.5)$$

On a le résultat

**Théorème 13 (Premier théorème de la projection).** Soit  $C$  un ensemble convexe, fermé et non vide. Alors le problème (1.5) admet une unique solution.

**Preuve.** La fonction  $y \mapsto \|y - x\|^2$  est coercive, et comme  $C$  est fermé et non vide, le théorème Théorème 7 garantit l'existence d'au moins une solution. De plus, la fonction  $y \mapsto \|y - x\|^2$  est strictement convexe, ce qui assure l'unicité du minimiseur d'après Théorème 11.  $\square$

Grâce au Théorème 12, on obtient la caractérisation géométrique suivante de la projection.

**Théorème 14 (Deuxième théorème de la projection).** Soit  $C \subset \mathbb{R}^n$  un ensemble convexe, fermé et non vide, et soit  $x \in \mathbb{R}^n$ . Alors  $z = P_C(x)$  si et seulement si

$$\langle x - z, y - z \rangle \leq 0, \quad \forall y \in C.$$

**Preuve.** Soit  $x \in \mathbb{R}^n$ . On sait que  $P_C(x)$  est la solution du problème convexe

$$\min_{y \in C} \phi(y) := \|y - x\|^2.$$

Comme  $\nabla \phi(P_C(x)) = 2(P_C(x) - x)$ , il résulte du Théorème 12 que  $z = P_C(x)$  si et seulement si

$$\langle \nabla \phi(P_C(x)), y - P_C(x) \rangle \geq 0, \quad \forall y \in C,$$

ce qui revient à écrire

$$\langle x - z, y - z \rangle \leq 0, \quad \forall y \in C.$$

$\square$

Le Théorème 14 affirme que, pour tout  $x \in \mathbb{R}^n$ , l'angle entre  $x - P_C(x)$  et  $y - P_C(x)$ , pour tout  $y \in C$ , est supérieur ou égal à  $\pi/2$ . Cette propriété est illustrée sur la Fig. 1.3.

On termine ce chapitre avec le résultat suivant, qui rassemble quelques propriétés fondamentales du projecteur orthogonal sur un convexe fermé.

**Théorème 15.** Soit  $C \subset \mathbb{R}^n$  un ensemble convexe fermé. Alors le projecteur orthogonal  $P_C$  est 1-cocoercif, c'est-à-dire que pour tout  $x, y \in \mathbb{R}^n$ ,

$$\langle P_C(x) - P_C(y), x - y \rangle \geq \|P_C(x) - P_C(y)\|^2. \quad (1.6)$$

En particulier,  $P_C$  est une application 1-lipschitzienne :

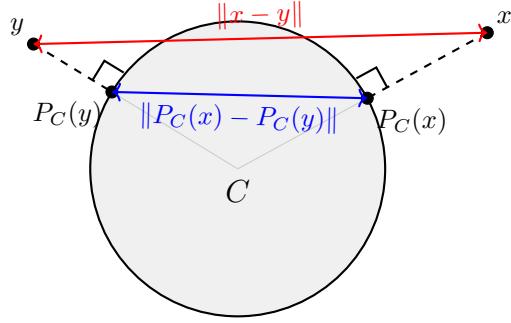
$$\|P_C(x) - P_C(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

**Preuve.** Par le Théorème 14, on a

$$\langle x - P_C(x), y - P_C(x) \rangle \leq 0, \quad \forall y \in C.$$

En particulier, en prenant  $y = P_C(y)$ , on obtient

$$\langle x - P_C(x), P_C(y) - P_C(x) \rangle \leq 0.$$

Figure 1.4: L'opérateur  $P_C$  est non-expansive.

De même,

$$\langle y - P_C(y), P_C(x) - P_C(y) \rangle \leq 0.$$

En sommant les deux inégalités, on obtient

$$\langle P_C(y) - P_C(x), x - y + P_C(y) - P_C(x) \rangle \leq 0.$$

En réarrangeant, cela donne

$$\langle P_C(x) - P_C(y), x - y \rangle \geq \|P_C(x) - P_C(y)\|^2,$$

ce qui prouve la 1-cocoercivité. Pour établir la propriété de 1-Lipschitz continuité, le résultat est évident si  $P_C(x) = P_C(y)$ . Sinon, l'inégalité précédente et l'inégalité de Cauchy–Schwarz donnent

$$\|P_C(x) - P_C(y)\| \|x - y\| \geq \langle P_C(x) - P_C(y), x - y \rangle \geq \|P_C(x) - P_C(y)\|^2.$$

En divisant des deux côtés par  $\|P_C(x) - P_C(y)\| > 0$ , on obtient

$$\|P_C(x) - P_C(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

□