

Fonctions convexes

Dans tout ce chapitre, C désigne un convexe non vide de $\mathbb{R}^n$.

1.1 Définitions et exemples

Définition 1. On dit qu'une fonction $f : C \rightarrow \mathbb{R}$ est convexe si

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

On dit que f est strictement convexe si

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y), \text{ pour tous } x, y \in C \text{ avec } x \neq y, \lambda \in (0, 1).$$

On dit que f est concave (resp. strictement concave) si $-f$ est convexe (resp. strictement convexe).

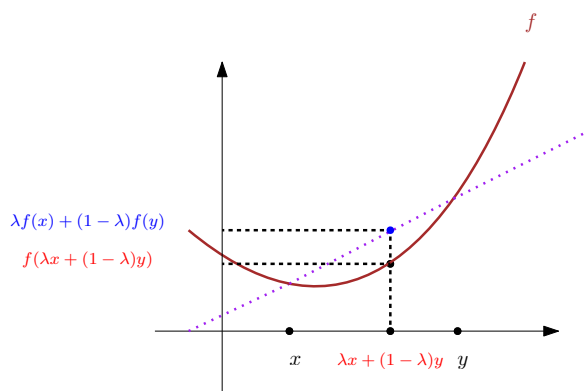


Figure 1.1: Illustration de l'inégalité de convexité de la Définition 1 : la corde liant les deux points $(x, f(x))$ et $(y, f(y))$ est au-dessus du graphe de f .

Remarque 1. 1. Une fonction est à la fois convexe et concave si et seulement si elle est affine sur C (i.e., de la forme $x \mapsto \langle a, x \rangle + b$ lorsque $C = \mathbb{R}^n$).

2. La convexité est une propriété "unidimensionnelle": f est convexe sur C si et seulement si, pour tous $x \in C$ et $d \in \mathbb{R}^n$, la fonction d'une variable $\varphi(t) = f(x + td)$ est convexe sur l'intervalle $I = \{t \in \mathbb{R} : x + td \in C\}$. En effet, $\varphi(\lambda t + (1 - \lambda)s) = f(\lambda(x + td) + (1 - \lambda)(x + sd))$, et réciproquement la restriction de f au segment $[x, y]$ correspond à $d = y - x$ et $t \in [0, 1]$. Pour vérifier qu'une fonction est convexe, il suffit donc de la regarder le long des droites.

Exemple 1. Voici quelques exemples de fonctions convexes. Les justifications seront données au fil du chapitre (notamment à l'aide des caractérisations de la Section 1.3).

1. Les fonctions affines $f(x) = \langle a, x \rangle + b$, avec $a \in \mathbb{R}^n$ et $b \in \mathbb{R}$ (à la fois convexes et concaves).

2. Les normes : $f(x) = \|x\|$ sur $\mathbb{R}^n$. En effet, par l'inégalité triangulaire et l'homogénéité,

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

3. Les fonctions quadratiques $f(x) = \frac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle + c$ avec $A \in \mathbb{S}_+^n$, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$ (Exemple 8). En particulier $f(x) = \|x\|^2$, qui est même strictement convexe.
4. La fonction $f(x) = -\log(x)$ est strictement convexe sur $\mathbb{R}_+^*$ ($f''(x) = 1/x^2 > 0$).
5. La fonction $f(x) = e^{\alpha x}$ est convexe sur $\mathbb{R}$ pour tout $\alpha \in \mathbb{R}$ ($f''(x) = \alpha^2 e^{\alpha x} \geq 0$), strictement convexe si $\alpha \neq 0$.
6. La fonction $f(x) = |x|^p$ est convexe sur $\mathbb{R}$ pour tout $p \geq 1$, strictement convexe si $p > 1$ (Exemple 5).
7. La fonction $f(x) = x \log(x)$ (avec $f(0) = 0$) est strictement convexe sur $\mathbb{R}_+$ ("entropie négative").
8. La fonction $f(x) = \max_{i=1, \dots, n} x_i$ est convexe sur $\mathbb{R}^n$ (Théorème 2).

Exemple 2 (Contre-exemples). 1. La fonction $x \mapsto x^3$ n'est pas convexe sur $\mathbb{R}$: avec $x = -2$, $y = 0$ et $\lambda = \frac{1}{2}$, on a $f(-1) = -1 > \frac{1}{2}(-8) + \frac{1}{2} \cdot 0 = -4$. Elle est en revanche convexe sur $\mathbb{R}_+$.

2. La fonction $x \mapsto \sqrt{x}$ est concave sur $\mathbb{R}_+$, donc pas convexe.

3. Une norme n'est jamais strictement convexe : pour $x \neq 0$, $y = 2x$ et $\lambda \in (0, 1)$, on a $\|\lambda x + (1 - \lambda)2x\| = (2 - \lambda)\|x\| = \lambda\|x\| + (1 - \lambda)\|2x\|$: la norme est affine le long des demi-droites issues de 0.

D'après la Définition 1, 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 (avec les mêmes coefficients) des valeurs $f(x)$ et $f(y)$. Cette propriété s'étend à la combinaison convexe d'un nombre quelconque (fini) de points.

Théorème 1 (Inégalité de Jensen). Soit $f : C \rightarrow \mathbb{R}$ une fonction convexe. Alors pour tous $x_1, \dots, x_m \in C$ et tout $\lambda \in \Delta_m$,

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

Preuve. La preuve suit exactement celle du ?? du chapitre 1. On raisonne par récurrence sur m ; le cas $m = 1$ est évident et le cas $m = 2$ est la définition. Supposons le résultat vrai pour m points et soient $x_1, \dots, x_{m+1} \in C$ et $\lambda \in \Delta_{m+1}$. Si $\lambda_{m+1} = 1$, il n'y a rien à démontrer. Sinon, on pose $\mu_i = \frac{\lambda_i}{1 - \lambda_{m+1}}$ pour $i = 1, \dots, m$, de sorte que $\mu \in \Delta_m$ et $y := \sum_{i=1}^m \mu_i x_i \in C$ (?). Par convexité de f , puis par hypothèse de récurrence,

$$f\left(\sum_{i=1}^{m+1} \lambda_i x_i\right) = f((1 - \lambda_{m+1})y + \lambda_{m+1}x_{m+1}) \leq (1 - \lambda_{m+1})f(y) + \lambda_{m+1}f(x_{m+1}) \leq \sum_{i=1}^m \lambda_i f(x_i) + \lambda_{m+1}f(x_{m+1}),$$

$$\text{car } (1 - \lambda_{m+1})f(y) \leq (1 - \lambda_{m+1}) \sum_{i=1}^m \mu_i f(x_i) = \sum_{i=1}^m \lambda_i f(x_i). \quad \square$$

Remarque 2. L'inégalité de convexité de la Définition 1 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 $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$; le Théorème 1 correspond au cas où X prend la valeur x_i avec probabilité λ_i . Plus généralement,

si $(\Omega, \mathcal{M}, \mu)$ est un espace probabilisé (où μ est une mesure positive avec $\mu(\Omega) = 1$), I un intervalle non vide de $\mathbb{R}$, $\phi : I \rightarrow \mathbb{R}$ une fonction convexe et $g : \Omega \rightarrow I$ une fonction dans $\mathcal{L}_\mu^1$, alors

$$\phi\left(\int_{\Omega} g d\mu\right) \leq \int_{\Omega} \phi \circ g d\mu.$$

Le résultat suivant est une conséquence immédiate du [Théorème 1](#).

Corollaire 1. Soit $f : C \rightarrow \mathbb{R}$ convexe. Pour toute combinaison convexe $x = \sum_{i=1}^m \lambda_i x_i$, avec $x_1, \dots, x_m \in C$ et $\lambda \in \Delta_m$, on a

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

Preuve. Par l'inégalité de Jensen,

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

□

Remarque 3. En particulier, une fonction convexe sur un polytope $\text{Conv}(\{x_1, \dots, x_m\})$ est majorée par $\max_i f(x_i)$. Nous utiliserons ce fait pour montrer qu'une fonction convexe est continue ([Section 1.4](#)), et nous y reviendrons dans la [Section 1.7](#).

1.2 Opérations préservant la convexité

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

- Théorème 2.**
1. Soit f une fonction convexe sur C et $\alpha \geq 0$. 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 $f : C \rightarrow \mathbb{R}$ convexe, $A \in \mathbb{R}^{n \times m}$ et $b \in \mathbb{R}^n$. Alors $g(y) = f(Ay + b)$ 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, pour tout $i = 1, \dots, m$,

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

donc

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

soit

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

donc f est convexe.

3. On a $f(\lambda x + (1 - \lambda)y) = \max_i f_i(\lambda x + (1 - \lambda)y) \leq \max_i \left(\lambda f_i(x) + (1 - \lambda)f_i(y) \right)$. On utilise alors le fait que, si $(\alpha_i)_{i=1}^m$ et $(\beta_i)_{i=1}^m$ sont deux familles de réels, $\max_i (\alpha_i + \beta_i) \leq$

$\max_i \alpha_i + \max_i \beta_i$. Donc

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

4. L'ensemble D est convexe, comme image réciproque du convexe C par l'application affine $y \mapsto Ay + b$ (?? du chapitre 1). Soient $y_1, y_2 \in D$ et $\lambda \in [0, 1]$. Les points $Ay_1 + b$ et $Ay_2 + b$ sont dans C et, par convexité de f ,

$$g(\lambda y_1 + (1 - \lambda)y_2) = f(\lambda(Ay_1 + b) + (1 - \lambda)(Ay_2 + b)) \leq \lambda f(Ay_1 + b) + (1 - \lambda)f(Ay_2 + b) = \lambda g(y_1) + (1 - \lambda)g(y_2),$$

d'où la convexité de g .

□

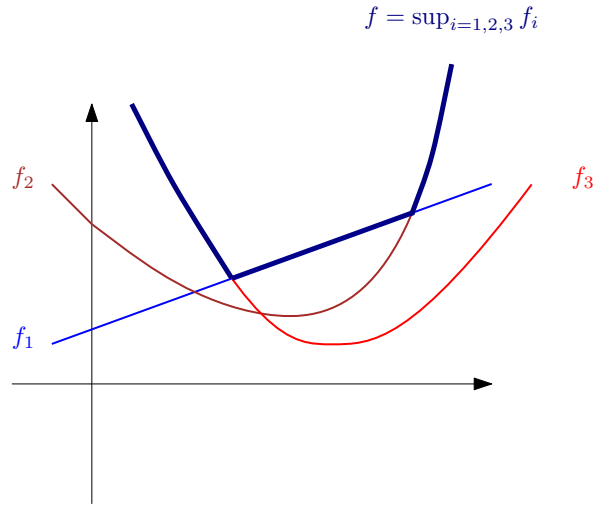


Figure 1.2: Le maximum de trois fonctions convexes f_1, f_2, f_3 est convexe.

Remarque 4. En revanche, le minimum de deux fonctions convexes n'est pas convexe en général : $\min(x^2, (x - 2)^2)$ n'est pas convexe sur $\mathbb{R}$ (elle vaut 0 en 0 et en 2, et 1 en 1). Le produit de deux fonctions convexes n'est pas convexe en général non plus : x et x^2 sont convexes, mais pas x^3 .

Exemple 3. 1. Pour $a_1, \dots, a_m \in \mathbb{R}^n$ et $b \in \mathbb{R}^m$, la fonction $f(x) = \max_{i=1, \dots, m} (\langle a_i, x \rangle + b_i)$ (maximum de fonctions affines, dite affine par morceaux) est convexe.

2. Pour $A \in \mathbb{R}^{m \times n}$ et $b \in \mathbb{R}^m$, les fonctions $x \mapsto \|Ax - b\|$ et $x \mapsto \|Ax - b\|^2$ sont convexes sur $\mathbb{R}^n$, comme composées d'une fonction convexe et d'une application affine.

3. Soit $f(x) = \frac{\|Ax + b\|^2}{\langle c, x \rangle + d}$ avec $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n \setminus \{0\}$ et $d \in \mathbb{R}$. Alors f est convexe sur $D = \{x \in \mathbb{R}^n : \langle c, x \rangle + d > 0\}$. En effet, on remarque que $f(x) = g(Ax + b, \langle c, x \rangle + d)$ avec $g(u, t) = \frac{\|u\|^2}{t}$, et $x \mapsto (Ax + b, \langle c, x \rangle + d)$ est affine. Or g est convexe sur $\mathbb{R}^m \times \mathbb{R}_+^*$ puisque $g(u, t) = \sum_{i=1}^m g_i(u, t)$ avec $g_i(u, t) = \frac{u_i^2}{t}$, et chaque g_i est convexe (Exemple 8).

Théorème 3. Soit $f : C \rightarrow \mathbb{R}$ une fonction convexe et $g : I \rightarrow \mathbb{R}$ une fonction convexe et croissante sur un intervalle I de $\mathbb{R}$. Supposons que $f(C) \subset I$. Alors $\phi(x) = g(f(x))$ est convexe sur C .

Preuve. Soient $x, y \in C$ et $\lambda \in [0, 1]$. Comme I est un intervalle, $\lambda f(x) + (1 - \lambda)f(y) \in I$, et

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

□

Exemple 4. 1. La fonction $\phi(x) = e^{\|x\|^2}$ est convexe sur $\mathbb{R}^n$. En effet, $\phi = g \circ f$ avec $g(t) = e^t$, qui est convexe et croissante, et $f(x) = \|x\|^2$, qui est convexe. Plus généralement, pour toute fonction convexe f , la fonction e^f est convexe.

2. Pour $p \geq 1$, la fonction $x \mapsto \|x\|^p$ est convexe : $g(t) = t^p$ est convexe et croissante sur $\mathbb{R}_+$.

3. L'hypothèse de croissance est nécessaire : $g(t) = e^{-t}$ est convexe (décroissante) et $f(x) = x^2$ est convexe, mais e^{-x^2} n'est pas convexe (elle admet un maximum strict en 0, cf. Section 1.7).

Exemple 5. Soit $p \geq 1$. La fonction $\varphi(t) = |t|^p$ est convexe sur $\mathbb{R}$, et strictement convexe si $p > 1$. En effet, $\varphi = g \circ |\cdot|$ avec $g(s) = s^p$ convexe et croissante sur $\mathbb{R}_+$: on conclut par le Théorème 3. Pour $p > 1$, la fonction g est strictement croissante et strictement convexe sur $\mathbb{R}_+$ (Proposition 2 : $g''(s) = p(p-1)s^{p-2} > 0$ pour $s > 0$). Soient $a \neq b$ et $\lambda \in (0, 1)$.

- Si $|a| \neq |b|$: $\varphi(\lambda a + (1 - \lambda)b) \leq g(\lambda|a| + (1 - \lambda)|b|) < \lambda g(|a|) + (1 - \lambda)g(|b|)$.
- Si $|a| = |b|$, alors $b = -a \neq 0$ et $|\lambda a + (1 - \lambda)b| = |2\lambda - 1||a| < |a|$, donc $\varphi(\lambda a + (1 - \lambda)b) < g(|a|) = \lambda \varphi(a) + (1 - \lambda)\varphi(b)$.

C'est ce résultat qui a été utilisé au chapitre 1 pour déterminer les points extrémaux des boules ℓ_p (??).

Soient $C \subset \mathbb{R}^n$ et $D \subset \mathbb{R}^k$ deux convexes et $f : C \times D \rightarrow \mathbb{R}$ une fonction convexe (sur le convexe $C \times D$ de $\mathbb{R}^{n+k}$). Le résultat suivant montre que la convexité est préservée par minimisation partielle.

Théorème 4. Supposons que $g(x) = \inf_{y \in D} f(x, y) > -\infty$ pour tout $x \in C$. Alors la fonction g est convexe sur C .

Preuve. Soient $x_1, x_2 \in C$ et $\varepsilon > 0$. Par définition de la borne inférieure (finie), il existe $y_1, y_2 \in D$ tels que $f(x_i, y_i) \leq g(x_i) + \varepsilon$ pour $i = 1, 2$. Soit $\lambda \in [0, 1]$. Comme $\lambda y_1 + (1 - \lambda)y_2 \in D$, on a

$$\begin{aligned}g(\lambda x_1 + (1 - \lambda)x_2) &= \inf_{y \in D} f(\lambda x_1 + (1 - \lambda)x_2, y) \\ &\leq f(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2) \\ &\leq \lambda f(x_1, y_1) + (1 - \lambda)f(x_2, y_2) \quad (\text{convexité de } f) \\ &\leq \lambda g(x_1) + (1 - \lambda)g(x_2) + \varepsilon.\end{aligned}$$

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. □

Exemple 6 (Distance à un convexe). Soit $S \subset \mathbb{R}^n$ un convexe non vide. La fonction distance à S ,

$$d_S(x) = \inf_{y \in S} \|x - y\|,$$

est convexe sur $\mathbb{R}^n$: c'est la minimisation partielle en y de $f(x, y) = \|x - y\|$, qui est convexe sur $\mathbb{R}^n \times S$ (composée de la norme et de l'application linéaire $(x, y) \mapsto x - y$), et $d_S \geq 0 > -\infty$.

L'hypothèse de convexité de S est essentielle : pour $S = \{-1, 1\} \subset \mathbb{R}$, $d_S(0) = 1 > \frac{1}{2}d_S(-1) + \frac{1}{2}d_S(1) = 0$.

On termine cette section par un lemme qui nous sera utile par la suite. Il traduit la convexité de f en termes de ce qui se passe au-delà d'un segment.

Lemme 1. Une fonction $f : C \rightarrow \mathbb{R}$ est convexe si et seulement si, pour tous $x, y \in C$ et tout $\alpha \geq 0$ tel que $y + \alpha(y - x) \in C$, on a

$$f(y + \alpha(y - x)) \geq f(y) + \alpha(f(y) - f(x)). \quad (1.1)$$

Preuve. Supposons (1.1) vérifiée. Soient $x, y \in C$, $t \in (0, 1]$ et $z = tx + (1 - t)y \in C$. On a $x = \frac{1}{t}z - \frac{1-t}{t}y = z + \alpha(z - y)$ avec $\alpha = \frac{1-t}{t} \geq 0$. En appliquant (1.1) au couple (y, z) , on obtient

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

i.e., $tf(x) \geq tf(z) + (1 - t)(f(z) - f(y)) = f(z) - (1 - t)f(y)$, soit $f(z) \leq tf(x) + (1 - t)f(y)$. Le cas $t = 0$ étant trivial, f est convexe.

Inversement, supposons f convexe, et soient $x, y \in C$ et $\alpha \geq 0$ tels que $z := y + \alpha(y - x) \in C$. Pour $t = \frac{\alpha}{1+\alpha} \in [0, 1]$, on a $y = (1 - t)z + tx$. Par convexité de f ,

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

En multipliant par $1 + \alpha > 0$, 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 (1.1). $\square$

Remarque 5. Géométriquement, le Lemme 1 dit qu'au-delà du segment $[x, y]$, le graphe de f est au-dessus de la droite passant par $(x, f(x))$ et $(y, f(y))$ (alors qu'il est en dessous entre x et y).

1.3 Caractérisations des fonctions convexes différentiables

Dans cette section, C est un *ouvert* convexe de $\mathbb{R}^n$, de sorte que les notions de différentiabilité du chapitre 0 ont un sens en tout point de C .

1.3.1 Caractérisations du premier ordre

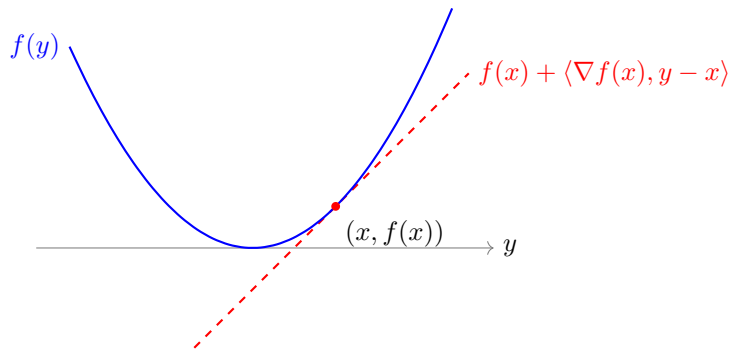


Figure 1.3: Illustration du Théorème 5 : le graphe d'une fonction convexe différentiable est au-dessus de chacune de ses tangentes.

Théorème 5. Soit $f : C \rightarrow \mathbb{R}$ une fonction de classe C^1 . Alors f est convexe si et seulement si

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \text{pour tous } x, y \in C. \quad (1.2)$$

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

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

ce qui donne

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

Quand $\lambda \rightarrow 0^+$, le membre de gauche tend vers la dérivée directionnelle $f'(x; y - x) = \langle \nabla f(x), y - x \rangle$ (?? du chapitre 0), d'où $f(x) + \langle \nabla f(x), y - x \rangle \leq f(y)$.

Supposons maintenant (1.2) vérifiée et montrons que f est convexe. Soient $u, v \in C$, $\lambda \in [0, 1]$ et $w = \lambda u + (1 - \lambda)v \in C$. En appliquant (1.2) au point w , avec $y = u$ puis $y = v$, on obtient

$$f(u) \geq f(w) + \langle \nabla f(w), u - w \rangle \quad \text{et} \quad f(v) \geq f(w) + \langle \nabla f(w), v - w \rangle.$$

En multipliant la première inégalité par λ , la seconde par $1 - \lambda$ et en sommant, on obtient

$$\lambda f(u) + (1 - \lambda)f(v) \geq f(w) + \underbrace{\langle \nabla f(w), \lambda u + (1 - \lambda)v - w \rangle}_{=0} = f(w),$$

ce qui est exactement l'inégalité de convexité. $\square$

Remarque 6. Dans la littérature, l'inégalité (1.2) est souvent appelée *the gradient inequality*. Elle affirme que les hyperplans tangents au graphe d'une fonction convexe sont situés sous ce graphe : une fonction convexe différentiable est minorée par chacune de ses approximations affines (Fig. 1.3). C'est une information *globale* déduite d'une quantité *locale*, le gradient en un point.

La conséquence suivante est fondamentale en optimisation et sera au cœur du chapitre suivant.

Corollaire 2. Soit $f : C \rightarrow \mathbb{R}$ convexe de classe C^1 et $x^* \in C$ tel que $\nabla f(x^*) = 0$. Alors x^* est un point de minimum global de f sur C : $f(y) \geq f(x^*)$ pour tout $y \in C$.

Preuve. Il suffit d'appliquer (1.2) avec $x = x^*$. $\square$

Remarque 7. Sans convexité, un point critique n'est pas forcément un minimum, même local : penser à $x \mapsto x^3$ en 0 ou à $(x, y) \mapsto x^2 - y^2$ en $(0, 0)$.

Pour la stricte convexité, on a la caractérisation analogue suivante.

Proposition 1. Soit $f : C \rightarrow \mathbb{R}$ une fonction de classe C^1 . Alors f est strictement convexe si et seulement si

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle \quad \text{pour tous } x, y \in C \text{ avec } x \neq y.$$

Preuve. Supposons f strictement convexe et soient $x \neq y$ dans C . Posons $z = \frac{x+y}{2}$. Par stricte convexité, puis par (1.2) appliquée en x ,

$$\frac{f(x) + f(y)}{2} > f(z) \geq f(x) + \langle \nabla f(x), z - x \rangle = f(x) + \frac{1}{2} \langle \nabla f(x), y - x \rangle.$$

En multipliant par 2 et en simplifiant, $f(y) > f(x) + \langle \nabla f(x), y - x \rangle$. La réciproque se démontre exactement comme dans le Théorème 5 : si $u \neq v$ et $\lambda \in (0, 1)$, alors $w \neq u$ et $w \neq v$, et les

deux inégalités sont strictes. □

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

Théorème 6 (Monotonie du gradient). Soit $f : C \rightarrow \mathbb{R}$ une fonction de classe C^1 . Alors f est convexe si et seulement si

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0 \quad \text{pour tous } x, y \in C. \quad (1.3)$$

Preuve. Supposons que f est convexe et soient $x, y \in C$. Par (1.2),

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

En sommant les deux inégalités, on obtient $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$.

Inversement, supposons que (1.3) est vérifiée et montrons que f est convexe. Soient $x, y \in C$ et posons $x_t = x + t(y - x) \in C$ pour $t \in [0, 1]$. D'après le théorème de Taylor avec reste intégral (?? du chapitre 0),

$$\begin{aligned} f(y) &= f(x) + \int_0^1 \langle \nabla f(x_t), y - x \rangle dt \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \langle \nabla f(x_t) - \nabla f(x), y - x \rangle dt \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \frac{1}{t} \underbrace{\langle \nabla f(x_t) - \nabla f(x), x_t - x \rangle}_{\geq 0} dt, \end{aligned}$$

car $x_t - x = t(y - x)$ (l'intégrande est bien défini et continu sur $(0, 1]$, et égal à $\langle \nabla f(x_t) - \nabla f(x), y - x \rangle$, qui se prolonge par continuité en $t = 0$). On obtient $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$, et f est convexe d'après le [Théorème 5](#). □

Remarque 8. En dimension 1, (1.3) s'écrit $(f'(x) - f'(y))(x - y) \geq 0$: on retrouve le résultat bien connu selon lequel une fonction dérivable sur un intervalle est convexe si et seulement si sa dérivée est croissante. De même, f est strictement convexe si et seulement si ∇f est strictement monotone, i.e., $\langle \nabla f(x) - \nabla f(y), x - y \rangle > 0$ pour $x \neq y$ (même preuve, en utilisant la [Proposition 1](#)).

Remarque 9 (Interprétation géométrique : un angle aigu). Soient $x, y \in C$ tels que $x \neq y$ et $\nabla f(x) \neq \nabla f(y)$. D'après l'inégalité de Cauchy-Schwarz (chapitre 0),

$$-1 \leq \frac{\langle \nabla f(x) - \nabla f(y), x - y \rangle}{\|\nabla f(x) - \nabla f(y)\| \|x - y\|} \leq 1,$$

il existe donc un unique $\theta = \theta(x, y) \in [0, \pi]$ tel que

$$\cos \theta = \frac{\langle \nabla f(x) - \nabla f(y), x - y \rangle}{\|\nabla f(x) - \nabla f(y)\| \|x - y\|}.$$

C'est l'angle entre le déplacement $x - y$ et la variation du gradient $\nabla f(x) - \nabla f(y)$ correspondante. La condition (1.3) s'écrit alors simplement

$$\cos \theta \geq 0, \quad \text{i.e.,} \quad \theta \in \left[0, \frac{\pi}{2}\right].$$

Autrement dit, f est convexe si et seulement si, lorsqu'on se déplace de y vers x , le gradient varie dans une direction qui fait un angle *aigu* (ou droit) avec la direction du déplacement : le gradient "suit" le déplacement et ne peut jamais revenir en arrière. En dimension 1, les seuls

angles possibles sont $\theta = 0$ et $\theta = \pi$, et on retrouve la croissance de f' . Pour une fonction strictement convexe, l'angle est strictement aigu : $\theta \in [0, \frac{\pi}{2})$.

On peut quantifier cet angle. Si f est μ -fortement convexe ([Proposition 3](#) ci-dessous), en écrivant l'inégalité de la [Proposition 3](#) en (x, y) puis en (y, x) et en sommant, on obtient

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2.$$

Si de plus ∇f est L -lipschitzien, i.e., $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$, alors

$$\cos \theta \geq \frac{\mu \|x - y\|^2}{L \|x - y\| \cdot \|x - y\|} = \frac{\mu}{L} :$$

l'angle reste uniformément éloigné de $\frac{\pi}{2}$. Le rapport $\frac{\mu}{L}$ jouera un rôle central dans la vitesse de convergence de la méthode du gradient.

Exemple 7 (Fonctions quadratiques). Soit $f(x) = \frac{1}{2} \langle Ax, x \rangle$ avec $A \in \mathbb{S}^n$. Alors $\nabla f(x) = Ax$, donc $\nabla f(x) - \nabla f(y) = A(x - y)$. En notant $d = x - y$, l'angle θ de la [Remarque 9](#) est l'angle entre d et Ad :

$$\cos \theta = \frac{\langle Ad, d \rangle}{\|Ad\| \|d\|}.$$

1. Si $A \succeq 0$, alors $\langle Ad, d \rangle \geq 0$ pour tout d : l'angle est toujours aigu ou droit, et on retrouve la convexité de f ([Exemple 8](#)). Par exemple, pour $A = \text{diag}(1, 4)$, i.e., $f(x) = \frac{1}{2}(x_1^2 + 4x_2^2)$, et $y = (0.4, 0.5)$, $x = (1.2, 0.2)$, on a $d = (0.8, -0.3)$ et $Ad = (0.8, -1.2)$, d'où

$$\cos \theta = \frac{0.64 + 0.36}{\sqrt{0.73} \sqrt{2.08}} \approx 0.81, \quad \text{soit } \theta \approx 36^\circ$$

([Fig. 1.4\(a\)](#)). Pour $A \succ 0$, on peut montrer (inégalité de Kantorovich) que $\cos \theta \geq \frac{2\sqrt{\lambda_{\min}(A)\lambda_{\max}(A)}}{\lambda_{\min}(A) + \lambda_{\max}(A)}$ pour tout $d \neq 0$; ici $\cos \theta \geq \frac{2 \cdot 2}{1+4} = 0.8$, soit $\theta \leq 37^\circ$ environ.

2. Si A a une valeur propre $\lambda < 0$, de vecteur propre u , alors pour $d = u$ on a $Ad = \lambda u$ et $\theta = \pi$: le gradient varie dans la direction *opposée* au déplacement. Par exemple, pour $f(x) = \frac{1}{2}(x_1^2 - x_2^2)$ ("selle de cheval") et $d = x - y = e_2$, on a $\nabla f(x) - \nabla f(y) = -e_2$ ([Fig. 1.4\(b\)](#)). La fonction f n'est pas convexe.

1.3.2 Caractérisations du second ordre

Quand la fonction est deux fois différentiable, la convexité de f est équivalente au fait que la matrice hessienne est semi-définie positive en tout point.

Théorème 7. Soit $f : C \rightarrow \mathbb{R}$ une fonction de classe C^2 sur l'ouvert convexe C . Alors f est convexe si et seulement si

$$\nabla^2 f(x) \succeq 0 \quad \text{pour tout } x \in C. \quad (1.4)$$

Preuve. Supposons que f est convexe et soient $x \in C$ et $d \in \mathbb{R}^n$. Comme C est ouvert, $x_t := x + td \in C$ pour $0 < t < \varepsilon$, avec $\varepsilon > 0$ suffisamment petit. Par monotonie de ∇f ([Théorème 6](#)),

$$0 \leq \frac{1}{t^2} \langle \nabla f(x_t) - \nabla f(x), x_t - x \rangle = \left\langle \frac{\nabla f(x + td) - \nabla f(x)}{t}, d \right\rangle.$$

Comme ∇f est différentiable de jacobienne $\nabla^2 f$, le quotient $\frac{1}{t}(\nabla f(x + td) - \nabla f(x))$ tend vers $\nabla^2 f(x)d$ quand $t \rightarrow 0^+$. On obtient $\langle \nabla^2 f(x)d, d \rangle \geq 0$. Comme d est arbitraire, $\nabla^2 f(x) \succeq 0$ pour tout $x \in C$.

Supposons maintenant (1.4) vérifiée et soient $x, y \in C$. D'après la formule de Taylor-Lagrange

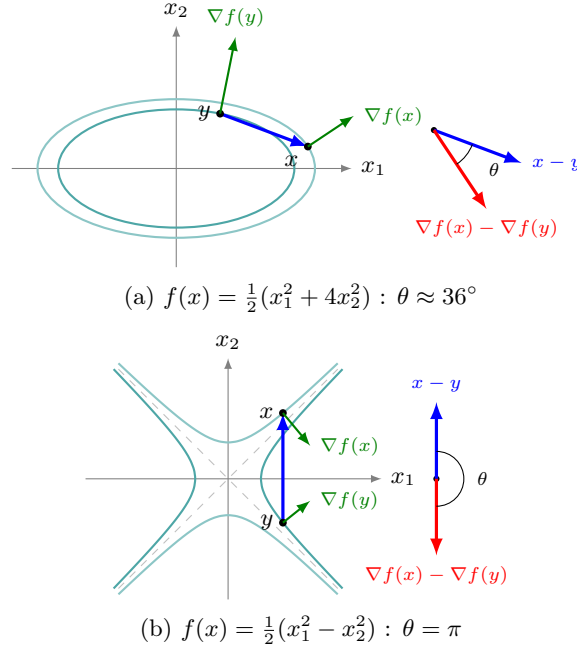


Figure 1.4: Interprétation géométrique de la monotonie du gradient (Remarque 9). En vert : les gradients en y et en x , orthogonaux aux lignes de niveau. Dans l'encart : le déplacement $x - y$ (bleu) et la variation du gradient $\nabla f(x) - \nabla f(y)$ (rouge). En (a), f est convexe et l'angle θ est aigu ; en (b), f n'est pas convexe et $\theta = \pi$.

(?? du chapitre 0), il existe $\xi \in [x, y]$ tel que

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \underbrace{\langle \nabla^2 f(\xi)(y - x), y - x \rangle}_{\geq 0} \geq f(x) + \langle \nabla f(x), y - x \rangle,$$

ce qui implique la convexité de f par le Théorème 5. □

Proposition 2. Soit $f : C \rightarrow \mathbb{R}$ une fonction de classe C^2 sur l'ouvert convexe C . Si $\nabla^2 f(x) \succ 0$ pour tout $x \in C$, alors f est strictement convexe.

Preuve. Même preuve que ci-dessus : pour $x \neq y$, le point ξ appartient à $[x, y] \subset C$ et $\langle \nabla^2 f(\xi)(y - x), y - x \rangle > 0$, donc $f(y) > f(x) + \langle \nabla f(x), y - x \rangle$, et on conclut par la Proposition 1. □

Remarque 10. 1. La réciproque de la Proposition 2 est fautive : $f(x) = x^4$ est strictement convexe sur $\mathbb{R}$, mais $f''(0) = 0$.

2. L'hypothèse C ouvert est utilisée dans le sens direct du Théorème 7, pour pouvoir dériver dans toutes les directions $d \in \mathbb{R}^n$. Elle est nécessaire : la fonction $f(x, y) = x^2 - y^2$, définie sur $\mathbb{R}^2$, est convexe sur la droite $C = \mathbb{R} \times \{0\}$, où elle vaut x^2 . Pourtant sa hessienne (calculée dans $\mathbb{R}^2$) en tout point de C est $\nabla^2 f(x, 0) = \text{diag}(2, -2)$, qui n'est pas semi-définie positive : le coefficient $-2 = \partial_{yy}^2 f$ correspond à la direction e_2 , qui sort de C , et sur laquelle la convexité de f sur C ne donne aucune information. Dans les directions de C , on a bien $\langle \nabla^2 f(x, 0)d, d \rangle = 2d_1^2 \geq 0$ pour $d = (d_1, 0)$.
3. En pratique, on utilise les critères du chapitre 0 (valeurs propres, critère de Sylvester) pour vérifier que $\nabla^2 f(x) \succeq 0$.

Exemple 8. 1. *Fonctions quadratiques.* Soit $f(x) = \frac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle + c$ avec $A \in \mathbb{S}^n$. On a $\nabla^2 f(x) = A$ pour tout x (chapitre 0). Donc f est convexe si et seulement si $A \succeq 0$, et strictement convexe si $A \succ 0$. Dans ce cas, on a même la réciproque : si $A \not\succeq 0$, il existe $d \neq 0$ avec $\langle Ad, d \rangle = 0$ et f est affine le long de la droite $\{x + td\}$, donc pas strictement convexe. Par exemple, $x \mapsto \frac{1}{2}\|Mx - y\|^2$ est convexe, de hessienne $M^T M \succeq 0$, et strictement convexe si et seulement si M est de rang n .

2. La fonction $g(u, t) = \frac{u^2}{t}$ sur $\mathbb{R} \times \mathbb{R}_+^*$. On calcule

$$\nabla^2 g(u, t) = \frac{2}{t^3} \begin{pmatrix} t^2 & -ut \\ -ut & u^2 \end{pmatrix} = \frac{2}{t^3} \begin{pmatrix} t \\ -u \end{pmatrix} \begin{pmatrix} t \\ -u \end{pmatrix}^T \succeq 0.$$

Donc g est convexe, mais pas strictement : $g(su, st) = s g(u, t)$ pour $s > 0$, donc g est linéaire le long de toute demi-droite issue de l'origine.

3. La fonction *log-somme-exp* $f(x) = \log(\sum_{i=1}^n e^{x_i})$ est convexe sur $\mathbb{R}^n$ (voir TD). C'est une approximation lisse de $\max_i x_i$: $\max_i x_i \leq f(x) \leq \max_i x_i + \log n$.

Remarque 11. La condition (1.4) est liée à la notion de courbure. Considérons la surface $\Gamma_f = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\}$ avec f de classe C^2 . La courbure de Gauss au point $(x, y, f(x, y))$ 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 des points de la surface : elliptiques ($\kappa > 0$), paraboliques ($\kappa = 0$) ou hyperboliques ($\kappa < 0$, points selle). Le graphe d'une fonction convexe n'a donc pas de point hyperbolique. La réciproque est fautive : le graphe d'une fonction concave a aussi une courbure de Gauss positive.

1.3.3 Convexité forte

Pour l'étude de la convergence des algorithmes (méthode du gradient, méthode de Newton), on aura besoin d'une notion plus forte que la stricte convexité.

Définition 2. Soit $\mu > 0$. On dit que $f : C \rightarrow \mathbb{R}$ est μ -fortement convexe si la fonction $x \mapsto f(x) - \frac{\mu}{2}\|x\|^2$ est convexe sur C .

Proposition 3. Soit $f : C \rightarrow \mathbb{R}$ et $\mu > 0$.

1. Si f est de classe C^1 , alors f est μ -fortement convexe si et seulement si

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2 \quad \text{pour tous } x, y \in C.$$

2. Si f est de classe C^2 , alors f est μ -fortement convexe si et seulement si $\nabla^2 f(x) \succeq \mu \text{Id}$ pour tout $x \in C$, i.e., $\lambda_{\min}(\nabla^2 f(x)) \geq \mu$ pour tout $x \in C$.

Preuve. On pose $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$, de sorte que $\nabla g(x) = \nabla f(x) - \mu x$ et $\nabla^2 g(x) = \nabla^2 f(x) - \mu \text{Id}$.

1. D'après le [Théorème 5](#), g est convexe si et seulement si $g(y) \geq g(x) + \langle \nabla f(x) - \mu x, y - x \rangle$ pour tous x, y , i.e.,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}(\|y\|^2 - \|x\|^2 - 2\langle x, y - x \rangle) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2.$$

2. D'après le [Théorème 7](#), g est convexe si et seulement si $\nabla^2 f(x) - \mu \text{Id} \succeq 0$ pour tout x , ce

qui équivaut à $\lambda_{\min}(\nabla^2 f(x)) \geq \mu$ (?? du chapitre 0).

□

- Remarque 12.** 1. Une fonction fortement convexe est strictement convexe (elle est somme d'une fonction convexe et de la fonction strictement convexe $\frac{\mu}{2} \|\cdot\|^2$). La réciproque est fautive : $x \mapsto e^x$ est strictement convexe sur $\mathbb{R}$, mais $f''(x) = e^x$ n'est minorée par aucun $\mu > 0$.
2. Une fonction quadratique $f(x) = \frac{1}{2} \langle Ax, x \rangle + \langle b, x \rangle + c$ avec $A \succ 0$ est μ -fortement convexe avec $\mu = \lambda_{\min}(A)$. Le premier point de la Proposition 3 dit qu'une fonction fortement convexe est minorée par une fonction quadratique (et non plus seulement affine) en chaque point.

1.4 Continuité et dérivées directionnelles des fonctions convexes

Dans cette section, C est de nouveau un convexe quelconque. On commence par un premier résultat quand $C = \mathbb{R}^n$.

Théorème 8. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction convexe. Alors f est continue en tout point $x \in \mathbb{R}^n$.

Preuve. Quitte à remplacer f par $y \mapsto f(x+y)$ (qui est convexe), on peut supposer que $x = 0$. Soit $(x_k)_k$ une suite de $\mathbb{R}^n \setminus \{0\}$ telle que $x_k \rightarrow 0$ quand $k \rightarrow \infty$; on peut supposer $\|x_k\| \leq 1$ pour tout k . Comme $x_k = (1 - \|x_k\|) \cdot 0 + \|x_k\| \frac{x_k}{\|x_k\|}$, on a par convexité de f

$$f(x_k) \leq (1 - \|x_k\|)f(0) + \|x_k\|f\left(\frac{x_k}{\|x_k\|}\right).$$

Les coordonnées de $\frac{x_k}{\|x_k\|}$ sont dans $[-1, 1]$, donc $\frac{x_k}{\|x_k\|}$ appartient au cube $[-1, 1]^n$, qui est l'enveloppe convexe de ses 2^n sommets $v_1, \dots, v_{2^n} \in \{-1, 1\}^n$ (?? et ?? du chapitre 1). On déduit du Corollaire 1 que $f(\frac{x_k}{\|x_k\|}) \leq \max_{j=1, \dots, 2^n} f(v_j) =: K$, et de même $f(-\frac{x_k}{\|x_k\|}) \leq K$. Il s'ensuit que

$$\limsup_{k \rightarrow \infty} f(x_k) \leq \limsup_{k \rightarrow \infty} \left((1 - \|x_k\|)f(0) + \|x_k\|K \right) = f(0).$$

De même, en remarquant que $0 = \frac{1}{1 + \|x_k\|}x_k + \frac{\|x_k\|}{1 + \|x_k\|}\left(-\frac{x_k}{\|x_k\|}\right)$, on a

$$f(0) \leq \frac{1}{1 + \|x_k\|}f(x_k) + \frac{\|x_k\|}{1 + \|x_k\|}f\left(-\frac{x_k}{\|x_k\|}\right) \leq \frac{1}{1 + \|x_k\|}f(x_k) + \frac{\|x_k\|}{1 + \|x_k\|}K,$$

soit $f(x_k) \geq (1 + \|x_k\|)f(0) - \|x_k\|K$. On en déduit que $f(0) \leq \liminf_{k \rightarrow \infty} f(x_k)$, et par suite $\lim_{k \rightarrow \infty} f(x_k) = f(0)$. D'où la continuité de f en 0. □

Sur un convexe $C \subsetneq \mathbb{R}^n$, on peut obtenir un résultat similaire (meilleur même) au Théorème 8 en tout point intérieur à C : une fonction convexe est lipschitzienne au voisinage de tout $x \in \text{int}(C)$. La raison de se restreindre aux points intérieurs est le comportement d'une fonction convexe au bord, qui peut présenter des discontinuités. Pour illustrer ceci, considérons la fonction $f : [0, 1] \rightarrow \mathbb{R}$ définie par $f(0) = 1$ et $f(x) = x^2$ pour $x \in (0, 1]$. Cette fonction est convexe (faites un dessin : toute corde reste au-dessus du graphe) mais n'est pas continue en 0.

On a le résultat suivant.

Théorème 9. Soit $f : C \rightarrow \mathbb{R}$ une fonction convexe 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

$$|f(x) - f(x_0)| \leq L\|x - x_0\| \quad \text{pour tout } x \in B(x_0, \varepsilon). \quad (1.5)$$

En particulier, f est continue sur $\text{int}(C)$.

Preuve. Soit $x_0 \in \text{int}(C)$. Il existe alors $\varepsilon > 0$ tel que le cube $B_\infty(x_0, \varepsilon) = x_0 + [-\varepsilon, \varepsilon]^n$ soit contenu dans C (les normes sont équivalentes). On commence par montrer que f est majorée sur $B_\infty(x_0, \varepsilon)$. Ce cube est un polytope, dont les points extrémaux sont

$$z_j = x_0 + \varepsilon \theta_j, \quad \theta_j \in \{-1, 1\}^n, \quad j = 1, 2, \dots, 2^n.$$

D'après la ?? du chapitre 1 (ou le théorème de Minkowski), tout $x \in B_\infty(x_0, \varepsilon)$ s'écrit $x = \sum_{j=1}^{2^n} \lambda_j z_j$ avec $\lambda \in \Delta_{2^n}$. Donc, par le [Corollaire 1](#), $f(x) \leq K := \max_{j=1, \dots, 2^n} f(z_j)$. Comme $B(x_0, \varepsilon) \subset B_\infty(x_0, \varepsilon)$, on en déduit que $f(x) \leq K$ pour tout $x \in B(x_0, \varepsilon)$; en particulier $K \geq f(x_0)$.

Soit $x \in B(x_0, \varepsilon)$ avec $x \neq x_0$. On pose $\alpha = \|x - x_0\|/\varepsilon \in (0, 1]$ et $z = x_0 + \alpha^{-1}(x - x_0)$. On a $\|z - x_0\| = \varepsilon$, 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

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

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 $f(x) - f(x_0) \geq -L\|x - x_0\|$. Posons $w = x_0 + \alpha^{-1}(x_0 - x)$ et remarquons que $\|w - x_0\| = \varepsilon$, donc $f(w) \leq K$. Comme $x = x_0 + \alpha(x_0 - w)$, on a par le [Lemme 1](#) (appliqué au couple (w, x_0))

$$f(x) = f(x_0 + \alpha(x_0 - w)) \geq f(x_0) + \alpha(f(x_0) - f(w)) \geq f(x_0) - \alpha(K - f(x_0)),$$

i.e., $f(x) - f(x_0) \geq -L\|x - x_0\|$. En conclusion, (1.5) est vérifiée pour tout $x \in B(x_0, \varepsilon)$ avec $L = (K - f(x_0))\varepsilon^{-1}$ (si $K = f(x_0)$, f est constante sur la boule et tout $L > 0$ convient). $\square$

Remarque 13. 1. Le [Théorème 8](#) est le cas particulier $C = \mathbb{R}^n$ du [Théorème 9](#). On peut montrer un peu mieux : avec les notations de la preuve, f est L' -lipschitzienne sur $B(x_0, \varepsilon/2)$, i.e., $|f(x) - f(y)| \leq L'\|x - y\|$ pour tous x, y de cette boule.

2. Dans la preuve, la convexité sert deux fois : pour majorer f sur un cube (à l'aide de ses sommets), puis pour transformer cette majoration en une estimation lipschitzienne. Une fonction convexe majorée au voisinage d'un point est donc automatiquement "régulière" en ce point.

Nous avons vu qu'en tout point intérieur à C , une fonction convexe $f : C \rightarrow \mathbb{R}$ vérifie (1.5). Montrons maintenant que les dérivées directionnelles de f existent en tout point intérieur. Rappelons (chapitre 0) qu'étant donnés $x \in \text{int}(C)$ et $d \in \mathbb{R}^n$, on dit que f admet une dérivée directionnelle en x dans la direction d si la limite

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

existe dans $\mathbb{R}$.

Théorème 10. Soit $f : C \rightarrow \mathbb{R}$ une fonction convexe et soit $x \in \text{int}(C)$. Alors $f'(x; d)$ existe pour tout $d \in \mathbb{R}^n$, et

$$f'(x; d) = \inf_{t \in T} \frac{f(x + td) - f(x)}{t}, \quad \text{où } T = \{t > 0 : x + td \in C\}.$$

De plus, $f(x + td) \geq f(x) + tf'(x; d)$ pour tout $t \geq 0$ tel que $x + td \in C$.

Preuve. Soit $x \in \text{int}(C)$ et $d \in \mathbb{R}^n$. L'ensemble T est un intervalle (car C est convexe) contenant $(0, \varepsilon]$ pour un certain $\varepsilon > 0$ (car $x \in \text{int}(C)$), et on pose $h(t) := \frac{f(x+td)-f(x)}{t}$ pour $t \in T$.

h est croissante. Soient $0 < s < t$ dans T et $\alpha = s/t \in (0, 1)$. Comme $x+sd = (1-\alpha)x + \alpha(x+td)$, on a par convexité

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

soit $f(x+sd) - f(x) \leq \frac{s}{t}(f(x+td) - f(x))$, i.e., $h(s) \leq h(t)$.

h est minorée. Comme $x \in \text{int}(C)$, il existe $t_0 > 0$ tel que $x - t_0d \in C$. Pour $t \in T$, on a $x = \frac{t}{t+t_0}(x - t_0d) + \frac{t_0}{t+t_0}(x + td)$, donc par convexité

$$f(x) \leq \frac{t}{t+t_0}f(x - t_0d) + \frac{t_0}{t+t_0}f(x + td),$$

soit, en multipliant par $\frac{t+t_0}{tt_0}$ et en réarrangeant, $h(t) \geq \frac{f(x)-f(x-t_0d)}{t_0}$.

Une fonction croissante et minorée admet une limite finie en 0^+ , égale à sa borne inférieure : $f'(x; d) = \lim_{t \rightarrow 0^+} h(t) = \inf_{t \in T} h(t)$. Enfin $h(t) \geq f'(x; d)$ donne $f(x+td) \geq f(x) + tf'(x; d)$. $\square$

Remarque 14. 1. Si f est de plus différentiable en x , alors $f'(x; d) = \langle \nabla f(x), d \rangle$, et la dernière inégalité du [Théorème 10](#) redonne l'inégalité (1.2) (avec $y = x + d$).

2. Les dérivées directionnelles existent toujours, mais f n'est pas forcément différentiable : pour $f(x) = \|x\|$, on a $f'(0; d) = \|d\|$, qui n'est pas linéaire en d . En dimension 1, une fonction convexe admet en tout point intérieur une dérivée à droite $f'_d(x) = f'(x; 1)$ et une dérivée à gauche $f'_g(x) = -f'(x; -1)$, avec $f'_g(x) \leq f'_d(x)$; pour $f(x) = |x|$ en 0 : $f'_g(0) = -1 < 1 = f'_d(0)$.

3. On peut montrer qu'une fonction convexe est différentiable "presque partout" sur $\text{int}(C)$ (théorème de Rademacher). Les points de non-différentiabilité seront étudiés en 2A à l'aide de la notion de sous-différentiel.

1.5 Sous-ensembles de niveau de fonctions convexes

Définition 3. Soit $f : S \rightarrow \mathbb{R}$ une fonction définie sur une partie S de $\mathbb{R}^n$. Le sous-ensemble de niveau $\alpha \in \mathbb{R}$ de f est l'ensemble

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

On a le résultat suivant.

Théorème 11. Soit $f : C \rightarrow \mathbb{R}$ une fonction convexe. Alors pour tout $\alpha \in \mathbb{R}$, l'ensemble $\text{Lev}(f, \alpha)$ est convexe.

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

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

i.e., $\lambda x + (1-\lambda)y \in \text{Lev}(f, \alpha)$. $\square$

Exemple 9. 1. Pour $f(x) = \|x\|$, les sous-ensembles de niveau sont les boules fermées $B(0, \alpha)$ (pour $\alpha \geq 0$) : on retrouve leur convexité.

2. Pour $f(x, y) = \frac{1}{2}(x^2 + 4y^2)$, les sous-ensembles de niveau $\alpha > 0$ sont des ellipses pleines ([Fig. 1.5](#)).

3. Si $g_1, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$ sont convexes, l'ensemble $\{x \in \mathbb{R}^n : g_1(x) \leq 0, \dots, g_m(x) \leq 0\} = \bigcap_i \text{Lev}(g_i, 0)$ est convexe. C'est sous cette forme qu'apparaîtront les ensembles de contraintes en optimisation.

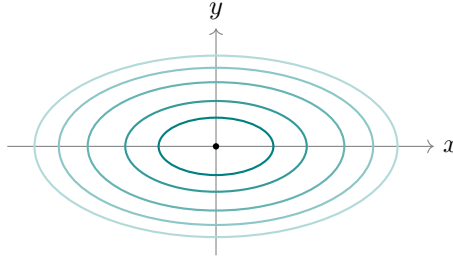


Figure 1.5: Lignes de niveau de la fonction $f(x, y) = \frac{1}{2}(x^2 + 4y^2)$, pour $\alpha \in \{0.2, 0.5, 1, 1.5, 2\}$. Les sous-ensembles de niveau $\text{Lev}(f, \alpha)$ sont les ellipses pleines qu'elles délimitent.

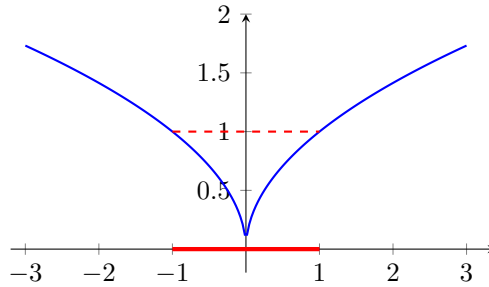


Figure 1.6: $f(x) = \sqrt{|x|}$: une fonction quasi-convexe non convexe. En rouge : $\text{Lev}(f, 1) = [-1, 1]$.

Remarque 15. La réciproque du [Théorème 11](#) est fautive, comme le montre la fonction $f(x) = \sqrt{|x|}$ ([Fig. 1.6](#)), qui n'est pas convexe mais dont tous les sous-ensembles de niveau sont convexes. En effet, pour $\alpha < 0$, $\text{Lev}(f, \alpha) = \emptyset$, et pour $\alpha \geq 0$, $\text{Lev}(f, \alpha) = [-\alpha^2, \alpha^2]$, qui est convexe. Une telle fonction est dite [quasi-convexe](#).

Définition 4. On dit qu'une fonction $f : C \rightarrow \mathbb{R}$ est quasi-convexe si pour tout $\alpha \in \mathbb{R}$, l'ensemble $\text{Lev}(f, \alpha)$ est convexe. On dit que f est quasi-concave si $-f$ est quasi-convexe.

Proposition 4. Une fonction $f : C \rightarrow \mathbb{R}$ est quasi-convexe si et seulement si

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

Preuve. Supposons f quasi-convexe et soient $x, y \in C$. Avec $\alpha = \max(f(x), f(y))$, les points x et y appartiennent au convexe $\text{Lev}(f, \alpha)$, donc $\lambda x + (1 - \lambda)y$ aussi, i.e., $f(\lambda x + (1 - \lambda)y) \leq \alpha$. Réciproquement, si $x, y \in \text{Lev}(f, \alpha)$, alors $f(\lambda x + (1 - \lambda)y) \leq \max(f(x), f(y)) \leq \alpha$. $\square$

Exemple 10. 1. Évidemment, une fonction convexe est quasi-convexe ($\lambda f(x) + (1 - \lambda)f(y) \leq \max(f(x), f(y))$).

2. L'application $x \in \mathbb{R}_+^n \mapsto \text{card}(x)$ (nombre de composantes non nulles de x) est quasi-concave sur $\mathbb{R}_+^n$: pour $x, y \in \mathbb{R}_+^n$ et $\lambda \in (0, 1)$, les composantes non nulles de $\lambda x + (1 - \lambda)y$ sont celles de x et celles de y , donc

$$\text{card}(\lambda x + (1 - \lambda)y) \geq \max\{\text{card}(x), \text{card}(y)\} \geq \min\{\text{card}(x), \text{card}(y)\}.$$

La positivité est importante : sur $\mathbb{R}^n$, $\text{card}(\frac{1}{2}e_1 + \frac{1}{2}(-e_1)) = 0$.

3. De même, l'application $M \mapsto \text{rang}(M)$ est quasi-concave sur $\mathbb{S}_+^n$: pour $M, N \in \mathbb{S}_+^n$ et $\lambda \in (0, 1)$, $\text{rang}(\lambda M + (1 - \lambda)N) \geq \min\{\text{rang}(M), \text{rang}(N)\}$ (voir TD).
4. Un exemple important est celui des fonctions continues sur $\mathbb{R}$: une fonction continue $f : \mathbb{R} \rightarrow \mathbb{R}$ est quasi-convexe si et seulement si l'une des propriétés suivantes est vérifiée :
 - (a) f est croissante ;
 - (b) f est décroissante ;
 - (c) il existe $x^* \in \mathbb{R}$ tel que f est décroissante sur $(-\infty, x^*]$ et croissante sur $[x^*, +\infty)$ (cf. Fig. 1.7).

En particulier, toute fonction monotone sur $\mathbb{R}$ est quasi-convexe.

5. Contrairement à la convexité, la quasi-convexité n'est pas stable par somme : $x \mapsto x^3$ et $x \mapsto -x$ sont quasi-convexes (monotones), mais $x \mapsto x^3 - x$ ne l'est pas, puisque $\text{Lev}(x^3 - x, 0) = (-\infty, -1] \cup [0, 1]$ n'est pas un intervalle.

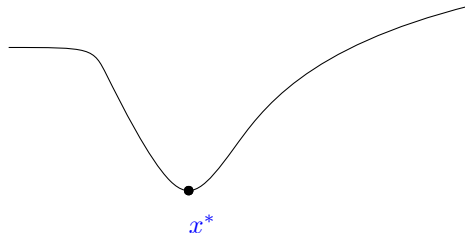


Figure 1.7: Exemple de fonction quasi-convexe sur $\mathbb{R}$.

1.6 Fonctions à valeurs réelles étendues

En pratique, on est amené à considérer des fonctions qui peuvent prendre des valeurs infinies, par exemple pour encoder une contrainte $x \in C$ dans la fonction elle-même. On considère donc des fonctions $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} = [-\infty, +\infty]$, avec les opérations arithmétiques classiques

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

ainsi que la règle moins usuelle $0 \cdot \infty = 0 \cdot (-\infty) = 0$. L'expression $\infty - \infty$ n'est pas définie. Dans ce cours, on s'intéressera surtout à des fonctions qui ne prennent pas la valeur $-\infty$ et dont le domaine (défini ci-dessous) est non vide.

Définition 5 (Domaine effectif). Soit $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$. Le domaine (ou domaine effectif) de f est

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

Exemple 11. Si $C \subset \mathbb{R}^n$, l'indicatrice de C (au sens de l'analyse convexe) est définie par

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

On a $\text{dom}(\delta_C) = C$. Elle permet d'écrire un problème sous contrainte comme un problème sans

contrainte : minimiser f sur C revient à minimiser $f + \delta_C$ sur $\mathbb{R}^n$.

Définition 6. On dit que $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ est propre si $f(x) > -\infty$ pour tout $x \in \mathbb{R}^n$ et $\text{dom}(f) \neq \emptyset$, i.e., s'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 7 (Épigraphe). Soit $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$. On définit l'épigraphe de f par

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

Pour une fonction $f : C \rightarrow \mathbb{R}$, on pose de même $\text{epi}(f) = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} : x \in C, f(x) \leq t \right\}$.

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

Théorème 12. 1. Une fonction $f : C \rightarrow \mathbb{R}$ est convexe si et seulement si $\text{epi}(f)$ est convexe.
2. Plus généralement, pour $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, l'inégalité de convexité de la Définition 1 (avec les règles de calcul ci-dessus) est vérifiée pour tous $x, y \in \mathbb{R}^n$ si et seulement si $\text{epi}(f)$ est convexe.

Pour une fonction à valeurs dans $\overline{\mathbb{R}}$, on définit la convexité de f par la convexité de $\text{epi}(f)$.

Preuve. On démontre le deuxième point, le premier s'en déduisant en prolongeant f par $+\infty$ en dehors de C (ce qui ne change pas l'épigraphe).

Supposons l'inégalité de convexité vérifiée et soient $\begin{pmatrix} x \\ s \end{pmatrix}, \begin{pmatrix} y \\ t \end{pmatrix} \in \text{epi}(f)$ et $\lambda \in [0, 1]$. Alors

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

donc $\lambda \begin{pmatrix} x \\ s \end{pmatrix} + (1 - \lambda) \begin{pmatrix} y \\ t \end{pmatrix} \in \text{epi}(f)$. Réciproquement, supposons $\text{epi}(f)$ convexe et soient $x, y \in \mathbb{R}^n$ et $\lambda \in (0, 1)$ (les cas $\lambda \in \{0, 1\}$ sont évidents). Si $x \notin \text{dom}(f)$ ou $y \notin \text{dom}(f)$, le membre de droite vaut $+\infty$ et il n'y a rien à démontrer. Sinon, $\begin{pmatrix} x \\ f(x) \end{pmatrix}$ et $\begin{pmatrix} y \\ f(y) \end{pmatrix}$ sont dans $\text{epi}(f)$, donc leur combinaison convexe aussi : $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$. $\square$

Remarque 16. En particulier, $f : C \rightarrow \mathbb{R}$ est convexe si et seulement si son prolongement par $+\infty$ en dehors de C est convexe sur $\mathbb{R}^n$, et $\text{dom}(f)$ est alors convexe. On pourra donc indifféremment travailler avec des fonctions convexes définies sur un convexe C ou avec des fonctions convexes $\mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$.

Exemple 12. 1. Si $f(x) = \langle a, x \rangle - \alpha$ avec $a \in \mathbb{R}^n$ et $\alpha \in \mathbb{R}$, alors

$$\text{epi}(f) = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \left\langle \begin{pmatrix} a \\ -1 \end{pmatrix}, \begin{pmatrix} x \\ t \end{pmatrix} \right\rangle \leq \alpha \right\}$$

est un demi-espace, donc convexe.

2. Si $f(x) = \frac{1}{2}\|x\|^2$, alors $\text{epi}(f)$ est la région de l'espace située au-dessus du paraboloïde $t = \frac{1}{2}\|x\|^2$ (Fig. 1.8 pour $n = 1$).
3. Pour l'indicatrice d'un ensemble C ,

$$\text{epi}(\delta_C) = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \delta_C(x) \leq t \right\} = C \times \mathbb{R}_+,$$

qui est convexe si et seulement si C est convexe : δ_C est convexe si et seulement si C est convexe.

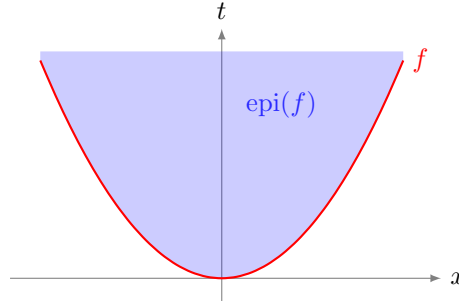


Figure 1.8: L'épigraph de $f(x) = \frac{1}{2}x^2$.

Le résultat suivant montre que le supremum d'une famille *quelconque* de fonctions convexes est convexe, et est à comparer avec le [Théorème 2](#).

Théorème 13. Soient $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$, $i \in I$, des fonctions convexes (la famille d'indices I est quelconque). Alors la fonction $f(x) = \sup_{i \in I} f_i(x)$ est convexe.

Preuve. On a $f(x) \leq t$ si et seulement si $f_i(x) \leq t$ pour tout $i \in I$, donc $\text{epi}(f) = \bigcap_{i \in I} \text{epi}(f_i)$. Comme les f_i sont convexes, les épigraphes $\text{epi}(f_i)$ sont convexes ([Théorème 12](#)), et une intersection quelconque de convexes est convexe (?? du chapitre 1). $\square$

Exemple 13. 1. *Fonction d'appui.* Pour $S \subset \mathbb{R}^n$ non vide, la fonction $\sigma_S(x) = \sup_{y \in S} \langle x, y \rangle$ est convexe (supremum de fonctions linéaires). Par exemple, $\sigma_{B(0,1)}(x) = \|x\|$ (Cauchy-Schwarz).
 2. *Plus grande valeur propre.* La fonction $A \in \mathbb{S}^n \mapsto \lambda_{\max}(A) = \max_{\|x\|=1} \langle Ax, x \rangle$ (?? du chapitre 0) est convexe sur $\mathbb{S}^n$, comme supremum des fonctions linéaires $A \mapsto \langle Ax, x \rangle$.
 3. *Conjuguée.* Pour $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ propre, la fonction $f^*(y) = \sup_{x \in \mathbb{R}^n} (\langle x, y \rangle - f(x))$ est convexe, même si f ne l'est pas (voir le cours de 2A).

Définition 8 (Fonction fermée). Une fonction $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ est dite fermée si $\text{epi}(f)$ est fermé.

Exemple 14. Revenons à l'exemple de la fonction indicatrice δ_C d'un sous-ensemble C de $\mathbb{R}^n$. Comme $\text{epi}(\delta_C) = C \times \mathbb{R}_+$, on a

$$\delta_C \text{ est fermée} \Leftrightarrow C \times \mathbb{R}_+ \text{ est fermé} \Leftrightarrow C \text{ est fermé.}$$

Définition 9 (Semi-continuité inférieure). Une fonction $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ est dite semi-continue inférieurement (sci) en $x \in \mathbb{R}^n$ si

$$f(x) \leq \liminf_{k \rightarrow \infty} f(x_k)$$

pour toute suite $(x_k)_k$ telle que $x_k \rightarrow x$ quand $k \rightarrow \infty$. On dit que f est sci si elle est sci en tout point de $\mathbb{R}^n$.

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

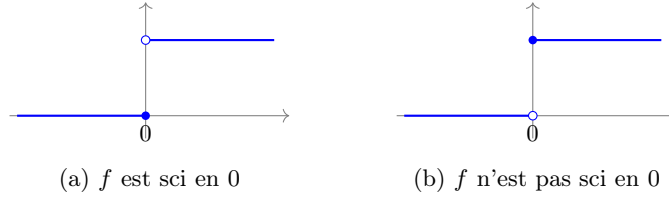


Figure 1.9: Une fonction est sci en un point si, à une discontinuité, elle prend la valeur "basse". À gauche, $\text{epi}(f)$ est fermé ; à droite, il ne l'est pas.

- Exemple 15.** 1. Toute fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est sci. En particulier, toute fonction convexe $f : \mathbb{R}^n \rightarrow \mathbb{R}$ appartient à $\Gamma_0(\mathbb{R}^n)$ ([Théorème 8](#)).
2. Si C est un convexe fermé non vide, $\delta_C \in \Gamma_0(\mathbb{R}^n)$.
3. La fonction convexe f égale à x^2 sur $(0, 1]$, à 1 en 0 et à $+\infty$ en dehors de $[0, 1]$ ([Section 1.4](#)) n'est pas sci en 0 : $f(0) = 1 > 0 = \lim_{k \rightarrow \infty} f(\frac{1}{k})$.

Le résultat suivant établit un lien entre la semi-continuité inférieure d'une fonction et la fermeture de son épigraphe et de ses sous-ensembles de niveau.

Théorème 14. Soit $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$. Les assertions suivantes sont équivalentes :

1. f est sci ;
2. f est fermée ;
3. $\text{Lev}(f, \alpha) = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ est fermé pour tout $\alpha \in \mathbb{R}$.

Preuve. (1) $\Rightarrow$ (2). Soit $\begin{pmatrix} x_k \\ t_k \end{pmatrix} \in \text{epi}(f)$ une suite convergeant vers $\begin{pmatrix} x \\ t \end{pmatrix}$. Alors $f(x_k) \leq t_k$ et, par semi-continuité inférieure et les propriétés de la limite inférieure (?? du chapitre 0),

$$f(x) \leq \liminf_{k \rightarrow \infty} f(x_k) \leq \liminf_{k \rightarrow \infty} t_k = t,$$

donc $\begin{pmatrix} x \\ t \end{pmatrix} \in \text{epi}(f)$.

(2) $\Rightarrow$ (3). Soit $\alpha \in \mathbb{R}$ et $(x_k)_k$ une suite de $\text{Lev}(f, \alpha)$ convergeant vers x . Alors $\begin{pmatrix} x_k \\ \alpha \end{pmatrix} \in \text{epi}(f)$ converge vers $\begin{pmatrix} x \\ \alpha \end{pmatrix}$, qui appartient donc à $\text{epi}(f)$: $f(x) \leq \alpha$, i.e., $x \in \text{Lev}(f, \alpha)$.

(3) $\Rightarrow$ (1). Supposons par l'absurde qu'il existe une suite $x_k \rightarrow x$ telle que $\liminf_k f(x_k) < f(x)$, et choisissons $\alpha \in \mathbb{R}$ tel que $\liminf_k f(x_k) < \alpha < f(x)$. Il existe alors une sous-suite $(x_{k_j})_j$ telle que $f(x_{k_j}) \leq \alpha$ pour tout j , i.e., $x_{k_j} \in \text{Lev}(f, \alpha)$. Comme $\text{Lev}(f, \alpha)$ est fermé et $x_{k_j} \rightarrow x$, on obtient $x \in \text{Lev}(f, \alpha)$, soit $f(x) \leq \alpha$: contradiction. $\square$

1.7 Maxima de fonctions convexes

Avant de passer au chapitre suivant sur l'optimisation convexe, dans lequel on s'intéressera essentiellement à la *minimisation* de fonctions convexes, on dégage quelques propriétés des *maxima* de fonctions convexes sur un convexe. Elles sont très différentes.

Théorème 15. Soit $f : C \rightarrow \mathbb{R}$ une fonction convexe. Si f atteint son maximum sur C en un point de $\text{int}(C)$, alors f est constante sur C .

Preuve. Supposons qu'il existe $x^* \in \text{int}(C)$ tel que $f(x^*) \geq f(x)$ pour tout $x \in C$, et que f ne soit pas constante : il existe $x_* \in C$ tel que $f(x_*) < f(x^*)$. Comme x^* est un point intérieur à C , il existe $\varepsilon > 0$ suffisamment petit tel que $z := x^* + \varepsilon(x^* - x_*) \in C$. Par le [Lemme 1](#) (appliqué

au couple (x_*, x^*) avec $\alpha = \varepsilon$,

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

ce qui contredit la maximalité de x^* . $\square$

Remarque 17. C'est un "principe du maximum" : une fonction convexe non constante n'atteint son maximum que sur le bord de C . À comparer avec la minimisation : $x \mapsto \|x\|^2$ atteint son minimum en 0, à l'intérieur de $B(0, 1)$. Le maximum peut aussi ne pas être atteint du tout : $f(x) = x$ sur $C = (0, 1)$.

En renforçant les hypothèses sur C , on obtient qu'au moins un des points de maximum de f est un point extrémal de C . C'est la généralisation aux fonctions convexes du principe du maximum du chapitre 1 (??), qui portait sur les fonctions linéaires.

Théorème 16. Soit $f : C \rightarrow \mathbb{R}$ une fonction convexe et continue, avec C un convexe compact. Alors f atteint son maximum sur C en (au moins) un point extrémal de C :

$$\max_{x \in C} f(x) = \max_{x \in \text{ext}(C)} f(x).$$

Preuve. La fonction f étant continue sur le compact non vide C , elle atteint son maximum en un point $x^* \in C$ (?? du chapitre 0). Si $x^* \in \text{ext}(C)$, il n'y a rien à démontrer. Sinon, d'après le théorème de Minkowski (?? du chapitre 1), on peut écrire $x^* = \sum_{i=1}^m \lambda_i x_i$ avec $x_i \in \text{ext}(C)$, $\lambda \in \Delta_m$ et $\lambda_i > 0$ pour tout $i = 1, \dots, m$. Par l'inégalité de Jensen, $f(x^*) \leq \sum_{i=1}^m \lambda_i f(x_i)$, donc, comme $\sum_i \lambda_i = 1$,

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

Il s'agit d'une somme positive ou nulle de quantités négatives ou nulles ; par conséquent $f(x_i) = f(x^*)$ pour tout $i = 1, \dots, m$, i.e., les points extrémaux $x_1, \dots, x_m$ sont des points de maximum de f . $\square$

Exemple 16. Considérons la fonction $f : x \in C \mapsto \langle Ax, x \rangle$ avec $A \in \mathbb{S}_+^n$ et $C = B_\infty(0, 1) = [-1, 1]^n$. La fonction f est convexe et continue, et $\text{ext}(C) = \{-1, 1\}^n$ (?? du chapitre 1). On en déduit qu'il existe un point de maximum de f sur C dont toutes les coordonnées valent -1 ou 1 :

$$\max_{x \in [-1, 1]^n} \langle Ax, x \rangle = \max_{x \in \{-1, 1\}^n} \langle Ax, x \rangle.$$

Remarque 18. Le problème ci-dessus est un problème combinatoire (2^n points à comparer), qui contient des problèmes réputés difficiles (comme le problème MAX-CUT en théorie des graphes). Maximiser une fonction convexe est donc en général difficile, alors que la minimiser est "facile" : c'est l'objet du chapitre suivant.

Exemple 17 (Distance maximale à un polytope, diamètre). 1. Soit $P = \text{Conv}(\{x_1, \dots, x_m\})$ un polytope et $a \in \mathbb{R}^n$. La fonction $x \mapsto \|x - a\|$ est convexe et continue, donc le point de P le plus éloigné de a peut être choisi parmi les sommets :

$$\max_{x \in P} \|x - a\| = \max_{i=1, \dots, m} \|x_i - a\|.$$

Par exemple, pour le triangle $T = \text{Conv}(\{(0, 0), (2, 0), (0, 1)\})$ et $a = (1, 1)$, il suffit de comparer $\sqrt{2}$, $\sqrt{2}$ et 1 : la distance maximale vaut $\sqrt{2}$. (À comparer avec le point de T le plus proche de a , qui n'est pas un sommet : c'est la projection de a sur l'arête $[(2, 0), (0, 1)]$.)

2. Le diamètre $\max_{x,y \in P} \|x - y\|$ d'un polytope est atteint en un couple de sommets. En effet, $(x, y) \mapsto \|x - y\|$ est convexe et continue sur le convexe compact $P \times P$, dont les points extrémaux sont les couples (x_i, x_j) (exercice : $\text{ext}(C \times D) = \text{ext}(C) \times \text{ext}(D)$).

Exemple 18 (Entropie sur le simplexe). Soit $f(x) = \sum_{i=1}^n x_i \log x_i$ sur Δ_n (avec $0 \log 0 = 0$), fonction convexe et continue (Exemple 1). Comme $\text{ext}(\Delta_n) = \{e_1, \dots, e_n\}$,

$$\max_{x \in \Delta_n} f(x) = \max_i f(e_i) = 0,$$

atteint aux sommets, i.e., aux lois de probabilité "déterministes". Le minimum, lui, est atteint au centre du simplexe : par l'inégalité de Jensen appliquée à la fonction concave $\log$ (sur les indices i tels que $x_i > 0$),

$$-f(x) = \sum_i x_i \log \frac{1}{x_i} \leq \log \left(\sum_i x_i \cdot \frac{1}{x_i} \right) \leq \log n,$$

avec égalité pour $x = (\frac{1}{n}, \dots, \frac{1}{n})$. Ainsi $-\log n \leq f \leq 0$ sur Δ_n : en théorie de l'information, $-f$ est l'entropie de Shannon, maximale pour la loi uniforme et nulle pour une loi déterministe.

Exemple 19 (Lien avec le quotient de Rayleigh). Soit $A \in \mathbb{S}_+^n$ et $f(x) = \langle Ax, x \rangle$ sur la boule euclidienne $B(0, 1)$. Comme $\text{ext}(B(0, 1))$ est la sphère unité, on a

$$\max_{\|x\| \leq 1} \langle Ax, x \rangle = \max_{\|x\|=1} \langle Ax, x \rangle = \lambda_{\max}(A),$$

la dernière égalité étant la ?? du chapitre 0. On retrouve donc que, sur la boule, le maximum est atteint sur la sphère, en un vecteur propre unitaire associé à $\lambda_{\max}(A)$. Ici l'ensemble des points extrémaux est infini, et le théorème ne réduit pas le problème à un nombre fini de comparaisons.

1.8 Inégalités et convexité

On présente ici, comme application, quelques inégalités classiques qui se démontrent grâce à la convexité.

Proposition 5 (Inégalité arithmético-géométrique). Pour tous réels $x_1, \dots, x_n \geq 0$, on a

$$\frac{1}{n} \sum_{i=1}^n x_i \geq \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

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

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

(avec la convention $0^0 = 1$).

Preuve. Le premier énoncé est le cas $\lambda = (\frac{1}{n}, \dots, \frac{1}{n})$ du second. Si l'un des x_i avec $\lambda_i > 0$ est nul, le membre de droite est nul et il n'y a rien à démontrer ; on peut donc supposer $x_i > 0$ pour tout i (les x_i avec $\lambda_i = 0$ n'intervenant pas). La fonction $\log$ est concave sur $\mathbb{R}_+^*$ (car $-\log$ est convexe), donc par l'inégalité de Jensen appliquée à $-\log$,

$$\log \left(\sum_{i=1}^n \lambda_i x_i \right) \geq \sum_{i=1}^n \lambda_i \log(x_i) = \log \left(\prod_{i=1}^n x_i^{\lambda_i} \right).$$

On conclut en appliquant la fonction croissante exp. □

Proposition 6 (Inégalité de Young). Soient $a, b \geq 0$ et $p, q > 1$ avec $\frac{1}{p} + \frac{1}{q} = 1$. On a

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

Preuve. On applique la Proposition 5 avec $n = 2$, $x_1 = a^p$, $x_2 = b^q$ et $\lambda = (\frac{1}{p}, \frac{1}{q}) \in \Delta_2$:

$$\frac{a^p}{p} + \frac{b^q}{q} \geq (a^p)^{1/p} (b^q)^{1/q} = ab.$$

□

Le résultat suivant est une généralisation de l'inégalité de Cauchy-Schwarz (qui correspond au cas $p = q = 2$).

Proposition 7 (Inégalité de Hölder). Soient $p, q \in [1, \infty]$ avec $\frac{1}{p} + \frac{1}{q} = 1$ (avec la convention $\frac{1}{\infty} = 0$). Pour tous $x, y \in \mathbb{R}^n$, on a

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

Preuve. La première inégalité est l'inégalité triangulaire. Si $p = 1$ et $q = \infty$, on a $\sum_i |x_i y_i| \leq (\max_i |y_i|) \sum_i |x_i| = \|x\|_1 \|y\|_\infty$ (et de même si $p = \infty$). Supposons $1 < p < \infty$. Si $x = 0$ ou $y = 0$, il n'y a rien à démontrer. Sinon, on pose $a_i = \frac{|x_i|}{\|x\|_p}$ et $b_i = \frac{|y_i|}{\|y\|_q}$, de sorte que $\sum_i a_i^p = \sum_i b_i^q = 1$. Par l'inégalité de Young,

$$\sum_{i=1}^n a_i b_i \leq \sum_{i=1}^n \left(\frac{a_i^p}{p} + \frac{b_i^q}{q} \right) = \frac{1}{p} + \frac{1}{q} = 1,$$

i.e., $\sum_i |x_i y_i| \leq \|x\|_p \|y\|_q$. □

Enfin, la convexité permet de démontrer que $\|\cdot\|_p$ vérifie l'inégalité triangulaire, et donc que c'est bien une norme.

Proposition 8 (Inégalité de Minkowski). Soit $p \geq 1$. Alors pour tous $x, y \in \mathbb{R}^n$,

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

Preuve. Si $x = 0$ ou $y = 0$, il n'y a rien à démontrer. Sinon, on pose $x' = \frac{x}{\|x\|_p}$, $y' = \frac{y}{\|y\|_p}$ et $\lambda = \frac{\|x\|_p}{\|x\|_p + \|y\|_p} \in (0, 1)$, de sorte que

$$\frac{x + y}{\|x\|_p + \|y\|_p} = \lambda x' + (1 - \lambda) y'.$$

La fonction $\varphi(t) = |t|^p$ étant convexe (Exemple 5), on a pour tout i , $\varphi(\lambda x'_i + (1 - \lambda) y'_i) \leq \lambda \varphi(x'_i) + (1 - \lambda) \varphi(y'_i)$. En sommant sur i ,

$$\left\| \frac{x + y}{\|x\|_p + \|y\|_p} \right\|_p^p \leq \lambda \|x'\|_p^p + (1 - \lambda) \|y'\|_p^p = \lambda + (1 - \lambda) = 1,$$

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

Remarque 19. 1. La preuve montre en fait que la boule unité $B_{\|\cdot\|_p}(0,1)$ est convexe, et c'est cette convexité qui fournit l'inégalité triangulaire. Plus généralement, une fonction $N : \mathbb{R}^n \rightarrow \mathbb{R}_+$ positivement homogène ($N(\alpha x) = |\alpha|N(x)$) et nulle seulement en 0 est une norme si et seulement si sa boule unité $\{N \leq 1\}$ est convexe.

2. Pour $0 < p < 1$, $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ n'est *pas* une norme : avec $x = e_1$ et $y = e_2$, $\|x + y\|_p = 2^{1/p} > 2 = \|x\|_p + \|y\|_p$. Sa boule unité n'est d'ailleurs pas convexe.