Module B1
Minimisation d'une fonction
Conditions d'optimalité

Dans ce module, sauf mention contraire, $\mathcal{X}$ désigne un espace de Hilbert, muni d'un produit scalaire noté $\langle \cdot, \cdot \rangle$ et de norme associée notée $\| \cdot \|$, tandis que E désigne un espace euclidien, que l'on identifiera à $\mathbb{R}^n$ avec $n \in \mathbb{N}^*$, muni du produit scalaire usuel noté $\langle \cdot, \cdot \rangle$ également et de norme associée la norme euclidienne, notée $\| \cdot \|_2$.

L'objectif de ce module est d'introduire le vocabulaire de la minimisation (sans contrainte) d'une fonction, puis d'établir des résultats généraux d'existence et d'unicité des minimiseurs. Enfin, on énoncera les conditions d'optimalité du premier et du second ordre pour les fonctions différentiables.

1 Minimisation d'une fonction

1.1 Optimum global d'une fonction

Commençons par donner les définitions suivantes :

Définition 1 (Minimiseur, minimum (globaux))

Soient $f : \mathcal{X} \to \mathbb{R}$ une fonction et $x^* \in \mathcal{X}$. On dit que x^* est un *minimiseur* de f si

$$\forall\, x \in \mathcal{X}, \quad f(x^*) \leq f(x)$$

Dans ce cas, la valeur $f(x^*) \in \mathbb{R}$ est appelée *minimum* de f, et est notée

$$\min_{x \in \mathcal{X}} f(x) \qquad \text{ou} \qquad \min_{\mathcal{X}} f$$

tandis que l'ensemble des minimiseurs de f est noté

$$\arg\min_{x \in \mathcal{X}} f(x) \qquad \text{ou} \qquad \arg\min_{\mathcal{X}} f$$

S'il existe, on dit aussi que x^* réalise le minimum de f ou que le minimum de f est atteint au point x^*. Noter dans les deux expressions précédentes l'article défini utilisé pour le mot *minimum*. On montre en effet que, s'il existe, le minimum est unique :

Proposition 1

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. On suppose que $\arg\min_{\mathcal{X}} f$ est non vide. Alors

$$x^*, \tilde{x} \in \arg\min_{\mathcal{X}} f \qquad \Longrightarrow \qquad f(x^*) = f(\tilde{x})$$

DÉMONSTRATION : Par définition d'un minimiseur, on a

$$\forall\, x \in \mathcal{X}, \qquad f(x^*) \leq f(x) \qquad \text{et} \qquad f(\tilde{x}) \leq f(x)$$

En particulier, en appliquant les inégalités précédentes à $x = \tilde{x}$ et $x = x^*$ respectivement, on obtient

$$f(x^*) \leq f(\tilde{x}) \qquad \text{et} \qquad f(\tilde{x}) \leq f(x^*)$$

ce qui entraîne l'égalité annoncée. ■

De la même manière, on peut définir

Définition 2 (Maximiseur, maximum)

Soient $f : \mathcal{X} \to \mathbb{R}$ une fonction et $x^* \in \mathcal{X}$. On dit que x^* est un *maximiseur* de f si

$$\forall\, x \in \mathcal{X}, \quad f(x^*) \geq f(x)$$

Dans ce cas, la valeur $f(x^*) \in \mathbb{R}$ est appelée *maximum* de f, et est notée

$$\max_{x \in \mathcal{X}} f(x) \qquad \text{ou} \qquad \max_{\mathcal{X}} f$$

tandis que l'ensemble des maximiseurs de f est noté

$$\arg\max_{x \in \mathcal{X}} f(x) \qquad \text{ou} \qquad \arg\max_{\mathcal{X}} f$$

et on dira de manière analogue que x^* réalise le maximum de f ou que le maximum de f est atteint au point x^*.

À ce stade, nous pouvons faire plusieurs remarques.

1. Tout d'abord, les définitions précédentes, parce qu'elles font intervenir une comparaison, nécessitent de considérer des fonctions f à **valeurs réelles**.
2. Lorsqu'ils existent, le minimum et le maxixum d'une fonction sont uniques ; ce n'est pas toujours le cas des minimiseurs et des maximiseurs.

1.2 Optimum local d'une fonction

Outre des minimiseurs / maximiseurs (globaux), une fonction peut également admettre une catégorie de points qui apparaîtront naturellement dans les modules qui suivent :

Définition 3 (Minimiseur, minimum, maximiseur, maximum locaux)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. On dit que $x^* \in \mathcal{X}$ est un *minimiseur local* (resp. *maximiseur local*) de f s'il existe un voisinage $\mathcal{V}(x^*)$ de x^* (c'est-à-dire un ouvert de $\mathcal{X}$ contenant x^*) tel que

$$\forall\, x \in \mathcal{V}(x^*), \quad f(x^*) \leq f(x)$$

(resp. $\quad \forall\, x \in \mathcal{V}(x^*), \quad f(x^*) \geq f(x) \quad$)

Dans ce cas, la valeur $f(x^*) \in \mathbb{R}$ est appelée *minimum local* (resp. *maximum local*) de f.

Contrairement aux optima globaux, une fonction peut admettre plusieurs minima ou maxima locaux, comme le montre l'exemple suivant :

EXEMPLE

Multiples minima et maxima locaux Considérons la fonction suivante :

$$f : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \dfrac{x}{2} + \sin x \end{array} \right.$$

Sa courbe représentative est montrée dans la figure 1. Une étude des variations de cette fonction dérivable montre que les points $x_1 = 4\pi/3$ et $x_2 = -2\pi/3$ (par exemple) sont des minimiseurs locaux tandis que les points $x_3 = -4\pi/3$ et $x_4 = 2\pi/3$ sont des maximiseurs locaux (on peut prendre par exemple comme voisinage $\mathcal{V}(x_i) = x_i + \,]-\pi/4\,;\pi/4\,[$). Or, on a

$$f(x_1) = \frac{2\,\pi}{3} - \frac{\sqrt{3}}{2} \neq f(x_2) = -\frac{\sqrt{3}}{2} - \frac{\pi}{3}$$

et
$$f(x_3) = -\frac{2\,\pi}{3} + \frac{\sqrt{3}}{2} \neq f(x_4) = \frac{\sqrt{3}}{2} + \frac{\pi}{3}$$

On en déduit qu'il n'y a pas unicité du minimum local et du maximum local. Par ailleurs, un minimum local peut être plus grand qu'un maximiseur local (ici $f(x_1) > f(x_3)$). Enfin, on peut montrer que la fonction f, malgré l'existence d'une infinité de minimiseurs locaux et de maximiseurs locaux, n'admet aucun minimiseur global, ni maximiseur global.

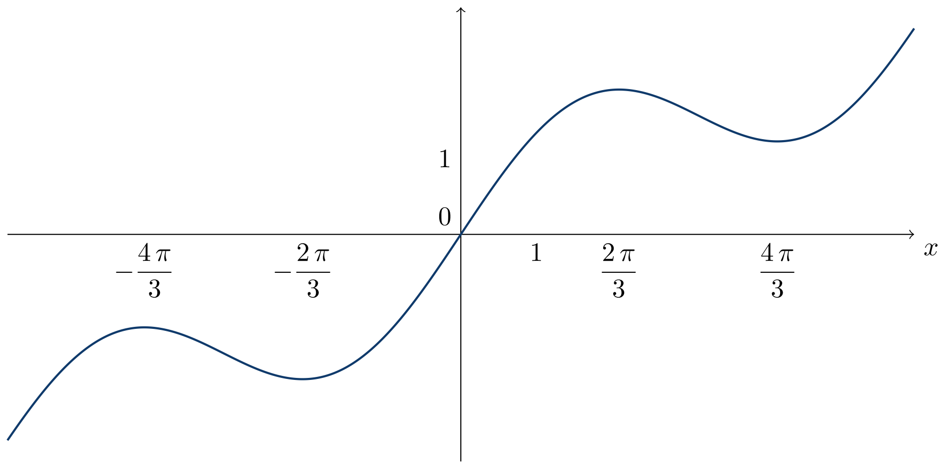


FIGURE 1 – Multiples minima et maxima locaux.

> Notons d'ores-et-déjà que, dans l'exemple ci-dessus, les minimiseurs et maximiseurs locaux de f sont des points qui annulent la dérivée de f. C'est un résultat que l'on va généraliser dans ce module.

Les fonctions convexes présentent l'intérêt de n'admettre que des minimiseurs globaux (quand elles en admettent) :

Proposition 2

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction *convexe*. Si $x^* \in \mathcal{X}$ est un minimiseur local de f, alors x^* est un minimiseur global de f.

DÉMONSTRATION : Soit $x^* \in \mathcal{X}$ un minimiseur local de f, c'est-à-dire qu'il existe un ouvert $\mathcal{V}(x^*)$ contenant x^* tel que

$$\forall\, x \in \mathcal{V}(x^*), \qquad f(x^*) \leq f(x)$$

Supposons que x^* n'est pas un minimiseur global de f, *i.e.* qu'il existe $y \in \mathcal{X}$ tel que

$$f(y) < f(x^*)$$

Posons $y_\lambda = \lambda\, y + (1-\lambda)\, x^*$ pour $\lambda \in \,]\,0\,;1\,[$. On a alors

$$\|y_\lambda - x^*\| = \lambda\, \|y - x^*\|$$

En particulier, si λ est suffisamment petit, alors $y_\lambda \in \mathcal{V}(x^*)$. Par conséquent, si λ est choisi suffisamment petit, alors $f(y_\lambda) \geq f(x^*)$. Par ailleurs, f est convexe donc

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

On en déduit que $$f(x^*) \leq \lambda\, f(y) + (1-\lambda)\, f(x^*)$$

soit, après simplification, $$\lambda\, f(x^*) \leq \lambda\, f(y)$$

Il suffit alors de diviser par λ, qui est strictement positif, pour établir une contradiction avec l'hypothèse initiale. ■

> Bien que ce résultat ne sera pas exploité dans ce cours, il constitue l'un des intérêts principaux de la convexité en optimisation, en empêchant les algorithmes de rester "bloqués" dans des minimiseurs locaux.

1.3 Opérations sur les ensembles de minimiseurs

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. On suppose que f admet un maximiseur x^*. Alors, par définition

$$\forall\, x \in \mathcal{X}, \quad f(x^*) \geq f(x) \qquad \text{soit} \qquad \forall\, x \in \mathcal{X}, \quad -f(x^*) \leq -f(x)$$

Autrement dit, x^* est un minimiseur de $-f$. De manière analogue, si x^* est un minimiseur de f, alors x^* est un maximiseur de $-f$. Supposons maintenant que la fonction f n'admet pas de minimiseur. Autrement dit, pour tout $x \in \mathcal{X}$, on peut trouver $y \in \mathcal{X}$ tel que

$$f(y) \leq f(x) \qquad \text{soit} \qquad -f(y) \geq -f(x)$$

Il s'ensuit que $-f$ n'admet pas de maximiseur. La réciproque se démontre de manière analogue. On en déduit donc finalement que

Proposition 3

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. Alors on a

$$\operatorname*{arg\,min}_{\mathcal{X}} f = \operatorname*{arg\,max}_{\mathcal{X}} (-f) \qquad \text{et} \qquad \min_{\mathcal{X}} f = -\max_{\mathcal{X}} -f$$

Ce résultat montre qu'au lieu de maximiser une fonction, on peut se contenter d'en minimiser l'opposée. Cela justifie le fait qu'à partir de maintenant, on ne s'intéresse plus qu'aux problèmes de minimisation. Si l'on souhaite utiliser les résultats de ce cours sur des problèmes de maximisation, il faut convertir le problème en un problème de minimisation équivalent.

> Attention : tous les auteurs ne font pas le choix de privilégier la minimisation sur la maximisation ! Il est donc essentiel de vérifier la convention adoptée dans les ouvrages consultés. En particulier, on signalera que le mot *optimisation* désigne indifféremment la minimisation ou la maximisation (de même que le terme *monotone* désigne à la fois la croissance et la décroissance). Comme on le verra dans la section 3, la convexité joue un rôle important en **minimisation** (en maximisation, c'est la concavité qui la remplace). Ainsi, la grande majorité des résultats d'optimisation énoncés dans ce cours s'applique à des problèmes de **minimisation**.

Lorsque l'on transforme un problème de maximisation d'une fonction f en un problème équivalent de minimisation de $-f$, on peut éventuellement le résoudre à l'aide de ces résultats, puis revenir au problème initial puisque les maximiseurs de f sont les minimiseurs trouvés, et le maximum de f vaut l'opposé du minimum trouvé. L'intérêt d'avoir procédé ainsi est de pouvoir utiliser les résultats de ce cours qui ne sont valables que pour les problèmes de minimisation (résultats d'existence de minimiseurs, caractérisation, algorithmes...). Il est essentiel pour cela de connaître le lien qui relie les solutions du problème initial à celles du problème équivalent. On peut étendre cette démarche à d'autres transformations.

Proposition 4

Soient $f : \mathcal{X} \to \mathbb{R}$ une fonction, $x^0 \in \mathcal{X}$ et $g : \mathbb{R} \to \mathbb{R}$ une fonction strictement croissante. Alors on a

$$\operatorname*{arg\,min}_{\mathcal{X}} f = x^0 + \operatorname*{arg\,min}_{x\in\mathcal{X}} g \circ f(x - x^0) \quad \text{et} \quad \min_{\mathcal{X}} f = g^{-1}\left(\min_{x\in\mathcal{X}} g \circ f(x - x^0)\right)$$

Démonstration :

- **Inclusion "$\subset$"** Soit x^* un minimiseur de f (s'il en existe). Par définition, on a
$$\forall\, x \in \mathcal{X}, \qquad f(x^*) \leq f(x)$$
Cette inégalité peut se réécrire (grâce à la croissance de g)
$$\forall\, x \in \mathcal{X}, \qquad g \circ f\big((x^* + x^0) - x^0\big) \leq g \circ f\big((x + x^0) - x^0\big)$$
Autrement dit, $x^* + x^0$ est un minimiseur de $x \mapsto g \circ f(x - x^0)$.
- **Inclusion "$\supset$"** On considère la fonction $\tilde{f} : x \mapsto g \circ f(x - x^0)$. En remarquant que g est inversible (car strictement croissante et que
$$\forall\, x \in \mathcal{X}, \qquad f(x) = g^{-1} \circ \tilde{f}(x + x^0)$$
avec g^{-1} strictement croissante, on peut appliquer le point précédent pour démontrer l'inclusion inverse, et donc l'identité entre les deux ensembles de minimiseurs (éventuellement vides).
- On vient de montrer que x^* est minimiseur de f si et seulement si $x^* + x^0$ est minimiseur de $\tilde{f}$. Par définition du minimum, on a donc
$$\min_{\mathcal{X}} f = f(x^*) = g^{-1}\big(g \circ f(x^* + x^0 - x^0)\big) = g^{-1}\big(\tilde{f}(x^* + x^0)\big) = g^{-1}\left(\min_{\mathcal{X}} \tilde{f}\right) \blacksquare$$

> Il existe de nombreuses raisons qui justifient de transformer un problème de minimisation en un problème de minimisation équivalent. On peut citer le fait de rendre la fonction à minimiser différentiable, ou de lui donner une forme connue ou plus simple à manipuler (par exemple, en supprimant les constantes additives ou multiplicatives). En pratique, il n'est pas utile de savoir inverser la transformation g, car l'inverse n'intervient pas pour le calcul des minimiseurs de f et que, pour calculer le minimum de f, il suffit d'appliquer f à un minimiseur de f.

2 Résultats d'existence et unicité des minimiseurs

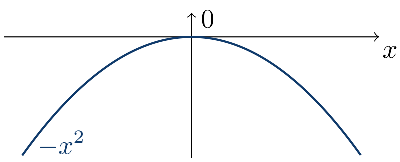
La minimisation d'une fonction comporte plusieurs étapes :

1. **Étude de l'existence de minimiseurs globaux :** il s'agit de savoir si la fonction considéré possède un minimiseur. Dans le cas contraire, il n'est évidemment pas nécessaire de poursuivre l'investigation.
2. **Recherche d'un ou des minimiseurs :** il s'agit généralement de la tâche principale. Il est parfois possible de déterminer le ou les minimiseurs de manière exacte, mais, en général, il faut passer par un algorithme qui calcule une suite de points dont il faut montrer qu'elle converge vers un minimiseur (global ou local). Dans le cas non convexe, il faut parfois déterminer par d'autres moyens si le point trouvé est un minimiseur global ou simplement local.
3. **Détermination du minimum :** si un minimiseur a été trouvé à l'étape suivante, il suffit d'appliquer la fonction au minimiseur.

2.1 Non existence de minimiseur

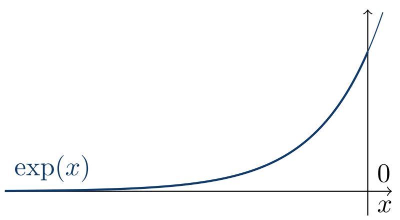
L'existence ou l'unicité des minimiseurs ne sont pas des propriétés acquises pour une fonction, comme en témoignent les exemples suivants, qui démontrent d'une grande variabilité dans les situations possibles :

Exemple



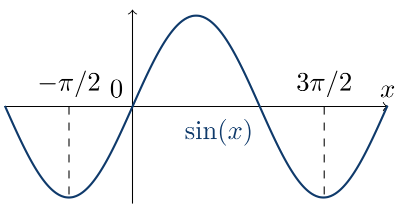
$$f : x \mapsto -x^2$$

Cette fonction ne possède pas de minimiseur. Remarque : f tend vers $-\infty$ lorsque x tend vers $+\infty$ et $-\infty$.



$$f : x \mapsto \exp(x)$$

Cette fonction ne possède pas de minimiseur. Remarque : f tend vers 0 lorsque x tend vers $-\infty$, mais sans jamais l'atteindre.

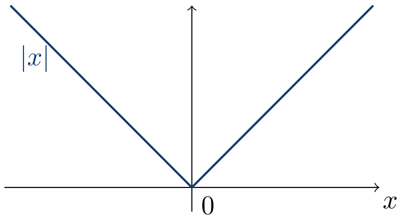


$$f : x \mapsto \sin(x)$$

Cette fonction possède une infinité de minimiseurs,

$$x^* \in \left\{ \frac{3\pi}{2} + 2\,k\,\pi \mid k \in \mathbb{Z} \right\}$$

Le minimum vaut alors -1.



$$f : x \mapsto |x|$$

Cette fonction possède un unique minimiseur.

$$x^* = 0$$

Le minimum vaut alors 0.

Notons que la première fonction n'admet aucun minimiseur car elle n'est pas minorée.

Définition 4 (Minorant d'une fonction)

Soient $\mathcal{C} \subset \mathcal{X}$ et $f : \mathcal{C} \to \mathbb{R}$ une fonction. On dit que m est un *minorant* de f sur $\mathcal{C}$ si

$$\forall\, x \in \mathcal{C}, \qquad m \leq f(x)$$

Lorsqu'une fonction f admet un minorant m sur $\mathcal{C}$, elle en admet une infinité, puisque tout $m' \leq m$ est alors un minorant. Dans ce cas, si $\mathcal{C}$ n'est pas vide, alors l'ensemble des minorants admet un plus grand élément, appelé *borne inférieure* de f sur $\mathcal{C}$ et noté

$$\inf_{\mathcal{X}} f \qquad \text{ou} \qquad \inf_{x \in \mathcal{X}} f(x)$$

Par définition de la borne inférieure, toute valeur inférieure est alors un minorant de la fonction. Si la fonction ne possède pas de minorant, alors, par convention, sa borne inférieure vaut $-\infty$ [1] et on dit que la fonction n'est pas minorée.

CONTRE-EXEMPLE

Fonctions non minorées Les fonctions $x \mapsto x^{2k+1}$ pour tout $k \in \mathbb{N}$ ne sont pas minorées.

Proposition 5

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. Si f n'admet pas de minorant, alors elle n'admet aucun minimiseur.

DÉMONSTRATION : Il suffit de revenir à la définition de minimiseur, qui assure que l'image par f de tout minimiseur est un minorant de la fonction f. ■

Pour montrer qu'un ensemble ou une fonction n'est pas minorée, on peut utiliser les suites minimisantes :

Définition 5 (Suite minimisante)

Soient $f : \mathcal{X} \to \mathbb{R}$ une fonction et $(x_k)_{k\in\mathbb{N}}$ une suite de $\mathcal{X}$. On dit que $(x_k)_{k\in\mathbb{N}}$ est une *suite minimisante* pour f si

$$\lim_{k\to+\infty} f(x_k) = \inf_{\mathcal{X}} f$$

1. Inversement, si la borne inférieure vaut $-\infty$, alors la fonction n'admet pas de minorant, car tout minorant est un nombre réel inférieur ou égal à la borne inférieure.

Proposition 6

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction et $(x_k)_{k\in\mathbb{N}}$ une suite de $\mathcal{X}$.

1. Si la suite $(f(x_k))_{k\in\mathbb{N}}$ converge vers $\ell \in \mathbb{R}$, alors
$$\inf_{\mathcal{X}} f \leq \ell$$
2. Si la suite $(f(x_k))_{k\in\mathbb{N}}$ diverge vers $-\infty$, alors
$$\inf_{\mathcal{X}} f = -\infty$$

Démonstration :

1. Par définition de la borne inférieure et de celle d'un minorant de f, on a
$$\forall\, k \in \mathbb{N}, \qquad \inf_{\mathcal{X}} f \leq f(x_k)$$
Il suffit alors de passer à la limite.
2. On raisonne par l'absurde, en supposant que la borne inférieure de f est finie. Alors, par définition de la borne inférieure et de celle d'un minorant, on a
$$\forall\, k \in \mathbb{N}, \qquad \inf_{\mathcal{X}} f \leq f(x_k)$$
On passe alors à nouveau à la limite pour obtenir la contradiction voulue. ■

On voit ainsi que, pour démontrer qu'une fonction n'admet pas de minimiseur, il suffit d'exhiber une suite de points $(x_k)_{k\in\mathbb{N}}$ telle que la suite des images par f diverge vers $-\infty$, ce qui démontrera que f n'est pas minorée. En revanche, une fonction minorée n'admet pas forcément de minimiseur !

Il est important de noter que, même si une fonction f admet un minimiseur, ses suites minimisantes ne convergent pas nécessairement vers un minimiseur (en général, elles ne convergent même pas). Pour s'en convaincre, on peut prendre l'exemple simple de la fonction réelle nulle, qui admet pour minimum 0 et pour minimiseurs l'ensemble de la droite réelle, de sorte que toute suite $(x_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^{\mathbb{N}}$ vérifie
$$\forall\, k \in \mathbb{N}, \qquad f(x_k) = 0$$
et donc est une suite minimisante pour f. Or, si l'on choisit $x_k = k$ pour tout $k \in \mathbb{N}$, cette suite diverge vers $+\infty$.

> Un contre-exemple remarquable de cette affirmation générale est cependant donnée par les fonctions fortement convexes.

2.2 Existence des minimiseurs d'une fonction

L'étude de l'existence de minimiseur pour une fonction repose généralement sur le résultat suivant :

Théorème 1 (Weierstrass)

Soient $\mathcal{K} \subset \mathcal{X}$ un **compact** non vide et $f : \mathcal{K} \to \mathbb{R}$ une fonction **continue**. Il existe $\tilde{x} \in \mathcal{K}$ tel que
$$\forall\, x \in \mathcal{K}, \qquad f(\tilde{x}) \leq f(x)$$

Démontrons le premier lemme suivant :

Lemme 1

Soit E un espace euclidien. Soit $f : E \to \mathbb{R}$ une fonction continue. Supposons qu'il existe $x_0 \in E$ tel que

$$\Big\{ x \in E \mid f(x) \leq f(x_0) \Big\}$$

soit borné. Alors f admet un minimiseur.

Démonstration : Par hypothèse, l'ensemble

$$\mathcal{K} = \{ x \in E \mid f(x) \leq f(x_0) \}$$

est un ensemble fermé et borné en dimension finie donc un ensemble compact. D'après le théorème 1 (puisque f est continue), il existe donc $x^* \in E$ tel que

$$\forall\, x \in \mathcal{K}, \qquad f(x) \geq f(x^*)$$

En particulier, puisque x_0 appartient évidemment à K, on a

$$f(x_0) \geq f(x^*)$$

Or, par définition de l'ensemble $\mathcal{K}$, on a

$$\forall\, x \notin \mathcal{K}, \qquad f(x) > f(x_0) > f(x^*)$$

ce qui prouve que x^* est un minimiseur de f. ■

Le résultat précédent est valable pour une classe de fonctions particulières, celles des fonctions continues *infinies à l'infini* :

Définition 6 (Fonction infinie à l'infini)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. On dit que f est *infinie à l'infini* si pour toute suite $(x_n)_{n\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$,

$$\lim_{n\to+\infty} \|x_n\| = +\infty \qquad \Longrightarrow \qquad \lim_{n\to+\infty} f(x_n) = +\infty$$

Petit point de vocabulaire. Dans la littérature consacrée à l'optimisation, on désigne (très) souvent cette notion sous le nom de *coercivité* (on parle alors de fonctions coercives). Cependant, la coercivité désigne une autre propriété dans le domaine des équations à dérivées partielles. Certains utilisent également l'épithète *propre* pour remplacer *infinie à l'infini*, mais en analyse convexe (et, par extension, en théorie de optimisation), une fonction propre est différente d'une fonction infinie à l'infini ! Il est donc impératif, si vous utilisez l'un de ces deux termes (propre ou coercive) de bien le définir en fonction de votre lecteur, et, de même, bien vérifier la définition choisie par l'auteur si vous le rencontrez dans un écrit (livre ou article).

Exercice

Montrer que la fonction $\left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \exp(\|x\|^2) \end{array} \right.$ est infinie à l'infini.

Certaines opérations préservent le caractère infini à l'infini :

Proposition 7

Soient $f : \mathcal{X} \to \mathbb{R}$ une fonction infinie à l'infini et $g : \mathcal{X} \to \mathbb{R}$ une fonction **minorée**. Soit $\alpha > 0$. Alors $\alpha\, f + g$ est infinie à l'infini.

DÉMONSTRATION : Il suffit de remarquer que, pour tout $x \in \mathcal{X}$, par définition de la borne inférieure, on a

$$\alpha\, f(x) + g(x) \geq \alpha\, f(x) + \inf_{\mathcal{X}} g$$

Puisque f est supposée infinie à l'infini, toute suite $(x_k)_{k\in\mathbb{N}}$ de $\mathcal{X}$ telle que les $\|x_k\|$ tendent vers ∞ implique que la suite des $f(x_k)$ tend également vers $+\infty$. Par comparaison, c'est également le cas de la suite des $\alpha\, f(x_k) + g(x_k)$ (car α est strictement positif). ■

Dans la démonstration de la proposition précédente, on voit se dégager une propriété plus générale :

Proposition 8

Soient $f : \mathcal{X} \to \mathbb{R}$ une fonction infinie à l'infini et $g : \mathcal{X} \to \mathbb{R}$ une fonction telle que

$$\forall\, x \in \mathcal{X}, \qquad g(x) \geq f(x)$$

Alors g est infinie à l'infini.

DÉMONSTRATION : Laissée en exercice.

On peut maintenant démontrer le résultat suivant concernant l'existence de minimiseurs chez les fonctions infinies sur l'infini :

Proposition 9

Soit E un espace euclidien. Soit $f : E \to \mathbb{R}$ une fonction **continue** et infinie à l'infini. Alors f admet un minimiseur.

REMARQUE : L'hypothèse de continuité, ainsi que celle de la dimention finie, sont importantes et ne doivent pas être oubliées !

DÉMONSTRATION : Il suffit de montrer que, dans ce cas, la fonction f satisfait les hypothèses du lemme 1. On l'établit par l'absurde, en supposant que pour tout $x_0 \in E$, l'ensemble

$$\Big\{x \in E \mid f(x) \leq f(x_0)\Big\}$$

n'est pas borné. Soit $x_0 \in E$. Il existe donc une suite $(x_n)_{n\in\mathbb{N}}$ de E telle que

$$\lim_{n\to+\infty} \|x_n\|_2 = +\infty \qquad \text{et} \qquad \forall\, n \in \mathbb{N}, \quad f(x_n) \leq f(x_0)$$

La suite $(f(x_n))_{n\in\mathbb{N}}$ est donc bornée, en particulier, elle ne peut pas tendre vers $+\infty$, ce qui contredit le fait que f soit infinie à l'infini. ■

2.3 Unicité du minimiseur d'une fonction

On termine cette section en énonçant quelques résultats sur l'unicité des minimiseurs. Celle-ci repose généralement sur l'hypothèse de stricte convexité :

Proposition 10

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction strictement convexe. Alors f admet au plus un minimiseur.

Démonstration : Supposons qu'il existe deux minimiseurs $x_1 \neq x_2$ de f. La fonction f étant strictement convexe, on a

$$f\left(\frac{x_1+x_2}{2}\right) < \frac{f(x_1)+f(x_2)}{2} = f(x_1)$$

ce qui est absurde car x_1 est un minimiseur de f. ■

L'unicité du minimiseur n'est utile que si l'existence a été démontrée. Dans ce cours, l'unique résultat assurant l'existence d'un minimiseur est le cas des fonctions continues sur un espace euclidien et infinies à l'infini. On va voir que c'est le cas des fonctions fortement convexes différentiables :

Proposition 11 (Fonctions fortement convexes Gateaux-différentiables)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction fortement convexe Gateaux-différentiable, de module $\alpha > 0$. Alors f est infinie à l'infini.

Démonstration : D'après la proposition 14 du module **A3 : Fonctions convexes différentiables**, il existe une fonction convexe $g : \mathcal{X} \to \mathbb{R}$ et un point $x^0 \in \mathcal{X}$ tel que

$$\forall\, x \in \mathcal{X}, \qquad f(x) = g(x) + \frac{\alpha}{2}\,\|x - x^0\|^2$$

En isolant g dans cette écriture, on voit que g est une fonction différentiable ; donc on peut écrire (voir module **A3 : Fonctions convexes différentiables**)

$$\forall\, x \in \mathcal{X}, \qquad g(x) \geq g(x^0) + \langle \nabla_G g(x^0), x - x^0 \rangle$$

On vérifie alors que

$$\langle \nabla_G g(x^0), x - x^0 \rangle + \frac{\alpha}{2}\,\|x - x^0\|^2 = \frac{\alpha}{2}\left\| x - x^0 + \frac{\nabla_G g(x^0)}{\alpha} \right\|^2 - \frac{\alpha}{2}\left\| \frac{\nabla_G g(x^0)}{\alpha} \right\|^2$$

(il suffit de développer et de simplifier). On en déduit donc que

$$\forall\, x \in \mathcal{X}, \qquad f(x) \geq g(x^0) + \frac{\alpha}{2}\left\| x - x^0 + \frac{\nabla_G g(x^0)}{\alpha} \right\|^2 - \frac{\alpha}{2}\left\| \frac{\nabla_G g(x^0)}{\alpha} \right\|^2$$

On conclut alors en utilisant le résultat de l'exercice précédent et des propositions précédentes. ■

> Ce résultat utilise le fait que la fonction convexe g se situe au-dessus de ses tangentes, c'est-à-dire qu'en particulier, il existe $p \in \mathcal{X}$ tel que
>
> $$\forall\, x \in \mathcal{X}, \qquad g(x) \geq g(x^0) + \langle p, x - x^0 \rangle$$
>
> (ici, $p = \nabla_G g(x^0)$). Il est possible de montrer que c'est le cas de toute fonction convexe définie sur $\mathcal{X}$. Aussi, le résultat précédent ainsi que le corollaire suivant reste valable lorsque f n'est pas Gateaux-différentiable.

On peut enfin établir le résultat suivant :

Corollaire 1

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable fortement convexe. Alors f admet un unique minimiseur.

Démonstration : Il suffit d'observer que f est strictement convexe, donc elle admet au plus un minimiseur ; par ailleurs, elle est infinie à l'infini et continue (car différentiable), donc elle admet au moins un minimiseur. ■

3 Conditions d'optimalité dans le cas différentiable

3.1 Condition nécessaire d'optimalité du premier ordre

Commençons par introduire la définition suivante :

Définition 7 (Point critique ou stationnaire)

Soient $\mathcal{U} \subset \mathcal{X}$ un ouvert et $f : \mathcal{U} \to \mathbb{R}$ une fonction différentiable. On appelle *point critique* ou *point stationnaire* de f tout point $x^* \in \mathcal{U}$ tel que

$$\nabla f(x^*) = 0$$

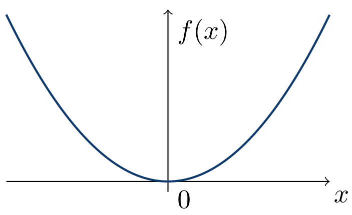
On note $\operatorname{crit} f \subset \mathcal{U}$ l'ensemble des points critiques de f.

Dans le cas d'une fonction réelle, les points critiques sont les points qui annulent la dérivée. Ainsi, si celle-ci est continue, il existe quatre types de points critiques isolés :

- **minimiseurs (locaux)** : la dérivée passe du signe négatif au signe positif ;
- **maximiseurs (locaux)** : la dérivée passe du signe positif au signe négatif ;
- **points d'inflexion montants** : la dérivée reste positive au voisinage du point ;
- **points d'inflexion descendants** : la dérivée reste négative au voisinage du point.

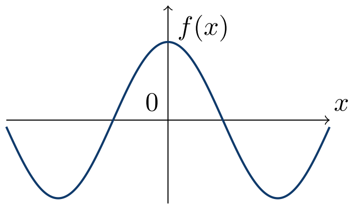
Exemple

Points critiques d'une fonction réelle.



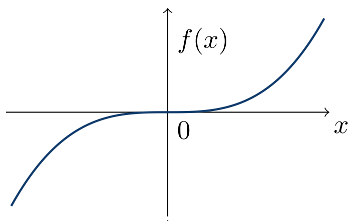
$$f : x \mapsto x^2 \quad \text{et} \quad f' : x \mapsto 2\,x$$

$x^* = 0$ est le minimiseur (global) de f.



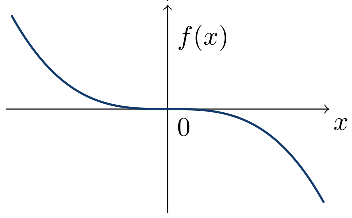
$$f : x \mapsto \cos x \quad \text{et} \quad f' : x \mapsto -\sin x$$

$x^* = 0$ est un maximiseur (global) de f.



$$f : x \mapsto x^3 \qquad \text{et} \qquad f' : x \mapsto 3\,x^2$$

$x^* = 0$ est un point d'inflexion montant de f (la fonction est croissante au voisinage de x^*).



$$f : x \mapsto -x^3 \quad \text{et} \quad f' : x \mapsto -3\,x^2$$

$x^* = 0$ est un point d'inflexion descendant de f (la fonction est décroissante au voisinage de x^*).

Notons qu'il existe des exemples de points critiques non isolés, pour lesquels la dérivée reste nulle sur un voisinage (il suffit par exemple de considérer la fonction nulle).

Énonçons un premier résultat central en optimisation :

Proposition 12 (Condition nécessaire d'optimalité du premier ordre)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction différentiable. Soit $x^* \in \mathcal{X}$ un minimiseur **local** de f. Alors

$$\nabla f(x^*) = 0$$

Autrement dit, tout minimiseur local ou global de f est un point critique de f. En particulier, si f n'admet pas de point critique (et donc, si $\operatorname{crit} f = \emptyset$), alors f n'admet pas de minimiseur.

> Cette condition est également connue sous le nom de *règle / théorème de* Fermat. On parle aussi parfois d'*équation d'*Euler–Lagrange.

Démonstration : Puisque f est différentiable, on a, pour h voisin de 0

$$f(x^* + h) = f(x^*) + \langle \nabla f(x^*), h \rangle + \|h\|\, \varepsilon(\|h\|)$$

avec $\varepsilon(t) \to 0$ lorsque $t \to 0$. Par définition d'un minimiseur local, on sait que, pour h dans un voisinage de 0, on a

$$f(x^* + h) \geq f(x^*)$$

En combinant ces deux relations, on en déduit que, pour h dans un voisinage de 0

$$\langle \nabla f(x^*), h \rangle + \|h\|\, \varepsilon(\|h\|) \geq 0$$

Supposons qu'il existe $h_0 \in \mathcal{X}$ tel que

$$\langle \nabla f(x^*), h_0 \rangle \neq 0$$

Notons que, nécessairement, $h_0 \neq 0$. Soit $h = t\, h_0$. On a donc pour t assez petit

$$t\, \langle \nabla f(x^*), h_0 \rangle + t\, \|h_0\|\, \varepsilon(\|t\, h_0\|) \geq 0$$

Si $t > 0$, on peut simplifier par t et faire tendre t vers 0, ce qui donne

$$\langle \nabla f(x^*), h_0 \rangle \geq 0$$

Si $t < 0$, on peut simplifier par t (ce qui change le sens de l'inégalité) et à nouveau faire tendre t vers 0, ce qui donne cette fois

$$\langle \nabla f(x^*), h_0 \rangle \leq 0$$

Autrement dit, on a nécessairement $\langle \nabla f(x^*), h_0 \rangle = 0$. On aboutit donc à une contradiction, ce qui assure que $\langle \nabla f(x^*), h \rangle = 0$ pour tout $h \in \mathcal{X}$. Par conséquent, on a bien $\nabla f(x^*) = 0$. ■

Si x^* est un minimiseur local de f, alors on a pour tout $v \in \mathcal{X}$

$$\forall\, h \in \mathbb{R}, \qquad f(x^*) \leq f(x^* + h\, v)$$

Il s'ensuit que 0 est un minimiseur local de $h \mapsto f(x^* + h\, v)$. Si cette fonction est dérivable, alors on peut appliquer la règle de FERMAT. En particulier, si f est GATEAUX-différentiable, cela donne :

Proposition 13 (Règle de FERMAT pour le cas GATEAUX-différentiable)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction GATEAUX-différentiable. Soit $x^* \in \mathcal{X}$ un minimiseur **local** de f. Alors

$$\nabla_{\mathrm{G}} f(x^*) = 0$$

3.2 Règles de FERMAT dans le cas convexe

Dans le cas convexe, la règle de FERMAT est une condition suffisante :

Proposition 14 (Condition nécessaire et suffisante d'optimalité du premier ordre)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** différentiable. Soit $x^* \in \mathcal{X}$. Alors on a l'équivalence entre les deux énoncés suivants :

(i) x^* est un minimiseur de f ;

(ii) $\nabla f(x^*) = 0$.

D'après cette proposition, toute fonction convexe f présente donc l'intéressante propriété d'avoir l'identité entre les trois ensembles suivants (possiblement vides) :

- l'ensemble $\operatorname{arg\,min}_{\mathcal{X}} f$ de ses minimiseurs ;
- l'ensemble de ses minimiseurs locaux ;
- l'ensemble $\operatorname{crit} f$ de ses points critiques.

DÉMONSTRATION : En vertu de la proposition 13, il suffit de démontrer la réciproque. Ainsi, on suppose que $\nabla f(x^*) = 0$. On va utiliser ici la caractérisation de la convexité pour les fonctions différentiables, qui permet d'écrire

$$\forall\, x \in \mathcal{X}, \qquad f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*)$$

Autrement dit, x^* est un minimiseur (global) de f. ■

Ainsi, dans le cas de l'optimisation convexe différentiable non contrainte, on a équivalence entre le problème d'optimisation

$$\text{Minimiser } f \text{ sur } \mathcal{X}$$

et le problème Trouver les zéros de l'application ∇f (sur $\mathcal{X}$)

On a donc transformé un problème d'optimisation sur f en un problème équivalent (c'est-à-dire que les deux problèmes ont mêmes solutions) sur ∇f. Or, il existe des méthodes numériques pour résoudre une telle équation. Lorsqu'elles sont appliquées au gradient ∇f, elles permettent de minimiser f. On parle alors de *méthodes du premier ordre* [2].

Proposition 15 (Règle de Fermat dans le cas Gateaux-différentiable convexe)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** Gateaux-différentiable. Soit $x^* \in \mathcal{X}$. Alors on a l'équivalence entre les deux énoncés suivants :

(i) x^* est un minimiseur de f ;

(ii) $\nabla_{\mathrm{G}} f(x^*) = 0$.

> En optimisation, le principe de la règle de Fermat (et sa conséquence $\operatorname{crit} f = \operatorname{arg\,min}_{\mathcal{X}} f$) est une propriété qui guide le choix de la définition de la différentiabilité (c'est-à-dire la généralisation de la dérivabilité en dimension supérieure à 1) et celle de la sous-différentiabilité (c'est-à-dire la généralisation de la dérivabilité aux fonctions non dérivables).

3.3 Conditions d'optimalité du second ordre

On suppose à présent que la fonction objectif f est deux fois différentiable et que $\mathcal{X} = \mathbb{R}^n$ est un espace euclidien.

Proposition 16 (Condition nécessaire d'optimalité du second ordre)

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable. Soit $x^* \in \mathbb{R}^n$ un minimiseur local de f. Alors $\operatorname{Hess} f(x^*)$ est semi-définie positive.

Démonstration : On écrit la formule de Taylor à l'ordre 2 : pour h voisin de 0

$$f(x^* + h) = f(x^*) + \langle \nabla f(x^*), h \rangle + \frac{1}{2} \langle \operatorname{Hess} f(x^*)\, h, h \rangle + \|h\|^2\, \varepsilon(h)$$

avec $\varepsilon(h) \to 0$ quand $h \to 0$. Puisque x^* est un minimiseur local de f, on a d'après la condition nécessaire d'optimalité du premier ordre $\nabla f(x^*) = 0$, donc pour h voisin de 0, on a

$$f(x^*) \leq f(x^* + h) = f(x^*) + \langle \operatorname{Hess} f(x^*)\, h, h \rangle + \|h\|^2\, \varepsilon(h)$$

En simplifiant la relation précédente, on obtient

$$0 \leq \langle \operatorname{Hess} f(x^*)\, h, h \rangle + \|h\|^2\, \varepsilon(h)$$

Soit $x \in \mathbb{R}^n$. Pour t assez petit, on peut poser $h = t\, x$, ce qui donne alors

$$0 \leq t^2 \langle \operatorname{Hess} f(x^*)\, x, x \rangle + t^2 \|x\|^2\, \varepsilon(t\, x)$$

Si $t \neq 0$, on peut simplifier par t^2, puis faire tendre t vers 0. Finalement, on démontre ainsi le résultat annoncé. ■

2. Plus de détails dans le cours de M2 **Méthodes du premier ordre pour l'optimisation non lisse et non convexe.**

On voit tout de suite qu'il ne peut s'agir d'une condition suffisante, car les fonctions convexes deux fois différentiables présentent cette propriété en tout point (or, tous ces points ne sont évidemment pas des minimiseurs).

Proposition 17 (Condition suffisante d'optimalité du second ordre)

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable. Soit $x^* \in \mathbb{R}^n$ un **point critique** de f. Si $\operatorname{Hess} f(x^*)$ est définie positive, alors x^* est un minimiseur local de f.

Attention ! Cette condition s'applique à un point critique de f, c'est-à-dire qu'il faut que

$$\nabla f(x^*) = 0$$

Sinon, elle se réduirait au fait que la matrice hessienne soit définie positive en x^*, ce qui impliquerait que tout point d'une fonction strictement convexe deux fois différentiable serait minimiseur (ce qui n'est bien évidemment pas le cas). Par ailleurs, il faut bien noter que seule la minimalité locale est assurée.

DÉMONSTRATION : On écrit à nouveau la formule de TAYLOR à l'ordre 2 : pour h voisin de 0, on a

$$f(x^* + h) = f(x^*) + \langle 0, h\rangle + \frac{1}{2}\langle \operatorname{Hess} f(x^*)\, h, h\rangle + \|h\|^2\, \varepsilon(h)$$

avec $\varepsilon(h) \to 0$ quand $h \to 0$. On va commencer par démontrer un résultat préliminaire. On pose $\mathcal{S}(0,1)$ la sphère unité dans $\mathbb{R}^n$, c'est-à-dire l'ensemble fermé borné suivant :

$$\mathcal{S}(0,1) = \left\{x \in \mathbb{R}^n \mid \|x\| = 1\right\}$$

Puisque $\mathbb{R}^n$ est de dimension finie, la sphère est compacte ; on sait donc que, grâce à sa continuité, la fonction $x \mapsto \langle \operatorname{Hess} f(x^*)\, x, x\rangle$ y atteint ses bornes. Autrement dit, il existe $x^0 \in \mathcal{S}(0,1)$ tel que

$$\langle \operatorname{Hess} f(x^*)\, x^0, x^0\rangle = \min_{x\in\mathcal{S}(0,1)} \langle \operatorname{Hess} f(x^*)\, x, x\rangle > 0$$

On en déduit que, pour tout $h \in \mathbb{R}^n \setminus \{0\}$,

$$\left\langle \operatorname{Hess} f(x^*)\, \frac{h}{\|h\|}, \frac{h}{\|h\|}\right\rangle \geq \min_{x\in\mathcal{S}(0,1)} \langle \operatorname{Hess} f(x^*)\, x, x\rangle$$

Cette minoration se réécrit

$$\langle \operatorname{Hess} f(x^*)\, h, h\rangle \geq \|h\|^2 \min_{x\in\mathcal{S}(0,1)} \langle \operatorname{Hess} f(x^*)\, x, x\rangle$$

On revient à la formule de TAYLOR ; on a pour h voisin de 0

$$f(x^* + h) \geq f(x^*) + \|h\|^2 \left(\frac{1}{2} \min_{x\in\mathcal{S}(0,1)} \langle \operatorname{Hess} f(x^*)\, x, x\rangle + \varepsilon(h)\right)$$

Par définition de la fonction ε, il existe $\delta > 0$ tel que pour h assez petit, la parenthèse soit strictement positif, et donc que $f(x^* + h) \geq f(x^*)$. ■