Chapitre 1.

Fonctions convexes

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

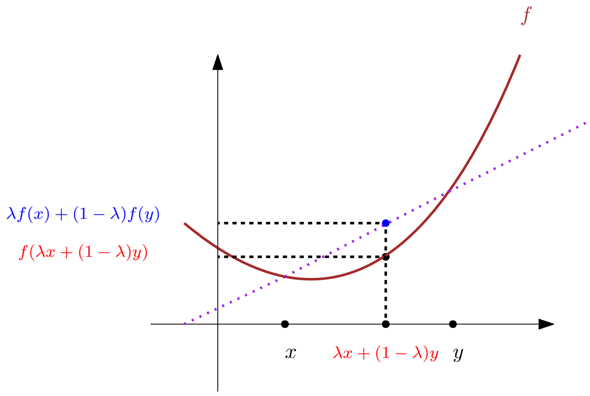
Définition 1. On dit qu'une fonction $f : C \to \mathbb{R}$ est convexe si

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

On dira que f est strictement convexe si

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

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



(a)

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

Exemple 1. Voici quelques examples de fonctions convexes:

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

D'après la Définition-1, une fonction f est convexe si l'image par f, d'une combinaison convexe de deux points x et y, par f est plus petite que la combinaison convexe des valeurs $f(x)$ et $f(y)$. Cette propriété s'étant à la combinaison convexe de n'importe quel nombre de vecteurs.

Théorème 1 (Inégalité de Jensen)**.** Soit $f : C \to \mathbb{R}$ une fonction convexe. Alors pour tout

$x_1, \dots, x_m \in C$ et $\lambda \in \Delta_m$:
$$f\Big(\sum_{i=1}^m \lambda_i x_i\Big) \le \sum_{i=1}^m \lambda_i f(x_i).$$

Preuve. Par récurrence. Laissée en exercice. □

Remarque 1. L'inégalité de convexité dans Définition 1 est parfois appelée inégalité de Jensen. Historiquement, le résultat démontré par Jensen était $f(\frac{x+y}{2}) \le \frac{f(x)+f(y)}{2}$. On rencontre d'autres versions de l'inégalité de Jensen en probabilités et théorie de la mesure et 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 $f(\mathbb{E}(X)) \le \mathbb{E}(f(X))$. Un autre exemple est le suivant. Si $(X, \mathcal{M}, \mu)$ est un espace probabilisé (où μ est une mesure positive avec $\mu(X) = 1$), I un intervalle non vide de $\mathbb{R}$, $\phi : I \to \mathbb{R}$ une fonction convexe, $f : X \to I$ une fonction dans $\mathcal{L}^1_\mu$ alors
$$\phi\Big(\int_X f\mathbf{d}\mu\Big) \le \int_X \phi \circ f\mathbf{d}\mu$$

Le résultat suivant est une conséquence immédiate de Théorème 1.

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

Preuve. On a par l'inégalité de Jensen
$$f(x) = f(\sum_{i=1}^m \lambda_i x_i) \le \sum_{i=1}^m \lambda_i f(x_i) \le \max_{i=1,\dots,m} f(x_i) \sum_{i=1}^m \lambda_i = \max_{i=1,\dots,m} f(x_i).$$
□

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

Théorème 2.
1. Soit f une fonction convexe sur C et $\alpha \in \mathbb{R}^+$. Alors αf est convexe sur C.
2. Soient $f_1, \dots, f_m$ des fonctions convexes sur C. Alors $f = \sum_{i=1}^m f_i$ est convexe sur C.
3. Soient $f_1, \dots, f_m$ des fonctions convexes sur C. Alors $f = \sup_{i=1,\dots,m} f_i$ est convexe sur C.
4. Si $f : C \to \mathbb{R}$ est convexe et $A \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n$. Alors $g(x) = f(Ax + b)$ 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, alors pour tout $i = 1, \dots, m$
$$f_i(\lambda x + (1-\lambda)y) \le \lambda f_i(x) + (1-\lambda) f_i(y),$$
donc
$$\sum_{i=1}^m f_i(\lambda x + (1-\lambda)y) \le \sum_{i=1}^m \Big(\lambda f_i(x) + (1-\lambda) f_i(y)\Big) = \lambda \sum_{i=1}^m f_i(x) + (1-\lambda) \sum_{i=1}^m f_i(y),$$
soit
$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$$
donc f est convexe.

3. On a $f(\lambda x+(1-\lambda)y) = \max_{i=1,\dots,m} f_i(\lambda x+(1-\lambda)y) \leq \max_{i=1,\dots,m} \Big(\lambda f_i(x)+(1-\lambda)f_i(y)\Big)$. En utilisant le fait que, si $(\alpha_i)_{i=1}^m, (\beta_i)_{i=1}^m$ sont deux suites de réels, alors $\max_i(\alpha_i+\beta_i) \leq \max_i \alpha_i + \max_i \beta_i$. Donc

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

4. Tout d'abord, remarquons que D est convexe. Soient $y_1, y_2 \in D$ et définissons $x_i = Ay_i + b$ avec $i = 1, 2$. Par définition, $x_1, x_2 \in D$. Soit donc $\lambda \in [0,1]$. On a par convexité de f

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

soit

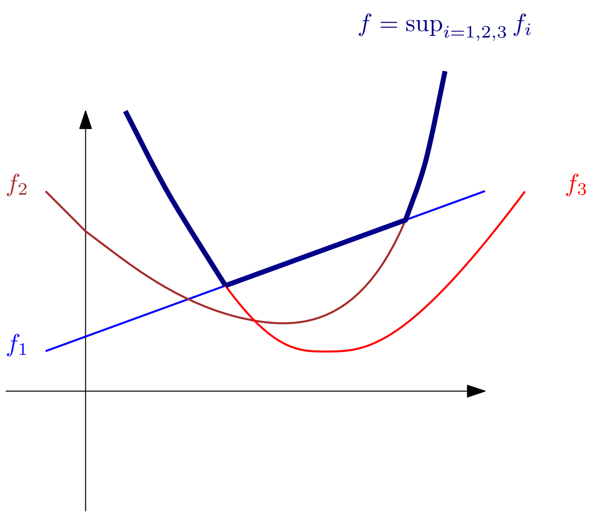
$$f(A(\lambda y_1 + (1-\lambda)) + b) \leq \lambda g(y_1) + (1-\lambda)g(y_2),$$

donc

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

d'où la convexité de g.

□



(a)

Figure 1.2: Convexité du la fonction sup de trois fonctions f_1, f_2, f_3.

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

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

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

$$\begin{aligned} \phi(\lambda x + (1-\lambda)y) &= g(f(\lambda x + (1-\lambda)y) \\ &\leq g(\lambda f(x) + (1-\lambda)f(y)) \text{ par convexité de f et monotonie de g} \\ &\leq \lambda g(f(x)) + (1-\lambda)g(f(y)) \text{ par convexité de g} \\ &= \lambda\phi(x) + (1-\lambda)\phi(y). \end{aligned} \tag{1.1}$$

□

Exemple 3. La fonction $\phi(x) = e^{\|x\|^2}$ est convexe sur $\mathbb{R}^n$. En effet $\phi(x) = g(f(x))$ avec $f(x) = e^x$ qui est croissante et $f(x) = \|x\|^2$ qui est convexe. Plus généralement, pour toute fonction convexe f, $e^{f(x)}$ est convexe.

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

Théorème 4. La fonction $g(x) = \inf_{y \in D} f(x, y)$ est convexe sur C.

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

$$\begin{aligned} g(\lambda x_1 + (1-\lambda)x_2) & \inf_{y \in D} f(\lambda x_1 + (1-\lambda)x_2, y) \\ & f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\ & \lambda f(x_1, y_1) + (1-\lambda) f(x_2, y_2) \\ & \lambda g(x_1) + (1-\lambda) g(x_2) + \varepsilon. \end{aligned} \tag{1.2}$$

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

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

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

$$f(y + \alpha(y - x)) \geq f(y) + \alpha(f(y) - f(x)). \tag{1.3}$$

Preuve. Soient $x, y \in C, t \in (0, 1]$ et considérons $z = tx + (1-t)y$. On a $x = \frac{1}{t}z - \frac{1-t}{t}y = z + \alpha(z - y)$ avec $\alpha = \frac{1-t}{t}$. Par (1.3)

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

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

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

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

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

soit $f(z) = f(y + \alpha(y - x)) \geq f(y) + \alpha(f(y) - f(x))$, qui n'est rien d'autre que (1.3) □

1.1 Caractérisations des fonctions convexes différentiables

1.1.1 Caractérisations du premier ordre

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

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \text{ pour tout } x, y \in C. \tag{1.4}$$

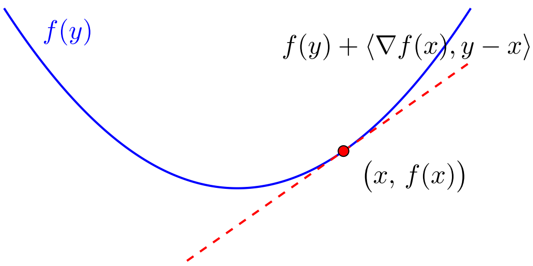


Figure 1.3: Illustration du Théorème 5.

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

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

ce qui donne

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

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

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

En appliquant (1.4) pour u, w et ensuite v, w on obtient

$$f(w)+\langle\nabla f(w), u-w\rangle \leq f(u), \tag{1.5}$$

et

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

soit

$$(1-\lambda)f(w)-\lambda\langle\nabla f(w), u-w\rangle \leq (1-\lambda)f(v). \tag{1.6}$$

En multipliant (1.5) par λ et sommant avec (1.6), on obtient le résultat. □

Remarque 2. Dans la littérature, l'inégalité (1.4) est souvent appelée *the gradient inequality*. Elle affirme que les hyperplans tangents à une fonction convexe minorent la fonction.

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

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

$$\langle\nabla f(x)-\nabla f(y), x-y\rangle \geq 0, \text{ pour tout } x, y \in C. \tag{1.7}$$

Preuve. Supposons que f est convexe et soient $x, y \in C$. Par (1.4) on a

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

En sommant les deux inégalités on obtient que $\langle\nabla f(x)-\nabla f(y), x-y\rangle \geq 0$. Inversement, supposons que (1.7) est vérifiée et montrant que f est convexe. Définitions la fonction $\phi : t \in$

$(0,1) \mapsto f(x+t(y-x))$. On a

$$\begin{aligned} f(y) = \phi(1) = \phi(0) + \int_0^1 \phi^{'}(t)\mathbf{d}t \\ &= f(x) + \int_0^1 \langle \nabla f(x+t(y-x)), y-x\rangle \mathbf{d}t \\ &= f(x) + \langle \nabla f(x), y-x\rangle + \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x\rangle \mathbf{d}t \\ &= f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{t}\int_0^1 \underbrace{\langle \nabla f(x+t(y-x)) - \nabla f(x), x+t(y-x)-x\rangle}_{\geq 0} \mathbf{d}t, \end{aligned} \tag{1.8}$$

soit $f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle$, et par Théorème 5 f est convexe. □

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

1.1.2 Caractérisations du second ordre

Théorème 7. Soit $f : C \to \mathbb{R}$ une fonction de classe C^2 avec C un ouvert convexe de $\mathbb{R}^n$ alors f est convexe si et seulement si la Hessienne $\nabla^2 f(x)$ est semi-définie positive, i.e.,

$$\nabla^2 f(x) \succeq 0 \text{ pour tout } x \in C. \tag{1.9}$$

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

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

En faisant tendre $t \to 0^+$ on obtient que $\langle \nabla^2 f(x)d, d\rangle \geq 0$. Comme d est arbitraire, il s'en suit que $\nabla^2 f(x) \succeq 0$ pour tout $x \in C$. Maintenant, supposons que (1.9) est vérifiée, on a pour tout $x, y \in C$

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x), y-x\rangle + \int_0^1 \Big(\int_0^s \underbrace{\langle \nabla^2 f(x_\tau)(y-x), y-x\rangle}_{\geq 0} \mathbf{d}\tau \Big) \mathbf{d}s \\ &\geq f(x) + \langle \nabla f(x), y-x\rangle, \end{aligned} \tag{1.10}$$

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

Remarque 3. La condition (1.9) est liée à la notion de courbure. En effet, considérons la surface $\Gamma_f = \{(x,y,z) \in \mathbb{R}^2 : \ z = f(x,y)\}$ avec f de classe C^2. La courbure de Gauss au point (x,y) est égale à

$$\kappa = \frac{\det(\nabla^2 f(x,y))}{(1 + \|\nabla f(x,y)\|^2)^2}.$$

Le signe de $\det(\nabla^2 f(x,y))$ (et donc de κ) donne une classification de la surface: elliptique, parabolique ou hyperbolique.

1.2 Continuité et différentiabilité des fonctions convexes

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

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

Preuve. Sans perte de généralité, on suppose que $x = 0$. Soit $(x_k)_k \geq 0$ telle que $x_k \to x^* = 0$ quand $k \to \infty$. On a, par convexité de f

$$f(x_k) \leq (1 - \|x_k\|)f(0) + \|x_k\| f(x_k/\|x_k\|).$$

Comme $(x_k/\|x_k\|)_i \in [-1, 1]$ pour tout $i = 1, \ldots, n$, il vient que $x_k/\|x_k\| \in [-1, 1]^n$, donc $x_k/\|x_k\| = \sum_{i=1}^{2^n} \lambda_i e_i$ avec $\lambda \in \Delta_{2^n}$. On déduit par le Corollaire 1 que $f(x_k/\|x_k\|) \leq \max_{i=1,\ldots,n} f(\pm e_i) := K$. Il s'en suit que

$$\limsup_{k\to\infty} f(x_k) \leq (1 - \|x^*\|)f(0) + \|x^*\|K = f(0).$$

De même, en remarquant que

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

on déduit que $f(0) \leq \liminf_{k\to\infty} f(x_k)$, et par la suite $\lim\limits_{k\to\infty} f(x_k) = f(0)$. D'où la continuité de f en 0. □

Sur un convexe $C \subsetneq \mathbb{R}^n$, on peut obtenir un résultat similaire (meilleur même) à Théorème 8 pour les point intérieurs à C. En effet, on peut démontrer qu'une fonction convexe est localement Lipschitzienne en tout $x \in \text{int}(C)$. La raison de se restreindre aux points intérieurs est le comportement d'une fonction convexe au bords qui peut créer des discontinuités. Pour illustrer ceci, considérons la fonction $f : [0, 1] \to \mathbb{R}$ définie par $f(0) = 1$ et $f(x) = x^2$ pour $x \in (0, 1]$. Cette fonction est évidemment convexe (faites un dessin) mais n'est pas continue.

On a le résultat suivant.

Théorème 9. Soit $f : C \to \mathbb{R}$ une fonction convexe définie sur un convexe C de $\mathbb{R}^n$ et soit $x_0 \in \text{int}(C)$. Alors il existe $L > 0$ et $\varepsilon > 0$ tels que $B(x_0, \varepsilon) \subset C$ et

$$|f(x) - f(x_0)| \leq L\|x - x_0\|, \ \forall x \in B(x_0, \varepsilon). \tag{1.11}$$

Preuve. Soit $x_0 \in \text{int}(C)$, il existe alors $\varepsilon > 0$ tel que $B_\infty(x_0, \varepsilon) \subset C$. On commence par montrer que f est bornée supérieurement sur $B_\infty(x_0, \varepsilon)$. Comme $B_\infty(x_0, \varepsilon)$ est convexe et compact, avec

$$\text{ext}(B_\infty(x_0, \varepsilon)) = \{z_i = x_0 + \varepsilon\theta_i, \text{ avec } \theta_i \in \{\pm 1\}, i = 1, 2, \ldots, 2^n\},$$

on déduit par le théorème de Krein-Milman que tout $x \in B_\infty(x_0, \varepsilon)$ s'écrit de la forme $\sum_{i=1}^{2^n} \lambda_i z_i$ avec $\lambda \in \Delta_{2^n}$. Donc par Corollaire 1 $f(x) \leq K := \max_{i=1,2,\ldots,2^n} f(z_i)$. Comme $B(x_0, \varepsilon) \subset B_\infty(x_0, \varepsilon)$ on en déduit que $f(x) \leq K$ pour tout $x \in B(x_0, \varepsilon)$.

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

$$f(x) \leq \alpha f(z) + (1 - \alpha)f(x_0) = f(x_0) + \alpha(f(z) - f(x_0)) \leq f(x_0) + \alpha(K - f(x_0)), \tag{1.12}$$

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

$$\begin{aligned} f(x) = f(x_0) + \alpha(x_0 - w)) &\geq f(x_0) + \alpha(f(x_0) - f(w)) \\ &\geq f(x_0) - \alpha(K - f(x_0)), \end{aligned} \tag{1.13}$$

i.e., $f(x) - f(x_0) \geq -L\|x - x_0\|$. En conclusion, on obtient que (1.11) pour tout $x \in B(x, \varepsilon)$ avec $L = (K - f(x_0))\varepsilon^{-1}$. □

Nous avons vu qu'en les points intérieurs à C, une fonction convexe $f : C \to \mathbb{R}$ est localement Lipschitzienne, i.e., vérifie (1.11). Maintenant, on montrera que les dérivées directionnelles de f en tout point intérieur existent. Rappelons qu'étant donnés $x \in \operatorname{int}(C)$ et $0 \neq d \in \mathbb{R}^n$, on dira que f est différentiable en x dans la direction d si la limite

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

existe.

Théorème 10. Soit $f : C \to \mathbb{R}$ une fonction convexe définie sur un convexe $C \subset \mathbb{R}^n$ et soit $x \in \operatorname{int}(C)$. Alors $f'(x; d)$ existe pour tout $0 \neq d \in \mathbb{R}^n$.

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

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

soit

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

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

1.3 Sous-ensembles de niveau de fonctions convexes:

Définition 2. Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ une fonction. Le sous-ensemble de f de niveau $\alpha \in \mathbb{R}$ est l'ensembles

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

On a le résultat suivant

Théorème 11. Soit $f : C \to \mathbb{R}$ une fonction convexe et avec C un ensemble convexe. Alors pour tout $\alpha \in \mathbb{R}$, l'ensembles $\mathbf{Lev}(f, \alpha)$ est convexe.

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

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

□

Remarque 4. La réciproque dans le Théorème-11 est fausse comme le montre la fonction $f(x) = \sqrt{|x|}$ (voir Figure-1.4) qui n'est pas convexe mais dont tout les sous-ensembles de niveau sont convexe. En effet, d'une part, pourt $\alpha < 0, \mathbf{Lev}(f, \alpha) = \emptyset$. D'autre part, pour $\alpha \geq 0$, on $\mathbf{Lev}(f, \alpha) = [-\alpha^2, \alpha^2]$ qui est convexe. Une telle fonction est dite quasi-convexe.

Définition 3. Soit $C \subset \mathbb{R}$ un convexe. On dit qu'une fonction $f : C \to \mathbb{R}$ est quasi-convexe si pour tout $\alpha \in \mathbb{R}$, l'ensemble $\mathbf{Lev}(f, \alpha)$ est convexe. De même, f est dite quasi-concave si $-f$ est quasi-convexe.

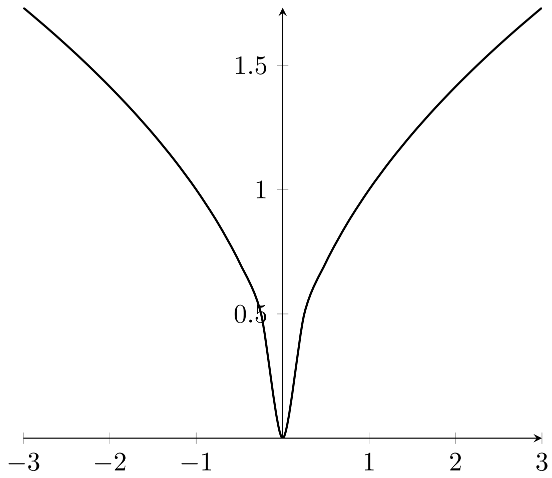


Figure 1.4: $f(x) = \sqrt{|x|}$ comme exemple de fonction quasi-convexe.

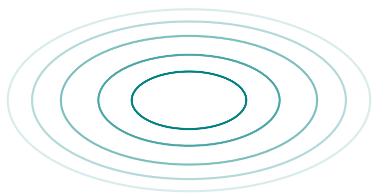
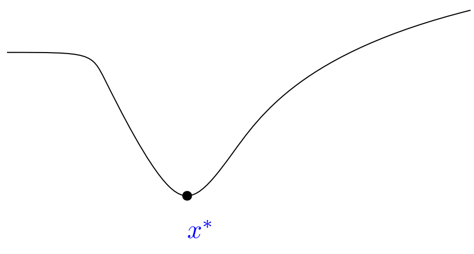


Figure 1.5: Sous-ensembles de niveau de la fonction $f(x, y) = 1/2(x^2 + 4y^2)$.

Exemple 4.

1. Évidemment, une fonction convexe est quasi-convexe.
2. L'application $x \in \mathbb{R}^n \mapsto \text{card}(x) = \{\text{nombre des composantes non nuls de } x\}$ est quasi-concave:
$$\text{card}(x + y) \geq \min\{\text{card}(x), \text{card}(y)\}.$$
3. L'application $M \in \mathbb{R}^{n\times n} \mapsto \text{rang}(M)$ est quasi-concave sur $\mathcal{S}^n_+$:
$$\text{rang}(M + N) \geq \min\{\text{rang}(M), \text{rang}(N)\}.$$
4. Un exemple important est celui des fonctions continue sur $\mathbb{R}$. En effet, si $f : \mathbb{R} \to \mathbb{R}$ est continue alors f est quasi-convexe ssi l'une des propriétés suivantes est vérifée
   (a) f est croissante;
   (b) f est décroissante;
   (c) il existe $x^* \in \mathbb{R}$ tel que pour tout $x \leq x^*$ f est décroissante, et pour tout $x \geq x^*$ est croissante (cf. Fig. 1.6).



(a)

Figure 1.6: Exemple de fonction quasi-convexe sur $\mathbb{R}$.

1.4 Fonctions à valeurs réelles étendues

Dans la pratique on peut être amené à travailler et considérer des fonctions qui peuvent prendre des valeurs infinies. En effet, considérons une fonction $f : \mathbb{R}^n \to \bar{\mathbb{R}} = [-\infty, \infty]$. La définition de convexité (1) reste valable pour de telles fonctions tenant compte des opération arithmétiques classiques

$$\begin{aligned} a + \infty = \infty + a &= \infty && (-\infty < a < \infty), \\ a - \infty = -\infty + a &= -\infty && (-\infty < a < \infty), \\ a \cdot \infty = \infty \cdot a &= \infty && (0 < a < \infty), \\ a \cdot (-\infty) = (-\infty) \cdot a &= -\infty && (0 < a < \infty), \\ a \cdot \infty = \infty \cdot a &= -\infty && (-\infty < a < 0), \\ a \cdot (-\infty) = (-\infty) \cdot a &= \infty && (-\infty < a < 0), \end{aligned} \tag{1.14}$$

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

Définition 4 (Domaine effectif)**.** Soit $f : \mathbb{R}^n \to [-\infty, \infty]$, le domaine (ou domaine effectif) de f est définit par

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

Exemple 5. Si $C \subset \mathbb{R}^n$, l'indicatrice de C au sens d'analyse convexe est définie par

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

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

Définition 5. On dit que $f : \mathbb{R}^n \to [-\infty, \infty]$ est propre si $-\infty \notin f(\mathbb{R}^n)$ et $\mathrm{dom}(f) \neq \emptyset$, i.e., il existe $x_0 \in \mathbb{R}^n$ tel que $f(x_0) < \infty$.

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

Définition 6 (Epigraphe)**.** Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, on définit l'épigraphe de f par

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

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

Théorème 12 (et définition)**.** f est convexe $\Leftrightarrow$ $\mathrm{epi}(f)$ est convexe.

Exemple 6.

1. Si $f(x) = a^T x - \alpha$ avec $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$. On a

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

est un demi-espace donc convexe.

2. Si $f(x) = 1/2\|x\|^2$, alors $\mathrm{epi}(f)$ est la région du l'espace au dessus de la parabole (cf. Fig. 1.7).

3. Pour l'indicatrice d'un ensemble

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

qui est donc convexe si et seulement si C est convexe.

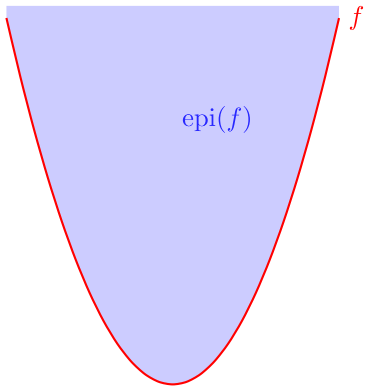


Figure 1.7: L'épigraphe de $\frac{1}{2}\|x\|^2$.

Le résultat suivant montre la préservation de la convexité du supremum de fonctions convexe et est à comparer avec Théorème 2.

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

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

Définition 7 (Fonction fermée)**.** Une fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est dite fermée si $\text{epi}(f)$ est fermé.

Exemple 7. 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

$$\delta_C \text{ et fermée} \Leftrightarrow C \times \mathbb{R}^+ \text{ et fermé} \Leftrightarrow C \text{ et fermé}.$$

Définition 8 (Semi-continuité inférieure)**.** Une fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est dite semi-continue inférieurement (sci) en $x \in \mathbb{R}^n$ si

$$f(x) \leq \liminf f(x_n),$$

pour toute suite $(x_n)_n$ telle que $x_n \to x$ quand $n \to \infty$.

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

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

Théorème 14. Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. On a

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

Preuve. Exercice. □

1.4.1 Maxima de fonctions convexes

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

Théorème 15. Soit $f : C \to \mathbb{R}$ une fonction convexe non constante avec C un convexe. Alors f n'atteint pas son maximum à l'intérieur de C.

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

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

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

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

cela contredit la maximalité de x^*. □

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

Théorème 16. Soit $f : C \to \mathbb{R}$ une fonction convexe continue avec C un convexe compact. Alors il existe au moins un maximum de f qui est un point extrémal de C.

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

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

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

Exemple 8. Considérons la fonction $f : x \in C \mapsto x^T Ax$ avec $A \in \mathcal{S}_n^+$ et $C = B_\infty(0, 1)$. Comme $\text{ext}(C) = \{\pm 1\}^n$ (voir TD), on déduit qu'il existe un maximum de f sur C qui appartient à $\{\pm 1\}^n$, i.e., toutes ses coordonnées sont soit -1 ou 1.