Bases d'analyse convexe et optimisation

Hamza Ennaji
Univ. Grenoble Alpes, CNRS, Grenoble INP*, LJK
hamza.ennaji@univ-grenoble-alpes.fr

December 24, 2025

Contents

Ces notes constituent un support de cours pour l'UE *Bases d'analyse convexe et optimisation* dispensée en 1$^{\text{re}}$ année à l'ENSIMAG. L'équipe pédagogique est composée de : Hamza Ennaji, Sylvain Meignen, Chloé Gergely, Anderson Augusma et Maxime Louis.

Ce polycopié s'appuie principalement sur les références bibliographiques suivantes (détails en fin de document) :

- [1] et [2] pour les fondements géométriques et l'analyse convexe;
- [3, 4] et [5] pour l'optimisation convexe;
- [4] pour les algorithmes et les aspects numériques;
- [6] pour l'analyse variationnelle.

Les notes sont régulièrement mises à jour. N'hésitez pas à me signaler toute coquille, erreur ou suggestion d'amélioration.

Chapitre 1.

Rappels et préliminaires

Tout au long de ce cours, on travaillera sur $\mathbb{E} = \mathbb{R}^n$ avec $n \in \mathbb{N}$.

1.1 Normes et distances

Définition 1 (Norme). Une norme sur $\mathbb{R}^n$ est une application $N : \mathbb{R}^n \to \mathbb{R}^+$ vérifiant

1. $N(x) = 0$ ssi $x = 0$,
2. $N(\lambda x) = |\lambda| N(x)$ pour tout $\lambda \in \mathbb{R}$ et $x \in \mathbb{R}^n$,
3. $N(x + y) \leq N(x) + N(y)$ pour tout $x, y \in \mathbb{R}^n$.

Exemple 1. On définit la norme euclidienne sur $\mathbb{R}^n$ (notée $\|\cdot\|$ ou $\|\cdot\|_2$) par

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}.$$

Plus généralement, pour $1 \leq p < \infty$, on définit la norme l_p par

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

Pour $p = \infty$:

$$\|x\|_\infty = \max_{i=1,\cdots,n} |x_i|.$$

Définition 2 (Métrique). Soit N une norme sur $\mathbb{R}^n$. On note par $d_N : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ la distance induite par N, qui est définie par

$$d_N(x, y) = N(x - y), \ \forall x, y \in \mathbb{R}^n,$$

et vérifie

1. $d_N(x, y) = 0$ ssi $x = y$,
2. $d_N(x, y) = d_N(y, x), \ \forall x, y \in \mathbb{R}^n$,
3. $d_N(x, y) \leq d_N(x, z) + d_N(z, y), \forall x, y, z \in \mathbb{R}^n$.

Définition 3 (Produit scalaire). Soient $x, y \in \mathbb{R}^n$. On note par $\langle x, y \rangle$ le produit scalaire usuel de x et y, défini par

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

On dira que x et y sont orthogonaux si $\langle x, y \rangle = 0$.

Remarque 1. On notera que $\|x\|_2 = \langle x, x\rangle^{1/2}$.

Définition 4 (Boules)**.** Soit N une norme sur $\mathbb{R}^n$ et $r \geq 0$. On note $B_N(x, r)$ la boule fermée de centre x et de rayon r pour la norme N.

Définition 5 (Ensemble borné)**.** On dira que $\Omega \subset \mathbb{R}^n$ est borné s'il existe $R > 0$ tel que $\Omega \subset B_N(0, R)$.

Remarque 2. Dans toute la suite, sauf mention contraire, les boules sont définies pour la norme euclidienne.

Définition 6 (Somme de Minkowski)**.** Soit $A, B \subset \mathbb{R}^n$ et $\lambda \in \mathbb{R}$ on définit

$$\lambda A := \{\lambda a : \ a \in A\}.$$

La somme de Minkowski de A et B est définie par

$$A + B := \{a + b : \ a \in A, b \in B\}.$$

1.2 Topologie

Définition 7 (Ensemble ouvert, Ensemble fermé)**.** Soient N une norme sur $\mathbb{R}^n$. Un ensemble $U \subset \mathbb{R}^n$ est un ouvert par rapport à la norme N si pour tout $x \in U$ il existe $r > 0$ tel que $B_N(x, r) \subset U$.
Un ensemble $F \subset \mathbb{R}^n$ est fermé si son complémentaire $F^c := \mathbb{R}^n \setminus F$ est un ouvert.

Théorème 1.
1. Les ensembles $\emptyset$ et $\mathbb{R}^n$ sont à la fois ouverts et fermés.
2. Toute union d'ouverts est toujours un ouvert. Toute intersection finie d'ouverts est un ouvert.
3. Toute union finie de fermés est un fermé. Toute intersection de fermés est un fermé.

Proposition 1. Un ensemble $F \subset \mathbb{R}^n$ est fermé si et seulement si $\forall (x_n)_n \in F^{\mathbb{N}}$

$$\left(\lim_{n\to\infty} x_n = x\right) \Rightarrow x \in F.$$

Définition 8. Soit $\Omega \subset \mathbb{R}^n$.

1. On dit que $x \in \Omega$ est un point intérieur à Ω s'il existe $\varepsilon > 0$ tel que $B_N(x, \varepsilon) \subset \Omega$. L'ensemble des points intérieurs à Ω est noté $\mathrm{int}(\Omega)$ ou $\overset{\circ}{\Omega}$.
2. On dira que Ω est ouvert si $\mathrm{int}(\Omega) = \Omega$.
3. L'adhérence de Ω, noté $\mathrm{cl}(\Omega)$ ou $\overline{\Omega}$ est l'ensemble $\mathbb{R}^n \setminus \mathrm{int}(\mathbb{R}^n \setminus \Omega)$.
4. Le bord de Ω est $\partial\Omega = \overline{\Omega} \setminus \mathrm{int}(\Omega)$.

Définition 9 (Ensemble compact)**.** On dira que $\Omega \subset \mathbb{R}^n$ est compact s'il est fermé borné dans $\mathbb{R}^n$.

Théorème 2. Soit $\Omega \subset \mathbb{R}^n$ un compact. Alors de toute suite $(x_n)_n \in \Omega^{\mathbb{N}}$ on peut extraire une sous-suite convergente.

Théorème 3 (Weierstrass)**.** Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ une fonction continue avec Ω un compact. Alors f est bornée et atteint ses bornes, i.e., $\exists x_*, x^* \in \Omega$ tels que

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

1.3 Calcul différentiel

1.3.1 Fonctions dérivables, fonctions différentiables

Définition 10. Une fonction $f : \mathbb{R} \to \mathbb{R}$ est dite dérivable en un point $a \in \mathbb{R}$ s'il existe $l_a \in \mathbb{R}$ tel que

$$\frac{f(a+h) - f(a)}{h} \to l_a, \text{ lorsque } h \to 0.$$

Cela revient à dire que

$$f(a+h) = f(a) + hl_a + \varepsilon_a(h)|h|,$$

avec $\varepsilon_a(h) \to 0$ quand $h \to 0$.

Définition 11. Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ avec U un ouvert. On dit que f est différentiable en $a \in U$ s'il existe une application linéaire $L_a : \mathbb{R}^n \to \mathbb{R}^p$ telle que, pour tout $h \in \mathbb{R}^n$ tel que $a+h \in U$

$$f(a+h) = f(a) + L_a(h) + \varepsilon_a(h)\|h\|_\infty.$$

De façon équivalente

$$\forall\varepsilon > 0, \exists\delta_\varepsilon > 0 : \ \|h\|_\infty \leq \delta_\varepsilon \Rightarrow \|f(a+h) - f(a) - L_a(h)\| \leq \varepsilon\|h\|_\infty$$

Remarque 3. Rappelons les résultats suivants:

1. Une fonction différentiable en un point $a \in U$ est continue en a.
2. La différentiabilité d'une fonction vectorielle en un point équivaut à celle de toutes ses fonctions composantes.
3. Quand $n = p = 1$, les notions de différentiabilité et dérivabilité coincident.

1.3.2 Dérivées partielles

Définition 12. Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ avec U un ouvert et fixons $k \in \{1, \cdots, n\}$. On dit que f admet une dérivée partielle par rapport à sa $k$ème variable en $a \in U$ si

$$\frac{f(a+te_k) - f(a)}{t}$$

admet une limite quand $t \to 0$. On note cette limite $\frac{\partial f}{\partial x_k}(a)$. Ici, $\{e_1, \cdots, e_n\}$ est la base canonique de $\mathbb{R}^n$.

1.3.3 Jacobienne et Gradient

Définition 13. Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ avec U un ouvert. Supposons que f admet en $a \in U$ des dérivées partielles par rapport à toutes ses variables. La matrice Jacobienne de f en a s'écrit

$$J_f(a) = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)\right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & & & \vdots \\ \frac{\partial f_p}{\partial x_1}(a) & \frac{\partial f_p}{\partial x_2}(a) & \cdots & \frac{\partial f_p}{\partial x_n}(a) \end{pmatrix}.$$

Si $p = 1$, on définit le gradient de f en a par

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)\right)^T.$$

1.3.4 Fonctions de classe C^1

Définition 14. Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ avec U un ouvert. On dit que f est de classe C^1 sur U, et on note $f \in C^1(U;\mathbb{R}^p)$ si pour tout $a \in U$, f admet en a des dérivées partielles par rapport à toutes ses variables et toutes les applications $a \mapsto \frac{\partial f}{\partial x_k}(a)$ sont continues sur U pour tout $k = 1, \dots, n$.

On rappelle le résultat fondamental suivant.

Théorème 4. Soit $f \in C^1(U;\mathbb{R}^p)$ avec $U \subset \mathbb{R}^n$ un ouvert. Soit $a \in U$, $r > 0$ tel que $B(a,r) \subset U$. Pour tout $h \in B(a,r)$

$$f(a+h) = f(a) + J_f(a)h + \varepsilon_a(h)\|h\|_\infty,$$

avec $\varepsilon_a : B(a,r) \to \mathbb{R}^p$ telle que $\varepsilon_a(h) \to 0$ quand $h \to 0$.

Remarque 4. Rappelons que $f \in C^1$ implique que f est différentiable. La différentiabilité de f implique la continuité de f et l'existence des dérivées partielles.

Remarque 5. Rappelons que pour tout $a \in U$ et $h \in \mathbb{R}^n$

$$J_f(a)h = \langle \nabla f(a), h \rangle,$$

i.e., $\nabla f(a) = J_f^T(a)$.

1.3.5 Composition

Théorème 5. Soit $U \subset \mathbb{R}^n, V \subset \mathbb{R}^m$ deux ouverts et $f : U \to V, g : V \to \mathbb{R}^p$ deux fonctions de classe C^1. Alors $g \circ f \in C^1(U;\mathbb{R}^p)$ et on a, pour tout $a \in U$ et $j \in \{1, \dots, n\}$

$$\frac{\partial}{\partial x_j}(g \circ f)(a) = \sum_{k=1}^{m} \frac{\partial g}{\partial y_k}(f(a)) \frac{\partial f_k}{\partial x_j}(a).$$

Cela s'exprime aussi de manière plus compacte comme

$$\forall a \in U,\ J_{g\circ f}(a) = J_g(f(a))J_f(a).$$

1.3.6 Dérivées partielles d'ordre deux:

Définition 15. Soit $f \in C^1(U;\mathbb{R})$ avec U un ouvert de $\mathbb{R}^n$. On dit que f admet des dérivées partielles d'ordre deux en $a \in U$ si $\partial_{x_1} f, \dots, \partial_{x_n} f$ admettent des dérivées partielles en a. On a

$$\frac{\partial^2}{\partial x_i \partial x_j} f(a) := \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f\right)(a).$$

Supposons que f est de classe C^2. On appelle matrice Hessienne de f en a, notée $H_f(a)$ ou $\nabla^2 f(a)$ la matrice des dérivées partielles de f en a. On a

$$\nabla^2 f(a) := \left(\frac{\partial^2}{\partial x_i \partial x_j} f(a)\right)_{i,j=1,\dots,n}.$$

Définition 16. On dit que $f : U \to \mathbb{R}$ est de classe C^2 si f est C^1 admettant des dérivées partielles d'ordre deux et pour tout $i, j = 1, \dots, n$, l'application

$$a \in U \mapsto \frac{\partial^2}{\partial x_i \partial x_j} f(a),$$

est continue.

On rappelle le résultat fondamental suivant.

Théorème 6. Soit $f \in C^2(U;\mathbb{R})$. Alors pour tout $i, j = 1, \dots, n$ et $a \in U$

$$\frac{\partial^2}{\partial x_i \partial x_j} f(a) = \frac{\partial^2}{\partial x_j \partial x_i} f(a).$$

Corollaire 1. Soit $f \in C^2(U;\mathbb{R})$. Alors pour tout $a \in U$, $\nabla^2 f(a)$ est symétrique, i.e.,

$$\nabla^2 f(a) = \left(\nabla^2 f(a)\right)^T.$$

1.4 Structure affine de $\mathbb{R}^n$

1.4.1 Plans

Définition 17 (Plan). On dit que $A \subset \mathbb{R}^n$ est un plan s'il s'écrit de la forme

$$A = a + V = \{a + v :\ v \in V\},$$

avec $a \in \mathbb{R}^n$ et V un sous-espace de $\mathbb{R}^n$.

Remarque 6. La Définition 17 nous dit qu'une partie A de $\mathbb{R}^n$ est un plan si elle est une translation d'un sous-espace vectoriel. Évidemment $\mathbb{R}^n$ est un plan. Par convention, $\emptyset$ est un plan. Dans la littérature, on rencontre souvent d'autres terminologies comme, ensemble affine, espace affine, variété affine etc.

Si $A \subset \mathbb{R}^n$ est un plan non-vide, et $a \in \mathbb{R}^n$. On considère

$$V = \{x - y : \; x, y \in A\}.$$

Théorème 7. Tout plan non-vide $A \subset \mathbb{R}^n$ est la translation d'un unique sous-espace vectoriel $V \subset \mathbb{R}^n$ donné par

$$V = \{x - y : \; x, y \in A\} := A - A.$$

De plus, pour $a \in \mathbb{R}^n$, $A = a + V$ si et seulement si $a \in A$. En particulier, A est un sous-espace si et seulement si $0 \in A$.

Notation. On notera par o, 0_n, 0_{R^n} (ou tout simplement 0 s'il n'y a pas de confusion) le vecteur nul $\mathbb{R}^n$.

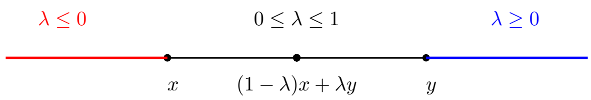
Définition 18 (Espace vectoriel associé et dimension)**.** Soit $A \subset \mathbb{R}^n$ un plan non-vide. Le sous-espace $V := A - A$ est appelé le sous-espace vectoriel associé à A (ou sous-espace caractéristique de A) et est noté $\mathrm{lin}(A)$. La dimension de A, notée $\dim(A)$ est définie comme la dimension (au sens des espaces vectoriels) de son espace associé V.

Remarque 7. Par convention $\dim(A) = -1$ si $A = \emptyset$. On appellera points (resp. droites) les plans de dimension 0 (resp. 1).

1.4.2 Droites et segments

Définition 19 (Droite)**.** Soient $x, y \in \mathbb{R}^n$ distincts. La droite passant par x et y, notée par $D(x, y)$ ou $< x, y >$ est définie par

$$< x, y >= \{(1 - \lambda)x + \lambda y : \; \lambda \in \mathbb{R}\}. \tag{1.1}$$



(a)

Figure 1.1: Droite passant par x et y.

D'après Théorème 7 et Remarque 7, une droite de $\mathbb{R}^n$ est la translation d'un sous-espace de dimension 1. En vue de (1.1), on voit que

$$< x, y >= x + \{\lambda(y - x) : \; \lambda \in \mathbb{R}\},$$

i.e., <x,y> est la translation du sous-espace engendré par $y - x$.

Définition 20 (Segments)**.** Soient $x, y \in \mathbb{R}^n$ distincts. Le segment fermé $[x, y]$, ouvert (x, y) et les segments semi-ouverts $(x, y]$ et $[x, y)$ sont définis par

$$\begin{aligned}
[x, y] &= \{(1 - \lambda)x + \lambda y : 0 \leq \lambda \leq 1\},\\
[x, y) &= \{(1 - \lambda)x + \lambda y : 0 \leq \lambda < 1\},\\
(x, y] &= \{(1 - \lambda)x + \lambda y : 0 < \lambda \leq 1\},\\
(x, y) &= \{(1 - \lambda)x + \lambda y : 0 < \lambda < 1\}.
\end{aligned}$$

1.4.3 Combinaisons affines

Définition 21. Une combinaison affine des points $x_1, \dots, x_m \in \mathbb{R}^n$ est une combinaison linéaire de la forme $\sum_{i=1}^m \lambda_i x_i$ avec $\sum_{i=1}^m \lambda_i = 1$.

Exemple 2. L'ensemble des combinaisons affines de deux points distincts x et y est la droite passant par x et y. En effet, l'application $\lambda \mapsto (1-\lambda)x + \lambda y$ est une bijection de $\mathbb{R}$ vers $< x, y >$.

Théorème 8. Une partie non vide A de $\mathbb{R}^n$ est un plan si et seulement si elle contient tous les points de la forme $(1-\lambda)x + \lambda y$ pour tout $x, y \in A$ et $\lambda \in \mathbb{R}$.

Preuve. Laissée en exercice. □

Plus généralement, on a

Théorème 9. Une partie non vide A de $\mathbb{R}^n$ est un plan si et seulement si elle contient toutes les combinaisons affines de points de A.

1.4.4 Espace affine engendré ou Enveloppe affine

Définition 22. Soit $A \subset \mathbb{R}^n$. On appelle sous-espace affine engendré par A, note $\text{aff}(A)$, l'intersection de tous les plans contenant A.

Le résultat suivant décrit $\text{aff}(A)$ en terme des combinaisons affines de points de A.

Théorème 10. Soit A une partie non-vide de $\mathbb{R}^n$.

1. L'enveloppe affine de A est l'ensemble de toutes les combinaisons affines de point de A, i.e.,
$$\text{aff}(A) = \{\sum_{i=1}^m \lambda_i x_i : \ m \geq 1, x_1, \dots, x_m \in A, \ \text{et} \ \sum_{i=1}^m \lambda_i = 1\}.$$
2. $\text{aff}(A)$ est le plus petit sous-espace affine contenant A.
3. Pour $x_1, \dots, x_m \in \mathbb{R}^n$:
$$\text{aff}(\{x_1, \dots, x_m\}) = \{\sum_{i=1}^m \lambda_i x_i : \ \sum_{i=1}^m \lambda_i = 1\}$$

1.4.5 Indépendance affine

Définition 23. On dit que des points $x_1, \dots, x_m \in \mathbb{R}^n$ avec $m \geq 2$, sont affinement dépendants s'il existe des scalaires $\lambda_1, \dots, \lambda_m$ non tous nuls tel que
$$\sum_{i=1}^m \lambda_i x_i = 0 \text{ et } \sum_{i=1}^m \lambda_i = 0.$$
On dira que la famille $\{x_1, \dots, x_m\}$ est affinement indépendante si elle n'est pas affinement dépendante.

1.5 Vecteurs et valeurs propres

Définition 24. Soit $A \in \mathbb{R}^{n\times n}$ et $0 \neq v \in \mathbb{R}^n$. On dit que v est un vecteur propre de A s'il existe $\lambda \in \mathbb{C}$ tel que $Av = \lambda v$. On appelle λ la valeur propre associée à v. L'ensemble des valeurs propres de A est appelé spectre de A et est souvent noté par $\sigma(A)$.

Soit $A \in \mathbb{R}^{n\times n}$, en général $\sigma(A) \subset \mathbb{C}$. Si A est symétrique, alors ses valeurs propres sont réelles. On note alors

$$\lambda_{\max}(A) := \lambda_1(A) \geq \ldots \geq \lambda_{\min}(A) := \lambda_n(A),$$

les valeurs propres de A où $\lambda_{\max}$ (respectivement $\lambda_{\min}$) est la plus grande (respectivement la plus petite) valeur propre de A.

Théorème 11 (Décomposition spectrale). Soit $A \in \mathbb{S}^n$. Il existe alors $U \in O(n)$ (i.e., $U^tU = UU^t = \mathrm{Id}$) et une matrice diagonale $D = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ tels que $U^tAU = D$.

Remarque 8. Les colonnes de U représentent les vecteurs propres de A tandis que les éléments diagonaux de D sont les valeurs propres associées.

Corollaire 2. Soit $A \in \mathbb{S}_n$, alors $\mathrm{trace}(A) = \sum_{i=1}^n \lambda_i(A)$ et $\det(A) = \prod_{i=1}^n \lambda_i(A)$.

1.6 Classification des matrices

On rappelle ici quelques résultats concernant la classification de matrices qui seront utiles pour les conditions d'optimalité dans la Section 4.1.

Définition 25. Soit $A \in \mathbb{S}^n$.

1. A est dite semi-définie positive, et on note $A \succeq 0$, si $\langle Ax, x\rangle \geq 0$ pour tout $x \in \mathbb{R}^n$.
2. A est dite définie positive, et on note $A \succ 0$, si $\langle Ax, x\rangle > 0$ pour tout $x \in \mathbb{R}^n \setminus \{0\}$.
3. A est dite semi-définie négative, et on note $A \preceq 0$, si $\langle Ax, x\rangle \leq 0$ pour tout $x \in \mathbb{R}^n$.
4. A est dite définie négative, et on note $A \prec 0$, si $\langle Ax, x\rangle < 0$ pour tout $x \in \mathbb{R}^n \setminus \{0\}$.
5. A est dite non-définie s'il existe $x, y \in \mathbb{R}^n$ tels que

$$\langle Ax, x\rangle < 0 \text{ et } \langle Ay, y\rangle > 0.$$

Proposition 2.
1. Soit $A \in \mathbb{S}^n_+$ (respectivement dans $\mathbb{S}^n_{++}$). Alors les éléments diagonaux de A sont positifs (respectivement strictement positifs).
2. Soit $A \in \mathbb{S}^n_-$ (respectivement dans $\mathbb{S}^n_{--}$). Alors les éléments diagonaux de A sont négatifs (respectivement strictement négatifs).

Preuve. Exercice. □

Pour les matrices non-définies, on a le résultat suivant.

Proposition 3. Soit $A \in \mathbb{S}^n$. Si A possède des éléments diagonaux strictement négatifs et positifs, alors A est non-définie.

Preuve. Exercice. □

Le résultat suivant permet de classifier les matrices à partir de l'information spectrale.

Théorème 12. Soit $A \in \mathbb{S}^n$. Alors

1. $A \in \mathbb{S}^n_{++}$ ssi $\sigma(A) \subset \mathbb{R}^*_+$.
2. $A \in \mathbb{S}^n_+$ ssi $\sigma(A) \subset \mathbb{R}_+$.
3. $A \in \mathbb{S}^n_{--}$ ssi $\sigma(A) \subset \mathbb{R}^*_-$.
4. $A \in \mathbb{S}^n_-$ ssi $\sigma(A) \subset \mathbb{R}_-$.
5. $A \in \mathbb{S}^n_\pm$ ssi il existe $\lambda, \mu \in \sigma(A)$ tels que $\lambda > 0$ et $\mu < 0$.

Preuve. Laissée en exercice (utiliser Théorème 11). □

Remarque 9. Dans toute la suite, on s'intéressera principalement aux matrices symétriques réelles.

1.7 Limites supérieures et inférieures de suites

Étant donnée une suite $(u_n)_n$ de réels, la quantité $\lim\limits_{n\to\infty} u_n$ n'a aucune raison d'exister. L'exemple typique est celui de $u_n = (-1)^n$. Quand la suite $(u_n)_n$ est bornée, le théorème de Bolzano-Weierstrass assure l'existence d'une sous-suite convergente $(u_{n_k})_k$, i.e., $\lim\limits_{k\to\infty} u_{n_k} = l \in \mathbb{R}$.

Définition 26. Soit $(u_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\overline{\mathbb{R}}$. Les suites $(\inf_{k\geq n} u_k)_{n\in\mathbb{N}}$ et $\left(\sup_{k\geq n} u_k\right)_{n\in\mathbb{N}}$ sont monotones (la première est croissante, la seconde décroissante). On définit alors

$$\liminf u_n = \lim_{n\to+\infty} \left(\inf_{k\geq n} u_k\right) = \sup_{n\in\mathbb{N}} \left(\inf_{k\geq n} u_k\right),$$
$$\limsup u_n = \lim_{n\to+\infty} \left(\sup_{k\geq n} u_k\right) = \inf_{n\in\mathbb{N}} \left(\sup_{k\geq n} u_k\right).$$

Après cette définition formelle, la proposition suivante donne une caractérisation plus intuitive de ces deux quantités $\limsup u_n$ et $\liminf u_n$.

Proposition 4. Soit $(u_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\overline{\mathbb{R}}$. Alors $\liminf u_n$ est la plus petite valeur d'adhérence de $(u_n)_n$ et $\limsup u_n$ est la plus grande valeur d'adhérence de $(u_n)_n$. De plus

$$\liminf u_n \leq \limsup u_n,$$

avec égalité si et seulement si $(u_n)_n$ converge vers $l = \liminf u_n = \limsup u_n$.

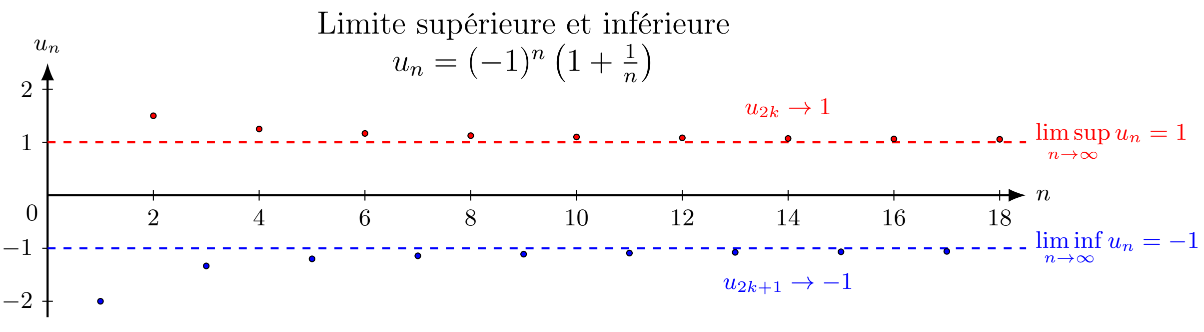


Figure 1.2: Illustration des limites supérieures et inférieures de la suite $u_n = (-1)^n \left(1 + \frac{1}{n}\right)$

Exemple 3. On considère les exemples suivants

1. Si $x_n = 1 + \frac{1}{n}$, alors $\liminf x_n = \limsup x_n = 1$.
2. Si $x_n = (-1)^n$, alors
$$\liminf x_n = -1, \qquad \limsup x_n = 1.$$
3. Si $x_n = (-1)^n + \frac{1}{n}$, alors
$$\liminf x_n = -1, \qquad \limsup x_n = 1.$$
4. Si $x_n = (-1)^n n$, alors
$$\liminf x_n = -\infty, \qquad \limsup x_n = +\infty.$$
5. Si $x_n = n$ alors $\liminf x_n = \limsup x_n = +\infty$.
6. Si
$$x_n = \begin{cases} 1/k & \text{si } n = 2k, \\ 2 - 1/k & \text{si } n = 2k+1, \end{cases}$$
alors
$$\liminf x_n = 0, \qquad \limsup x_n = 2.$$

1.8 Exercices

Exercice 1. Calculer le gradient des fonctions suivantes:

- $f : x \in \mathbb{R}^n \mapsto u^T x$.
- $f : x \in \mathbb{R}^n \mapsto Ax + b \in \mathbb{R}^m$ avec $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.
- $f : x \in \mathbb{R}^n \mapsto \frac{1}{2}(x^T Ax) + b^T x + c$, avec $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Même question pour $A \in \mathcal{S}^n$.
- $f : \mathbb{R}^n \mapsto \|Ax - b\|_2^2$.
- $f(x) = \|x\|_2$.
- $f(x) = \log\left(\sum_{i=1}^m \exp(a_i^T x + b_i)\right)$, avec $a_1, \cdots, a_m \in \mathbb{R}^n, b_1, \cdots, b_m \in \mathbb{R}$.
- $f(X) = \log\det(X)$ avec $X \in \mathcal{S}_{++}^n$.

Exercice 2. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction de classe C^1 et $g : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$g(x, y) = f(x + y, x^2 + y^2).$$

Exprimer les dérivées partielles de g en $(1, 2)$ en fonction de celles de f.

Exercice 3.

1. Soit $A \in \mathcal{S}_{++}^n$. Peut-on dire que les composantes de A sont positives ?
2. Soit $A \in \mathcal{S}_+^n$ (respectivement dans $\mathcal{S}_{++}^n$). Montrer que les éléments diagonaux de A sont positifs (respectivement strictement positifs).
3. Soit $A \in \mathcal{S}_-^n$ (respectivement dans $\mathcal{S}_{--}^n$). Montrer que les éléments diagonaux de A sont négatifs (respectivement strictement négatifs). Soit $A \in \mathcal{S}^n$. Montrer que si A possède des éléments diagonaux strictement négatifs et positifs, alors A est non-définie.

Exercice 4.

1. Soient $A, B \in \mathcal{S}_+^n$. Montrer que $A + B \in \mathcal{S}_+^n$.
2. Soit $A \in \mathbb{R}^{n \times k}$ et posons $B = AA^T$. Montrer que $B \in \mathcal{S}_+^n$. Montrer que $B \in \mathcal{S}_{++}^n$ si et seulement si $\mathrm{rg}(A) = n$.
3. Soient $A \in \mathcal{S}^n$ et $B \in \mathcal{S}^m$. Montrer que $A \in \mathcal{S}_+^n$ et $B \in \mathcal{S}_+^m$ si et seulement si $\begin{pmatrix} A & 0_{n \times m} \\ 0_{m \times n} & B \end{pmatrix} \in \mathcal{S}_+^{n+m}$

Exercice 5. Soit $M \in \mathcal{S}_{++}^n$. Montrer que $\|x\|_M = \sqrt{x^T M x}$ est une norme.

Chapitre 2.

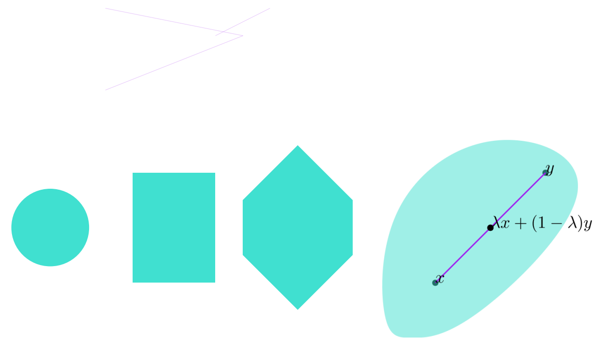
Ensembles convexes

2.1 Définitions

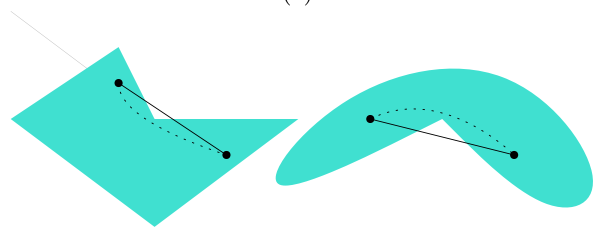
2.1.1 Ensembles convexes

Définition 27. Soit $C \subset \mathbb{R}^n$. On dit que C est convexe si pour tout $x, y \in C$ et tout $\lambda \in [0,1]$, $\lambda x + (1-\lambda)y \in C$.

Remarque 10. Cette définition est équivalente à dire que le segment $[x, y]$ est inclus dans C (i.e., $[x, y] \subset C$). La Fig. 2.1 illustre quelques exemples d'ensembles convexes et non convexes.



(a)



(b)

Figure 2.1: La Figure-2.1a: des ensembles convexes de $\mathbb{R}^2$. Figure-2.1b: ensembles non convexes.

Exemple 4. Voici quelques exemples d'ensembles convexes :

- Dans $\mathbb{R}$, les seuls exemples d'ensembles convexes sont les intervalles : (a, b), $[a, b]$, $(-\infty, a]$, etc.
- Les boules (ouvertes, fermées) :

$$B(a, r) = \{x \in \mathbb{R}^n : \ \|x - a\| < r\} \quad \text{et} \quad \bar{B}(a, r) = \{x \in \mathbb{R}^n : \ \|x - a\| \leq r\},$$

  avec $a \in \mathbb{R}^n$ et $r > 0$.
- Les hyperplans :

$$H = \{x \in \mathbb{R}^n : \ a^T x = b\}, \quad \text{avec } a \in \mathbb{R}^n, b \in \mathbb{R}.$$

- Les demi-espaces (ou demi-plans), les ellipsoïdes.
- Soit $\mathcal{S}^n = \{A \in \mathbb{R}^{n\times n} : A^T = A\}$ l'ensemble des matrices symétriques. On rappelle que $\mathcal{S}^n$ est un espace vectoriel de dimension $n(n+1)/2$. Les ensembles $\mathcal{S}^n_+ = \{A \in \mathcal{S}^n : A \succeq 0\}$ (matrices semi-définies positives) et $\mathcal{S}^n_{++} = \{A \in \mathcal{S}^n : A \succ 0\}$ (matrices définies positives) sont des ensembles convexes.

2.1.2 Ensembles affines

Définition 28. Soit $C \subset \mathbb{R}^n$. On dit que C est affine si pour tout $x, y \in C$ et tout $\lambda \in \mathbb{R}$, $\lambda x + (1-\lambda)y \in C$.

Évidemment, tout ensemble affine est convexe. La réciproque n'est pas vraie.

Exemple 5.
- L'ensemble des solutions $S = \{x \in \mathbb{R}^n : \; Ax = b\}$ d'un système linéaire est un ensemble affine.
- Soit E un sous-espace vectoriel de $\mathbb{R}^n$ et $a \in \mathbb{R}^n$. L'ensemble
$$a + E := \{a + x : \; x \in E\}$$
est un sous-espace affine (ou ensemble affine).

2.2 Opérations et propriétés topologiques préservant la convexité

Le résultat suivant récapitule quelques opérations préservant la convexité.

Théorème 13.
1. Soit $(C_i)_{i\in I}$ une famille d'ensembles convexes de $\mathbb{R}^n$. Alors $C := \bigcap_{i\in I} C_i$ est convexe.
2. Soient $C_1, \dots, C_m$ des convexes de $\mathbb{R}^n$ et $\lambda_1, \dots, \lambda_m \in \mathbb{R}$. Alors les ensembles
$$\sum_{i=1}^m \lambda_i C_i = \left\{ \sum_{i=1}^m \lambda_i x_i : \; x_i \in C_i, \; i = 1, \dots, m \right\}$$
et
$$\prod_{i=1}^m C_i = \left\{ (x_1, \dots, x_m) : \; x_i \in C_i, \; i = 1, \dots, m \right\}$$
sont convexes.
3. Si $C \subset \mathbb{R}^n$ est convexe et $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ est une application affine définie par $\Phi(x) = Ax + b$ (avec $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$), alors l'image de C
$$\Phi(C) = \left\{ Ax + b : \; x \in C \right\}$$
est convexe. De même, si $D \subset \mathbb{R}^m$ est convexe, alors l'image réciproque
$$\Phi^{-1}(D) = \left\{ x \in \mathbb{R}^n : \; \Phi(x) \in D \right\}$$
est convexe.

Preuve.
1. Soient $x, y \in C$ et $\lambda \in [0,1]$. Alors $x, y \in C_i$ pour tout $i \in I$. Par convexité des C_i, on a $z := \lambda x + (1-\lambda)y \in C_i$ pour tout $i \in I$. Donc $z \in C$.
2. Voir TD.

3. Soient $y_1, y_2 \in \Phi(C)$ et $\lambda \in [0,1]$. Montrons que $y := \lambda y_1 + (1-\lambda)y_2 \in \Phi(C)$. Par définition, il existe $x_1, x_2 \in C$ tels que $y_1 = \Phi(x_1)$ et $y_2 = \Phi(x_2)$ (i.e., $y_1 = Ax_1 + b$ et $y_2 = Ax_2 + b$). Donc $y = A(\lambda x_1 + (1-\lambda)x_2) + b$. Par convexité de C, le point $x_\lambda = \lambda x_1 + (1-\lambda)x_2$ appartient à C, et donc $y = \Phi(x_\lambda) \in \Phi(C)$.

   De même, soient $x_1, x_2 \in \Phi^{-1}(D)$. Alors $\Phi(x_1) \in D$ et $\Phi(x_2) \in D$. Soit $x = \lambda x_1 + (1-\lambda)x_2$ avec $\lambda \in [0,1]$. On a

$$\Phi(x) = A(\lambda x_1 + (1-\lambda)x_2) + b = \lambda(Ax_1 + b) + (1-\lambda)(Ax_2 + b) = \lambda\Phi(x_1) + (1-\lambda)\Phi(x_2).$$

   Comme D est convexe, cette combinaison convexe appartient à D. Ainsi $\Phi(x) \in D$, ce qui signifie que $x \in \Phi^{-1}(D)$.

$\square$

Remarque 11. Dans le Théorème 13-1, l'ensemble d'indices I est quelconque, i.e., il peut être infini.

Exemple 6. Soient $A \in \mathbb{R}^{m\times n}$ et $b \in \mathbb{R}^m$. L'ensemble suivant, appelé polyèdre (ou polytope),

$$P = \{x \in \mathbb{R}^n : \; Ax \le b\},$$

est convexe. En effet, on peut voir que

$$P = \bigcap_{i=1}^{m} H_i, \quad \text{avec} \quad H_i = \{x \in \mathbb{R}^n : \; l_i(A)x \le b_i\},$$

où $l_i(A)$ est la i-ème ligne de A. Les ensembles H_i sont des demi-espaces fermés, donc ils sont convexes (pourquoi ?). Ainsi, P est convexe comme intersection d'ensembles convexes.

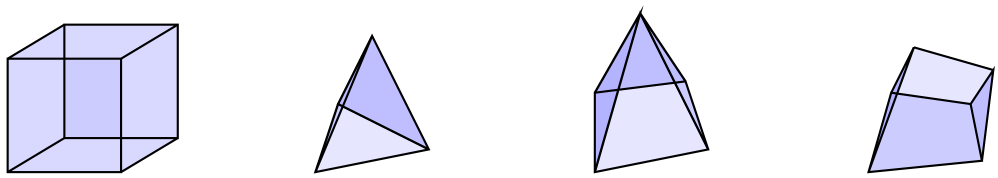


Figure 2.2: Exemples de polytopes.

Théorème 14. Soit $C \subset \mathbb{R}^n$ un ensemble convexe. Alors $\overline{C}$ est convexe.

Preuve. Soient $x, y \in \overline{C}$ et $\lambda \in [0,1]$. Montrons que $\lambda x + (1-\lambda)y \in \overline{C}$. Il existe deux suites $(x_n)_n$ et $(y_n)_n$ d'éléments de C telles que $x_n \to x$ et $y_n \to y$ quand $n \to \infty$. Comme C est convexe, il vient que $\lambda x_n + (1-\lambda)y_n \in C$. En passant à la limite quand $n \to \infty$, on obtient $\lambda x_n + (1-\lambda)y_n \to \lambda x + (1-\lambda)y$, i.e., $\lambda x + (1-\lambda)y \in \overline{C}$. $\square$

Preuve. (bis) Il suffit de remarquer que

$$\overline{C} = \bigcap_{\varepsilon>0} \big(C + B(0,\varepsilon)\big),$$

et de conclure par le Théorème 13-1 (sachant que la somme de deux convexes est convexe). $\square$

Théorème 15. Soit $C \subset \mathbb{R}^n$ un convexe. Alors $\mathrm{int}(C)$ est convexe.

Avant de démontrer ce résultat, on introduit le lemme suivant.

Lemme 1. Soit $C \subset \mathbb{R}^n$ un convexe d'intérieur non vide. Soit $x \in \text{int}(C)$ et $y \in \overline{C}$. Alors $\lambda x + (1-\lambda) \in C$ pour tout $\lambda \in (0,1)$.

Preuve. Comme $x \in \text{int}(C)$, il existe $\varepsilon > 0$ tel que $B(x,\varepsilon) \subset C$. Soit $\lambda \in (0,1)$ et posons $z = (1-\lambda)x + \lambda y$. Montrons que $B(z,\eta) \subset C$ pour un certain $\eta > 0$. Soit W tel que $\|z_1 - z\| < \varepsilon(1-\lambda)$. Comme $y \in \overline{C}$, il existe $u \in C$ tel que $\|u - y\| < (\varepsilon(1-\lambda) - \|w - z\|)\, \lambda^{-1}$. On pose $v = (w - \lambda u)/(1-\lambda)$. Comme $(1-\lambda)x = z - \lambda y$, il vient que

$$\begin{aligned} \|v - x\| &= \frac{1}{1-\lambda}\|(w-z) + \lambda(y-u)\| \\ &\leq \frac{1}{1-\lambda}\Big(\|w-z\| + \varepsilon(1-\lambda) - \|w-z\|\Big) < \varepsilon. \end{aligned} \tag{2.1}$$

On en déduit que $v \in C$ car $B(x,\varepsilon) \subset C$ et par la suite $w = (1-\lambda)v + \lambda u \in C$. Finalement, pour $\eta = (1-\lambda)\varepsilon$, on a $B(z,\eta) \subset C$, ce qui termine le preuve. □

Preuve. du Théorème 15. Si $\text{int}(C) = \emptyset$ rien à démontrer. Sinon, soient $x, y \in \text{int}(C)$ et soit $z = (1-\lambda)x + \lambda y$ avec $\lambda \in (0,1)$. Il existe $\varepsilon > 0$ tel que $B(x,\varepsilon) \subset C$. On a

$$B(z,(1-\lambda)\varepsilon) = \lambda y + (1-\lambda)B(x,\varepsilon) \subset C,$$

et donc $z \in \text{int}(C)$. □

Le résultat suivant se démontrer facilement grâce au Lemme 1.

Proposition 5. Soit $C \subset \mathbb{R}^n$ un convexe d'intérieur non vide. Alors

1. $\overline{\text{int}(C)} = \overline{C}$.
2. $\text{int}(\overline{C}) = \text{int}(C)$.

On termine cette partie par le résultat suivant.

Proposition 6. Soit $X \subset \mathbb{R}^n$ un compact. Alors $\text{Conv}(X)$ est compact.

Remarque 12. En général, le passage par l'enveloppe convexe ne préserve pas le caractère fermé. Cela peut être vu avec l'exemple suivant. Soit

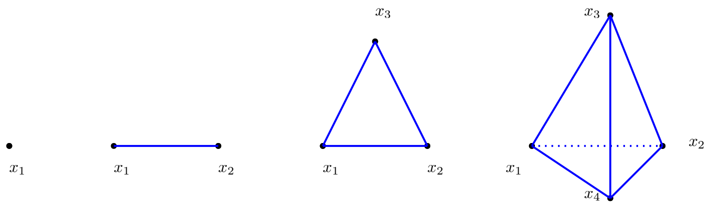
$$X = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2_+ : xy \geq 1 \right\}.$$

On vérifie facilement que X est fermé, tandis que $\text{Conv}(X) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \cup \mathbb{R}^2_{++}$ n'est ni ouvert ni fermé.

2.3 Enveloppe convexe

Définition 29. Soient $x_1, \ldots, x_m \in \mathbb{R}^n$. Une combinaison convexe des x_i est un vecteur de la forme $\sum_{i=1}^m \lambda_i x_i$ avec $\lambda_i \geq 0$ pour tout $i = 1, \ldots, m$ et $\sum_{i=1}^m \lambda_i = 1$. On écrit $\lambda = (\lambda_1, \ldots, \lambda_m) \in \Delta_m$ où

$$\Delta_m = \Big\{ \lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m : \lambda_i \geq 0,\ \forall i = 1, \ldots, m \text{ et } \sum_{i=1}^m \lambda_i = 1 \Big\}.$$



(a)

Figure 2.3: Simplexes dans $\mathbb{R}^3$.

Théorème 16. Soit $C \subset \mathbb{R}^n$ un ensemble convexe et $x_1, \ldots, x_m \in C$. Alors pour tout $\lambda \in \Delta_m$, on a $\sum_{i=1}^m \lambda_i x_i \in C$.

Preuve. On procède par récurrence. Pour $m = 1$, il n'y a rien à démontrer. Supposons que pour tous vecteurs $x_1, \ldots, x_m \in C$ et $\lambda \in \Delta_m$, on a $\sum_{i=1}^m \lambda_i x_i \in C$.

Soient maintenant $x_1, \ldots, x_{m+1} \in C$ et $\lambda \in \Delta_{m+1}$. Montrons que $z := \sum_{i=1}^{m+1} \lambda_i x_i \in C$.

Si $\lambda_{m+1} = 1$, on a $\lambda_1 = \cdots = \lambda_m = 0$, donc $z = x_{m+1} \in C$.

Supposons donc que $\lambda_{m+1} < 1$. On peut écrire :

$$z = \sum_{i=1}^{m+1} \lambda_i x_i = \sum_{i=1}^{m} \lambda_i x_i + \lambda_{m+1} x_{m+1} = (1 - \lambda_{m+1}) \underbrace{\sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} x_i}_{y} + \lambda_{m+1} x_{m+1}.$$

On remarque que les coefficients $\mu_i = \frac{\lambda_i}{1-\lambda_{m+1}}$ sont positifs et vérifient $\sum_{i=1}^m \mu_i = 1$ (car $\sum_{i=1}^m \lambda_i = 1 - \lambda_{m+1}$). Par hypothèse de récurrence, il s'ensuit que $y \in C$.

Enfin, comme $z = (1 - \lambda_{m+1}) y + \lambda_{m+1} x_{m+1}$ est une combinaison convexe de deux points de C (y et x_{m+1}), on conclut par définition de la convexité que $z \in C$. $\square$

Définition 30. Soit $S \subset \mathbb{R}^n$. On appelle enveloppe convexe de S, notée $\mathrm{Conv}(S)$, l'ensemble des combinaisons convexes de points de S :

$$\mathrm{Conv}(S) = \Big\{ \sum_{i=1}^{m} \lambda_i x_i : \; m \geq 1, \; x_i \in S, \; \lambda \in \Delta_m \Big\}.$$

Remarque 13.

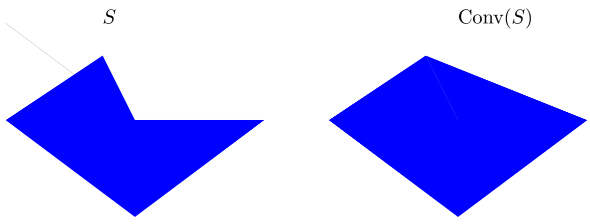
1. Dans la Définition 30, l'entier m est quelconque.
2. L'enveloppe convexe d'un ensemble S est le plus petit ensemble convexe contenant S : si $S \subset U$ avec U convexe, alors $\mathrm{Conv}(S) \subset U$.

Proposition 7. Soit $S \subset \mathbb{R}^n$. Si $S \subset U$ avec U convexe, alors $\mathrm{Conv}(S) \subset U$.

Preuve. La preuve est immédiate. Si $z \in \mathrm{Conv}(S)$, alors $z = \sum_{i=1}^m \lambda_i x_i$ avec $x_i \in S$ et $\lambda = (\lambda_i)_i \in \Delta_m$. Comme $S \subset U$, alors les x_i appartiennent aussi à U. Ce dernier étant convexe, il contient $\sum_{i=1}^m \lambda_i x_i$, i.e., $z \in U$. Donc $\mathrm{Conv}(S) \subset U$. $\square$

Le résultat suivant affirme que tout vecteur x dans l'enveloppe convexe d'un sous-ensemble S de $\mathbb{R}^n$ peut s'écrire comme combinaison convexe d'au plus $n+1$ vecteurs de S.

Théorème 17 (Carathéodory). Soit $S \subset \mathbb{R}^n$ et $x \in \mathrm{Conv}(S)$. Alors il existe $x_1, \ldots, x_{n+1} \in S$ tels que $x \in \mathrm{Conv}(\{x_1, \ldots, x_{n+1}\})$.



(a)

Figure 2.4: Exemple d'un ensemble non-convexe et son enveloppe convexe.

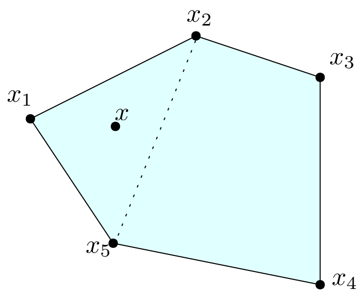


Figure 2.5: Dans cet example on a $x \in \mathrm{Conv}(\{x_1, \dots, x_5\})$. Par le théorème de Carathéordory x peut s'exprimer comme combinaison d'au plus 3 vecteur. Ici, on voit bien que $x \in \mathrm{Conv}(\{x_1, x_2, x_5\})$.

Preuve. Soit $x \in \mathrm{Conv}(S)$. Il existe, par définition, $x_1, \dots, x_m \in S$ et $\lambda \in \Delta_m$ tels que $x = \sum_{i=1}^m \lambda_i x_i$. Supposons que $\lambda_i > 0$ pour tout $i = 1, \dots, m$. Si $m \leq n+1$, il n'y a rien à démontrer. Sinon, supposons $m > n + 1$.

Considérons la famille de vecteurs $(v_i)_{i=2,\dots,m}$ définie par $v_i = x_i - x_1$. On remarque que cette famille est liée puisqu'elle contient $m - 1$ vecteurs et que $m - 1 > n$. Il existe donc des réels $\mu_2, \dots, \mu_m$ non tous nuls tels que $\sum_{i=2}^m \mu_i v_i = 0$, i.e., $\sum_{i=2}^m \mu_i (x_i - x_1) = 0$.

On définit $\mu_1 = -\sum_{i=2}^m \mu_i$. On a par construction :

$$\sum_{i=1}^m \mu_i x_i = 0 \quad \text{et} \quad \sum_{i=1}^m \mu_i = 0,$$

où les μ_i ne sont pas tous nuls. En particulier, il existe $i \in \{1, \dots, m\}$ tel que $\mu_i < 0$. Pour tout $\eta \geq 0$, on a :

$$x = \sum_{i=1}^m \lambda_i x_i = \sum_{i=1}^m \lambda_i x_i + \eta \sum_{i=1}^m \mu_i x_i = \sum_{i=1}^m \beta_i x_i,$$

avec $\beta_i = \lambda_i + \eta \mu_i$ et $\sum_{i=1}^m \beta_i = 1$. L'écriture $\sum_{i=1}^m \beta_i x_i$ est une combinaison convexe si et seulement si $\beta_i \geq 0$ pour tout $i = 1, \dots, m$. Comme les $\lambda_i > 0$, cela est vrai pour $\eta \in [0, \varepsilon]$ avec $\varepsilon = \min_{\{i:\ \mu_i < 0\}} \frac{-\lambda_i}{\mu_i}$.

Pour $\eta = \varepsilon$, on obtient que $\lambda_{i^*} + \eta \mu_{i^*} = 0$ pour un certain $i^* \in \{1, \dots, m\}$. Cela annule le coefficient devant x_{i^*} et fournit une combinaison convexe de x avec $m - 1$ vecteurs. On répète le processus jusqu'à obtenir une combinaison convexe avec au plus $n + 1$ vecteurs. $\square$

Exemple 7. On considère $S = \{x_1, x_2, x_3, x_4\} \subset \mathbb{R}^2$ avec

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ x_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \ x_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ x_4 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

Soit $x \in \mathrm{Conv}(S)$ donné par $x = 1/8 x_1 + 1/4 x_2 + 1/2 x_3 + 1/8 x_4 = \frac{1}{8} \begin{pmatrix} 13 \\ 11 \end{pmatrix}$. Par le Théorème 17, on peut représenter x par trois vecteurs de S (car $n = 2$). On procède comme dans la preuve du théorème.

Les vecteurs $u = (x_2 - x_1)$, $v = (x_3 - x_1)$ et $w = (x_4 - x_1)$ sont linéairement dépendants (trois vecteurs dans $\mathbb{R}^2$). On a $\mu_2 u + \mu_3 v + \mu_4 w = 0$ avec $\mu_2 = \mu_3 = 1$ et $\mu_4 = -1$. Donc $\mu_1 = -(\mu_2 + \mu_3 + \mu_4) = -1$. La relation de dépendance affine s'écrit $-x_1 + x_2 + x_3 - x_4 = 0$. On a, pour $\eta \geq 0$:

$$\begin{aligned} x &= (1/8 + \eta\mu_1)x_1 + (1/4 + \eta\mu_2)x_2 + (1/2 + \eta\mu_3)x_3 + (1/8 + \eta\mu_4)x_4 \\ &= (1/8 - \eta)x_1 + (1/4 + \eta)x_2 + (1/2 + \eta)x_3 + (1/8 - \eta)x_4, \end{aligned} \tag{2.2}$$

et pour assurer que l'écriture ci-dessus est une combinaison convexe, on impose que $1/8 - \eta \geq 0$, $1/4 + \eta \geq 0$ et $1/2 + \eta \geq 0$, soit $\eta \in [0, 1/8]$. Pour $\eta = 1/8$, on obtient

$$x = \frac{3}{8}x_2 + \frac{5}{8}x_3 \quad (cf.\,Fig.\ 2.6).$$

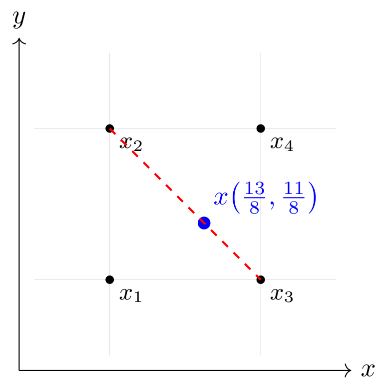


Figure 2.6: Illustration de l'Exemple 7.

2.4 Cônes convexes

Définition 31 (Cône). Soit $S \subset \mathbb{R}^n$. On dit que S est un cône si pour tout $x \in S$ et $\lambda \geq 0$, on a $\lambda x \in S$.

Le résultat suivant donne une caractérisation d'un cône convexe.

Proposition 8. Un ensemble $S \subset \mathbb{R}^n$ est un cône convexe si et seulement si

1. $x \in S,\ \lambda \geq 0 \Rightarrow \lambda x \in S$,
2. $x, y \in S \Rightarrow\ x + y \in S$.

Preuve. Supposons que S vérifie 1 and 2. Soient $x, y \in S$ et $\lambda \in [0, 1]$. On sait déjà par 1 que S est un cône. D'autre part, comme $\lambda \geq 0$ et $(1 - \lambda) \geq 0$, les vecteurs λx et $(1 - \lambda)y$ appartiennent à S (toujours d'après 1). On déduit alors par 2 que leur somme $\lambda x + (1 - \lambda)y$ appartient à S, d'où la convexité.

Réciproquement, supposons que S est un cône convexe. La propriété 1 découle directement de la définition d'un cône. Soient $x, y \in S$. Par convexité de S, on a $\frac{1}{2}x + \frac{1}{2}y = \frac{1}{2}(x + y) \in S$. Comme S est un cône, il est stable par multiplication scalaire positive, donc $x + y = 2\left(\frac{1}{2}(x + y)\right) \in S$, d'où 2. $\square$

Voici quelques exemples de cônes convexes.

Exemple 8. 1. Tout sous-espace vectoriel de $\mathbb{R}^n$ (en particulier, l'espace $\mathbb{R}^n$ lui-même et $\{0\}$).

2. L'orthant positif $\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ x_1 \geq 0, \ldots, x_n \geq 0\}$. Ou plus généralement, un ensemble de la forme
$$S = \{x \in \mathbb{R}^n : \ Ax \leq 0\},$$
avec $A \in \mathbb{R}^{m \times n}$ (c'est un cône polyédrique). On remarque que pour $A = -I_n$, on obtient $\mathbb{R}^n_+$.

3. Le cône de Lorentz (ou cône du second ordre) :
$$L^n = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \ \|x\|_2 \leq t\}.$$

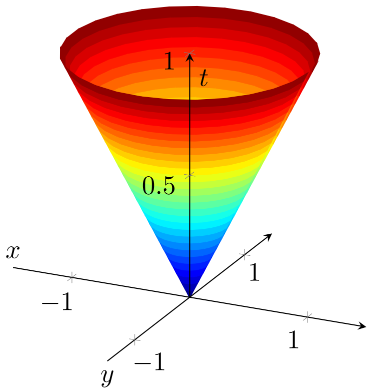


Figure 2.7: Le cône de Lorentz L^2.

2.4.1 Points extrémaux (ou sommets)

Définition 32. Soit $S \subset \mathbb{R}^n$. On dit que $x \in S$ est un point extrémal de S s'il n'existe pas de $y, z \in S$ avec $y \neq z$ et de $\lambda \in (0, 1)$ tels que $x = \lambda y + (1 - \lambda) z$. On note $\mathrm{ext}(S)$ l'ensemble des points extrémaux de S.

Remarque 14. Une autre définition équivalente est la suivante. Soit $x \in S$. Alors $x \in \mathrm{ext}(S)$ si et seulement s'il n'existe pas de $y, z \in S$ distincts tels que $x = \frac{y+z}{2}$.

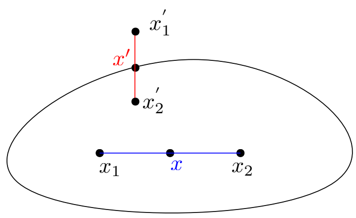


Figure 2.8: Exemple d'un point extrémal x' car il n'existe pas $x_1' \neq x_2' \in C$ tels que $x' = \frac{x_1' + x_2'}{2}$, tandis que x n'est pas un point extrémal.

Étant donné un convexe $C \subset \mathbb{R}^n$, on cherche en pratique un ensemble $S \subset \mathbb{R}^n$ tel que $C = \mathrm{Conv}(S)$. On a le résultat suivant.

Théorème 18. Soit $C = \mathrm{Conv}(S)$ avec $S \subset \mathbb{R}^n$ et soit $x \in S$. Alors $x \in \mathrm{ext}(C)$ si et seulement si $x \notin \mathrm{Conv}(S \setminus \{x\})$.

Preuve. Soit $x \in \mathrm{Conv}(S \setminus \{x\})$. Montrons que $x \notin \mathrm{ext}(C)$, i.e., qu'il existe $y, z \in C$ distincts tels que $x \in [y, z]$. Comme $x \in \mathrm{Conv}(S \setminus \{x\})$, il vient que $x = \sum_{i=1}^{k} \lambda_i x_i$ avec $x_i \in S \setminus \{x\}$ pour tout $i = 1, \dots, k$ et $\lambda \in \Delta_k$ (avec $\lambda_i > 0$). Comme $x \neq x_i$ pour tout i, on a nécessairement $k > 1$.

Posons $y = x_k$ et $z = \sum_{i=1}^{k-1} \frac{\lambda_i}{1-\lambda_k} x_i$. On remarque que z est une combinaison convexe bien définie (car $\sum_{i=1}^{k-1} \lambda_i = 1 - \lambda_k$). Avec ce choix, on a bien

$$x = \lambda_k y + (1 - \lambda_k) z.$$

Puisque $\lambda_k \in (0, 1)$, x est une combinaison convexe stricte de y et z, donc $x \notin \mathrm{ext}(C)$.

Inversement, supposons que $x \notin \mathrm{ext}(C)$, i.e., $x = (y + z)/2$ avec $y, z \in C$ et $y \neq z$. On peut écrire $y = \sum_{i=1}^{k} \lambda_i x_i$ et $z = \sum_{i=1}^{k} \mu_i x_i$ avec $x_i \in S$ pour tout $i = 1, \dots, k$ (quitte à compléter avec des coefficients nuls pour avoir les mêmes indices).

Si $x \neq x_i$ pour tout i, on a

$$x = \frac{y+z}{2} = \sum_{i=1}^{k} \frac{\lambda_i + \mu_i}{2} x_i.$$

Les coefficients sont positifs et leur somme vaut 1, donc $x \in \mathrm{Conv}(S \setminus \{x\})$.

Sinon, supposons que $x = x_i$ pour un certain i, disons $x = x_k$. Comme $x \neq y$ et $x \neq z$, il vient que $\lambda_k < 1$ et $\mu_k < 1$ (sinon, par exemple si $\lambda_k = 1$, alors $y = x_k = x$, ce qui impliquerait $z = x$ et donc $y = z$, une contradiction). On a

$$\sum_{i=1}^{k-1} \frac{\lambda_i + \mu_i}{2} x_i = \frac{y+z}{2} - \frac{\lambda_k + \mu_k}{2} x_k = \left(1 - \frac{\lambda_k + \mu_k}{2}\right) x.$$

En divisant par le coefficient de droite (qui est non nul car $\lambda_k + \mu_k < 2$), on obtient :

$$x = \sum_{i=1}^{k-1} \frac{\lambda_i + \mu_i}{2 - \lambda_k - \mu_k} x_i.$$

Ceci est une combinaison convexe d'éléments de $S \setminus \{x\}$, donc $x \in \mathrm{Conv}(S \setminus \{x\})$. □

Le résultat ci-dessus affirme en particulier que si $C = \mathrm{Conv}(S)$ et $x \in \mathrm{ext}(C)$, alors nécessairement $x \in S$ (i.e., S contient $\mathrm{ext}(C)$). Mais en général, on ne peut pas affirmer que $C = \mathrm{Conv}(\mathrm{ext}(C))$, comme le montrent les exemples suivants : $C = \mathbb{R}^n_+$ dont le seul point extrémal est 0, ou encore la boule ouverte $C = B(0, 1)$ qui n'a pas de points extrémaux. Le premier exemple n'est pas borné, tandis que le deuxième n'est pas fermé. Le résultat suivant, dit de Krein-Milman (ou Minkowski), assure qu'un compact convexe est égal à l'enveloppe convexe de ses points extrémaux.

Théorème 19 (Krein-Milman)**.** Soit $C \subset \mathbb{R}^n$ un compact convexe. Alors $C = \mathrm{Conv}(\mathrm{ext}(C))$.

Preuve. Admise. □

2.5 Exercices

Exercice 6. Soient $C_1, C_2 \subset \mathbb{R}^n$ deux ensembles convexes. Montrer que les ensembles $C_1 \cap C_2, C_1 + C_2, C_1 \times C_2$ sont convexes.

Exercice 7. Montrer que les ensembles suivants sont convexes:

1. $L = \{x + td : \ t \in \mathbb{R}\}$ avec $x, d \in \mathbb{R}^n$ et $d \neq 0$.
2. Les boules ouvertes et fermées: $B(a,r), B_f(a,r)$ avec $a \in \mathbb{R}^n, r > 0$.
3. $H = \{x \in \mathbb{R}^n : \ a^T x = b\}, \ a \in \mathbb{R}^n, b \in \mathbb{R}$.
4. $H^- = \{x \in \mathbb{R}^n : \ a^T x \leq b\}, \ a \in \mathbb{R}^n, b \in \mathbb{R}$.

Exercice 8. On note par

$$\mathcal{S}^n_{++} = \{A \in \mathcal{S}_n : \ x^T A x > 0 : \ \text{pour tout } x \in \mathbb{R}^n \setminus \{0\}\}$$

l'ensemble des matrice symétriques définies positives et

$$\mathcal{S}^n_{+} = \{A \in \mathcal{S}^n : \ x^T A x \geq 0 : \ \text{pour tout } x \in \mathbb{R}^n\}$$

l'ensemble des matrice symétriques semi-définies positives.

- Montrer que $\mathcal{S}^n_{+}$ et $\mathcal{S}^n_{++}$ sont convexes. Montrer que $\mathcal{S}^n_{+}$ est un cône, i.e., pour tout $\lambda \geq 0$ et $A \in \mathcal{S}^n_{+}$, on a $\lambda A \in \mathcal{S}^n_{+}$.
- Montrer que $\mathcal{S}^n_{+}$ est fermé. Qu'elle est l'adhérence de $\mathcal{S}^n_{++}$?

Exercice 9. Soient $S, T \subset \mathbb{R}^n$. Montrer que:

- Si $S \subset T$ alors $\text{Conv}(S) \subset \text{Conv}(T)$.
- $\text{Conv}(S + T) = \text{Conv}(S) + \text{Conv}(T)$
- $\text{Conv}(\text{Conv}(S)) = \text{Conv}(S)$.
- Montrer que $\text{Conv}(S \times T) = \text{Conv}(S) \times \text{Conv}(T)$.

Exercice 10. Soit $C \subset \mathbb{R}^n$ vérifiant la propriété dite de demi-somme:

$$x, y \in C \Rightarrow \frac{x + y}{2} \in C.$$

C est-il convexe ? Que dire si C est supposé être fermé ?

Exercice 11. Montrer que

$$\{x \in \mathbb{R}^2 : \ |x_1| + |x_2| \leq 1\} := B_{\|\cdot\|_1}(0,1) = \text{Conv}(e_1, -e_1, e_2, -e_2),$$

e_1 et e_2 étant les vecteurs de la base canonique de R^2. Quels sont les points extrémaux de $B_{\|\cdot\|_1}(0,1)$.

Exercice 12. Soient $C_1, C_2 \subset \mathbb{R}^m \times \mathbb{R}^n$ deux ensembles convexes. Montrer que la somme partielle de C_1 et C_2 définie par

$$C := \{(x, y_1 + y_2) : \ x \in \mathbb{R}^m, y_1, y_2 \in \mathbb{R}^n, (x, y_1) \in C_1, (x, y_2) \in C_2\},$$

est convexe

Exercice 13. Soit $a \in \mathbb{R}^n$ et considérons les deux hyperplans parallèles suivants

$$H_i = \{x \in \mathbb{R}^n : \ a^T x = b_i\}, \text{pour } i = 1, 2.$$

Calculer la distance entre H_1 et H_2.

Exercice 14. Soient $a \neq b \in \mathbb{R}^n$ et $\theta \in [0, 1]$. On considère l'ensemble

$$V_\theta = \{x \in \mathbb{R}^n : \ \|x - a\| \leq \theta \|x - b\|\}.$$

Montrer que V_1 est un demi-plan, i.e., s'écrit de la forme $\{x : u^T x \leq r\}$ pour $u \in \mathbb{R}^n, r \in \mathbb{R}$ à préciser. Que devient V_θ pour $\theta < 1$?. Pour $\theta > 1$?

Exercice 15. Soit $C \subseteq \mathbb{R}^n$ un cône. Le cône dual de C est définie par

$$C^* = \{y \in \mathbb{R}^n : \ \langle y, x \rangle \geq 0 \text{ pour tout } x \in C\}.$$

1. Montrer que C^* est un cône convexe fermé (même si C n'est pas convexe).
2. Montrer que si C_1, C_2 sont des cônes convexes tels que $C_1 \subseteq C_2$, alors $C_2^* \subseteq C_1^*$.
3. Montrer que $(L^n)^* = L^n$ avec

$$L^n = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \ \|x\| \leq t, \ x \in \mathbb{R}^n, t \in \mathbb{R} \right\}.$$

4. Montrer que le cône dual de $K = \left\{ \binom{\mathbf{x}}{t} : \|\mathbf{x}\|_1 \leq t \right\}$ est

$$K^* = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} : \|x\|_\infty \leq t \right\}.$$

Chapitre 3.

Fonctions convexes

Dans cette section, on considère un ensemble convexe C de $\mathbb{R}^n$.

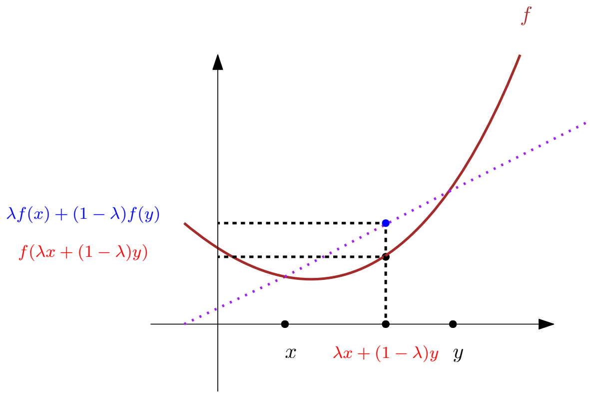
Définition 33. On dit qu'une fonction $f : C \to \mathbb{R}$ est convexe si

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall x, y \in C, \forall \lambda \in [0,1].$$

On dit que f est strictement convexe si

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y), \quad \forall x, y \in C, x \neq y, \forall \lambda \in (0,1).$$

On dit que f est concave si $-f$ est convexe.



(a)

Figure 3.1: Illustration de l'inégalité de convexité dans la Définition 33: la corde liant les deux points $(x, f(x))$ et $(y, f(y))$ est au dessus du graphe de f.

Exemple 9. Voici quelques exemples de fonctions convexes :

- Les normes : $f(x) = \|x\|$ sur $\mathbb{R}^n$.
- Les fonctions affines : $f(x) = a^T x + b$, avec $a \in \mathbb{R}^n$ et $b \in \mathbb{R}$.
- La fonction $f(x) = -\log(x)$ est convexe sur $\mathbb{R}_+^*$.
- La fonction $f(x) = e^{\lambda x}$ est convexe sur $\mathbb{R}$ pour tout $\lambda \in \mathbb{R}$.
- La fonction $f(x) = |x|^p$ est convexe sur $\mathbb{R}$ pour tout $p \geq 1$.

D'après la Définition 33, une fonction f est convexe si l'image par f d'une combinaison convexe de deux points x et y est plus petite que la combinaison convexe des valeurs $f(x)$ et $f(y)$. Cette propriété s'étend à la combinaison convexe de n'importe quel nombre de vecteurs.

Théorème 20 (Inégalité de Jensen)**.** Soit $f : C \to \mathbb{R}$ une fonction convexe. Alors pour tout

$x_1, \dots, x_m \in C$ et $\lambda \in \Delta_m$:

$$f\Big(\sum_{i=1}^{m} \lambda_i x_i\Big) \leq \sum_{i=1}^{m} \lambda_i f(x_i).$$

Preuve. Par récurrence. Laissée en exercice. □

Remarque 15. L'inégalité de convexité dans la Définition 33 est parfois appelée inégalité de Jensen. Historiquement, le résultat démontré par Jensen était $f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$. On rencontre d'autres versions de l'inégalité de Jensen en probabilités et en théorie de la mesure et de l'intégration. Typiquement, si f est une fonction convexe définie sur un convexe C de $\mathbb{R}^n$ et X une variable aléatoire à valeurs dans C telle que l'espérance $\mathbb{E}(X)$ existe, alors $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$. Un autre exemple est le suivant : si $(X, \mathcal{M}, \mu)$ est un espace probabilisé (où μ est une mesure positive avec $\mu(X) = 1$), I un intervalle non vide de $\mathbb{R}$, $\phi : I \to \mathbb{R}$ une fonction convexe et $f : X \to I$ une fonction dans $\mathcal{L}^1_\mu$, alors

$$\phi\Big(\int_X f \mathbf{d}\mu\Big) \leq \int_X \phi \circ f \mathbf{d}\mu.$$

Le résultat suivant est une conséquence immédiate du Théorème 20.

Corollaire 3. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe. Pour toute combinaison convexe $x = \sum_{i=1}^m \lambda_i x_i$, avec $\lambda_i \geq 0$, $\sum_{i=1}^m \lambda_i = 1$ et $x_1, \dots, x_m \in \mathbb{R}^n$, on a

$$f(x) \leq \max_{i=1,\dots,m} f(x_i).$$

Preuve. D'après l'inégalité de Jensen, on a

$$f(x) = f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i) \leq \max_{i=1,\dots,m} f(x_i) \sum_{i=1}^{m} \lambda_i = \max_{i=1,\dots,m} f(x_i).$$

□

Le résultat suivant récapitule quelques opérations préservant la convexité de fonctions.

Théorème 21.

1. Soit f une fonction convexe sur C et $\alpha \in \mathbb{R}^+$. Alors αf est convexe sur C.
2. Soient $f_1, \dots, f_m$ des fonctions convexes sur C. Alors $f = \sum_{i=1}^m f_i$ est convexe sur C.
3. Soient $f_1, \dots, f_m$ des fonctions convexes sur C. Alors $f = \max_{i=1,\dots,m} f_i$ est convexe sur C.
4. Soit $f : C \to \mathbb{R}$ une fonction convexe, $A \in \mathbb{R}^{n\times m}$ et $b \in \mathbb{R}^n$. Alors la fonction g définie par $g(y) = f(Ay + b)$ est convexe sur

$$D = \{y \in \mathbb{R}^m : \ Ay + b \in C\}.$$

Preuve.

1. Immédiat.
2. Soient $x, y \in C$ et $\lambda \in [0, 1]$. Comme les f_i sont convexes, on a pour tout $i = 1, \dots, m$:

$$f_i(\lambda x + (1-\lambda)y) \leq \lambda f_i(x) + (1-\lambda) f_i(y).$$

En sommant ces inégalités, on obtient :

$$\sum_{i=1}^{m} f_i(\lambda x + (1-\lambda)y) \leq \sum_{i=1}^{m} \Big(\lambda f_i(x) + (1-\lambda) f_i(y)\Big) = \lambda \sum_{i=1}^{m} f_i(x) + (1-\lambda) \sum_{i=1}^{m} f_i(y),$$

soit

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y),$$

donc f est convexe.

3. Soient $x, y \in C$ et $\lambda \in [0,1]$. On a :

$$f(\lambda x + (1-\lambda)y) = \max_{i=1,\ldots,m} f_i(\lambda x + (1-\lambda)y) \leq \max_{i=1,\ldots,m} \Big(\lambda f_i(x) + (1-\lambda) f_i(y)\Big).$$

Rappelons que si $(\alpha_i)_{i=1}^m$ et $(\beta_i)_{i=1}^m$ sont deux suites de réels, alors $\max_i(\alpha_i+\beta_i) \leq \max_i \alpha_i + \max_i \beta_i$. Donc :

$$f(\lambda x + (1-\lambda)y) \leq \lambda \max_{i=1,\ldots,m} f_i(x) + (1-\lambda) \max_{i=1,\ldots,m} f_i(y) = \lambda f(x) + (1-\lambda) f(y).$$

4. Tout d'abord, remarquons que D est convexe (c'est l'image réciproque du convexe C par une application affine). Soient $y_1, y_2 \in D$ et définissons $x_i = Ay_i + b$ pour $i = 1, 2$. Par définition de D, on a $x_1, x_2 \in C$. Soit $\lambda \in [0,1]$. Par convexité de f, on a :

$$f\Big(\lambda(Ay_1 + b) + (1-\lambda)(Ay_2 + b)\Big) \leq \lambda f(Ay_1 + b) + (1-\lambda) f(Ay_2 + b),$$

ce qui s'écrit aussi (par linéarité de l'application $y \mapsto Ay$) :

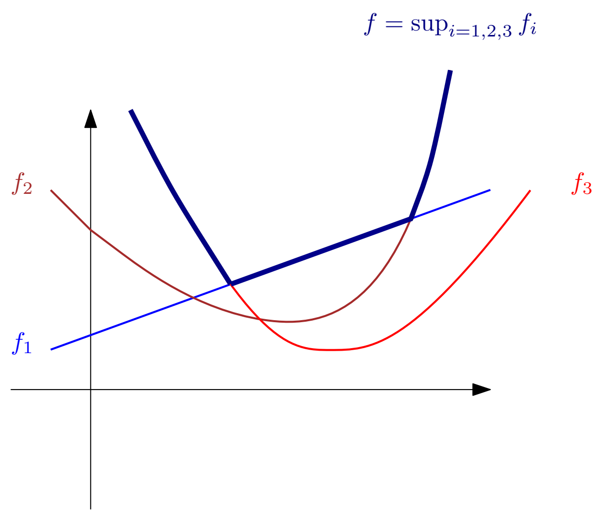
$$f\Big(A(\lambda y_1 + (1-\lambda)y_2) + b\Big) \leq \lambda g(y_1) + (1-\lambda) g(y_2),$$

c'est-à-dire

$$g(\lambda y_1 + (1-\lambda)y_2) \leq \lambda g(y_1) + (1-\lambda) g(y_2),$$

d'où la convexité de g.

□



(a)

Figure 3.2: Convexité du la fonction sup de trois fonctions f_1, f_2, f_3.

Exemple 10. Soit $f(x) = \frac{\|Ax+b\|^2}{c^T x + d}$ avec $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^n, c \in \mathbb{R}^n \setminus \{0\}$ et $d \in \mathbb{R}$. Alors f est convexe sur $D = \{x : c^T x + d > 0\}$. En effet, on remarque que $f(x) = g(Ax+b, c^T x + d)$ avec $g(x,t) = \frac{\|x\|^2}{t}$ qui est convexe sur $\{\binom{x}{t} \in \mathbb{R}^{m+1} : \ y \in \mathbb{R}^m, t > 0\}$ puisque $g(x,t) = \sum_{i=1}^m g_i(x,t)$ avec $g_i(x,t) = \frac{y_i^2}{t}$ qui est convexe sur $\mathbb{R} \times \mathbb{R}_+^*$ (voir TD).

Théorème 22. Soit $f : C \to \mathbb{R}$ une fonction convexe et $g : I \subset \mathbb{R} \to \mathbb{R}$ une fonction convexe croissante. Supposons que $f(C) \subset I$. Alors $\phi(x) = g(f(x))$ est convexe sur C.

Preuve. Soient $x, y \in C$ et $\lambda \in [0,1]$. On a

$$\begin{aligned}\phi(\lambda x + (1-\lambda)y) &= g(f(\lambda x + (1-\lambda)y) \\ &\leq g(\lambda f(x) + (1-\lambda)f(y)) \text{ par convexité de f et monotonie de g} \\ &\leq \lambda g(f(x)) + (1-\lambda)g(f(y)) \text{ par convexité de g} \\ &= \lambda\phi(x) + (1-\lambda)\phi(y).\end{aligned} \tag{3.1}$$

□

Exemple 11. La fonction $\phi(x) = e^{\|x\|^2}$ est convexe sur $\mathbb{R}^n$. En effet $\phi(x) = g(f(x))$ avec $f(x) = e^x$ qui est croissante et $f(x) = \|x\|^2$ qui est convexe. Plus généralement, pour toute fonction convexe f, $e^{f(x)}$ est convexe.

Soient $C \subset \mathbb{R}^d$ et $D \subset \mathbb{R}^k$ deux convexe et $f : C \times D \to \mathbb{R}$ une fonction convexe. Le résultat suivant montre la préservation de la convexité de f avec la minimisation partielle.

Théorème 23. La fonction $g(x) = \inf_{y \in D} f(x,y)$ est convexe sur C.

Preuve. Soient $x_1, x_2 \in C$ et $\varepsilon > 0$. Il existe $y_1, y_2 \in D$ tel que $f(x_i, y_i) \leq g(x_i) + \varepsilon, i = 1, 2$. Soit $\lambda \in [0,1]$. On a

$$\begin{aligned}g(\lambda x_1 + (1-\lambda)x_2) & \inf_{y \in D} f(\lambda x_1 + (1-\lambda)x_2, y) \\ & f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\ & \lambda f(x_1, y_1) + (1-\lambda)f(x_2, y_2) \\ & \lambda g(x_1) + (1-\lambda)g(x_2) + \varepsilon.\end{aligned} \tag{3.2}$$

L'inégalité étant vraie pour tout $\varepsilon > 0$, on obtient $g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2)$. Ce qui termine la preuve. □

On termine cette section par ce lemme qui nous sera utile par la suite.

Lemme 2. Une fonction $f : C \to \mathbb{R}$ est convexe, avec C convexe, si et seulement si pour tout $x, y \in C$ et $\alpha \geq 0$ tel que $y + \alpha(y - x) \in C$ on a

$$f(y + \alpha(y-x)) \geq f(y) + \alpha(f(y) - f(x)). \tag{3.3}$$

Preuve. Soient $x, y \in C, t \in (0,1]$ et considérons $z = tx + (1-t)y$. On a $x = \frac{1}{t}z - \frac{1-t}{t}y = z + \alpha(z - y)$ avec $\alpha = \frac{1-t}{t}$. Par (3.3)

$$f(x) \geq f(y) + \alpha(f(z) - f(y)),$$

i.e., $tf(x) \geq tf(z) + (1-t)(f(z) - f(y))$, soit $f(z) \leq tf(x) + (1-t)f(y)$, d'où la convexité de f. Inversement, pour $t = \alpha/(1+\alpha)$ et $z = y + \alpha(y - x)$, on a $y = (1-t)z + tx$, comme f est convexe, on a par l'inégalité de Jensen

$$f(y) \leq tf(x) + (1-t)f(z) = \frac{\alpha}{1+\alpha}f(x) + \frac{1}{1+\alpha}f(z),$$

multipliant par $\alpha + 1$ dans l'inégalité précédente, on obtient

$$(1+\alpha)f(y) \leq \alpha f(x) + f(z),$$

soit $f(z) = f(y + \alpha(y-x)) \geq f(y) + \alpha(f(y) - f(x))$, qui n'est rien d'autre que (3.3) □

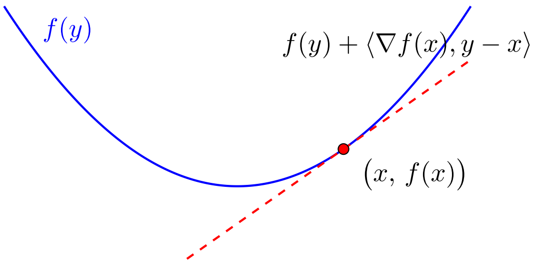


Figure 3.3: Illustration du Théorème 24.

3.1 Caractérisations des fonctions convexes différentiables

3.1.1 Caractérisations du premier ordre

Théorème 24. Soit $f : C \to \mathbb{R}$ une fonction de classe C^1. Alors f est convexe si et seulement si

$$f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle \text{ pour tout } x, y \in C. \tag{3.4}$$

Preuve. Supposons que f est convexe et soient $x \neq y \in C$ et $\lambda \in (0, 1)$. On a par définition

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y),$$

ce qui donne

$$\frac{f(\lambda x + (1-\lambda)y) - f(y)}{\lambda} \leq f(x) - f(y).$$

Quand $\lambda \to 0^+$, on obtient $f^{'}(y; x-y) \leq f(x)-f(y)$. Comme f est C^1, $f^{'}(y; x-y) = \langle \nabla f(y), x-y\rangle$ et donc $f(y)+\langle \nabla f(y), x-y\rangle \leq f(x)$. Supposons maintenant que $f(x) \geq f(y)+\langle \nabla f(y), x-y\rangle$ pour tout $x, y \in C$ et montrons que f est convexe. Soient $u, v \in C$ et $\lambda \in (0, 1)$. Posons $w = \lambda u + (1-\lambda)v$ et montrons que $f(w) \leq \lambda f(u) + (1-\lambda)f(v)$. On a

$$u - w = \frac{w - (1-\lambda)v}{\lambda} - w = \frac{\lambda - 1}{\lambda}(v - w).$$

En appliquant (3.4) pour u, w et ensuite v, w on obtient

$$f(w) + \langle \nabla f(w), u - w\rangle \leq f(u), \tag{3.5}$$

et

$$f(w) + \langle \nabla f(w), v - w\rangle = f(w) - \frac{\lambda}{1-\lambda}\langle \nabla f(w), u - w\rangle \leq f(v),$$

soit

$$(1-\lambda)f(w) - \lambda\langle \nabla f(w), u - w\rangle \leq (1-\lambda)f(v). \tag{3.6}$$

En multipliant (3.5) par λ et sommant avec (3.6), on obtient le résultat. $\square$

Remarque 16. Dans la littérature, l'inégalité (3.4) est souvent appelée *the gradient inequality*. Elle affirme que les hyperplans tangents à une fonction convexe minorent la fonction.

Le résultat suivant est une caractérisation de la convexité en terme de la monotonie du gradient "au sens des opérateurs".

Théorème 25 (Monotonie du gradient)**.** Soit $f : C \to \mathbb{R}$ une fonction de classe C^1. Alors f est convexe si et seulement si

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0, \text{ pour tout } x, y \in C. \tag{3.7}$$

Preuve. Supposons que f est convexe et soient $x, y \in C$. Par (3.4) on a

$$f(x) + \langle \nabla f(x), y - x \rangle \leq f(y), \text{ et } f(y) + \langle \nabla f(y), x - y \rangle \leq f(x).$$

En sommant les deux inégalités on obtient que $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$. Inversement, supposons que (3.7) est vérifiée et montrant que f est convexe. Définitions la fonction $\phi : t \in (0, 1) \mapsto f(x + t(y - x))$. On a

$$\begin{aligned} f(y) = \phi(1) = \phi(0) + \int_0^1 \phi^{'}(t)\mathbf{d}t \\ &= f(x) + \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle \mathbf{d}t \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle \mathbf{d}t \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{t} \int_0^1 \underbrace{\langle \nabla f(x + t(y - x)) - \nabla f(x), x + t(y - x) - x \rangle}_{\geq 0} \mathbf{d}t, \end{aligned} \tag{3.8}$$

soit $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$, et par Théorème 24 f est convexe. □

Quand la fonction est deux fois différentiable, alors la convexité de f est équivalente au fait que la matrice Hessienne est semi-définie positive.

3.1.2 Caractérisations du second ordre

Théorème 26. Soit $f : C \to \mathbb{R}$ une fonction de classe C^2 avec C un ouvert convexe de $\mathbb{R}^n$ alors f est convexe si et seulement si la Hessienne $\nabla^2 f(x)$ est semi-définie positive, i.e.,

$$\nabla^2 f(x) \succeq 0 \text{ pour tout } x \in C. \tag{3.9}$$

Preuve. Supposons que f est convexe de classe C^2 et soient $x \in C$ et $d \in \mathbb{R}^n$. Comme C est ouvert, il vient que $x_t := x + td \in C$ pour $t < \varepsilon$ avec $\varepsilon > 0$ suffisamment petit. Comme f est convexe, on a, par monotonie de ∇f

$$0 \leq t^{-1} \langle \nabla f(x_t) - \nabla f(x), x_t - x \rangle = \langle \nabla f(x_t) - \nabla f(x), d \rangle = \int_0^t \langle \nabla^2 f(x_\tau), d, d \rangle \mathbf{d}\tau$$

En faisant tendre $t \to 0^+$ on obtient que $\langle \nabla^2 f(x) d, d \rangle \geq 0$. Comme d est arbitraire, il s'en suit que $\nabla^2 f(x) \succeq 0$ pour tout $x \in C$. Maintenant, supposons que (3.9) est vérifiée, on a pour tout $x, y \in C$

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \Big(\int_0^s \underbrace{\langle \nabla^2 f(x_\tau)(y - x), y - x \rangle}_{\geq 0} \mathbf{d}\tau \Big) \mathbf{d}s \\ &\geq f(x) + \langle \nabla f(x), y - x \rangle, \end{aligned} \tag{3.10}$$

ce qui implique par Théorème 24 la convexité de f. □

Remarque 17. La condition (3.9) est liée à la notion de courbure. En effet, considérons la surface $\Gamma_f = \{(x, y, z) \in \mathbb{R}^2 : \ z = f(x, y)\}$ avec f de classe C^2. La courbure de Gauss au point (x, y)

est égale à
$$\kappa = \frac{\det(\nabla^2 f(x,y))}{(1+\|\nabla f(x,y)\|^2)^2}.$$
Le signe de $\det(\nabla^2 f(x,y))$ (et donc de κ) donne une classification de la surface: elliptique, parabolique ou hyperbolique.

3.2 Continuité et différentiabilité des fonctions convexes

On commence par un premier résultat quand $C = \mathbb{R}^n$.

Théorème 27. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe. Alors f est continue en tout point en tout $x \in \mathbb{R}^n$.

Preuve. Sans perte de généralité, on suppose que $x = 0$. Soit $(x_k)_k \geq 0$ telle que $x_k \to x^* = 0$ quand $k \to \infty$. On a, par convexité de f
$$f(x_k) \leq (1 - \|x_k\|)f(0) + \|x_k\| f(x_k/\|x_k\|).$$
Comme $(x_k/\|x_k\|)_i \in [-1,1]$ pour tout $i = 1,\dots,n$, il vient que $x_k/\|x_k\| \in [-1,1]^n$, donc $x_k/\|x_k\| = \sum_{i=1}^{2^n} \lambda_i e_i$ avec $\lambda \in \Delta_{2^n}$. On déduit par le Corollaire 3 que $f(x_k/\|x_k\|) \leq \max_{i=1,\dots,n} f(\pm e_i) := K$. Il s'en suit que
$$\limsup_{k\to\infty} f(x_k) \leq (1 - \|x^*\|)f(0) + \|x^*\|K = f(0).$$
De même, en remarquant que
$$f(0) \leq \frac{\|x_k\|}{1+\|x_k\|} f(-x_k/\|x_k\|) + \frac{1}{1+\|x_k\|} f(x_k),$$
on déduit que $f(0) \leq \liminf_{k\to\infty} f(x_k)$, et par la suite $\lim\limits_{k\to\infty} f(x_k) = f(0)$. D'où la continuité de f en 0. □

Sur un convexe $C \subsetneq \mathbb{R}^n$, on peut obtenir un résultat similaire (meilleur même) à Théorème 27 pour les point intérieurs à C. En effet, on peut démontrer qu'une fonction convexe est localement Lipschitzienne en tout $x \in \text{int}(C)$. La raison de se restreindre aux points intérieurs est le comportement d'une fonction convexe au bords qui peut créer des discontinuités. Pour illustrer ceci, considérons la fonction $f : [0,1] \to \mathbb{R}$ définie par $f(0) = 1$ et $f(x) = x^2$ pour $x \in (0,1]$. Cette fonction est évidemment convexe (faites un dessin) mais n'est pas continue.

On a le résultat suivant.

Théorème 28. Soit $f : C \to \mathbb{R}$ une fonction convexe définie sur un convexe C de $\mathbb{R}^n$ et soit $x_0 \in \text{int}(C)$. Alors il existe $L > 0$ et $\varepsilon > 0$ tels que $B(x_0,\varepsilon) \subset C$ et
$$|f(x) - f(x_0)| \leq L\|x - x_0\|, \ \forall x \in B(x_0,\varepsilon). \tag{3.11}$$

Preuve. Soit $x_0 \in \text{int}(C)$, il existe alors $\varepsilon > 0$ tel que $B_\infty(x_0,\varepsilon) \subset C$. On commence par montrer que f est bornée supérieurement sur $B_\infty(x_0,\varepsilon)$. Comme $B_\infty(x_0,\varepsilon)$ est convexe et compact, avec
$$\text{ext}(B_\infty(x_0,\varepsilon)) = \{z_i = x_0 + \varepsilon\theta_i, \text{ avec } \theta_i \in \{\pm 1\}, i = 1,2,\dots,2^n\},$$
on déduit par le théorème de Krein-Milman que tout $x \in B_\infty(x_0,\varepsilon)$ s'écrit de la forme $\sum_{i=1}^{2^n} \lambda_i z_i$ avec $\lambda \in \Delta_{2^n}$. Donc par Corollaire 3 $f(x) \leq K := \max_{i=1,2,\dots,2^n} f(z_i)$. Comme $B(x_0,\varepsilon) \subset B_\infty(x_0,\varepsilon)$ on en déduit que $f(x) \leq K$ pour tout $x \in B(x_0,\varepsilon)$.

Soit $x \in B(x_0,\varepsilon)$ avec $x \neq x_0$. On définit $z = x_0 + \alpha^{-1}(x - x_0)$ avec $\alpha = \|x - x_0\|/\varepsilon$. On voit que $0 < \alpha \leq 1$ et $\|z - x_0\| = \varepsilon$, i.e., $z \in B(x_0,\varepsilon)$ et donc $f(z) \leq K$. Comme $x = \alpha z + (1-\alpha)x_0$, on a par convexité de f

$$f(x) \leq \alpha f(z) + (1-\alpha) f(x_0) = f(x_0) + \alpha(f(z) - f(x_0)) \leq f(x_0) + \alpha(K - f(x_0)), \quad (3.12)$$

i.e., $f(x) - f(x_0) \leq L\|x - x_0\|$, avec $L = (K - f(x_0))\varepsilon^{-1}$. Pour finir la preuve, montrons que $f(x) - f(x_0) \geq -L\|x - x_0\|$. Définissons $w = x_0 + (x_0 - x)\alpha^{-1}$ et remarquons que $\|w - x_0\| = \varepsilon$ et donc $f(w) \leq K$. Comme $x = x_0 + \alpha(x_0 - w)$, on a par Lemme 2

$$\begin{aligned} f(x) = f(x_0) + \alpha(x_0 - w)) &\geq f(x_0) + \alpha(f(x_0) - f(w)) \\ &\geq f(x_0) - \alpha(K - f(x_0)), \end{aligned} \quad (3.13)$$

i.e., $f(x) - f(x_0) \geq -L\|x - x_0\|$. En conclusion, on obtient que (3.11) pour tout $x \in B(x, \varepsilon)$ avec $L = (K - f(x_0))\varepsilon^{-1}$. □

Nous avons vu qu'en les points intérieurs à C, une fonction convexe $f : C \to \mathbb{R}$ est localement Lipschitzienne, i.e., vérifie (3.11). Maintenant, on montrera que les dérivées directionnelles de f en tout point intérieur existent. Rappelons qu'étant donnés $x \in \text{int}(C)$ et $0 \neq d \in \mathbb{R}^n$, on dira que f est différentiable en x dans la direction d si la limite

$$\lim_{t \to 0^+} t^{-1}(f(x + td) - f(x)) := f'(x; d),$$

existe.

Théorème 29. Soit $f : C \to \mathbb{R}$ une fonction convexe définie sur un convexe $C \subset \mathbb{R}^n$ et soit $x \in \text{int}(C)$. Alors $f'(x; d)$ existe pour tout $0 \neq d \in \mathbb{R}^n$.

Preuve. Soit $x \in \text{int}(C)$, $d \in \mathbb{R}^n, 0 < \alpha \leq 1$ et $\varepsilon > 0$ tel que $x + td \in C$ pour tout $0 < t \leq \varepsilon$. On note $h(t) := t^{-1}(f(x + td) - f(x))$. Comme $\alpha t \leq t$ et $x + td = (1-\alpha)x + \alpha(x + td)$, on a par l'inégalité de Jensen

$$f(x + \alpha td) \leq (1-\alpha) f(x) + \alpha f(x + td),$$

soit

$$h(t) = t^{-1}(f(x + td) - f(x)) \geq (\alpha) t^{-1}(f(x + t\alpha d) - f(x)) = h(\alpha t).$$

Ce qui prouve que h est décroissante. De même, on montre que pour tout $t_0 >$ tel que $x - t_0 d \in C$, on a que $h(t) \geq t_0^{-1}(f(x) - f(x + t_0 d))$, i.e., h est bornée inférieurement. Par conséquence $\lim_{t \to 0^+} h(t)$ existe, i.e., la dérivée directionnelle $f'(x; d)$ existe. □

3.3 Sous-ensembles de niveau de fonctions convexes:

Définition 34. Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ une fonction. Le sous-ensemble de f de niveau $\alpha \in \mathbb{R}$ est l'ensembles

$$\textbf{Lev}(f, \alpha) = \{x \in S : \ f(x) \leq \alpha\}.$$

On a le résultat suivant

Théorème 30. Soit $f : C \to \mathbb{R}$ une fonction convexe et avec C un ensemble convexe. Alors pour tout $\alpha \in \mathbb{R}$, l'ensembles $\textbf{Lev}(f, \alpha)$ est convexe.

Preuve. Soient $x, y \in \textbf{Lev}(f, \alpha)$ et $\lambda \in [0, 1]$ avec $\alpha \in \mathbb{R}$. Comme $f(x), f(y) \leq \alpha$, il vient par convexité de f que

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

□

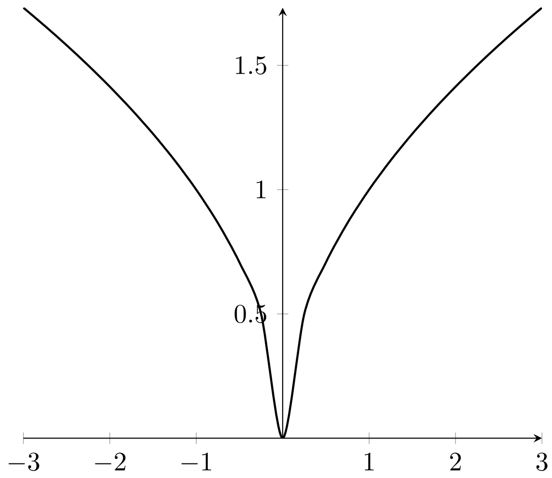


Figure 3.4: $f(x) = \sqrt{|x|}$ comme exemple de fonction quasi-convexe.

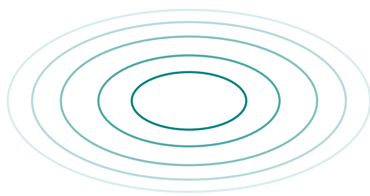
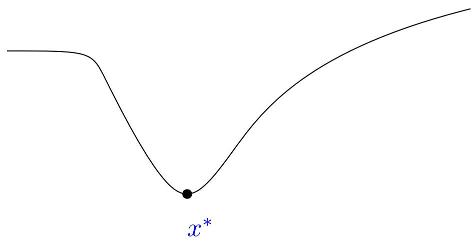


Figure 3.5: Sous-ensembles de niveau de la fonction $f(x, y) = 1/2(x^2 + 4y^2)$.

Remarque 18. La réciproque dans le Théorème-30 est fausse comme le montre la fonction $f(x) = \sqrt{|x|}$ (voir Figure-3.4) qui n'est pas convexe mais dont tout les sous-ensembles de niveau sont convexe. En effet, d'une part, pourt $\alpha < 0, \mathbf{Lev}(f, \alpha) = \emptyset$. D'autre part, pour $\alpha \geq 0$, on $\mathbf{Lev}(f, \alpha) = [-\alpha^2, \alpha^2]$ qui est convexe. Une telle fonction est dite quasi-convexe.

Définition 35. Soit $C \subset \mathbb{R}$ un convexe. On dit qu'une fonction $f : C \to \mathbb{R}$ est quasi-convexe si pour tout $\alpha \in \mathbb{R}$, l'ensemble $\mathbf{Lev}(f, \alpha)$ est convexe. De même, f est dite quasi-concave si $-f$ est quasi-convexe.

Exemple 12.
1. Évidemment, une fonction convexe est quasi-convexe.
2. L'application $x \in \mathbb{R}^n \mapsto \text{card}(x) = \{\text{nombre des composantes non nuls de } x\}$ est quasi-concave:
$$\text{card}(x + y) \geq \min\{\text{card}(x), \text{card}(y)\}.$$
3. L'application $M \in \mathbb{R}^{n \times n} \mapsto \text{rang}(M)$ est quasi-concave sur $\mathcal{S}_+^n$:
$$\text{rang}(M + N) \geq \min\{\text{rang}(M), \text{rang}(N)\}.$$
4. Un exemple important est celui des fonctions continue sur $\mathbb{R}$. En effet, si $f : \mathbb{R} \to \mathbb{R}$ est continue alors f est quasi-convexe ssi l'une des propriétés suivantes est vérifée
   (a) f est croissante;
   (b) f est décroissante;
   (c) il existe $x^* \in \mathbb{R}$ tel que pour tout $x \leq x^*$ f est décroissante, et pour tout $x \geq x^*$ est croissante (cf. Fig. 3.6).



(a)

Figure 3.6: Exemple de fonction quasi-convexe sur $\mathbb{R}$.

3.4 Fonctions à valeurs réelles étendues

Dans la pratique on peut être amené à travailler et considérer des fonctions qui peuvent prendre des valeurs infinies. En effet, considérons une fonction $f : \mathbb{R}^n \to \bar{\mathbb{R}} = [-\infty, \infty]$. La définition de convexité (33) reste valable pour de telles fonctions tenant compte des opération arithmétiques classiques

$$\begin{aligned} a + \infty = \infty + a &= \infty && (-\infty < a < \infty), \\ a - \infty = -\infty + a &= -\infty && (-\infty < a < \infty), \\ a \cdot \infty = \infty \cdot a &= \infty && (0 < a < \infty), \\ a \cdot (-\infty) = (-\infty) \cdot a &= -\infty && (0 < a < \infty), \\ a \cdot \infty = \infty \cdot a &= -\infty && (-\infty < a < 0), \\ a \cdot (-\infty) = (-\infty) \cdot a &= \infty && (-\infty < a < 0), \end{aligned} \tag{3.14}$$

ainsi qu'avec la règle "moins usuelle" $0 \cdot \infty = 0 \cdot (-\infty) = 0$. Néanmoins, on s'intéresse plutôt à des fonctions qui ne prennent pas la valeur $-\infty$ et dont le domaine est non vide.

Définition 36 (Domaine effectif). Soit $f : \mathbb{R}^n \to [-\infty, \infty]$, le domaine (ou domaine effectif) de f est définit par

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n : \; f(x) < \infty\}.$$

Exemple 13. Si $C \subset \mathbb{R}^n$, l'indicatrice de C au sens d'analyse convexe est définie par

$$\delta_C(x) = \begin{cases} 0, & \text{si } x \in C, \\ \infty & \text{sinon}, \end{cases}$$

alors $\mathrm{dom}(\delta_C) = C$.

Définition 37. On dit que $f : \mathbb{R}^n \to [-\infty, \infty]$ est propre si $-\infty \notin f(\mathbb{R}^n)$ et $\mathrm{dom}(f) \neq \emptyset$, i.e., il existe $x_0 \in \mathbb{R}^n$ tel que $f(x_0) < \infty$.

Une autre caractérisation géométrique des fonctions convexes est donnée par l'ensemble suivant.

Définition 38 (Epigraphe). Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, on définit l'épigraphe de f par

$$\mathrm{epi}(f) = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \; f(x) \leq t \right\}.$$

Clairement, si $\binom{x}{t} \in \mathrm{epi}(f)$, alors $x \in \mathrm{dom}(f)$. Contrairement aux sous-ensembles de niveaux, la convexité de l'épigraphe est équivalente à celle de f.

Théorème 31 (et définition)**.** f est convexe $\Leftrightarrow$ $\mathrm{epi}(f)$ est convexe.

Exemple 14.

1. Si $f(x) = a^T x - \alpha$ avec $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$. On a
$$\mathrm{epi}(f) = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \left\langle \begin{pmatrix} a \\ -1 \end{pmatrix}, \begin{pmatrix} x \\ t \end{pmatrix} \right\rangle \leq \alpha \right\},$$
est un demi-espace donc convexe.
2. Si $f(x) = 1/2\|x\|^2$, alors $\mathrm{epi}(f)$ est la région du l'espace au dessus de la parabole (cf. Fig. 3.7).
3. Pour l'indicatrice d'un ensemble
$$\mathrm{epi}(\delta_C) = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \delta_C(x) \leq t \right\} = C \times \mathbb{R}^+,$$
qui est donc convexe si et seulement si C est convexe.

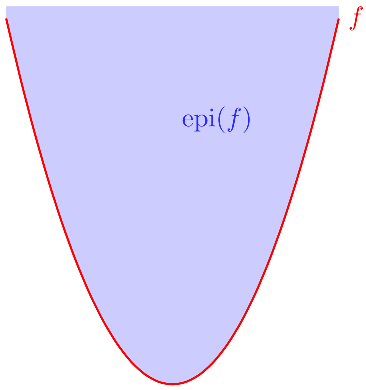


Figure 3.7: L'épigraphe de $\frac{1}{2}\|x\|^2$.

Le résultat suivant montre la préservation de la convexité du supremum de fonctions convexe et est à comparer avec Théorème 21.

Théorème 32. Soient $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ des fonctions convexes avec $i \in I$ (la famille d'indices I est quelconque). Alors la fonction $f(x) = \sup_{i \in I} f_i(x)$ est convexe.

Preuve. Comme les f_i sont convexes, les épigraphes $\mathrm{epi}(f_i)$ sont aussi convexes pour tout $i \in I$. On conclue on remarquant que $\mathrm{epi}(f) = \cap_{i \in I} \mathrm{epi}(f_i)$. $\square$

Définition 39 (Fonction fermée)**.** Une fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est dite fermée si $\mathrm{epi}(f)$ est fermé.

Exemple 15. Revenons à l'exemple de la fonction indicatrice δ_C d'un sous-ensemble C de $\mathbb{R}^n$. Comme $\mathrm{epi}(\delta_C) = C \times \mathbb{R}^+$, on a
$$\delta_C \text{ et fermée} \Leftrightarrow C \times \mathbb{R}^+ \text{ et fermé} \Leftrightarrow C \text{ et fermé.}$$

Définition 40 (Semi-continuité inférieure)**.** Une fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est dite semi-continue inférieurement (sci) en $x \in \mathbb{R}^n$ si
$$f(x) \leq \liminf f(x_n),$$
pour toute suite $(x_n)_n$ telle que $x_n \to x$ quand $n \to \infty$.

Notation. On note par $\Gamma_0(\mathbb{R}^n)$ la classe de fonctions convexes, propres et semi-continuité inférieurement à valeurs dans $\mathbb{R} \cup \{\infty\}$.

Le résultat suivant établit un lien entre la semi-continuité inférieure et la fermeture de son épigraphe et sous-ensembles de niveaux.

Théorème 33. Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. On a

$$f \text{ est sci} \Leftrightarrow f \text{ est fermée} \Leftrightarrow \mathbf{Lev}(f, \alpha) \text{ est fermé } \forall \alpha \in \mathbb{R}.$$

Preuve. Exercice. □

3.5 Maxima de fonctions convexes

Avant de commencer le nouveau chapitre sur l'optimisation convexe; dans lequel on s'intéressera essentiellement à la minimisation de fonctions convexes, on essayera de dégager quelques propriétés des maximas de fonctions convexes sur un convexe.

Théorème 34. Soit $f : C \to \mathbb{R}$ une fonction convexe non constante avec C un convexe. Alors f n'atteint pas son maximum à l'intérieur de C.

Preuve. Supposons qu'il existe $x^* \in \int(C)$ tel que $f(x^*) \geq f(x)$ pour tout $x \in C$. Comme f est non constante, il existe $x_* \in C$ tel que $f(x_*) < f(x^*)$. Comment x^* est un point intérieur à C, il existe $\varepsilon > 0$, suffisamment petit tel que $z := x^* + \varepsilon(x^* - x_*) \in C$. En particulier $x^* = \frac{\varepsilon}{1+\varepsilon} x_* + \frac{1}{1+\varepsilon} z$, et par convexité de f

$$f(x^*) \leq \frac{\varepsilon}{1+\varepsilon} f(x_*) + \frac{1}{1+\varepsilon} f(z),$$

en multipliant des deux cotés de l'inégalité pas $1 + \varepsilon$ et en réarrangeant les termes on obtient

$$f(x^*) < f(x^*) + \varepsilon(\underbrace{f(x^*) - f(x_*)}_{>0}) \leq f(z),$$

cela contredit la maximalité de x^*. □

En renforçant les hypothèses sur C, on obtient qu'au moins un des maxima de f est un point extrémal de C.

Théorème 35. Soit $f : C \to \mathbb{R}$ une fonction convexe continue avec C un convexe compact. Alors il existe au moins un maximum de f qui est un point extrémal de C.

Preuve. Soit x^* un point maximum de f (dont l'existence sera admise pour le moment et sera traitée dans le chapitre suivant). Si $x^* \in \text{ext}(C)$ rien à démontrer. Sinon, par Krein-Milman, écrivons $x = \sum_{i=1}^m \lambda_i x_i$ avec $x_i \in \text{ext}(C), \lambda_i > 0$ pour tout $i = 1, \ldots, met$ $\lambda \in \Delta_m$. Par l'inégalité de Jensen $f(x^*) \leq \sum_{i=1}^m \lambda_i f(x_i)$, et donc

$$\sum_{i=1}^m \underbrace{\lambda_i}_{>0} \underbrace{(f(x_i) - f(x^*))}_{\leq 0} \geq 0,$$

il s'agit donc d'une somme positive ou nulle de quantités négatives ou nulles, et par conséquence $f(x_i) = f(x^*)$ pour tout $i = 1, \ldots, m$, i.e., les points extrémaux $x_1, \ldots, x_m$ sont des maxima de f. □

Exemple 16. Considérons la fonction $f : x \in C \mapsto x^T Ax$ avec $A \in \mathcal{S}_n^+$ et $C = B_\infty(0,1)$. Comme $\mathrm{ext}(C) = \{\pm 1\}^n$ (voir TD), on déduit qu'il existe un maximum de f sur C qui appartient à $\{\pm 1\}^n$, i.e., toutes ses coordonnées sont soit -1 ou 1.

3.6 Inégalités et convexité

On présente ici, comme applications, quelques inégalités qui peuvent se démontrer grâce à la convexité.

Proposition 9 (Inégalité arithmético-géométrique)**.** Pour tous réels $x_1, \ldots, x_n \geq 0$, on a

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \sqrt[n]{\prod_{i=1}^{n} x_i}.$$

Plus généralement, pour tout $\lambda \in \Delta_n$,

$$\sum_{i=1}^{n} \lambda_i x_i \geq \prod_{i=1}^{n} x_i^{\lambda_i}.$$

Proposition 10 (Inégalité de Young)**.** Soient $a, b \geq 0$ et $p, q > 1$ avec $\frac{1}{p} + \frac{1}{q} = 1$. On a

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Le résultat suivant est une généralisation de l'inégalité de Cauchy-Schwarz.

Proposition 11 (Inégalité de Hölder)**.** Pour tous $x, y \in \mathbb{R}^n$ et $p, q > 1$ avec $\frac{1}{p} + \frac{1}{q} = 1$, on a

$$|\langle x, y\rangle| \leq \|x\|_p \|y\|_q.$$

Proposition 12 (Inégalité de Minkowski)**.** Soit $p \geq 1$. Alors pour tous $x, y \in \mathbb{R}^n$, on a

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p.$$

3.7 Exercices

Exercice 16. Montrer que les fonctions suivantes sont convexes:

1. $f_1(x) = \langle a, x\rangle + b$, pour $x \in \mathbb{R}^n$, avec $a \in \mathbb{R}^n$ et $b \in \mathbb{R}$.
2. $f_2(x) = \|x\|$, $x \in \mathbb{R}^n$.
3. $f_3(x) = x^p$, pour $x \in \mathbb{R}^+$ avec $p \geq 1$.
4. $f_4(x) = x\ln(x)$, pour $x > 0$.
5. $f_5(x) = e^{\alpha x}$, $x \in \mathbb{R}$ et $\alpha \in \mathbb{R}$.
6. $f_6(x) = \max\limits_{i=1,\ldots,n} x_i$ avec $x \in \mathbb{R}^n$.
7. $f_7(x, y) = \frac{x^2}{y}$ avec $(x, y) \in \mathbb{R} \times \mathbb{R}_+^*$.

Exercice 17. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction convexe et $a < b$ deux réels.

1. Montrer que
$$f(x) \leq \frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b), \forall x \in [a, b].$$
2. Montrer que
$$\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x}, \ \forall x \in (a, b).$$
3. Supposons que f est différentiable. En utilisant ce qui précède, montrer que
$$f'(a) \leq \frac{f(b)-f(a)}{b-a} \leq f'(b).$$
En déduire que si f est deux fois différentiable, on a $f''(b) \ \geq 0$ et $f''(a) \geq 0$.

Exercice 18. Montrer qu'une fonction $f : \mathbb{R} \to \mathbb{R}$ est convex si et seulement si
$$\Delta := \begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ f(x) & f(y) & f(z) \end{vmatrix} \geq 0$$
pour tout x, y, z tels que $x < y < z$.

Exercice 19. Montrer qu'une fonction continue $f : \mathbb{R}^n \to \mathbb{R}$ est convexe ssi pour tout $x, y \in \mathbb{R}^n$
$$\int_0^1 f(x + t(y-x))dt \leq \frac{f(x)+f(y)}{2}. \tag{3.15}$$

Exercice 20. Soient $x_1, \ldots, x_n \geq 0$ et $\lambda \in \Delta_n$. Montrer que
$$\sum_{i=1}^{n} \lambda_i x_i \geq \Pi_{i=1}^n x_i^{\lambda_i}.$$

En déduire que

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \sqrt[n]{\Pi_{i=1}^{n} x_i}.$$

Exercice 21. Soient $a, b \geq 0$ et $p, q > 1$ avec $\frac{1}{p} + \frac{1}{q} = 1$. Montrer que

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Exercice 22. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction convexe différentiable. On définit, pour $x > 0$

$$F(x) = \frac{1}{x}\int_0^x f(t)dt.$$

Montrer que F est convexe.

Exercice 23. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe bornée supérieurement. Montrer que f est constante.

Exercice 24. Soit $f : C \to \mathbb{R}$ avec $C \subset \mathbb{R}$ convexe. Montrer que f est quasi-convexe ssi

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\},\ \forall x, y \in C, \lambda \in [0,1].$$

Exercice 25. Soit $C \subset \mathbb{R}^n$ un convexe et $\mu > 0$. On rappelle qu'une fonction $f : C \to \mathbb{R}$ est dite $\mu-$fortement convexe (ou fortement convexe de module μ) si $g(x) := f(x) - \frac{\mu}{2}\|x\|^2$ est convexe sur C.

1. Montrer que f est $\mu-$fortement convexe sur C ssi
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2,\ \forall x, y \in C \text{ et } \lambda \in [0,1].$$
2. Montrer si f est fortement convexe alors f est strictement convexe.
3. Supposons que f est C^1. Montre que f est $\mu-$fortement convexe sur C ssi
$$f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle + \frac{\mu}{2}\|x-y\|^2,\ \forall x, y \in C.$$
4. Supposons que f est C^1. Montre que f est $\mu-$fortement convexe sur C ssi
$$\langle \nabla f(x) - \nabla f(y), x-y\rangle \geq \mu\|x-y\|^2,\ \forall x, y \in C.$$
5. Supposons que f est C^2. Montre que f est $\mu-$fortement convexe sur C ssi
$$\nabla^2 f(x) \succeq \mu \mathrm{I},\ \forall x \in C.$$
6. Soit $h = f + g$ avec f fortement convexe et g convexe. Montrer que h est fortement convexe.
7. Montrer que $f(x) = \sqrt{1 + \|x\|^2}$ est strictement convexe mais pas fortement convexe.
8. Soit $f(x) = x^T Ax + 2b^T x + c$ avec $A \in \mathcal{S}^n, b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Montrer que f est μ-fortement convexe ssi $A \in \mathcal{S}^n_{++}$. Dans ce cas le module de forte convexité est $\mu = 2\lambda_{\min}(A)$.

Exercice 26. On rappelle qu'une application $F : \mathbb{R}^n \to \mathbb{R}^n$ est monotone si

$$\langle F(x) - F(y), x - y \rangle \geq 0, \ \forall x, y \in \mathbb{R}^n.$$

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe. Montrer que ∇f est monotone. Peut-on dire que toute application monotone est le gradient d'une fonction convexe ?

Chapitre 4.

Introduction à l'optimisation convexe

4.1 Conditions d'optimalité pour la minimisation sans contraintes

Dans cette section, nous introduisons quelques définitions pour 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 $\mathbb{R}^n$.

4.1.1 Minimum/Maximum global

Définition 41 (Minimum/Maximum global)**.** Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ 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 42. On note par $\operatorname{argmin}_S f$ l'ensemble des minimas globaux de f sur S. De même, on note par $\operatorname{argmax}_S f$ l'ensemble des maximas globaux de f sur S.

Exemple 17. 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 $\operatorname{argmax}_S f = \{(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})\}$. On verra plus loin (ou l'année prochaine :)) comment trouver résoudre ce type de problèmes grâce aux conditions KKT.

Définition 43.

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 $x^* \neq x \in S \cap B(x^*, \varepsilon)$

4.1.2 Condition d'optimalité du premier ordre

On rappelle le résultat suivant, souvent connu sous la **règle de Fermat** (c.f Fig. 4.1).

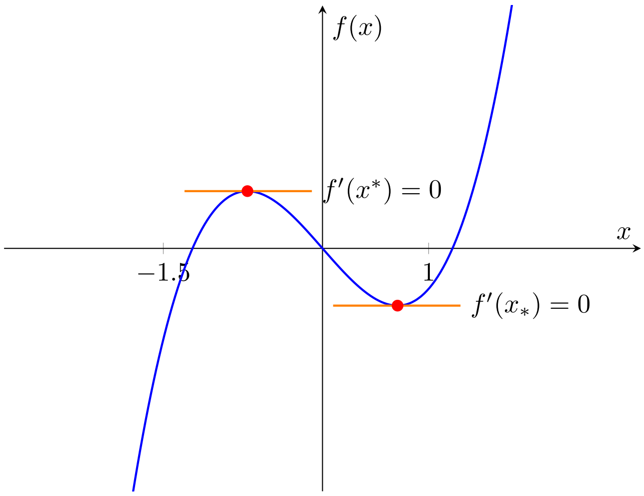


Figure 4.1: Illustration de la règle de Fermat: la fonction f a un minimum (respectivement maximum) local en x_* (respectivement x^*). La dérivée de f s'annule en ces deux points. Géométriquement, la tangente de f en ces deux points est horizontale.

Proposition 13 (Règle de Fermat). Soit $f : (a, b) \subset \mathbb{R} \to \mathbb{R}$ 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 :

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

avec $\varepsilon(h) \to 0$ quand $h \to 0$. 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 $h \to 0^+$ puis $h \to 0^-$, on obtient $f'(x^*) \geq 0$ et $f'(x^*) \leq 0$, d'où $f'(x^*) = 0$. □

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

Théorème 36. Soit $f : S \to \mathbb{R}$ une fonction définie sur un ensemble $S \subset \mathbb{R}^n$. Supposons que $x^* \in \mathrm{int}(S)$ est un extremum local et que toutes les dérivées partielles de f en x^* existent. Alors $\nabla f(x^*) = 0$.

Preuve. On définit, pour $i \in [\![1, n]\!]$, 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 ϕ est différentiable en 0 et $\phi^{'}(0) = \partial_{x_i} f(x^*)$. Comme x^* est un extremum local de f, il vient que $t = 0$ est un extremum local de ϕ, et par la suite $\phi^{'}(0) = 0$. Il s'en suit que $\partial_{x_i} f(x^*) = 0$, pour tout $i \in [\![1, n]\!]$, i.e., $\nabla f(x^*) = 0$. □

Le Théorème 36 est donne 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 ses points est nul. Cela motive la définition suivante.

Définition 44 (Point critique). Soit $f : S \to \mathbb{R}$ une fonction définie sur un ensemble $S \subset \mathbb{R}^n$. Supposons que $x^* \in \operatorname{int}(S)$ 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 19. Le Théorème 36 affirme que les extrema locaux sont des points critiques.

Exemple 18. 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.

4.1.3 Conditions d'optimalité du second ordre

Théorème 37 (Conditions nécessaires). Soit $f : S \to \mathbb{R}$ 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 38 (Condition suffisante). Soit $f : S \to \mathbb{R}$ une fonction définie sur un ouvert $S \subset \mathbb{R}^n$. Supposons que f est deux fois différentiable sur S et soit $x^* \in \operatorname{int}(S)$ 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 $\nabla^2 f(x^*) \succeq \lambda_{\min}\mathrm{Id}$. D'autre part, comme f est deux fois différentiable et $\nabla f(x^*) = 0$, on a par **??**

$$\begin{aligned} f(x) &= f(x^*) + \frac{1}{2}\langle \nabla^2 f(x^*)(x - x^*), x - x^*\rangle + o(\|x - x^*\|^2) \\ &\geq f(x^*) + \frac{1}{2}\lambda_{\min}\|x - x^*\|^2 + o(\|x - x^*\|^2) \end{aligned} \tag{4.1}$$

Comme $o(r)/r \to 0$, 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

$$\begin{aligned} f(x) &= f(x^*) + \frac{1}{2}\langle \nabla^2 f(x^*)(x - x^*), x - x^*\rangle + o(\|x - x^*\|^2) \\ &\geq f(x^*) + \frac{1}{2}\lambda_{\min}\|x - x^*\|^2 + o(\|x - x^*\|^2) \\ &\geq f(x^*) + \frac{1}{4}\lambda_{\min}\|x - x^*\|^2 > f(x^*). \end{aligned} \tag{4.2}$$

□

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

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

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 39 (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 18 et Fig. 4.2.

Définition 45 (Point selle)**.** Soit $f : S \to \mathbb{R}$ 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 40 (Condition suffisante pour un point selle)**.** Soit $f : S \to \mathbb{R}$ 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. □

4.1.4 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

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

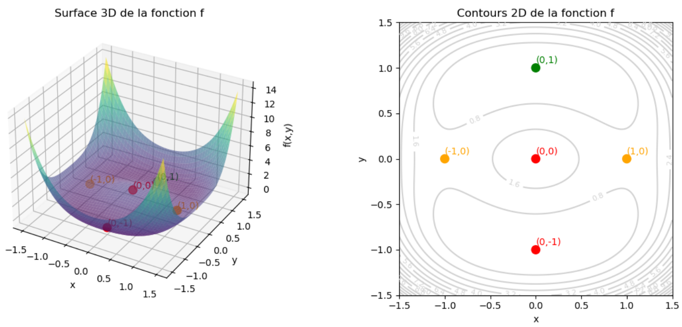


Figure 4.2: Points critiques de la fonction $f(x,y) = (x^2+y^2-1)^2 + (y^2-1)^2$. Le point $(0,0)$ est un maximum local, les points $(0,1)$ et $(0,-1)$ sont des minimas locaux stricts et aussi des minimas globaux tandis que $(1,0)$ et $(-1,0)$ sont des points-selles.

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

Théorème 41 (Weirestrass)**.** Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ 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 coercitivité de f.

Définition 46 (Coercivité)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue. On dit que f est coercive si

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

Théorème 42. Soit $f : \mathbb{R}^n \to \mathbb{R}$ 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 41 assure l'existence d'un minimum global de f sur $S \cap B(0,K)$, et donc sur S. □

En combinant la coercitivité (Définition 46) et la semi-continuité inférieure (Définition 40) on peut étendre le résultat de Weierstrass:

Théorème 43 (Existence d'un minimum)**.** Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction **propre**, **semi-continue inférieurement** et **coercive**. Alors f admet un minimiseur, c'est-à-dire qu'il existe $x^* \in \mathbb{R}^n$ tel que

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

4.2 Optimisation convexe

4.2.1 Définitions et premières propriétés

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

Définition 47 (Problème convexe). Le problème d'optimisation

$$\min_{x \in C} f(x) \tag{4.3}$$

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 $g_i : \mathbb{R}^n \to \mathbb{R}$ sont des fonctions **affines**, et $h_j : \mathbb{R}^n \to \mathbb{R}$ 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 44. Soit $f : C \to \mathbb{R}$ 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 $\|x^* - x\| \leq r$. Pour montre 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 sent 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 45. Soit $f : C \to \mathbb{R}$ 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 stricte de f sur C.

Notation. On note par $\underset{x \in C}{\operatorname{argmin}}\, f(x)$ (des fois $\mathcal{S}$) l'ensemble des minimiseurs de f, i.e., les solutions de (4.3).

Théorème 46. Soit $f : C \to \mathbb{R}$ une fonction convexe, alors $\mathcal{S}$ est convexe. Si de f est strictement convexe, alors $\mathcal{S}$ contient au plus un élément.

Preuve. Si $\mathcal{S} = \emptyset$, rient 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 20. Pour la convexité de $\mathcal{S}$, on peut remarquer que $\mathcal{S} = \{x : f(x) \leq f^*\} \cap C = \mathbf{Lev}(f, f^*) \cap C$ qui est convexe comme intersection de convexes.

4.2.2 Stationnarité

Nous avons introduit, dans la Définition 44, 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

$$\min_{x \in C} f(x), \tag{4.4}$$

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

Définition 48 (Point stationnaire pour (4.4))**.** 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 (4.4) si

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

Remarque 21. 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

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

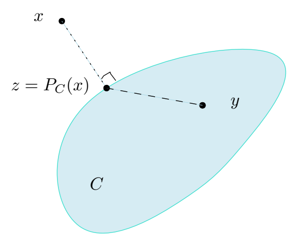
ce qui impose $\nabla f(x^*) = 0$. Autrement dit, on retrouve la condition de stationnarité classique du cas non contraint, introduite en Définition 44.

La notion de stationnarité introduite en Définition 48 constitue en fait une **condition nécessaire d'optimalité** pour le problème (4.4).

Théorème 47 (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 (4.4). Alors x^* est un point stationnaire de (4.4).

De plus, si f est **convexe**, alors x^* est un point stationnaire de (4.4) si et seulement si x^* est une solution optimale du problème.

Preuve. Exercice. $\square$



(a)

Figure 4.3: Projection orthogonale.

4.3 Projection orthogonale

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

Définition 49. L'opérateur de projection orthogonale sur C est l'application $P_C : \mathbb{R}^n \to C$ qui à $x \in \mathbb{R}^n$ associe

$$P_C(x) = \underset{y \in C}{\operatorname{argmin}} \; \|y - x\|^2 \tag{4.5}$$

On a le résultat

Théorème 48 (Premier théorème de la projection)**.** Soit C un ensemble convexe, fermé et non vide. Alors le problème (4.5) 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 42 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 46. $\square$

Grâce au Théorème 47, on obtient la caractérisation géométrique suivante de la projection.

Théorème 49 (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 47 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 49 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. 4.3.

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

Théorème 50. 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 $x, y \in \mathbb{R}^n$,

$$\langle P_C(x) - P_C(y),\, x - y\rangle \geq \|P_C(x) - P_C(y)\|^2. \tag{4.6}$$

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 49, on a

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

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

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

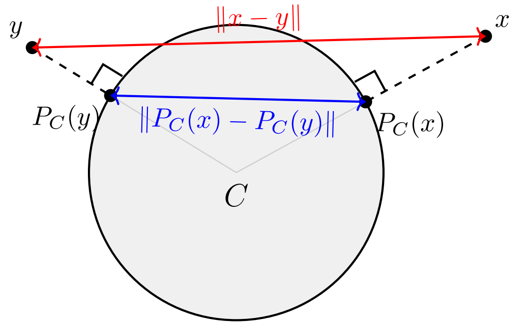


Figure 4.4: L'opérateur P_C est non-expansive.

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

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

$\square$

4.4 Exercices

Exercice 27. Soit $C \subset \mathbb{R}^n$ un compact non vide et $f : C \to \mathbb{R}$ une fonction continue. Montrer qu'il existe $x^* \in C$ tel que $f(x^*) = \min_{x \in C} f(x)$.

Exercice 28. Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ on rappelle que f est dite propre si $-\infty \notin f(\mathbb{R}^n)$ et $\mathrm{dom}(f) \neq \emptyset$. On dira que f est coercive si $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$. Finalement, f est dite semi-continue inférieurement (sci) ssi pour toute suite $(x_n)_n$ convergeant vers x on a $f(x) \leq \liminf_{n \to \infty} f(x_n)$.

1. Montrer que f est coercive si et seulement si $\mathbf{Lev}(f, \alpha)$ est borné pour tout $\alpha \in \mathbb{R}$.
2. Supposons que $f : \mathbb{R}^n \to (-\infty, +\infty]$ est propre, semi-continue inférieurement et coercive. Monter qu'il existe $x^* \in \mathbb{R}^n$ tel que $f(x^*) = \min_{x \in \mathbb{R}^n} f(x)$.
3. Supposons que f est continue et coercive et soit $S \subset \mathbb{R}^n$ un fermé non vide. Montrer que $x^* \in S$ tel que $f(x^*) = \min_{x \in S} f(x)$.
4. Montrer que $f(x, y) = x^2 + y^2$ admet un minimum global sur $C = \{(x, y) \in \mathbb{R}^2 : \ x + y \leq -1\}$.

Exercice 29. Trouver et déterminer la nature des points critique des fonctions $f : \mathbb{R}^d \to \mathbb{R}$, $d = 2, 3$.

1. $f(x, y) = 2x^3 + 3y^2 + 3x^2y - 24y$.
2. $f(x, y) = f(x, y) = x^2 + y^2 - 2x - 4y$.
3. $f(x, y) = f(x, y) = x^2 - y^2$.
4. $f(x, y) = x^3 + y^3 - 3x - 3y$.
5. $f(x, y, z) = -3z^2 + 2y^2 + 2xz + 2y + 1$.

Exercice 30. Soit $n \geqslant 2$ et $p : \mathbb{R}^n \to \mathbb{R}$ la fonction polynomiale définie par

$$x = (x_1, \ldots, x_n) \in \mathbb{R}^n \longmapsto p(x) := (1 + x_n)^3 \sum_{i=1}^{n-1} x_i^2 + x_n^2.$$

1. Montrer que $0_{\mathbb{R}^n}$ est le seul point critique de p.
2. Montrer que $0_{\mathbb{R}^n}$ est minimum local strict de p, mais qu'il n'est pas minimum global de p.

Exercice 31. Soit $f(x) = x^T Ax + 2b^T x + c$ avec $A \in \mathcal{S}^n, b \in \mathbb{R}^n, c \in \mathbb{R}$.

1. Montrer que x^* est un point stationnaire si et seulement si $Ax^* = -b$.
2. Si $A \succeq 0$, alors x^* est un minimum global si et seulement si $A^* x = -b$.
3. Si $A \succ 0$, alors $x^* = -A^{-1}b$ est un minium global strict de f.
4. Montre que f est coercive ssi $A \succ 0$.
5. Montrer que

$$f(x) \geq 0, \ \forall x \in \mathbb{R}^n \Leftrightarrow \begin{pmatrix} A & b \\ b^\top & c \end{pmatrix} \succeq 0.$$

Exercice 32. Soit $g : \mathbb{R}^n \to \mathbb{R}$ convexe et différentiable sur $\mathbb{R}^n$, et $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexe, fini en au moins un point. On pose

$$f(x) := (g+h)(x) = g(x) + h(x), \qquad x \in \mathbb{R}^n,$$

et on considère le problème $\min_{x\in\mathbb{R}^n} f(x)$.

1. Montrer que $\bar{x}$ minimise f sur $\mathbb{R}^n$ si et seulement si

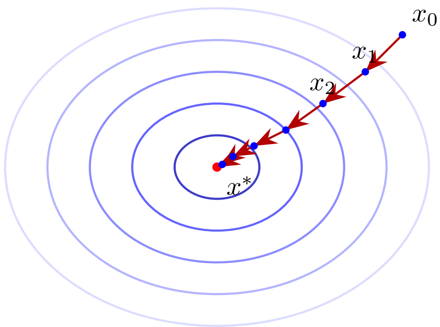
$$\langle \nabla g(\bar{x}), x - \bar{x}\rangle + h(x) - h(\bar{x}) \geq 0 \quad \text{pour tout } x \in \mathbb{R}^n. \tag{E}$$

2. Vérifier que (E) est équivalente à

$$\langle \nabla g(x), x - \bar{x}\rangle + h(x) - h(\bar{x}) \geq 0 \quad \text{pour tout } x \in \mathbb{R}^n. \tag{F}$$

Chapitre 5.

Algorithmes d'optimisation



Descente de gradient
Minimisation de $f(x)$ par itérations successives

Considérons un problème de la forme

$$\min_{x\in\mathbb{R}^n} f(x), \tag{P}$$

avec $f : \mathbb{R}^n \to \mathbb{R}$. L'objectif est de construire une suite $(x_k)_k$ telle que $f(x_k) \to f^* := \min_{\mathbb{R}^n} f$ quand $k \to \infty$ et, si possible, $x_k \to x^* \in \operatorname{argmin}_{\mathbb{R}^n} f$ dans un sens à préciser. Une façon de générer une telle suite est de considérer

$$x_{k+1} = x_k + t_k\mathbf{d}_k,$$

où $\mathbf{d}_k \in \mathbb{R}^n$ est un vecteur de $\mathbb{R}^n$ et $t_k \geq 0$ est un pas de déplacement. Dans un premier temps, on s'intéressera donc à $(t_k, \mathbf{d}_k)$.

5.1 Directions de descente

Définition 50. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^1. On dit qu'un vecteur non nul $\mathbf{d}$ de $\mathbb{R}^n$ est une direction de descente de f en x si

$$f^{'}(x;\mathbf{d}) = \nabla f(x)^T\mathbf{d} < 0.$$

Il se trouve qu'en suivant des directions de descente avec un pas suffisamment petit, on arrive à décroître la fonction objectif. On a le résultat suivant :

Lemme 3. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^1 et $\mathbf{d}$ une direction de descente de f en $x \in \mathbb{R}^n$. Alors, il existe $\eta > 0$ tel que

$$f(x + t\mathbf{d}) < f(x),$$

pour tout $t \in (0, \eta]$.

Preuve. Soit $x \in \mathbb{R}^n$ et $\mathbf{d}$ une direction de descente en x. On a $f(x+t\mathbf{d}) = f(x)+t\langle\nabla f(x), \mathbf{d}\rangle + o(t\|\mathbf{d}\|)$. On sait que pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que $\frac{|o(t\|\mathbf{d}\|)|}{t} \leq \varepsilon$ pour tout $t \in (0, \eta]$. En prenant $\varepsilon = -\langle\nabla f(x), \mathbf{d}\rangle$ qui est strictement positif puisque $\mathbf{d}$ est une direction de descente en x, on obtient que

$$o(t\|\mathbf{d}\|)/t \leq |o(t\|\mathbf{d}\|)|/t \leq -\langle\nabla f(x), \mathbf{d}\rangle,$$

donc $f(x + t\mathbf{d}) < f(x), \forall t \in (0, \eta]$. □

D'une façon générale, une méthode de descente s'écrit comme suit :

Méthode de descente

- Initialisation : On se donne un $x_0 \in \mathbb{R}^n$.
- Pour $k \geq 0$
  - Choisir une direction de descente $\mathbf{d}_k$.
  - Choisir un pas t_k tel que : $f(x_k + t_k\mathbf{d}_k) < f(x_k)$.
  - Prendre $x_{k+1} = x_k + t_k\mathbf{d}_k$.
  - Si un critère d'arrêt est vérifié, renvoyer x_{k+1}.

Remarque 22.

- Le choix de la direction de descente donne lieu à des méthodes différentes de descente (e.g., gradient, gradient conjugué, Newton, quasi-Newton etc).
- Il existe plusieurs critères d'arrêt ; le plus utilisé est $\|\nabla f(x_{k+1})\| \leq \varepsilon$ pour une tolérance $\varepsilon > 0$ donnée.
- Le procédé permettant de trouver le pas t_k s'appelle recherche linéaire (line search en anglais). L'appellation vient du fait que trouver un tel t_k revient à minimiser la restriction de f sur la ligne $x_k + t\mathbf{d}_k$, i.e., la fonction $g(t) = f(x_k + t\mathbf{d}_k)$. On discute ici quelques choix populaires.

  1. **Pas constant** : dans ce cas $t_k = \bar{t}$ pour tout k. Bien que ce soit un choix simple, un pas trop petit peut donner lieu à une convergence lente de l'algorithme tandis qu'avec un pas trop grand on peut perdre la décroissance de la fonction objectif (cf. TP).
  2. **Recherche linéaire exacte** : dans ce cas on cherche à minimiser exactement la fonction $g(t)$ définie ci-dessus. On choisit donc $t_k \in \operatorname{argmin}_{t\geq 0} f(x_k + t\mathbf{d}_k)$. Il se trouve qu'en général, on ne peut pas trouver le minimiseur exact de g.
  3. **Rebroussement (ou Backtracking)** : cette approche permet d'éviter une recherche linéaire exacte en choisissant convenablement un pas assurant la décroissance de la fonction objectif. L'approche requiert trois paramètres $s > 0$ et $\alpha, \beta \in (0, 1)$. Dans un premier temps on prend $t_k = s$ et tant que

     $$f(x_k) - f(x_k + t_k\mathbf{d}_k) < -\alpha t_k \nabla f(x_k)^T \mathbf{d}_k,$$

     (i.e., $f(x_{k+1}) > f(x_k)$) on met à jour t_k par $t_k\beta$. Dans ce cas, le pas choisi serait $t_k = s\beta^{i_k}$ où i_k est le plus petit entier tel que

     $$f(x_k) - f(x_k + t_k\mathbf{d}_k) \geq -\alpha s\beta^{i_k} \nabla f(x_k)^T \mathbf{d}_k. \tag{5.1}$$

     Comme $\mathbf{d}_k$ est une direction de descente de f en x_k, l'inégalité (5.1) assure la décroissance de la suite $(f(x_k))_k$.

Il s'avère que (5.1) est toujours vérifiée pour des pas t_k suffisamment petits.

Lemme 4. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^1 et $\mathbf{d}$ une direction de descente de f en $x \in \mathbb{R}^n$, et soit $\alpha \in (0, 1)$. Alors, il existe $\eta > 0$ tel que

$$-\alpha t\langle \nabla f(x), \mathbf{d}\rangle < f(x) - f(x + t\mathbf{d}),$$

pour tout $t \in (0, \eta]$.

Preuve. Soit $x \in \mathbb{R}^n$ et $\mathbf{d}$ une direction de descente en x. On a $f(x+t\mathbf{d}) = f(x)+t\langle\nabla f(x),\mathbf{d}\rangle + o(t\|\mathbf{d}\|)$. Par définition :

$$\forall\varepsilon > 0, \exists\eta > 0 : \frac{|o(t\|\mathbf{d}\|)|}{t} \leq \varepsilon, \forall t \in (0,\eta].$$

En prenant $\varepsilon = -(\alpha-1)\langle\nabla f(x),\mathbf{d}\rangle$ qui est strictement positif puisque $\mathbf{d}$ est une direction de descente en x, on obtient que

$$o(t\|\mathbf{d}\|)/t \leq |o(t\|\mathbf{d}\|)|/t \leq (\alpha-1)\langle\nabla f(x),\mathbf{d}\rangle,$$

soit

$$o(t\|\mathbf{d}\|) + t\langle\nabla f(x),\mathbf{d}\rangle < \alpha t\langle\nabla f(x),\mathbf{d}\rangle$$

donc $-\alpha t\langle\nabla f(x),\mathbf{d}\rangle < f(x) - f(x+t\mathbf{d}), \forall t \in (0,\eta]$. □

L'exemple suivant présente le cas d'une recherche linéaire exacte pour une fonction quadratique.

Exemple 19. On considère $f(x) = x^TAx + 2b^Tx + c$ avec $A \in \mathcal{S}_n^{++}, b \in \mathbb{R}^n, c \in \mathbb{R}$. Soit $x \in \mathbb{R}^n$ et $\mathbf{d} \in \mathbb{R}^n \setminus \{0\}$ une direction de descente de f en x. Effectuer une recherche linéaire exacte revient à résoudre

$$\min_{t\geq 0} f(x+t\mathbf{d}) := g(t).$$

On a $g'(t) = \mathbf{d}^T\nabla f(x+t\mathbf{d})$ et comme $\nabla f(x) = 2(Ax+b)$, on trouve que $g'(t) = 2(\mathbf{d}^TA\mathbf{d})t + 2\mathbf{d}^T\nabla f(x)$. Donc

$$g^{'}(t) = 0 \Leftrightarrow t = -\frac{\mathbf{d}^T\nabla f(x)}{2\mathbf{d}^TA\mathbf{d}} > 0.$$

Comme $g^{''}(t) = 2(\mathbf{d}^TA\mathbf{d}) > 0$ alors le pas donné par la méthode de recherche linéaire exacte est $t^* = -\frac{\mathbf{d}^T\nabla f(x)}{2\mathbf{d}^TA\mathbf{d}}$.

5.2 Descente de gradient

Comme son nom l'indique, le choix de la direction de descente dans la méthode du gradient est $\mathbf{d}_k = -\nabla f(x_k)$. Ce choix est bien une direction de descente au sens de la Définition 50. En effet, pour $\nabla f(x_k) \neq 0$:

$$f^{'}(x_k; -\nabla f(x_k)) = -\nabla f(x_k)^T\nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0.$$

Au-delà du fait que $\mathbf{d} = -\nabla f(x_k)$ est une direction de descente, le vecteur unitaire $\mathbf{d}^* = \mathbf{d}/\|\mathbf{d}\|$ est celui de la plus forte pente, i.e., il minimise la dérivée directionnelle $f'(x_k;\mathbf{d})$ parmi tous les vecteurs unitaires $\mathbf{d}$. Plus formellement, on a le résultat suivant.

Proposition 14. Soit $f \in C^1(\mathbb{R}^n)$ et $x \in \mathbb{R}^n$ tel que $\nabla f(x) \neq 0$. Alors une solution du problème

$$\min_{\mathbf{d}\in\mathbb{R}^n:\ \|\mathbf{d}\|=1} f'(x;\mathbf{d}), \tag{5.2}$$

est $\mathbf{d}^* = -\nabla f(x)/\|\nabla f(x)\|$.

Preuve. Le problème (5.2) s'écrit aussi

$$\min_{\mathbf{d}:\|\mathbf{d}\|=1} \langle\nabla f(x),\mathbf{d}\rangle,$$

et grâce à l'inégalité de Cauchy-Schwarz, on a

$$\langle\nabla f(x),\mathbf{d}\rangle \geq -\|\mathbf{d}\|\|\nabla f(x)\| = -\|\nabla f(x)\|,$$

donc $-\|\nabla f(x)\|$ est une borne inférieure de la valeur optimale de (5.2). Cette valeur est atteinte en particulier en $\mathbf{d}^*$. En effet

$$f'(x;\mathbf{d}^*) = \langle \nabla f(x), \frac{-\nabla f(x)}{\|\nabla f(x)\|}\rangle = -\|\nabla f(x)\|,$$

d'où le résultat. □

On présente ci-dessous les étapes de la méthode, avec comme critère d'arrêt $\|\nabla f(x_{k+1})\| \leq \varepsilon$ pour une tolérance ε donnée.

Méthode du gradient

- Initialisation : On se donne un $x_0 \in \mathbb{R}^n$.
- Pour $k \geq 0$
  - Choisir un pas t_k par recherche linéaire sur $g(t) = f(x_k - t\nabla f(x_k))$.
  - Prendre $x_{k+1} = x_k - t_k \nabla f(x_k)$.
  - Si $\|\nabla f(x_{k+1})\| \leq \varepsilon$, s'arrêter et renvoyer x_{k+1}.

5.3 Analyse de convergence

Avant de commencer notre analyse, rappelons qu'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dans $C_L^{1,1}(\mathbb{R}^n)$ si elle est de classe C^1 et que son gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ est L-Lipschitz, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \text{ pour tout } x, y \in \mathbb{R}^n.$$

Exemple 20.
- On vérifie facilement que $f(x) = a^T x + b \in C_0^{1,1}(\mathbb{R}^n)$.
- Une fonction quadratique $f(x) = x^T Ax + 2b^T x + c$ avec $A \in \mathcal{S}_n, b \in \mathbb{R}^n$ et $c \in \mathbb{R}$ est $2\|A\|$-Lipschitz. En effet, comme $\nabla f(x) = 2Ax + 2b$ on a

$$\|\nabla f(x) - \nabla f(y)\| = 2\|(Ax + b) - (Ay + b)\| \leq 2\|A\|\|x - y\|.$$

Si la fonction est C^2, alors la régularité Lipschitz du gradient est équivalente au fait que la Hessienne est bornée :

Théorème 51. Soit $f \in C^2(\mathbb{R}^n)$. Alors

$$f \in C_L^{1,1}(\mathbb{R}^n) \Leftrightarrow \|\nabla^2 f(x)\| \leq L \text{ pour tout } x \in \mathbb{R}^n.$$

Le résultat suivant joue un rôle important dans la suite.

Lemme 5 (Lemme de descente). Soit $f \in C_L^{1,1}(\mathbb{R}^n)$. Alors pour tout $x, y \in \mathbb{R}^n$

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|x - y\|^2.$$

Preuve. Soient $x, y \in \mathbb{R}^n$. Par le théorème fondamental de l'analyse

$$f(y) = f(x) + \int_0^1 \langle \nabla f(x + t(y - x)), y - x\rangle \mathbf{d}t,$$

donc

$$f(y) - f(x) = \langle \nabla f(x), y - x\rangle + \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle \mathbf{d}t.$$

Il s'ensuit que

$$\begin{aligned}\Big|f(y)-f(x)-\langle\nabla f(x),y-x\rangle\Big| &= \Big|\int_0^1 \langle\nabla f(x+t(y-x))-\nabla f(x),y-x\rangle \mathbf{d}t\Big|,\\ &\le \int_0^1 \Big|\langle\nabla f(x+t(y-x))-\nabla f(x),y-x\rangle\Big|\mathbf{d}t,\\ &\overset{\text{C.S}}{\le} \int_0^1 \|\nabla f(x+t(y-x))-\nabla f(x)\|\|y-x\|\mathbf{d}t,\\ &\overset{f\in C_L^{1,1}}{\le} \int_0^1 Lt\|y-x\|^2\mathbf{d}t = \frac{L}{2}\|x-y\|^2.\end{aligned} \tag{5.3}$$

□

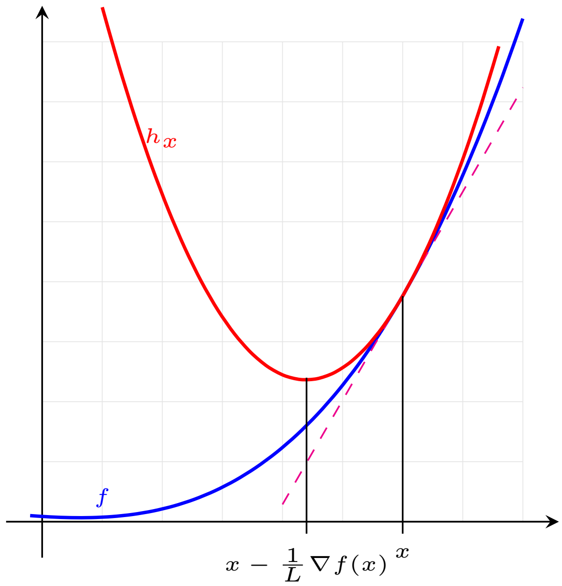


Figure 5.1: La fonction quadratique h_x majore f et la touche en x.

Fixons $x \in \mathbb{R}^n$ et considérons $h_x(y) = f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|x-y\|^2$ qui est quadratique en y et majore f : $f(y) \le h_x(y)$ pour tout $y \in \mathbb{R}^n$. De plus, le minimum de h_x est atteint en $x^* = x - \frac{1}{L}\nabla f(x)$ (voir Fig. 5.1).

Comme conséquence du Lemme 5, on a

Lemme 6. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$. Alors pour tout $x \in \mathbb{R}^n$ et $t > 0$

$$f(x) - f(x - t\nabla f(x)) \ge t\left(1 - \frac{tL}{2}\right)\|\nabla f(x)\|^2. \tag{5.4}$$

Preuve. On applique le Lemme de descente avec $y = x - t\nabla f(x)$ pour un $x \in \mathbb{R}^n$. On a donc

$$f(x - t\nabla f(x)) \le f(x) - t\|\nabla f(x)\|^2 + \frac{Lt^2}{2}\|\nabla f(x)\|^2,$$

d'où le résultat en réarrangeant. □

Si nous optons pour une stratégie avec un pas constant, i.e., $t_k = t^* \in (0, 2/L)$ pour tout k, on a alors d'après le Lemme 6

$$f(x_k) - f(x_{k+1}) \ge t^*\left(1 - \frac{t^*L}{2}\right)\|\nabla f(x_k)\|^2.$$

La fonction $t \mapsto t\left(1 - \frac{tL}{2}\right)$ sur $(0, 2/L)$ atteint un maximum en $t^* = 1/L$. Pour ce choix on a $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$ et

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Pour une recherche linéaire exacte, i.e., $x_{k+1} = x_k - t_k \nabla f(x_k)$ avec

$$t_k \in \operatorname*{argmin}_{t\geq 0} f\Big(x_k - t\nabla f(x_k)\Big),$$

on remarque que $f(x_k - t_k\nabla f(x_k)) \leq f\Big(x_k - \frac{1}{L}\nabla f(x_k)\Big)$ soit

$$f(x_k) - f(x_{k+1}) \geq f(x_k) - f\Big(x_k - \frac{1}{L}\nabla f(x_k)\Big) \geq \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Dans le cadre de backtracking, on cherche pour $\alpha \in (0,1)$, un pas t_k suffisamment petit tel que

$$f(x_k) - f(x_{k+1}) \geq \alpha t_k \|\nabla f(x_k)\|^2.$$

Dans ce cas deux options se présentent : prendre $t_k = s$ où $s > 0$ est la valeur initiale du pas, soit prendre t_k avec la méthode de backtracking comme décrit dans la Remarque 22, et dans ce cas, le choix t_k/β ne serait pas admissible, i.e.,

$$f(x_k) - f\Big(x_k - \frac{t_k}{\beta}\nabla f(x_k)\Big) < \frac{\alpha t_k}{\beta}\|\nabla f(x_k)\|^2. \tag{5.5}$$

En appliquant (5.4) avec $x = x_k$ et $t = \frac{t_k}{\beta}$ on obtient

$$f(x_k) - f\Big(x_k - \frac{t_k}{\beta}\nabla f(x_k)\Big) \geq \frac{t_k}{\beta}\left(1 - \frac{t_k L}{2\beta}\right)\|\nabla f(x_k)\|^2. \tag{5.6}$$

Les équations (5.5)–(5.6) impliquent que $\frac{t_k}{\beta}\left(1 - \frac{t_k L}{2\beta}\right) < \frac{\alpha t_k}{\beta}$, soit $t_k > \frac{2(\alpha-1)\beta}{L}$. Finalement, pour la méthode de backtracking, on a

$$f(x_k) - f(x_k - t_k\nabla f(x_k)) \geq \alpha \min\left(s, \frac{2(\alpha-1)\beta}{L}\right)\|\nabla f(x_k)\|^2.$$

On récapitule cette discussion dans le résultat suivant :

Proposition 15. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$ et $(x_k)_k$ la suite générée par la méthode du gradient pour résoudre

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

avec pas constant, recherche linéaire exacte ou backtracking de paramètres $s > 0, \alpha, \beta \in (0,1)$. Alors

$$f(x_k) - f(x_{k+1}) \geq M\|\nabla f(x_k)\|^2, \tag{5.7}$$

avec

$$M = \begin{cases} t^*\left(1 - \frac{t^* L}{2}\right) & \text{pas constant,} \\ \frac{1}{2L} & \text{recherche linéaire exacte,} \\ \alpha \min\left(s, \frac{2(\alpha-1)\beta}{L}\right) & \text{backtracking.} \end{cases}$$

On démontre maintenant que $\|\nabla f(x_k)\| \to 0$ quand $k \to \infty$.

Théorème 52. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$ et $(x_k)_k$ la suite générée par la méthode du gradient pour résoudre

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

avec pas constant, recherche linéaire exacte ou backtracking de paramètres $s > 0, \alpha, \beta \in (0,1)$. Supposons que f est minorée, i.e., il existe $c \in \mathbb{R}$ tel que $f(x) \geq c$ pour tout $x \in \mathbb{R}^n$. Alors

1. La suite $(f(x_k))_k$ est décroissante, avec $f(x_{k+1}) < f(x_k)$ sauf si $\|\nabla f(x_k)\| = 0$.
2. $\|\nabla f(x_k)\| \to 0$ quand $k \to \infty$.

Preuve. 1. On a grâce à (5.7)

$$f(x_k) - f(x_{k+1}) \geq M\|\nabla f(x_k)\|^2 \geq 0, \tag{5.8}$$

pour $M > 0$. Donc $(f(x_k))_k$ est décroissante et $f(x_k) = f(x_{k+1})$ ne peut avoir lieu que si $\nabla f(x_k) = 0$.

2. Comme la suite $(f(x_k))_k$ est décroissante et f minorée, alors elle est convergente, i.e., $f(x_k) - f(x_{k+1}) \underset{k\to\infty}{\to} 0$. Cela implique avec (5.8) que $\|\nabla f(x_k)\| \to 0$ quand $k \to \infty$.

□

On termine cette section avec le résultat suivant permettant d'avoir des taux de convergence de la norme du gradient.

Théorème 53. Sous les mêmes hypothèses, on a pour tout $n \geq 0$

$$\min_{k=0,\dots,n} \|\nabla f(x_k)\| \leq \sqrt{\frac{f(x_0) - f^*}{M(n+1)}}, \tag{5.9}$$

où f^* est la limite de la suite $(f(x_k))_k$ et la constante M est donnée dans Proposition 15.

Preuve. En sommant de 0 à n dans l'inégalité (5.7), on obtient

$$f(x_0) - f(x_{n+1}) = \sum_{k=0}^{n}(f(x_k) - f(x_{k+1})) \geq M \sum_{k=0}^{n} \|\nabla f(x_k)\|^2.$$

Par définition, $f^* \leq f(x_{n+1})$, et par suite

$$f(x_0) - f^* \geq M \sum_{k=0}^{n} \|\nabla f(x_k)\|^2 \geq M \min_{k=0,\dots,n} \|\nabla f(x_k)\|^2 \sum_{k=0}^{n} 1,$$

soit

$$f(x_0) - f^* \geq M(n+1) \min_{k=0,\dots,n} \|\nabla f(x_k)\|^2.$$

Finalement $\min_{k=0,\dots,n} \|\nabla f(x_k)\| \leq \sqrt{\frac{f(x_0)-f^*}{M(n+1)}}$, comme voulu. □

Remarque 23. On note $\nabla f_n^* = \min_{k=0,\dots,n} \|\nabla f(x_k)\|$, on vient de démontrer que $\nabla f_n^* = \mathcal{O}(1/\sqrt{n})$. Au-delà du fait qu'il s'agisse d'un taux de convergence peu satisfaisant, le résultat ne donne aucune information sur les quantités $f(x_k) - f^*$ et $\|x_k - x^*\|$. Nous verrons plus tard que dans le cas convexe, le taux démontré précédemment peut être amélioré en $\mathcal{O}(1/n)$ et des taux de convergence de la fonction objectif et des itérées $(x_k)_k$ peuvent être démontrés.

5.4 Méthode de Newton

Avant de présenter la méthode de Newton comme une méthode de descente, rappelons la méthode de Newton pour trouver les zéros d'une fonction réelle. Soit donc $f : \mathbb{R} \to \mathbb{R}$, on cherche à résoudre l'équation $f(t) = 0$. L'idée est de remplacer f par son approximation de premier ordre en un point t_0, i.e., résoudre $\tilde{f}(t) = f(t_0) + (t - t_0)f'(t_0) = 0$ au lieu de $f(t) = 0$. Si $f'(t_0) \neq 0$, on obtient $t = t_0 - \frac{f(t_0)}{f'(t_0)}$. Les itérations de Newton s'écrivent

$$t_{k+1} = t_k - \frac{f(t_k)}{f'(t_k)}.$$

Maintenant, soit $g : \mathbb{R}^n \to \mathbb{R}^n$. On s'intéresse au système $g(x) = 0$. De la même manière, en approchant g à l'ordre un autour d'un point $x_0 \in \mathbb{R}^n$, i.e., $g(x_0) + \nabla g(x_0)(x - x_0) = 0$. On obtient

$$\tilde{g}(x) := g(x_0) + \nabla g(x_0)(x - x_0) = 0 \Rightarrow x = x_0 - \Big(\nabla g(x_0)\Big)^{-1} g(x_0).$$

L'équation ci-dessus suppose évidemment que $\det\Big(\nabla g(x_0)\Big) \neq 0$. De même, on peut construire une suite $(x_k)_k$ qui approchera, a priori, le zéro de g, en définissant

$$x_{k+1} = x_k - \Big(\nabla g(x_k)\Big)^{-1} g(x_k). \tag{5.10}$$

Maintenant, revenons à notre problème d'optimisation

$$\min_{x \in \mathbb{R}^n} \; f(x). \tag{5.11}$$

On s'intéresse ici au système $\nabla f(x) = 0$. Rappelons que lorsque f est convexe, la condition $\nabla f(x^*) = 0$ est nécessaire et suffisante pour dire que x^* minimise f. On applique (5.10) avec $g = \nabla f$, pour trouver

$$x_{k+1} = x_k - \Big(\nabla^2 f(x_k)\Big)^{-1} \nabla f(x_k). \tag{5.12}$$

Ceci est une façon de faire le lien entre Newton et Newton...

Une autre façon de retrouver (5.12) est la suivante. Supposons que f est C^2 et que $\nabla^2 f(x) \in \mathcal{S}^n_{++}$ pour tout $x \in \mathbb{R}^n$. On cherchera donc à minimiser l'approximation d'ordre deux de f autour d'un certain $x_0 \in \mathbb{R}^n$, i.e., remplacer f par

$$\tilde{f}(x) = f(x_0) + (x - x_0)^T \nabla f(x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0) \tag{5.13}$$

dans le problème (5.11). Donc étant donnée une itérée x_k, x_{k+1} est obtenue comme suit

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \Big\{ f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k) \Big\}. \tag{5.14}$$

L'unique minimiseur de (5.14) est en fait l'unique point stationnaire, et on a

$$\nabla \tilde{f}(x_{k+1}) = 0 \Leftrightarrow \nabla f(x_k) + \nabla^2 f(x_k)(x_{k+1} - x_k) = 0,$$

ce qui donne

$$x_{k+1} = x_k - \Big(\nabla^2 f(x_k)\Big)^{-1} \nabla f(x_k). \tag{5.15}$$

La méthode de Newton est donc une méthode de descente avec comme direction $\mathbf{d}_k = -\Big(\nabla^2 f(x_k)\Big)^{-1} \nabla f(x_k)$.

Méthode de Newton

- Initialisation : On se donne un $x_0 \in \mathbb{R}^n$ et une tolérance $\varepsilon > 0$.
- Pour $k \geq 0$
  - On calcule la direction de Newton $\mathbf{d}_k$ comme solution du système linéaire

    $$\nabla^2 f(x_k)\mathbf{d}_k = -\nabla f(x_k).$$

  - Prendre $x_{k+1} = x_k + \mathbf{d}_k$.
  - Si $\|\nabla f(x_{k+1})\| \leq \varepsilon$, s'arrêter et renvoyer x_{k+1}.

Il se trouve que sous certaines conditions de régularité sur la Hessienne, on peut obtenir autour de la solution optimale un taux de convergence quadratique.

Théorème 54. Soit $f \in C^2(\mathbb{R}^n)$ et supposons que :

1. il existe $m > 0$ tel que $\nabla^2 f(x) \succeq m I_n$ pour tout $x \in \mathbb{R}^n$,
2. il existe $L > 0$ tel que
$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|$$
pour tout $x, y \in \mathbb{R}^n$.

Alors
$$\|x_{k+1} - x^*\| \leq \frac{L}{2m}\|x_k - x^*\|^2,$$
où $(x_k)_k$ est la suite générée par la méthode de Newton et x^* l'unique minimiseur de f sur $\mathbb{R}^n$. De plus, si $\|x_0 - x^*\| \leq m/L$, alors
$$\|x_k - x^*\| \leq \frac{2m}{L}\left(\frac{1}{2}\right)^{2^k}, \quad k \geq 0.$$

Preuve. On a pour tout $k \geq 0$
$$\begin{aligned} x_{k+1} - x^* &= x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k) - x^* \\ &= x_k - x^* + \left(\nabla^2 f(x_k)\right)^{-1} \left(\nabla f(x^*) - \nabla f(x_k)\right) \\ &= x_k - x^* + \left(\nabla^2 f(x_k)\right)^{-1} \int_0^1 \left[\nabla^2 f(x_k + t(x^* - x_k))\right](x^* - x_k)\,\mathbf{d}t \qquad (5.16) \\ &= \left(\nabla^2 f(x_k)\right)^{-1} \int_0^1 \left[\nabla^2 f(x_k + t(x^* - x_k)) - \nabla^2 f(x_k)\right](x^* - x_k)\mathbf{d}t. \end{aligned}$$

Comme $\nabla^2 f(x) \succeq m I_n$, il vient que $\|\left(\nabla^2 f(x_k)\right)^{-1}\| \leq 1/m$. En utilisant la Lipschitz-continuité de la Hessienne, on obtient que
$$\begin{aligned} \|x_{k+1} - x^*\| &\leq \|\left(\nabla^2 f(x_k)\right)^{-1}\| \left\| \int_0^1 \left[\nabla^2 f(x_k + t(x^* - x_k)) - \nabla^2 f(x_k)\right](x^* - x_k)\mathbf{d}t \right\| \\ &\leq \frac{1}{m}\int_0^1 \|\nabla^2 f(x_k + t(x^* - x_k)) - \nabla^2 f(x_k)\|\|x^* - x_k\|\mathbf{d}t \qquad (5.17) \\ &\leq \frac{1}{m}\int_0^1 Lt\|x^* - x_k\|^2\mathbf{d}t = \frac{L}{2m}\|x^* - x_k\|^2. \end{aligned}$$

On conclut par récurrence. □

5.5 Méthode de gradient projeté

5.5.1 Cas général

On s'intéresse ici à des problèmes du type
$$\min_{x \in C} f(x), \qquad (5.18)$$
où $C \subset \mathbb{R}^n$ est un convexe fermé et $f : C \to \mathbb{R}$ de classe C^1. Il se trouve qu'on peut caractériser les points stationnaires de (5.18) en termes de l'opérateur de projection orthogonale.

Théorème 55. Soit $f : C \to \mathbb{R}$ de classe C^1 où C est un convexe fermé et $s > 0$. Alors x^* est un point stationnaire de (5.18) si et seulement si

$$x^* = P_C(x^* - s\nabla f(x^*)). \tag{5.19}$$

Preuve. Laissée en exercice. □

L'équation (5.19) affirme que x^* est point fixe de l'application $x \mapsto P_C(x - s\nabla f(x))$. Cela suggère des algorithmes du type point fixe pour résoudre (5.19), et donc (5.18).

Méthode du gradient projeté
On se donne une tolérance $\varepsilon > 0$.

- Initialisation : On se donne un $x_0 \in C$.
- Pour $k \geq 0$
  - Choisir un pas t_k par recherche linéaire.
  - Prendre $x_{k+1} = P_C(x_k - t_k \nabla f(x_k))$.
  - Si $\|x_k - x_{k+1}\| \leq \varepsilon$, renvoyer x_{k+1}.

On démontre un résultat similaire au Lemme 6.

Lemme 7. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$. Alors pour tout $x \in \mathbb{R}^n$ et $t > 0$

$$f(x) - f(P_C(x - t\nabla f(x))) \geq t\left(1 - \frac{tL}{2}\right)\left\|\frac{1}{t}(x - P_C(x - t\nabla f(x)))\right\|^2. \tag{5.20}$$

Preuve. Il suffit d'appliquer le lemme de descente Lemme 5 avec $y = P_C(x - t\nabla f(x))$ et utiliser le fait que

$$\langle x - t\nabla f(x) - y, x - y\rangle \leq 0.$$

□

Remarque 24. Quand $C = \mathbb{R}^n$, la méthode du gradient projeté n'est rien d'autre que la méthode de descente de gradient classique présentée en Section 5.2.

Notation. Par la suite, on note, pour $M > 0$

$$G_M(x) = M\left(x - P_C\left(x - \frac{1}{M}\nabla f(x)\right)\right).$$

On note que pour $C = \mathbb{R}^n$, $G_M(x) = \nabla f(x)$. L'opérateur servira comme une mesure d'optimalité. Avec cette notation, le Lemme 7 se réécrit

$$f(x) - f(P_C(x - t\nabla f(x))) \geq t\left(1 - \frac{tL}{2}\right)\|G_{1/t}(x)\|^2.$$

On peut démontrer sans difficultés un résultat similaire à la Proposition 15 et au Théorème 52 pour la méthode du gradient projeté ; on changera $\nabla f(x)$ par $G_M(x)$.

5.5.2 Cas convexe

On s'intéresse ici à des problèmes du type

$$\min_{x\in C} f(x), \tag{5.21}$$

où $C \subset \mathbb{R}^n$ est un convexe fermé et $f \in C_L^{1,1}(C)$ est **convexe**. Dans ce cas, on peut démontrer à la fois la convergence de $(f(x_k))_k$ vers f^* ainsi que celle de la suite $(x_k)_k$.

Théorème 56. Soit $f \in C_L^{1,1}(C)$ une fonction convexe avec C un convexe fermé de $\mathbb{R}^n$. Soit $(x_k)_k$ la suite générée par la méthode du gradient projeté avec un pas $t_k = t^* \in (0, 1/L]$. Supposons que $\operatorname{argmin}_C f \neq \emptyset$. Alors

1. Pour tout $k \geq 0$ et $x^* \in \operatorname{argmin}_C f$
$$f(x_k) - f^* \leq \frac{\|x_0 - x^*\|^2}{2t^* k}.$$
2. $x_k \to x^* \in \operatorname{argmin}_C f$ lorsque $k \to \infty$.

Remarque 25.

- **Amélioration du taux de convergence :** Dans le cas convexe, on obtient un taux de convergence $f(x_k) - f^* = \mathcal{O}(1/k)$ pour la fonction objectif, ce qui est significativement meilleur que le taux $\|\nabla f(x_k)\| = \mathcal{O}(1/\sqrt{k})$ obtenu dans le cas général non-convexe (voir Théorème 52). De plus, on obtient ici une information directe sur la décroissance de la fonction objectif, et non seulement sur son gradient.
- **Choix optimal du pas :** Le pas optimal est $t^* = 1/L$, qui maximise la vitesse de convergence tout en garantissant la décroissance. Ce choix donne
$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|^2}{2k}.$$
Notons que contrairement au cas non contraint où on pouvait prendre $t^* \in (0, 2/L)$, ici la contrainte $t^* \leq 1/L$ assure que la projection reste bien définie et que l'algorithme converge.
- **Convergence des itérées :** Le point (2) du théorème garantit la convergence de la suite $(x_k)_k$ elle-même vers un minimiseur, et pas seulement la convergence de $(f(x_k))_k$ vers f^*. Ceci est une conséquence directe de la convexité de f et de la structure du problème contraint. Si $\operatorname{argmin}_C f$ est un singleton, alors la convergence est vers l'unique minimiseur. Sinon, la suite $(x_k)_k$ converge vers un élément de $\operatorname{argmin}_C f$, qui dépend du point initial x_0.
- **Comparaison avec le cas non contraint :** Lorsque $C = \mathbb{R}^n$, la méthode du gradient projeté se réduit à la méthode du gradient classique, et on retrouve $G_M(x) = \nabla f(x)$. Le théorème ci-dessus généralise donc les résultats de convergence de la Section 5.2 au cas contraint, avec les mêmes garanties de convergence sous l'hypothèse de convexité.
- **Dépendance en la condition initiale :** La borne sur $f(x_k) - f^*$ dépend de $\|x_0 - x^*\|^2$, c'est-à-dire de la distance entre le point initial et un minimiseur. Cela signifie qu'un bon choix de x_0 (proche d'un minimiseur) peut accélérer significativement la convergence. En pratique, on ne connaît évidemment pas x^*, mais cette borne théorique donne une estimation du nombre d'itérations nécessaires pour atteindre une précision donnée.
- **Complexité algorithmique :** Pour obtenir une solution ε-optimale, i.e., $f(x_k) - f^* \leq \varepsilon$, il suffit de prendre
$$k \geq \frac{\|x_0 - x^*\|^2}{2t^*\varepsilon} = \frac{L\|x_0 - x^*\|^2}{2\varepsilon}.$$
Le nombre d'itérations nécessaires est donc linéaire en $1/\varepsilon$, ce qui caractérise les méthodes de gradient du premier ordre pour les fonctions convexes lisses.

5.6 Exercices

Exercice 33. On considère la fonction $f(x,y) = \frac{1}{2}(x^2 + \gamma y^2)$, avec $\gamma > 0$. En appliquant une descente de gradient avec recherche linéaire exacte partant de $(x_0, y_0) = (\gamma, 1)$, trouver l'expression des itérations (x_k, y_k).

Exercice 34. Soit $A \in \mathcal{S}^n_{++}$. Montrer que $\forall x \in \mathbb{R}^n \setminus \{0\}$:

$$\langle Ax, x\rangle \langle A^{-1}x, x\rangle \leq \frac{\|x\|^4}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2,$$

où λ_1 (resp. λ_n) est la plus grande (resp. plus petite) valeur propre de A.

Exercice 35. On considère le problème d'optimisation suivant :

$$\min_{x\in\mathbb{R}^n} f(x), \tag{P}$$

où $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$, avec $A \in \mathcal{S}^n_{++}(\mathbb{R}), b \in \mathbb{R}^n, c \in \mathbb{R}$.

1. Justifier pourquoi (P) admet une unique solution, que l'on notera x^*, qui est l'unique solution de l'équation $\nabla f(x) = 0$.
2. On se propose d'approcher x^* en utilisant une descente de gradient à pas optimal, i.e., générer une suite $(x_k)_{k\geq 0}$ où $x_{k+1} = x_k - t_k \nabla f(x_k)$ avec

$$t_k = \text{argmin}_t\, f(x_k - t\nabla f(x_k)).$$

   Montrer les propriétés suivantes :

   (a) $t_k = \frac{\|\nabla f(x_k)\|^2}{\langle A\nabla f(x_k), \nabla f(x_k)\rangle}$,

   (b) $\nabla f(x_{k+1}) = \nabla f(x_k) - t_k A\nabla f(x_k)$,

   (c) $\langle \nabla f(x_{k+1}), \nabla f(x_k)\rangle = 0$,

   (d) $f(x_{k+1}) - f^* = (f(x_k) - f^*)\left(1 - \frac{\|\nabla f(x_k)\|^4}{\langle A\nabla f(x_k), \nabla f(x_k)\rangle \langle A^{-1}\nabla f(x_k), \nabla f(x_k)\rangle}\right)$.

3. Soient $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ les valeurs propres de A rangées dans l'ordre décroissant. On rappelle que le conditionnement de A est donné par $c(A) = \lambda_1/\lambda_n$. Montrer que:

   (a) $f(x_k) - f^* \leq (f(x_0) - f^*)\left(\frac{c(A)-1}{c(A)+1}\right)^{2k}$.

   (b) $\|x_k - x^*\| \leq \sqrt{\frac{2(f(x_0)-f^*)}{\lambda_n}}\left(\frac{c(A)-1}{c(A)+1}\right)^k$.

Références Bibliographiques

[1] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of Convex Analysis*. Springer, 2001. (p. 1)

[2] Valeriu Soltan. *Lectures on Convex Sets*. World Scientific, 2022. (p. 1)

[3] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. (p. 1)

[4] Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. MOS-SIAM Series on Optimization. SIAM, 2014. (p. 1)

[5] Yurii Nesterov. *Lectures on Convex Optimization*. Springer, 2018. (p. 1)

[6] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*. Springer, 1998. (p. 1)