OPTIMISATION DÉTERMINISTE
ING2 MATHS-INFO CYTECH

LOUIS GARRIGUE

Soit un ensemble $K \subset \mathbb{R}^d$ une application $f : K \to \mathbb{R}$ bornée inférieurement, c'est-à-dire qu'il existe $c \in \mathbb{R}$ tel que $f(x) \geqslant c$ pour tout $x \in K$. Optimiser f, c'est trouver le ou les minimums globaux de f sur k, c'est-à-dire trouver les solutions du problème

$$\inf_{x \in K} f(x),$$

s'il en existe. Ce problème très général a un ensemble indénombrable d'applications dans tous les domaines de l'ingénierie et des sciences, en intelligence artificielle, en statistiques, en finance, en physique, etc.

Nous ne nous concentrerons pas seulement sur des problèmes abstraits et théoriques, mais nous donnerons les concepts suffisants pour le travail d'un ingénieur, et nous présenterons les méthodes permettant de résoudre très concrètement ces problèmes. Les TPs associés auront de nombreuses applications numériques. Un bon compris entre théorie et application sera suivi pour une formation ingénieur.

Pour réviser et aller plus loin dans le cours, nous conseillons la lecture du livre

"Introduction à l'Optimisation", Jean-Christophe Culioli, éditions Ellipses [?]

Contents

1. Rappels sur la différentiation

Nous aurons besoin de manipuler les dérivées. En effet, par exemple si $f(x) = 3x^2 - x + 1$ et si on veut la minimiser sur $\mathbb{R}$, il faut d'abord trouver les points $x \in \mathbb{R}$ tels que $f'(x) = 0$. Sur $\mathbb{R}^d$, la différentiation est plus délicate.

Généralement, on nomme "application" un élément qui à chaque élément d'un espace de départ, associe un élément d'un espace d'arrivée. Quand l'espace d'arrivée est $\mathbb{R}$, on parle aussi de "fonction".

Date: September 12, 2025.

Dans toute cette section, $\Omega \subset \mathbb{R}^d$ sera ouvert de $\mathbb{R}^d$.

1.1. **Gradient et différentielle.**

1.1.1. *Gradient.* Rappelons que le gradient d'une application $f : \Omega \to \mathbb{R}$, $\mathcal{C}^1$ est l'application notée $\nabla f : \Omega \to \mathbb{R}^d$ donnée par

$$\forall x = (x_1, \dots, x_d) \in \Omega, \qquad (\nabla f)(x) := \begin{pmatrix} (\partial_{x_1} f)(x) \\ \vdots \\ (\partial_{x_d} f)(x) \end{pmatrix}. \tag{1}$$

Par exemple si $d = 2$ et $f(x) = x_1^2 + 3x_2$, alors $(\nabla f)(x) = \begin{pmatrix} 2x_1 \\ 3 \end{pmatrix}$.

Nous noterons aussi $\nabla f(x) := (\nabla f)(x)$ pour alléger le texte.

1.1.2. *Différentielle.* Prenons une fonction $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}^n$, donc plus générale que pour la définition du gradient. On rappelle la définition de la différentielle de f, notée $\mathrm{d}f : \Omega \to \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$, où $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$ est l'ensemble des applications linéaires allant de $\mathbb{R}^d$ and $\mathbb{R}^n$.

Definition 1.1 (Différentielle en x)**.** *Prenons $x \in \Omega$. Comme Ω est ouvert, il existe $r > 0$ tel que $B(x, r) \subset \Omega$. Supposons qu'il existe un application linéaire $L : \mathbb{R}^d \to \mathbb{R}^n$ et $c > 0$ et $s \in [0, r]$ tels que pour tout $y \in \mathbb{R}^d$ tel que $\|y\| \leqslant s$,*

$$\|f(x + y) - (f(x) + Ly)\| \leqslant c \|y\|^2 .$$

Alors f est différentiable en x et on note $\mathrm{d}_x f = L$.

On dit que f est différentiable sur Ω si elle l'est en tout point de Ω.

La différentielle généralise les développements limités d'ordre 1, c'est-à-dire qu'on peut écrire que pour tout $x \in \Omega, y \in \mathbb{R}^d$ tel que $x + y \in \Omega$,

$$f(x + y) = f(x) + (\mathrm{d}_x f)\, y + O(\|y\|^2), \tag{2}$$

quand $\|y\| \to 0$. Pour un $x \in \Omega$ donné, $\mathrm{d}_x f$ est une application linéaire allant de $\mathbb{R}^d$ dans $\mathbb{R}^n$. Si f et g sont deux applications composables et $\mathcal{C}^1$, $g \circ f$ est aussi $\mathcal{C}^1$ et la règle de la chaîne donne

$$\mathrm{d}_x (f \circ g) y = \big(\mathrm{d}_{g(x)} f\big) \circ (\mathrm{d}_x g)\, y. \tag{3}$$

Exercice 1.2. *L'application exponentielle* $\exp : \mathcal{M}_d(\mathbb{R}) \to \mathcal{M}_d(\mathbb{R})$ *est définie par* $\exp(A) := \sum_{k=0}^{+\infty} \frac{A^k}{k!}$*. Calculer sa différentielle en 0.*

Solution. On prend $A = 0$. On a

$$\exp{(A + H)} = \exp H = 1 + H + \sum_{k=2}^{+\infty} \frac{H^k}{k!} = \exp A + H + \sum_{k=2}^{+\infty} \frac{H^k}{k!}.$$

On conjecture que $(\mathrm{d}_A \exp) H = H$. On pose $LH := H$, qui est une application linéaire. On a

$$\begin{aligned}\|\exp(A+H) - ((\exp A) + LH)\| = \left\| \sum_{k=2}^{+\infty} \frac{H^k}{k!} \right\| \leqslant \sum_{k=2}^{+\infty} \frac{\|H\|^k}{k!} = \|H\|^2 \sum_{k=0}^{+\infty} \frac{\|H\|^k}{(k+2)!} \\ \leqslant \|H\|^2 \sum_{k=0}^{+\infty} \frac{\|H\|^k}{k!} = \|H\|^2 e^{\|H\|}.\end{aligned}$$

On peut donc écrire

$$\exp(A+H) = \exp(A) + LH + O_{\|H\|\to 0}\left(\|H\|^2\right),$$

donc exp est différentiable en 0 et $(\mathrm{d}_0 \exp)H = LH = H$, i.e. $\mathrm{d}_0 \exp = \mathbb{1}_{\mathcal{M}_d(\mathbb{R})}$. □

1.1.3. *Dérivée.* Prenons un intervalle I. Pour les applications $\xi : I \to \mathbb{R}^d$, c'est-à-dire quand l'espace de départ est plongé dans $\mathbb{R}$, on définit la dérivée de ξ par

$$\xi'(t) := \frac{(\mathrm{d}_t \xi)\, s}{s} \in \mathbb{R}^d$$

pour tout $t \in I$ et tout $s \in \mathbb{R}$. Si $\xi : I \to \mathbb{R}$, la dérivée coïncide avec le gradient.

1.1.4. *Autre définition du gradient.* Reprenons une fonction $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$. Sur tout point $x \in \Omega$, le gradient $\nabla f(x)$ peut être en fait défini à partir de la différentielle par dualité (via le théorème de Riesz disant que les formes linéaires on un vecteur représentant) via

$$\forall y \in \mathbb{R}^d, \qquad (\mathrm{d}_x f)\, y = \langle \nabla f(x), y \rangle, \tag{4}$$

et on peut montrer qu'il est bien égal à celui donné en (1). La différentielle ne dépend pas du produit scalaire alors que le gradient si.

Lorsque ces quelques règles sont bien maitrisées, on ne se trompe plus dans les calculs de dérivation de fonctions définies sur $\mathbb{R}^d$.

1.2. **Hessienne.** Soit $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ deux fois différentiable. On note $\mathcal{S}_d(\mathbb{R})$ l'ensemble des matrices carrées réelles symétriques de taille $d \times d$. On appelle Hessienne de f l'application $\nabla^{\otimes 2} f : \Omega \to \mathcal{S}_d(\mathbb{R})$, où

$$\left(\nabla^{\otimes 2} f\right)(x) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)_{1 \leqslant i,j \leqslant d}.$$

Lemma 1.3. *Pour tout $x \in \Omega$, on a bien $\nabla^{\otimes 2} f(x) \in \mathcal{S}_d(\mathbb{R})$.*

Proof. Pour tout $i, j \in \{1, \ldots, d\}$, $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$, par le théorème de Schwartz. □

On ne le prouvera pas mais on a, pour $\|y\| \to 0$, le développement limité d'ordre 2 suivant, pour toute direction $y \in \mathbb{R}^d$,

$$f(x+y) = f(x) + (\mathrm{d}_x f)\, y + \left\langle y, \left(\nabla^{\otimes 2} f\right) y \right\rangle + O(\|y\|^3). \tag{5}$$

Comme $\nabla^{\otimes 2} f(x)$ est symétrique, on peut la diagonaliser. Il existe $P(x) \in GL_d(\mathbb{R})$ et $D(x) \in \mathcal{M}_d(\mathbb{R})$ diagonale telles que

$$\nabla^{\otimes 2} f(x) = P(x) D(x) P(x)^{-1}. \tag{6}$$

1.3. **Exemples de calculs.** Donnons quelques exemples simples.

1.3.1. $t \mapsto f(ta+b)$. Supposons $\Omega = \mathbb{R}^d$ et $a, b \in \mathbb{R}^d$, et prenons $f : \mathbb{R} \to \mathbb{R}$. On définit $w : \mathbb{R} \to \mathbb{R}$ par

$$w(t) := f(ta+b)$$

et on veut calculer sa dérivée. On peut poser $m : \mathbb{R} \to \mathbb{R}^d$ définie par $m(t) := ta+b$, on a alors $w = f \circ m$, donc la règle de la chaîne (3) donne, pour tout $t, s \in \mathbb{R}$,

$$(\mathrm{d}_t w)\, s = \left(\mathrm{d}_{m(t)} f\right) \circ (\mathrm{d}_t m)\, s.$$

Or, on peut calculer que $(\mathrm{d}_t m)s = as$ donc

$$\begin{aligned} w'(t) &= \frac{\left(\mathrm{d}_{m(t)} f\right) \circ (\mathrm{d}_t m)\, s}{s} = \frac{\left(\mathrm{d}_{m(t)} f\right) \circ as}{s} = \left(\mathrm{d}_{m(t)} f\right) \circ a = \langle \nabla f(m(t)), a\rangle \\ &= \langle \nabla f(ta+b), a\rangle . \end{aligned}$$

Le calcul est plus simple qu'en passant par (1). De même, on peut calculer

$$w''(t) = \left\langle a, \left(\left(\nabla^{\otimes 2} f\right)(ta+b)\right) a \right\rangle .$$

1.3.2. *Norme.* Calculons la différentielle de l'application $f : \mathbb{R}^d \to \mathbb{R}$ définie par $f(x) := \|x\|^2$.

On a $f(x) = f(x_1, \ldots, x_d) = \sum_{k=1}^d x_k^2$ donc

$$\left(\partial_{x_j} f\right)(x) = \sum_{k=1}^d \partial_{x_j} x_k^2 = \sum_{k=1}^d 2\delta_{k=j} x_k = 2x_j$$

et

$$\nabla f(x) = \begin{pmatrix} (\partial_{x_1} f)(x) \\ \vdots \\ (\partial_{x_d} f)(x) \end{pmatrix} = \begin{pmatrix} 2x_1 \\ \vdots \\ 2x_d \end{pmatrix} = 2x.$$

On calcule $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \sum_{k=1}^d \frac{\partial^2}{\partial x_i \partial x_j} x_k^2 = 2\delta_{i=j}$ donc $\nabla^{\otimes 2} f(x) = 2 \times \mathbb{1} = 2$ où $\mathbb{1}$ est la matrice identité.

1.3.3. *Formes quadratiques.*

Proposition 1.1: Gradient et Hessienne d'une forme quadratique

Soit $b \in \mathbb{R}^d$ et $c \in \mathbb{R}$, $A \in \mathcal{M}_n(\mathbb{R})$ et $f : \mathbb{R}^d \to \mathbb{R}$ telle que $\forall x \in \mathbb{R}^d$,

$$f(x) := \langle Ax, x\rangle + \langle b, x\rangle + c. \tag{7}$$

On a alors

$$\nabla f(x) = \left(A^T + A\right) x + b, \qquad \text{et} \qquad \nabla^{\otimes 2} f(x) = A^T + A.$$

Sous cette forme, f est appelée une forme quadratique.

Proof. On fait un calcul direct et un calcul via la différentielle. On verra que le calcul utilisant la différentielle est beaucoup plus simple et rapide.

- Calcul direct. On appelle $g(x) := \langle Ax, x\rangle$ et $h(x) := \langle b, x\rangle$. Alors $g(x) = \sum_{i=1}^d (Ax)_i x_i = \sum_{1\leqslant i,j\leqslant d} A_{ij}x_j x_i$ donc

$$
\begin{aligned}
(\partial_{x_k} g)(x) &= \sum_{1\leqslant i,j\leqslant d} A_{ij}\partial_{x_k} x_j x_i = \sum_{1\leqslant i,j\leqslant d} A_{ij}\left(2x_k\delta_{k=i=j} + x_i\delta_{k=j,i\neq k} + x_j\delta_{k=i,j\neq k}\right) \\
&= 2A_{kk}x_k + \sum_{\substack{1\leqslant i\leqslant d\\ i\neq k}} A_{ik}x_i + \sum_{\substack{1\leqslant j\leqslant d\\ j\neq k}} A_{kj}x_j = \sum_{i=1}^d A_{ik}x_i + \sum_{j=1}^d A_{kj}x_j \\
&= \left(A^t x\right)_k + (Ax)_k .
\end{aligned}
$$

De même, $h(x) = \sum_{i=1}^d b_i x_i$ donc $\partial_{x_k} h(x) = b_k$.

- En utilisant $\mathrm{d}_x f$. On a

$$
\begin{aligned}
f(x+y) &= \langle A(x+y), x+y\rangle + \langle b, x+y\rangle + c \\
&= f(x) + \langle Ax, y\rangle + \langle Ay, x\rangle + \langle Ay, y\rangle + \langle b, y\rangle \\
&= f(x) + \left\langle (A+A^T)x + b, y\right\rangle + O(\|y\|^2).
\end{aligned}
$$

Donc pour que la définition (2) soit respectée, on déduit la différentielle $(\mathrm{d}_x f)y = \left\langle (A+A^T)x + b, y\right\rangle$. Or, le gradient est défini via (4) donc $\nabla f(x) = (A+A^T)x + b$.

- Hessienne. On calcule

$$
\frac{\partial^2 g}{\partial x_i \partial x_j}(x) = \sum_{1\leqslant k,p\leqslant d} A_{kp}\frac{\partial^2}{\partial x_i\partial x_j}x_k x_p = \sum_{1\leqslant k,p\leqslant d} A_{kp}\delta_{\substack{k=i,p=j\\ \text{ou}\\ k=j,p=i}} = A_{ij} + A_{ji} = (A+A^T)_{ij}.
$$

donc $\nabla^{\otimes 2} f(x) = A + A^T$. □

1.3.4. *Forme quadratique particulière.* Si $f(x_1, x_2) := x_1^2 + \lambda x_2^2$ où $\alpha \in \mathbb{R}$, alors $\nabla f(x) = 2\begin{pmatrix} x_1 \\ \lambda x_2\end{pmatrix}$ and $\nabla^{\otimes 2} f(x) = \begin{pmatrix} 1 & 0 \\ 0 & \lambda\end{pmatrix}$.

2. Convexité

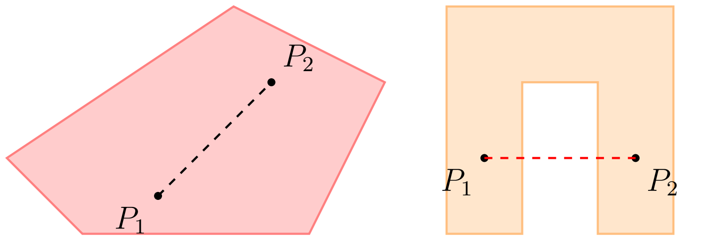
2.1. Ensembles convexes.

Definition 2.1 (Intervalle)**.** *Un intervalle $I \subset \mathbb{R}$ est un sous-ensemble connexe de $\mathbb{R}$ qui n'est pas réduit à un singleton. Il est donc de la forme $[b,+\infty[$ ou $]b,+\infty[$ ou $[a,b]$ ou $[a,b[$ ou $]a,b]$ ou $]a,b[$ ou $]-\infty,a]$ ou $]-\infty,a[$, où $a,b\in\mathbb{R}$ et $a<b$.*

Definition 2.2 (Ensemble convexe)**.** *Un ensemble $K \subset \mathbb{R}^d$ est dit convexe si*

$$
\forall x, y \in K,\ \forall t\in[0,1], \qquad tx + (1-t)y \in K.
$$

Autrement dit, un ensemble est convexe si quand on prend deux points lui appartenant, le segment reliant ces deux points y est également. En dimension $d = 1$, les ensembles convexes sont les intervalles et les singletons. En figure 1, en dimension $d = 2$, nous dessinons deux exemples.



Ensemble convexe Ensemble non convexe

FIGURE 1. Exemples sur la convexité en dimension 2

2.1.1. *Exemples.*

- $\mathbb{R}^d$ est convexe
- la boule ouverte de centre $a \in \mathbb{R}^d$ et de rayon $R > 0$, $B(a, R) \subset \mathbb{R}^d$, est convexe
- n'importe quel sous-espace affine de $\mathbb{R}^d$ est convexe
- pour $d = 1$, un ensemble $C \subset \mathbb{R}$ est convexe ssi c'est un intervalle
- les demi-espaces affines

$$C := \left\{ x \in \mathbb{R}^d \mid \langle a, x \rangle \leqslant \lambda \right\},$$

où $a \in \mathbb{R}^d \backslash \{0\}$ et $\lambda \in \mathbb{R}$, sont convexes.

2.2. **Fonctions convexes.**

Definition 2.3 (Convexité des fonctions)**.** *Soit $K \subset \mathbb{R}^d$ et une fonction $f : K \to \mathbb{R}$ une fonction. On dit que f est*

- *convexe si*

$$\forall x, y \in K,\ \forall t \in [0, 1], \qquad f\left(tx + (1-t)y\right) \leqslant tf(x) + (1-t)f(y).$$

- *strictement convexe si*

$$\forall x, y \in K \text{ tels que } x \neq y,\ \forall t \in [0, 1], \qquad f\left(tx + (1-t)y\right) < f(x) + (1-t)f(y).$$

- *fortement convexe s'il existe $\alpha > 0$ tel que*

$$\forall x, y \in K,\ \forall t \in [0, 1], \qquad f\left(tx + (1-t)y\right) \leqslant f(x) + (1-t)f(y) - \alpha t(1-t) \left\| x - y \right\|^2$$

- *concave si $-f$ est convexe.*

On a que fortement convexe $\implies$ strictement convexe $\implies$ convexe. En figure 2 nous donnons quelques exemples. Les fonctions affines sont les seules fonctions convexes et concaves. Visuellement, une fonction est convexe ssi son graph est au-dessus de tous ses plans tangents, ou encore ssi les arcs reliant deux points du graph sont au-dessus du graph.

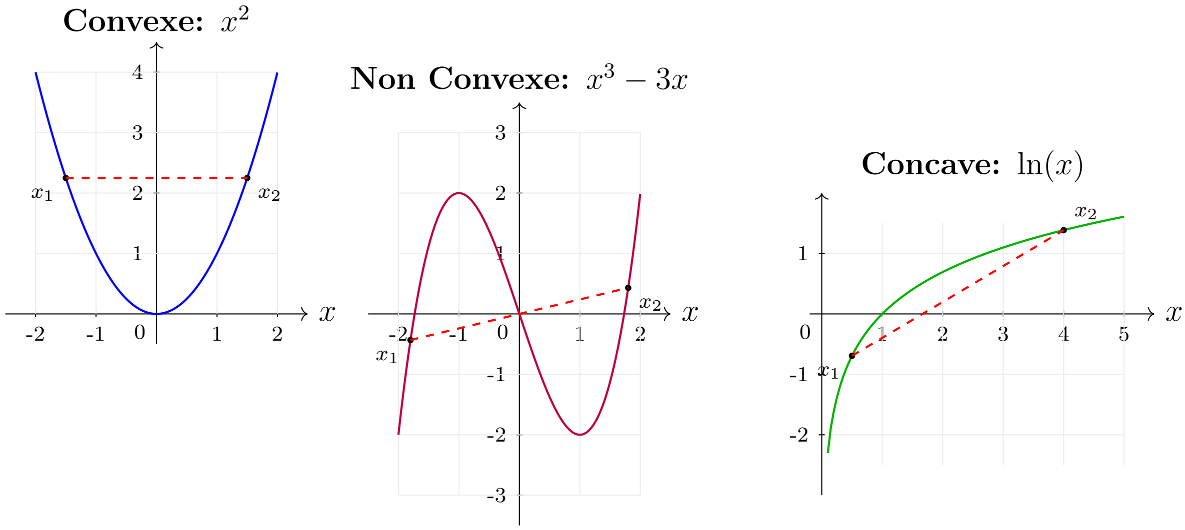


FIGURE 2. Exemples sur la convexité des fonctions en dimension un.

2.2.1. *Fonctions différentiables.* Lorsque f est $\mathcal{C}^1$, c'est-à-dire qu'elle est différentiable et que $x \mapsto \nabla f(x)$ est continue, la convexité peut s'exprimer plus simplement en utilisant le gradient.

Proposition 2.4 (Caractérisation de la convexité avec ∇f)**.** *Soit $K \subset \mathbb{R}^d$ un ensemble convexe et $f : K \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ dans un voisinage de K. Les assertions suivantes sont équivalentes :*

(1) *f est convexe sur K,*
(2) *$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle$, $\forall x, y \in K$,*
(3) *$\langle \nabla f(y) - \nabla f(x), y - x\rangle \geqslant 0$, $\forall x, y \in K$,*
(4) *(Si f est $\mathcal{C}^2$) $\nabla^{\otimes 2} f(x) \geqslant 0$ for all $x \in K$.*

La seconde assertion signifie que f est au-dessus de son développement d'ordre 1. On peut énoncer le même résultat en remplaçant les inégalités par des inégalités strictes dans les assertions précédentes.

Proof. Prouvons que (1) implique (2). On prend $x, y \in K$. On va se réduire au cas de dimension 1 en posant $\xi(t) := f(x + t(y - x))$. On a $\xi(0) = f(x)$, $\xi(1) = f(y)$. Comme f est convexe, $\xi(t) \leqslant t\xi(1)+(1-t)\xi(0)$, donc la courbe de ξ est en-dessous de la droite passant par $(0, \xi(0))$ et $(1, \xi(1))$. On a donc $\frac{1}{t}(\xi(t) - \xi(0) \leqslant \xi(1) - \xi(0)$ et en faisant $t \to 0$ on obtient $\xi'(0) \leqslant \xi(1)-\xi(0)$. Or, $\xi'(t) = \langle \nabla f(x + t(y - x)), y - x\rangle$ (voir Section (1.3.1)) donc $\xi(0) = \langle \nabla f(x), y - x\rangle$ et on obtient le résultat.

Prouvons que (1) implique (2). On écrit (1) en échangeant x et y puis avec (1) on obtient

$$\langle \nabla f(x), y - x\rangle \leqslant f(y) - f(x) \leqslant \langle \nabla f(y), y - x\rangle .$$

Prouvons que (2) implique (4). On prend $y = x + tz$, où $t \in \mathbb{R}$, on a

$$f(x + tz) = f(x) + t\,\langle \nabla f(x), y\rangle + t^2 \left\langle z, \left(\nabla^{\otimes 2} f(x)\right) z\right\rangle + O(t^3)$$

donc

$$0 \underset{(2)}{\leqslant} \frac{1}{t^2}\left(f(x+tz) - f(x) - t\langle \nabla f(x), y\rangle\right) = \left\langle z, \left(\nabla^{\otimes 2} f(x)\right) z\right\rangle + O(t)$$

on fait $t \to 0$ et on obtient $\langle z, (\nabla^{\otimes 2} f(x))\, z\rangle \geqslant 0$. C'est vrai pour tout $z \in \mathbb{R}^d$ donc on peut conclure. □

La stricte convexité a la même caractérisation en remplaçant les inégalités par des inégalités strictes.

Proposition 2.5 (Caractérisation de la forte convexité avec ∇f)**.** *Soit $K \subset \mathbb{R}^d$ un ensemble convexe et $f : K \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ dans un voisinage de K. Les assertions suivantes sont équivalentes :*

- *f est α-fortement convexe sur K, avec $\alpha > 0$*
- *$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}\|y-x\|^2, \quad \forall x, y \in K$*
- *$\langle \nabla f(y) - \nabla f(x), y - x\rangle \geqslant \alpha\|y-x\|^2, \quad \forall x, y \in K.$*
- *(Si f est $\mathcal{C}^2$) $\nabla^{\otimes 2} f(x) \geqslant \alpha$, $\forall x \in K$*

La condition $\nabla^{\otimes 2} f(x) \geqslant \alpha$ est équivalente à ce que les valeurs propres de $A^T + A$ soient supérieures à α, ou encore à ce que tous les éléments diagonaux de $D(x)$ (définie en (6)) soient supérieurs à α.

Definition 2.6 (Fonction elliptique)**.** *On appelle fonction elliptique une fonction $f : K \to \mathbb{R}$ de classe $\mathcal{C}^1$ et fortement convexe.*

2.2.2. *Dimension 1.* Les propriétés se simplifient en dimension 1. Soit I un intervalle et $f : I \to \mathbb{R}$ une fonction deux fois différentiable. Alors on a

$$\begin{aligned} &f \text{ convexe} \iff f' \text{ croissante} \iff f'' \geqslant 0, \\ &f \text{ strictement convexe} \iff f' \text{ strictement croissante} \iff f'' > 0, \\ &f \text{ fortement convexe} \iff f'' \geqslant \alpha > 0. \end{aligned}$$

Proposition 2.7 (Propriétés des fonctions convexes en dimension 1)**.** *Soit I un intervalle et $f : I \to \mathbb{R}$ convexe. Alors f est continue et localement lipschitzienne sur $\mathring{I}$. En tout point $a \in \mathring{I}$ (qui signifie l'intérieur de I), f admet une dérivée à gauche $f'_g(a)$ et une dérivée à droite $f'_d(a)$, pas forcément égales. De plus, pour tout $a, b \in \mathring{I}$, on a*

$$f'_g(a) \leqslant f'_d(a) \leqslant \frac{f(b) - f(a)}{b - a} \leqslant f'_g(b) \leqslant f'_d(b).$$

En particulier, les dérivées gauches et droites f'_g et f'_d sont croissantes. Nous ne donnons pas la preuve. La fonction f n'est pas forcément continue sur les bords de I.

2.3. **Exemples.**

Proposition 2.8 (Combinaisons linéaires et compositions)**.**

(1) *Soient $f_j : \mathbb{R}^d \to \mathbb{R}$ des fonctions convexes, et $\alpha_j \in \mathbb{R}_+$. Alors $\sum_{j=1}^n \alpha_j f_j$ est convexe.*

(2) *Soit $f : \mathbb{R}^d \to \mathbb{R}$ convexe et $g : \mathbb{R} \to \mathbb{R}$ convexe croissante. Alors $g \circ f$ est convexe.*

Preuve de (2). Comme $f(tx+(1-t)y) \leqslant tf(x)+(1-t)f(y)$ pour tout x, $y \in \mathbb{R}$, alors,

$$g \circ f(tx+(1-t)y) \underset{g \text{ croissante}}{\leqslant} g(tf(x)+(1-t)f(y)) \underset{g \text{ convexe}}{\leqslant} tg \circ f(x)+(1-t)g \circ f(y).$$

□

La composée de deux fonctions convexes n'est pas forcément convexe, par exemple $f(x) = -x$ et $g(x) = e^x$ sont convexes mais $g \circ f(x) = e^{-x}$ est concave.

(1) Les normes sont convexes

Proof. Soit $N : \mathbb{R}^d \to \mathbb{R}$ une norme. Pour $t \in [0,1]$ on a

$$N(tx+(1-t)y) \underset{\substack{\text{ineg.} \\ \text{triang.}}}{\leqslant} N(tx) + N((1-t)y) = tN(x) + (1-t)N(y).$$

□

(2) La fonction $f : \mathbb{R} \to \mathbb{R}$, $f(x) := x^2$ est fortement convexe avec $\alpha = 2$.

Proof. On a pour tout $x \in \mathbb{R}$, $\nabla f(x) = f'(x) = 2x$ et $f''(x) = 2 > 0$. □

(3) La fonction $f : \mathbb{R}_+^* \to \mathbb{R}$, $f(x) := x^p$ est convexe ssi $p \geqslant 1$ ou $p \leqslant 0$.

Proof. On a $f''(x) = p(p-1)x^{p-2}$. Or, $p(p-1) \geqslant 0$ ssi $p \geqslant 1$ ou $p \leqslant 0$. □

(4) La fonction $f : \mathbb{R} \to \mathbb{R}$, $f(x) := e^{\lambda x}$ avec $\lambda > 0$ est strictement convexe mais pas fortement convexe.

Proof. On a $f''(x) = \lambda^2 e^{\lambda x} > 0$ mais $f''(x) \to 0$ quand $x \to -\infty$. □

(5) La fonction $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) := \|x\|^2$ est fortement convexe avec $\alpha = 2$.

Proof. On a $\nabla f(x) = 2x$, on déduit $\langle \nabla f(y) - \nabla f(x), y-x \rangle = 2\|y-x\|^2$. □

(6) Plus généralement, soit f une forme quadratique avec les mêmes notations que la proposition 1.1. On a

$$\langle \nabla f(y) - \nabla f(x), y-x \rangle = \left\langle (A+A^T)(y-x), y-x \right\rangle = 2 \langle A(y-x), (y-x) \rangle .$$

Par conséquent,

(a) A est semi-définie positive $\Leftrightarrow$ f convexe.
(b) A est définie positive de plus petite valeur propre $\lambda_{\min} > 0 \Leftrightarrow f$ fortement convexe avec $\alpha = \lambda_{\min}$.

Laboratoire AGM - UMR 8088, CNRS - CY Cergy Paris Université, 2 avenue Adolphe Chauvin, Cergy-Pontoise, 95302 cedex, France

Email address: louis.garrigue@cyu.fr