Chp. 9. Convexité

Avertissement! *Dans tout ce chapître, C désigne une partie convexe de $\mathbb{R}^n$, et f une fonction numérique partout définie sur C.*

9.1 Fonctions affines, convexes, strictement convexes

Définition 16 *On dit qu'une fontion $\mathbf{a}$ de C dans dans $\mathbb{R}^m$ est affine sur C si :*
(1) $\forall x, y \in C,\ x \neq y, \quad 0 < t < 1 \ \Rightarrow\ \mathbf{a}\,[\,t\,x + (1-t)\,y\,] = t\ \mathbf{a}\,(x) + (1-t)\ \mathbf{a}\,(y)$

Exemple 9.1 *Toute fonction linéaire et toute fonction constante sur $\mathbb{R}^n$ sont affines sur $\mathbb{R}^n$.*

Exemple 9.2 *La relation : $\mathbf{a}\,(x,y) = x - 2\,y + 3$ définit une fonction affine $\mathbf{a}$ de $\mathbb{R}^2$ dans $\mathbb{R}$.*

Plus généralement, toute fonction affine $\mathbf{a}$ d'une partie convexe C de $\mathbb{R}^n$ dans $\mathbb{R}$ est définie sur C par une relation de la forme :

- $\mathbf{a}\,(x) = \sum_{i=1}^{n} a_i\,x_i + b = a^T x + b$

où : a est un vecteur donné de $\mathbb{R}^n$, et b un réel donné. C'est la somme d'une fonction *linéaire* et d'une fonction *constante.*

Définition 17 *On dit que f est convexe sur C si :*
(2) $\forall x, y \in C,\ x \neq y, \quad 0 < t < 1 \ \Rightarrow\ f\,[\,t\,x + (1-t)\,y\,] \leq t\,f(x) + (1-t)\,f(y)$

Exemple 9.3 *Toute fonction affine sur C est bien sur convexe sur C*

Exemple 9.4 *Toute norme sur $\mathbb{R}^n$ est convexe sur $\mathbb{R}^n$*

On notera que (1) équivaut à :
(3) $\forall x, y \in C, \quad 0 \leq t \leq 1 \ \Rightarrow\ \mathbf{a}\ [\,t\,x + (1-t)\,y\,] = t\ \mathbf{a}\,(x) + (1-t)\ \mathbf{a}\,(y)$
et que (2) équivaut à :
(4) $\forall x, y \in C, \quad 0 \leq t \leq 1 \ \Rightarrow\ f\,[\,t\,x + (1-t)\,y\,] \leq t\,f(x) + (1-t)\,f(y)$
On ne change donc pas les définitions 16 et 17 en remplaçant (1) par (3) et (2) par (4). Utiliser (1) plutôt que (3) et (2) plutôt que (4) permet cependant de conserver le parallèle avec la définition suivante :

Définition 18 *On dit que f est strictement convexe sur C si :*
(5) $\forall x, y \in C,\ x \neq y, \quad 0 < t < 1 \ \Rightarrow\ f\,[\,t\,x + (1-t)\,y\,] < t\,f(x) + (1-t)\,f(y)$

Exemple 9.5 *Aucune norme n'est strictement convexe sur $\mathbb{R}^n$, car sa restriction à toute droite vectorielle est affine, mais le carré de la norme Euclidienne, par exemple, définit une fonction strictement convexe sur $\mathbb{R}^n$.*

9.2 Caractérisation géométrique

Le caractère affine, convexe, ou strictement convexe d'une fonction φ d'une variable sur un intervalle I correspond à une propriété géométrique simple du graphe de φ au dessus de I :

- Elle est affine sur I si et seulement si son graphe sur I est un segment de droite.
- Elle est convexe sur I si et seulement si le segment joignant deux points quelconques de son graphe sur I est toujours situé « au dessus » de ce graphe (éventuellement partiellement ou tout entier contenu dans le graphe).
- Elle est strictement convexe sur I si et seulement si elle y est convexe, et son graphe ne contient aucun segment.

Théorème 9.1 *Les assertions suivantes sont équivalentes :*

(1) C est convexe et f est convexe (resp. affine, strictement convexe) sur C.

(2) Pour tout x dans C, et tout u dans $\mathbb{R}^n$, $I = \{\, t \in \mathbb{R} \mid x + t\,u \in C \,\}$ est un intervalle de $\mathbb{R}$, et : $\varphi(t) = f(x + t\,u)$ est convexe (resp. affine, strictement convexe) sur I.

Preuve :

- (1) $\Rightarrow$ (2) : Soient x un point de C et u un vecteur de $\mathbb{R}^n$ donnés :

Posons : $\varphi(t) = f(x + t\,u)$, et : $I = \{\, t \in \mathbb{R} \mid x + t\,u \in C \,\}$. Si t_1 et t_2 sont deux points distincts de I, et : $0 < s < 1$, on peut écrire : $x + [\, s\ t_1 + (1-s)\, t_2 \,]\, u = s\, x_1 + (1-s)\, x_2$, où : $x_1 = x + t_1\, u$, et : $x_2 = x + t_2\, u$ sont deux points *distincts* appartenant à C. Puisque C est convexe, et f est convexe (resp. affine, strictement convexe) sur C, le point : $x + s\,[\, t_1 + (1-s)\, t_2 \,]\, u$ appartient encore à C, donc : $s\, t_1 + (1-s)\, t_2$ appartient à I, et :

$$\varphi[\, s\, t_1 + (1-s)\, t_2 \,] = f[\, s\, x_1 + (1-s)\, x_2 \,] \leq s\ f(x_1) + (1-s)\ f(x_2) = s\ \varphi(t_1) + (1-s)\ \varphi(t_2) \qquad (\text{resp. } =, <)$$

Le résultat valant pour tout couple de points distincts t_1 et t_2 de I, et tout réel s de l'intervalle $]\,0, 1\,[$, I est un intervalle, et φ est convexe (resp. affine, strictement convexe) sur I.

- (2) $\Rightarrow$ (1) : Soient x_1 et x_2 deux points distincts donnés dans C, et : $0 < s < 1$:

Posons : $u = x_1 - x_2$, et : $\varphi(t) = f(x_2 + t\,u)$, de sorte que : $I = \{\, t \in \mathbb{R} \mid x + t\,u \in C \,\}$ contient : $t_1 = 1$ et : $t_2 = 0$, $\varphi(t_1) = f(x_1)$, et : $\varphi(t_2) = f(x_2)$.

Par hypothèse, I est un intervalle, donc : $s = s\, t_1 + (1-s)\, t_2$ appartient à I et φ est convexe (resp. affine, strictement convexe) sur I. Ainsi : $x + s\,u = s\, x_1 + (1-s)\, x_2$ appartient à C, et :

$$f[\, s\, x_1 + (1-s)\, x_2 \,] = \varphi(s) \leq s\ \varphi(t_1) + (1-s)\ \varphi(t_2) = s\ f(x_1) + (1-s)\ f(x_2) \quad (\text{resp. } =, <)$$

Le résultat valant pour tout couple de points distincts x_1 et x_2 de C, et tout réel s de l'intervalle $]\,0, 1\,[$, C est convexe, et f est convexe (resp. affine, strictement convexe) sur C.

□

Corollaire 9.2 *Si f est convexe, mais pas strictement convexe sur C, sa restriction à un segment $[\,x, y\,]$ de longueur non nulle contenu dans C est affine, et donc son graphe contient un segment de droite de longueur non nulle.*

9.3 Cas des fonctions de classe $\mathbf{C^2}$

Théorème 9.3 *Une fonction f de classe $\mathcal{C}^2$ sur un ouvert convexe Ω est convexe sur Ω si et seulement si sa Hessienne $\nabla^2 f(x)$ est SDP en tout point x de Ω.*

Si en outre l'ensemble : $\{\, x \in \Omega \mid det\ \nabla^2 f(x) = 0 \,\}$ ne contient aucun segment $[\, x, y\,]$ de longueur non nulle contenu dans Ω, f est strictement convexe sur Ω.

Preuve : On utilise le théorème 9.1. Soient x un point donné de Ω, et u un vecteur donné de $I\!R^n$. Posons : $\varphi(t) = f(x + t\,u)$, et : $I = \{\, t \in I\!R \mid x + t\,u \in \Omega \,\}$. Puisque Ω est un ouvert convexe, I est toujours un intervalle ouvert contenant zéro, φ est $\mathcal{C}^2$ sur I, et, pour tout t dans I :

$$\varphi''(t) = u^T \nabla^2 f(x + t\,u)\, u$$

Si f est convexe sur Ω, φ est toujours convexe sur I (thm 9.1). En particulier : $\varphi''(0) = u^T \nabla^2 f(x)\, u$ est toujours positive, et donc : $\nabla^2 f(x)$ est SDP en tout point x de Ω. Réciproquement, si $\nabla^2 f(x)$ est SDP en tout point x de Ω, φ'' est toujours positive sur I, donc φ est toujours convexe sur I, et f est convexe sur Ω (thm 9.1).

Enfin si f n'est pas strictement convexe sur Ω, il existe au moins un point x de Ω et un vecteur u de $I\!R^n$ tels que φ soit affine sur un intervalle : $J = [\, t_1, t_2\,]$ de longueur non nulle contenu dans : $I = \{\, t \in I\!R \mid x + t\,u \in \Omega \,\}$. Mais alors : $\varphi''(t) = u^T \nabla^2 f(x + t\,u)\, u$ est nulle en tout point t de J, et l'ensemble : $\{\, x \in \Omega \mid det\ \nabla^2 f(x) = 0 \,\}$ contient le segment : $[\, x + t_1\, u, x + t_2\, u\,]$.

□

Exemple 9.6 *Toute fonction elliptique sur un ouvert convexe Ω est strictement convexe sur Ω.*

Exemple 9.7 *La fonction : $f = \dfrac{1}{x\,y} + x + y$ est strictement convexe, mais pas elliptique, sur l'ouvert convexe : $\Omega = \{\, (x, y) \in I\!R^2 \mid x, y > 0 \,\}$ (le vérifier par le calcul de sa Hessienne).*

Attention! *Même si f est strictement convexe sur Ω, l'ensemble : $\{\, x \in \Omega \mid det\ \nabla^2 f(x) = 0 \,\}$ peut contenir un segment de longueur non nulle.*

Contre-exemple 9.8 *La fonction $f = x^4 + y^4$ est strictement convexe sur $I\!R^2$, bien que : det $\nabla^2 f(x, y)$ en tout point (x, y) vérifiant : $x\,y = 0$.*

9.4 Cas des fonctions de classe $\mathbf{C^1}$

Théorème 9.4 *Si f est convexe (resp. strictement convexe) sur Ω, et dérivable au point x de Ω :*

$$y \in \Omega,\ y \neq x, \ \Rightarrow\ f(y) \geq f(x) + \nabla f(x)^T (y - x) \qquad (\textit{resp.} >)$$

Preuve : Posons $u = y - x$. Par hypothèse, Ω est un ensemble convexe contenant $[\, x, y\,]$ donc : $I = \{\, t \in I\!R \mid x + t\,u \in \Omega \,\}$ est un intervalle contenant $[\, 0, 1\,]$. Puisque f est convexe (resp. strictement convexe) sur Ω, $\varphi(t) = f(x + t\,u)$ est convexe (resp. strictement convexe) sur I, donc, pour $0 < t < 1$:

(6) $\quad \varphi(t) \leq t\,\varphi(1) + (1 - t)\,\varphi(0) \qquad$ (resp. $<$)

et :

(7) $\quad 0 < s < 1 \ \Rightarrow\ \varphi(s\,t) \leq s\,\varphi(t) + (1 - s)\,\varphi(0) \qquad$ (resp. $<$)

En combinant (6) et (7), il vient :

$$(8) \qquad \frac{\varphi(s\,t)-\varphi(0)}{s\,t} \leq \frac{\varphi(t)-\varphi(0)}{t} \leq \varphi(1)-\varphi(0) \qquad \text{(resp. <)}$$

Puisque f est dérivable en x, φ est dérivable en 0, d'où, en faisant tendre s vers 0 dans (8) on déduit : $\varphi'(0) \leq \varphi(1)-\varphi(0)$, soit : $\nabla f(x)^T u \leq \nabla f(x)^T(y-x) \leq f(y)-f(x)$ (resp. <).

□

Théorème 9.5 *Une fonction f de classe $\mathcal{C}^1$ sur un ouvert convexe Ω est convexe (resp. strictement convexe) sur Ω si et seulement si :*

$$(9) \qquad x, y \in \Omega,\ x \neq y, \ \Rightarrow\ f(y) \geq f(x) + \nabla f(x)^T(y-x) \qquad \textit{(resp. >)}$$

Preuve : Si f est $\mathcal{C}^1$ sur Ω, elle est dérivable en tout point de Ω et on déduit alors du théorème 9.4 : $y \in \Omega,\ y \neq x \ \Rightarrow\ f(y) \geq f(x) + \nabla f(x)^T(y-x)$ (resp. >), donc (9) est vérifiée.

Réciproquemment, supposons (9) vérifiée. Pour montrer que f est convexe (resp. strictement convexe) sur Ω, il suffit de prouver que, pour tout couple (x, y) de points distincts de Ω :

$$\varphi(t) = t\,f(x) + (1-t)\,f(y) - f[\,t\,x + (1-t)\,y\,]$$

est positive (resp. strictement positive) sur l'intervalle $]\,0,1\,[$. Puisque Ω est convexe et f est $\mathcal{C}^1$ sur Ω, φ est toujours bien définie et $\mathcal{C}^1$ sur $[\,0,1\,]$. En outre : $\varphi(0) = \varphi(1) = 0$.

Si φ n'est pas strictement positive sur $]\,0,1\,[$, elle atteint nécessairement son minimum sur en un point $t^\star$ de $]\,0,1\,[$ pour lequel : $\varphi(t^\star) \leq 0$, et : $\varphi'(t^\star) = f(x) - f(y) - \nabla f(z^\star)^T(x-y) = 0$, où : $z^\star = t^\star x + (1-t^\star)\,y$ est un point de Ω, distinct de x et de y. De (9) on déduit alors :

$$(10) \quad \nabla f(z^\star)^T(y-z^\star) \leq f(y) - f(z^\star) \qquad \text{(resp. <)}$$

$$\begin{aligned} \text{Mais :}\quad \varphi(t^\star) &= t^\star f(x) + (1-t^\star)\,f(y) - f(z^\star) \\ &= t^\star\,[f(x)-f(y)] + f(y) - f(z^\star) \\ &= t^\star\,\varphi'(t^\star) + t^\star\,\nabla f(z^\star)^T(x-y) + f(y) - f(z^\star) \\ &= \nabla f(z^\star)(z^\star - y) + f(y) - f(z^\star) \end{aligned}$$

Si l'inégalité est toujours stricte dans (10), on déduit : $\varphi(t^\star) > 0$, une contradiction. Donc φ est nécessairement toujours strictement positive sur $]\,0,1\,[$, et f est strictement convexe sur Ω.

Si l'inégalité est seulement utilisée au sens large, on déduit : $\varphi(t^\star) = 0$, donc φ est toujours positive sur $]\,0,1\,[$, et f est convexe sur Ω.

□

9.5 Règles opératoires sur les fonctions convexes

Les règles suivantes et le « catalogue » des fonctions d'une ou plusieurs variable dont les propriétés de convexité sont connues (fonctions usuelles d'une variable, normes, fonctions affines, ...etc.) permettent éventuellement de reconnaître sans calculs le caractère convexe ou strictement convexe d'une fonction donnée. Elles se déduisent directement des définitions 16, 17 et 18, et leur vérification est abandonnée au lecteur à titre d'exercice :

- Stabilité par addition :

  Si les fonctions $f_1, \ldots, f_M$ sont toutes convexes sur C, leurs somme est convexe sur C. Si *l'une d'entre elles* au moins est *strictement* convexe sur C, la somme est *strictement* convexe sur C.

- Stabilité par composition à gauche :

  Si f est convexe sur C et φ est convexe *croissante* sur un intervalle I contenant $f(C)$, $\varphi \circ f$ est convexe sur C. Si en outre f est *strictement* convexe sur C, et φ *strictement* croissante sur I, $\varphi \circ f$ est *strictement* convexe sur C.

  Exemple 9.9 $f = e^{x^2+y^2}$ *est strictement convexe sur* $I\!R^2$.

- Stabilité par composition à droite :

  Si $\mathbf{a}$ est affine sur C, et f est convexe sur $\mathbf{a}(C)$, $f \circ \mathbf{a}$ est convexe sur C ($\mathbf{a}(C)$ est nécessairement convexe si $\mathbf{a}$ est affine).

  Attention! $f \circ \mathbf{a}$ *n'est pas nécessairement strictement convexe sur* C*, même si* f *est strictement convexe sur* $\mathbf{a}(C)$.

  Contre-exemple 9.10 $f = e^{x+y}$ *est convexe, mais pas strictement convexe sur* $I\!R^2$ *(le vérifier par le calcul de la Hessienne).*

  Cependant, si $\mathbf{a}$ est à la fois *affine* sur C et *injective*, et si f est *strictement convexe* sur $\mathbf{a}(C)$, $f \circ \mathbf{a}$ est encore *strictement convexe* sur $\mathbf{a}(C)$.

- Enveloppe supérieure d'une famille de fonctions convexes :

  Si $\mathcal{F}$ est une famille (finie ou infinie) de fonctions convexes sur C, leur borne supérieure point par point : $\max\{ f \in \mathcal{F} \}$ est convexe sur C

  Exemple 9.11 $f = \max(e^{x+y}, e^{x^2+y^2})$ *est convexe sur* $I\!R^2$.

Attention! *Le produit de deux fonctions convexes sur* C *n'est pas, en général, convexe sur* C.

Contre-exemple 9.12 *Les applications linéaires :* $(x, y) \mapsto x$*, et :* $(x, y) \mapsto y$ *sont convexes sur* $I\!R^2$*, mais pas leur produit :* $(x, y) \mapsto x\,y$ *(le vérifier par le calcul de la Hessienne).*

9.6 Inégalités de convexité généralisées

Théorème 9.6 *f est convexe (resp. affine, strictement convexe) sur C si et seulement si, pour tout entier $M \geq 2$, tout ensemble fini : $x_1, \ldots, x_M$ de points distincts de C, et toute suite : $t_1, \ldots, t_M$ de réls strictement positifs et de somme un :*

$$(11) \quad f\Big[\sum_{k=1}^{M} t_k\, x_k\Big] \leq \sum_{k=1}^{M} t_k\, f(x_k) \qquad \textit{(resp. =, <)}$$

Preuve : Notons $(\mathcal{P})$ la propriété : « Pour toute suite : $x_1, \ldots, x_M$ de M points distincts de C, et toute suite : $t_1, \ldots, t_M$ de M réels strictement positifs et de somme un, l'inégalité (11) est vérifiée ». Par définition, f est convexe (resp. affine, strictement convexe) si $(\mathcal{P})$ est vérifiée pour : $M = 2$. Pour démontrer le théorème 9.6, il suffit donc de prouver que, si f est convexe (resp. affine, strictement convexe) sur C, et donc si $(\mathcal{P})$ est vérifiée pour : $M = 2$, alors $(\mathcal{P})$ est nécessairement vérifiée pour tout entier : $M \geq 2$.

Supposons f convexe (resp. affine, strictement convexe) sur C, et $(\mathcal{P})$ vérifiée pour un entier : $M \geq 2$ donné, et considérons une suite : $x_1, \ldots, x_{M+1}$ de $M+1$ points distincts de C, et une suite : $t_1, \ldots, t_{M+1}$ de $M+1$ réels strictement positifs de somme un.

Posons : $t = t_{M+1}$, et : $s_k = \dfrac{t_k}{1-t}$ $(k = 1, \ldots, M)$, de sorte que : $0 < t < 1$, et : $s_1, \ldots, s_M$ est une suite de M réels strictement positifs de somme un. Puisque C est convexe et f convexe (resp. affine, strictement convexe) sur C, le point : $\sum_{k=1}^{M} s_k\, x_k$ appartient à C, et :

$$f\left[\sum_{k=1}^{M+1} t_k\, x_k\right] \leq f\left[t\, x_{M+1} + (1-t)\sum_{k=1}^{M} s_k\, x_k\right] \leq t\, f\left[x_{M+1}\right] + (1-t)\, f\left[\sum_{k=1}^{M} s_k\, x_k\right]$$

(resp. $=$ si f est affine) d'où, en utilisant l'hypothèse de récurrence :

$$f\left[\sum_{k=1}^{M+1} t_k\, x_k\right] \leq t\, f(x_{M+1}) + \sum_{k=1}^{M} (1-t)\, s_k\, f(x_k) = \sum_{k=1}^{M+1} t_k\, f(x_k) \qquad (\text{resp. } =, <)$$

Le résultat valant quelles que soient les suites : $x_1, \ldots, x_{M+1}$ de points distincts de C, et : $t_1, \ldots, t_{M+1}$ de réels strictement positifs de somme un, $(\mathcal{P})$ est vraie pour $M+1$.

Puisque $(\mathcal{P})$ est vraie pour $M = 2$, on conclut qu'elle est vraie pour tout entier M, d'où le résultat.

□

Exemple 9.13 *L'inégalité arithmético-géométrique :*

$$x_k > 0,\ r_k > 0,\ \sum_{k=1}^{M} r_k = 1 \quad \Rightarrow \quad \prod_{k=1}^{M} x_k^{r_k} \leq \sum_{k=1}^{M} r_k x_k$$

(Appliquer l'inegalite de convexité (9.6) à la fonction : $\varphi(t) = -\log t$, *convexe sur* $]\,0, +\infty\,[$*).*

Exemple 9.14 *L'inégalité de Jensen : si f est convexe sur C, et X est une variable aléatoire à valeurs dans C : $f\,[\,E(X)\,] \leq E\,[\,f(X)\,]$, où : $E(X)$ (resp. $E\,[\,f(X)\,]$) désigne l'espérance de X (resp. de : $Y = f(X)$).*

9.7 Utilisation de la convexité en optimisation

Théorème 9.7 *Si f est convexe sur C, les ensembles de niveau de f dans C sont convexes.*

Preuve : Notons : $S_\alpha = \{\, x \in C \mid f(x) \leq \alpha \,\}$ l'ensemble de niveau α de f dans C. Si x et y sont deux points distncts de S_α, et t un réel de l'intervalle $]\,0, 1\,[$, on a, par définition de S_α :

$$f[\,tx + (1-t)\,y\,] \leq t\,f(x) + (1-t)\,f(y) \leq t\,\alpha + (1-t)\alpha = \alpha$$

d'où le résultat.

□

Exemple 9.15 *L'ensemble :* $C = \left\{\, (x, y) \in \mathbb{R}^2 \mid x, y > 0, \ \text{ et : } \dfrac{1}{x\,y} + x + y \leq 3 \,\right\}$ *est une partie convexe de* $\mathbb{R}^2$.

En particulier, si f est convexe sur C et atteint son minimum sur C, *l'ensemble des points minimisant f sur C est toujours convexe* .

Théorème 9.8 *Si f est convexe sur C, et $x^\star$ est un point critique de f appartenant à C, $x^\star$ minimise f sur C. Si f est strictement convexe sur C, c'est l'unique minimum de f sur C.*

Preuve : C'est une conséquence directe du théorème 9.4

□

On notera que, dans cet énoncé, l'ensemble convexe C n'est pas nécessairement *ouvert*. Le résultat fournit cependant une réciproque partielle du principe de Fermat :

- Tout minimum de f sur un ouvert Ω est un point critique de f.
- Réciproquemment, si Ω est un ouvert *convexe* et f est *convexe* sur Ω, tout point critique de f appartenant à Ω minimise f sur Ω.

On dit d'un problème de minimisation, avec ou sans contraintes, qu'il est *convexe* lorsque son ensemble admissible $\mathcal{A}$ est convexe et son critère convexe sur $\mathcal{A}$.

Le théorème 9.8 montre que la résolution d'un problème de minimisation convexe *sans contrainte* se réduit à trouver un point critique admissible du critère.