FONCTIONS CONVEXES

CADRE : Les fonctions f considérées ici sont définies sur une partie convexe C de $\mathbb{R}^N$, $N \geq 1$, et à valeurs réelles.

1 Définitions, premières propriétés

1.1 Convexité

Définition 1.1 *(Convexité, convexité stricte, convexité forte)*
Soit $f : C \subset \mathbb{R}^N \to \mathbb{R}$.

- *f est dite convexe sur C si*

$$\forall \theta \in]0,1[, \forall x, y \in C, f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y).$$

- *f est dite strictement convexe sur C si*

$$\forall \theta \in]0,1[, \forall x, y \in C, x \neq y, f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y).$$

- *f est dite fortement convexe sur C s'il existe $\alpha > 0$ tel que*

$$\forall \theta \in]0,1[, \forall x, y \in C, x \neq y, f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y) - \frac{1}{2}\alpha\theta(1-\theta)\|x-y\|^2.$$

  On dit aussi que f est α-convexe.

Remarque 1.1 *1. α-convexe $\Longrightarrow$ strictement convexe $\Longrightarrow$ convexe.*

2. f est α-convexe sur C si et seulement si $f - 1/2\alpha\|\cdot\|^2$ est convexe sur C.

Remarque 1.2 *La forte convexité est le cadre agréable pour de nombreux problèmes d'optimisation, car elle donne facilement l'existence, l'unicité, et des algorihmes de calcul performants ; c'est une hypothèse forte, mais elle inclut le cas des fonctions quadratiques, très important en pratique (cf ex.3 section 2.2).*

On montre aisément, à partir de la définition, les propriétés suivantes :

Proposition 1.1 – *Une somme de fonctions convexes est convexe. Un supremum de fonctions convexes est convexe.*

Proposition 1.2 *Soit f une fonction continue sur un convexe C. Alors*

1. *f est convexe sur C si et seulement si $\forall x, y \in C, f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$;*
2. *f est α-convexe sur C si et seulement si*
$$\forall x, y \in C, f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} - \frac{\alpha}{8}\|x-y\|^2.$$

Preuve. – Il suffit de démontrer le deuxième point ; on obtiendra 1. en prenant $\alpha = 0$. Supposons donc que f vérifie la condition, et prenons $\theta \in]0,1[$, $x, y \in C$. En utilisant la décomposition diadique, on sait qu'il existe une suite $(\theta_n)_{n\in\mathbb{N}}$, avec $\theta_n = p_n/2^n$, $p_n \in \mathbb{N}$, qui converge vers θ quand n tend vers $+\infty$. On montre par récurrence sur n que

$$f(\theta_n x + (1-\theta_n)y) \leq \theta_n f(x) + (1-\theta_n) f(y) - \frac{\alpha}{2}\theta_n(1-\theta_n)\|x-y\|^2.$$

Il reste à faire tendre n vers l'infini et utiliser la continuité de f. □

Remarque 1.3 *L'hypothèse supplémentaire de continuité peut sembler restrictive. En réalité, on montre que toute fonction convexe est continue dans l'intérieur de son domaine. Voir section 3.*

1.2 Premiers exemples

1. Toute fonction affine est trivialement convexe.
2. Si f est convexe, et φ convexe croissante, alors $\varphi \circ f$ est convexe. (La preuve est immédiate à partir de la définition.)
3. Soit $\varphi : [a, b] \to \mathbb{R}$ une fonction croissante, alors la fonction Φ définie sur $[a, b]$ par
$$\Phi(x) = \int_a^x \varphi(u)du$$
est convexe.
Remarque : si φ est continue, alors Φ est dérivable sur $[a, b]$, et sa convexité est une conséquence immédiate des résultats de la section 2. Cependant une preuve directe est possible, qui s'étend au cas où φ n'est pas continue.

Preuve. – Soient $x, y \in [a,b]$, $\alpha \in]0,1[$, et supposons par exemple que $x < y$. Calculons la différence

$$\Delta = \Phi(\alpha x + (1-\alpha)y) - \alpha\Phi(x) - (1-\alpha)\Phi(y).$$

La relation de Chasles sur les intégrales donne

$$\begin{aligned}
\Delta &= -(1-\alpha)\int_x^y \varphi(u)du + \int_x^{x+(1-\alpha)(y-x)} \varphi(u)du \\
&= -(1-\alpha)\int_0^{y-x} \varphi(x+u)du + (1-\alpha)\int_0^{y-x} \varphi(x+(1-\alpha)v)dv \\
&= (1-\alpha)\int_0^{y-x} \left(\varphi(x+(1-\alpha)u) - \varphi(x+u)\right)du \le 0,
\end{aligned}$$

en utilisant l'hypothèse φ croissante. □

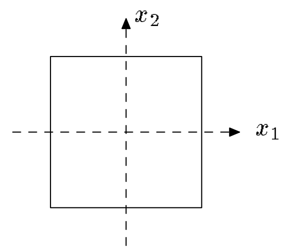
4. Une norme sur $\mathbb{R}^N$ (et même sur un evn quelconque) est toujours convexe :

$$\forall x, y \in E, \forall \alpha \in [0,1], \|\alpha x + (1-\alpha)y\| \le \alpha\|x\| + (1-\alpha)\|y\|.$$

mais n'est pas jamais strictement convexe, puisqu'elle est positivement homogène :

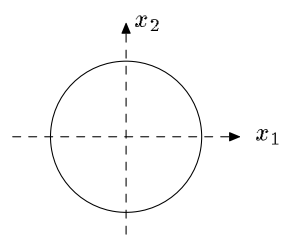
$$\forall x \in E, \forall \alpha \in [0,1], \forall \lambda > 0, \|\alpha x + (1-\alpha)\lambda x\| \le \alpha\|x\| + (1-\alpha)\|\lambda x\|.$$

5. En utilisant les points 2 et 4, on obtient que l'application $\|\cdot\|^2$ et toujours convexe ; par contre, elle n'est pas nécessairement strictement convexe.
Contre-exemple : $\mathbb{R}^2$, muni de la norme $\|\cdot\|_\infty$, cf dessin de la boule unité ; on a



$$\left\|\frac{(1,-1)+(1,1)}{2}\right\|^2 = 1 = \frac{1}{2}\left(\|(1,-1)\|^2 + \|(1,1)\|^2\right).$$

6. Par contre, si $\|\cdot\|$ est une norme dérivant d'un produit scalaire (même en dimension infinie), l'application $\|\cdot\|^2$ est strictement, et même fortement convexe ; en effet, elle est continue et vérifie :



$$\forall x, y, \left\|\frac{x+y}{2}\right\|^2 = \frac{1}{2}\left(\|x\|^2 + \|y\|^2\right) - \frac{1}{4}\|x-y\|^2.$$

(Cette propriété implique l'existence d'un projeté sur une partie convexe fermée non vide dans un Hilbert, propriété qui n'existe plus en général dans le cas d'un Banach.)

D'autres exemples importants seront listés en section 2 ; en effet, la caractérisation des fonctions convexes dans le cas différentiable permet de fournir de nombreux exemples.

1.3 Caractérisations et propriétés

Les caractérisations suivantes se démontrent facilement à partir de la définition.

Proposition 1.3 *(Critère de l'épigraphe)*
f convexe si et seulement si son épigraphe est convexe, où

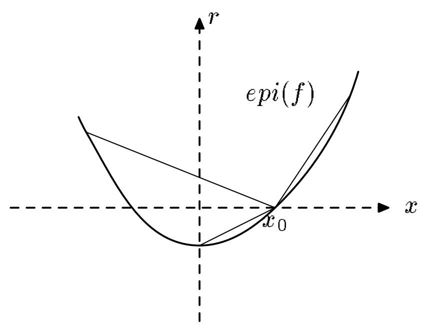
$$epi(f) = \{(x, r) \in C \times \mathbb{R}, f(x) \leq r\}.$$

La preuve est laissée au lecteur.

Proposition 1.4 *(Critère des pentes croissantes)*
Supposons $N = 1$; C est donc un intervalle de $\mathbb{R}$. Alors, f est convexe sur C si et seulement si, pour tout $x_0 \in C$, la fonction

$$x \mapsto \frac{f(x) - f(x_0)}{x - x_0}$$

est croissante sur $C \backslash \{x_0\}$.



Preuve. – Supposons d'abord f convexe. Soient $x_0 \in C, x, y \in C \backslash \{x_0\}$. Supposons $x < y$. Montrons l'inégalité :

$$\frac{f(x) - f(x_0)}{x - x_0} \leq \frac{f(y) - f(x_0)}{y - x_0}.$$

– Si $x_0 < x$, alors il existe $\theta \in]0, 1[$ tel que $x = \theta x_0 + (1 - \theta)y$. On a alors :

$$f(x) \leq \theta f(x_0) + (1 - \theta) f(y).$$

On obtient le résultat voulu en soustrayant $f(x_0)$ à chgaque membre, et en divisant par $x - x_0$, qui est positif.

– Le cas $x_0 > y$ se traite de façon analogue.
– Si par contre $x < x_0 < y$, alors il existe $\theta \in]0,1[$ tel que $x_0 = \theta x + (1-\theta)y$. On a alors :

$$f(x_0) \leq \theta f(x) + (1-\theta) f(y).$$

On soustrait $f(x_0)$ aux deux membres, et on divisse par $x - x_0$ qui est négatif. Le résultat en decoule alors, en remarquant que

$$y - x_0 = \theta(y - x) = -\frac{\theta}{1-\theta}(x - x_0).$$

Pour la réciproque, soient $x, y \in C$, $\theta \in]0,1[$. Supposons $x < y$. On note $x_\theta = \theta x + (1-\theta)y$. Alors, l'hypothèse assure que

$$\frac{f(x_\theta) - f(x)}{x_\theta - x} \leq \frac{f(y) - f(x)}{y - x}.$$

Après quelques calculs on en déduit que

$$f(x_\theta) \leq \theta f(x) + (1-\theta) f(y).$$

□

Remarque : ces propriétés sont de nature géométrique. Dans le cas d'une fonction de la variable réelle, la première caractérisation traduit simplement le fait que le graphe de f est toujours situé sous les cordes, tandis que la deuxième signifie que la pente des cordes reliant les points d'abscisse x_0 et d'abscisse x croît avec x.

Comme corollaire de la proposition 2, nous pouvons énoncer le résultat suivant, parfois utile, et pas si facile à démontrer directement :

Corollaire 1.5 – *Soit $f :]0,+\infty[\to \mathbb{R}$, on pose, pour $x > 0, g(x) = x f\left(\frac{1}{x}\right)$. Alors g est convexe si et seulement si f est convexe.*

Preuve. – Remarquons d'abord que $x g\left(\frac{1}{x}\right) = f(x)$, donc il suffit de démontrer une des implications.

Supposons donc f convexe. Soit $x_0 > 0$, alors la fonction

$$x \mapsto s_f(1/x) = \frac{f(1/x) - f(1/x_0)}{1/x - 1/x_0} = \frac{x x_0}{x - x_0}(f(1/x) - f(1/x_0))$$

est décroissante. Considérons la fonction pente de g définie sur $]0,+\infty[\setminus\{x_0\}$ par:

$$\begin{aligned}
s_g(x) &= \frac{g(x)-g(x_0)}{x-x_0} \\
&= \frac{xf(1/x)-x_0f(1/x_0)}{x-x_0} \\
&= \frac{x-x_0}{x-x_0}f(1/x_0)+\frac{x}{x-x_0}(f(1/x)-f(1/x_0)) \\
&= f(1/x_0)-\frac{1}{x_0}s_f(1/x).
\end{aligned}$$

Il apparaît ainsi que s_g est croissante, donc g est convexe. □

2 Caractérisations dans le cas différentiable

Dans cette section, on se place dans le cadre de fonctions définies au voisinage d'une partie convexe de $\mathbb{R}^N$. Soit donc $C \subset \mathbb{R}^N$, convexe, et Ω un ouvert contenant C. On suppose ici que f est différentiable sur Ω.

On va établir des résultats qui généralisent les critères bien connus en dimension 1 :

- une fonction dérivable est convexe si et seulement si sa dérivée est croissante ;
- une fonction dérivable est convexe si et seulement si son graphe est situé au-dessus des tangentes.

2.1 Encore une caractérisation géométrique

Théorème 2.1 – *Soit Ω un ouvert de $\mathbb{R}^N$, $C \subset \Omega$, convexe, et $f : \Omega \to \mathbb{R}$, différentiable. On a les équivalences suivantes :*

1. *f est convexe sur C si et seulement si*
$$\forall x, x_0 \in C,\ f(x) \geq f(x_0) + \langle \nabla f(x_0), x-x_0\rangle.$$
2. *f est strictement convexe sur C si et seulement si*
$$\forall x, x_0 \in C, x \neq x_0,\ f(x) > f(x_0) + \langle \nabla f(x_0), x-x_0\rangle.$$
3. *f est α-convexe sur C si et seulement si*
$$\forall x, x_0 \in C,\ f(x) \geq f(x_0) + \langle \nabla f(x_0), x-x_0\rangle + \frac{\alpha}{2}\|x-x_0\|^2.$$

Remarque 2.1 *Le membre de droite des inégalités 1 et 2 donne l'équation de l'hyperplan tangent au graphe en x_0, tandis que celui de l'inégalité 3 donne l'équation d'un paraboloïde tangent au graphe en x_0.*

Preuve. –

1. Montrons d'abord la première équivalence. Supposons donc f convexe. Soient $x, x_0 \in C$, on a donc, pour tout $\theta \in]0,1[$,

$$f(\theta x + (1-\theta)x_0) \leq \theta f(x) + (1-\theta) f(x_0),$$

soit :

$$\frac{f(\theta x + (1-\theta)x_0) - f(x_0)}{\theta} \leq f(x) - f(x_0).$$

En passant à la limite quand θ tend vers 0, on obtient le résultat voulu.

Réciproquement, supposons les inégalités vérifiées et montrons que f est convexe. Soient donc $x, y \in C, \theta \in]0,1[$. Posons $x_0 = \theta x + (1-\theta)y$. On a donc :

$$\begin{array}{rcl} f(x) & \geq & f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle \\ f(y) & \geq & f(x_0) + \langle \nabla f(x_0), y - x_0 \rangle \end{array}$$

En multipliant la première inégalité par θ, la seconde par $1-\theta$, et en les ajoutant, on obtient exactement :

$$\theta f(x) + (1-\theta) f(y) \geq f(x_0) = f(\theta x + (1-\theta)y).$$

2. Pour le point 2, la réciproque se montre comme ci-dessus en remplaçant les inégalités larges par des inégalités strictes. Par contre, on ne peut pas calquer la preuve du sens direct, car les inégalités strictes deviendraient larges à la limite $\theta \to 0$. On peut cependant appliquer le résultat du point 1. Soit donc f supposée strictement convexe, et deux points $x, x_0 \in C, x \neq x_0$. On a alors, pour tout $\theta \in]0,1[$,

$$f(x_\theta) < \theta f(x) + (1-\theta) f(x_0), \text{ où } x_\theta = \theta x + (1-\theta)x_0.$$

Mais on a aussi, d'après le résultat 1. qu'on vient de montrer,

$$f(x_\theta) \geq f(x_0) + \langle \nabla f(x_0), x_\theta - x_0 \rangle.$$

En comparant ces deux inégalités, et en simplifiant $f(x_0)$, on obtient :

$$\theta \langle \nabla f(x_0), x - x_0 \rangle < \theta (f(x) - f(x_0)).$$

Il reste à diviser par θ pour obtenir le résultat recherché.

3. Pour le troisième point, on utilise la remarque 1.1 :

$$\begin{array}{lcl} f\ \alpha - \text{convexe} & \Longleftrightarrow & f - \dfrac{\alpha}{2}\|\cdot\|^2 \text{convexe} \\ & \Longleftrightarrow & \forall x, x_0 \in C, \\ & & f(x) - \dfrac{\alpha}{2}\|x\|^2 \geq f(x_0) - \dfrac{\alpha}{2}\|x_0\|^2 + \langle \nabla f(x_0) - \alpha x_0, x - x_0 \rangle. \end{array}$$

Après calculs, on arrive à l'équivalence voulue.

□

2.2 Fonctions monotones

Définition 2.1 – *Soient $C \subset \mathbb{R}^n$, convexe, $\alpha \geq 0$ et soit une fonction $F : C \to \mathbb{R}^n$.*

- *F est dite monotone sur C si*

$$\forall x, y \in C, \langle F(x) - F(y), x - y \rangle \geq 0.$$

- *F est dite strictement monotone sur C si*

$$\forall x, y \in C, x \neq y, \langle F(x) - F(y), x - y \rangle > 0.$$

- *F est dite α-monotone sur C si*

$$\forall x, y \in C, \langle F(x) - F(y), x - y \rangle \geq \alpha \|x - y\|^2.$$

Théorème 2.2 – *Si f est différentiable sur Ω, on a les équivalences suivantes :*

1. *f est convexe sur C $\iff$ ∇f est monotone sur C.*
2. *f est strictement convexe sur C $\iff$ ∇f est strictement monotone sur C.*
3. *f est α-convexe sur C $\iff$ ∇f est α-monotone sur C.*

Preuve. – Montrons le troisième point. Le premier s'en déduira en prenant $\alpha = 0$, et le deuxième se montre de manière analogue en remplaçant les inégalités larges par des inégalités strictes.

Supposons d'abord f α-convexe. Soient $x, y \in C$. Alors on a, d'après le théorème 1,

$$\begin{aligned} f(y) &\geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2. \\ f(x) &\geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{\alpha}{2}\|y - x\|^2. \end{aligned}$$

En sommant ces deux inégalités, on obtient :

$$0 \geq \langle \nabla f(x) - \nabla f(y), y - x \rangle + \alpha \|y - x\|^2,$$

i.e. ∇f est α-monotone.

Supposons maintenant que ∇f est α-monotone. Posons $\varphi(t) = f(x + t(y - x))$, $t \in [0, 1]$. Alors φ est dérivable sur $[0, 1]$ et pour tout t,

$$\varphi'(t) = \langle \nabla f(x_t), y - x \rangle, \text{ où } x_t = x + t(y - x).$$

Soit $t \in [0, 1]$, alors

$$\varphi'(t) - \varphi'(0) = \frac{1}{t}\langle \nabla f(x_t) - \nabla f(x), x_t - x \rangle \geq \frac{\alpha}{t}\|x_t - x\|^2 = \alpha t \|y - x\|^2.$$

En appliquant l'inégalité des accroissements finis sur $[0,1]$, on obtient :

$$\varphi(1)-\varphi(0) \geq \varphi'(0) + \int_0^1 \alpha t \|y-x\|^2 dt = \varphi'(0) + \frac{\alpha}{2}\|y-x\|^2,$$

ce qui se réécrit :

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

Le théorème 1 permet de conclure que f est α-convexe. □

Corollaire 2.3 – *Soit maintenant f deux fois différentiable sur Ω. Alors f est convexe sur C si et seulement si, pour tout $x \in C$, $D^2 f(x)$ définit une forme quadratique positive, i.e.*

$$\forall x \in C, \forall v \in \mathbb{R}^N, \langle D^2 f(x) v, v \rangle \geq 0.$$

Preuve. – La preuve repose sur le théorème précédent et la formule de Taylor-Lagrange à l'ordre 2.

2.3 Encore des exemples

1. sur $\mathbb{R}$, les fonctions $x \mapsto |x|^\alpha$, $\alpha > 1$, $x \mapsto e^{\beta x}$, $\beta \in \mathbb{R}$, sont convexes.

2. sur $]0,+\infty[$, les fonctions $x \mapsto x^{-\gamma}$, $\gamma > 0$, $x \mapsto -\ln x$, sont convexes. D'après le corollaire 1.4, $x \mapsto x \ln x$ est convexe.

3. *Exemple important pour les applications :*
   Soient A une matrice symétrique réelle de taille N, $b \in \mathbb{R}^N$, et f une fonction définie sur $\mathbb{R}^N$ par :

   $$f(x) = \frac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle.$$

   Alors
   f est convexe $\iff$ A est positive : $\langle Ax, x \rangle \geq 0, \forall x \in \mathbb{R}^N$,
   $\iff$ toutes les valeurs propres de A sont positives ou nulles.
   f est strictement convexe $\iff$ A est définie positive : $\langle Ax, x \rangle > 0, \forall x \neq 0$,
   $\iff$ toutes les valeurs propres de A sont strictement positives.
   Dans ce cas, f est même fortement convexe.

Preuve. – f est différentiable, et $\nabla f(x) = Ax - b$. A est symétrique, donc ses valeurs propres sont réelles, notons les $\lambda_1 \geq \cdots \geq \lambda_N$ (avec multiplicité). S'il existe une valeur propre strictement négative, il est clair que ∇f n'est pas monotone. Supposons donc le contraire. En diagonalisant A (qui est symétrique) dans une base orthonormée, on montre facilement que

$$\forall x, y \in \mathbb{R}^n, \langle Ax - Ay, x - y \rangle \geq \lambda_N \|x - y\|^2,$$

et que l'égalité est atteinte lorsque $x - y$ est vecteur propre associé à λ_N. Ainsi, d'après le théorème 2.2, si $\lambda_N = 0$, f est convexe mais non strictement convexe ; si $\lambda_N > 0$, f est λ_N-convexe.

Remarque 2.2 – *Cet exemple est important car il fournit des algorithmes de résolution de systèmes linéaires. En effet, dans le cas où f est convexe et différentiable, on verra que son gradient s'annule uniquement lorsque f atteint un minimum (voir théorème 3.3). On peut donc appliquer des algorithmes de minimisation de f pour résoudre le système $Ax = b$ (en particulier, méthodes de gradient, gradient conjugué, etc.).*

4. Une fonction de deux variables (x, y) qui est convexe par rapport à x et convexe par rapport à y n'est pas nécessairement convexe par rapport au couple (x, y).
   Contre-exemple : la fonction f définie sur $\mathbb{R}^2$ par : $f(x, y) = (1 + x^2)(1 + y^2)$, n'est pas convexe sur $\mathbb{R}^2$. En effet, sa hessienne en (x, y) vaut :

   $$D^2 f(x, y) = 2 \begin{pmatrix} 1 + y^2 & 2xy \\ 2xy & 1 + x^2 \end{pmatrix}.$$

   Au point (2,2), elle vaut donc

   $$D^2 f(2, 2) = \begin{pmatrix} 10 & 16 \\ 16 & 10 \end{pmatrix},$$

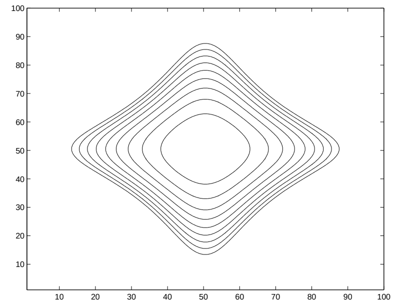
   dont le déterminant vaut $-156 < 0$, donc une valeur propre est négative strictement.

5. *Somme des plus grandes valeurs propres d'une matrice symétrique*
   Soit $E = S_n(\mathbb{R})$ l'espace des matrices symétriques réelles de taille $n \geq 1$, muni du produit scalaire

   $$\langle\langle A, B \rangle\rangle = tr(AB) = \sum_{i,j=1}^{n} A_{ij} B_{ij}.$$

   On associe à une matrice $A \in E$ ses valeurs propres, rangées par ordre décroissant :

   $$\lambda_1(A) \geq \cdots \geq \lambda_n(A),$$



(a) (b)

FIG. 1 – (a) Représentation de la fonction f sur le domaine $[-4, +4] \times [-4, +4]$. (b) courbes de niveau de f pour des valeurs entières de 2 à 10.

et une base orthonormée de vecteurs propres $(v_1, \cdots v_n)$. Pour tout entier m, $1 \leq m \leq n$, on note

$$f_m(A) = \sum_{j=1}^{m} \lambda_j(A).$$

On va montrer que f_m est convexe, en vérifiant l'égalité :

$$\forall A \in E, f_m(A) = \sup\{\langle\langle Q\ {}^tQ, A\rangle\rangle, Q \in \Omega_m\}, \tag{1}$$

où $\Omega_m = \{Q \in M_{n,m}(\mathbb{R}), {}^tQ.Q = I_m\}$.

1ère étape : Notons, pour $Q \in \Omega_m$, $\varphi_A(Q) = \langle\langle Q\ {}^tQ, A\rangle\rangle$. Remarquons que :

$$\forall Q \in \Omega_m, \quad \varphi_A(Q) = tr(Q^tQA) = tr({}^tQAQ) = \sum_{j=1}^{m} \langle AC_j, C_j\rangle, \tag{2}$$

où $(C_1, \cdots C_m)$ sont les vecteurs colonnes de la matrice Q, et $\langle \cdot, \cdot \rangle$ désigne le produit scalaire usuel de $\mathbb{R}^n$. (En particulier, l'égalité (1) est claire pour $m = 1$ et pour $m = n$.) De plus, la fonction φ_A est différentiable sur Ω_m, de différentielle

$$D\varphi_A(\bar{Q}) \cdot Q = 2 \sum_{j=1}^{m} \langle A\bar{C}_j, C_j\rangle$$

où $(C_1, \cdots C_m)$ sont les vecteurs colonnes de la matrice Q, $(\bar{C}_1, \cdots \bar{C}_m)$ sont les vecteurs colonnes de la matrice $\bar{Q}$.

Par ailleurs,

$$\Omega_m = \{Q \in M_{n,m}(\mathbb{R}), \forall i, j = 1 \cdots m, \langle C_i, C_j\rangle = \delta_{ij}\}.$$

En particulier, Ω_m est fermé et borné donc compact. Ainsi, le sup de φ_A sur Ω_m est bien défini et est atteint.

2ème étape : Notons encore, pour $1 \leq i, j \leq m$:

$$\forall Q \in M_{n,m}(\mathbb{R}), \quad h_{i,j}(Q) = \langle C_i, C_j \rangle - \delta_{ij} \in \mathbb{R}.$$

Chaque fonction h_{ij} est différentiable, de différentielle définie par :

$$Dh_{i,j}(\bar{Q}) \cdot Q = 2\langle \bar{C}_i, C_j \rangle.$$

En particulier, on voit qu'en tout point $\bar{Q} \in \Omega_m$, ces différentielles sont linéairement indépendantes. En effet, si $(\lambda_{i,j})_{1 \leq i,j \leq m}$ sont tels que

$$\forall Q \in M_{n,m}(\mathbb{R}), \quad \sum_{i,j=1}^{m} \lambda_{i,j} Dh_{i,j}(\bar{Q}) \cdot Q = 0,$$

alors en choisissant une matrice Q dont les vecteurs colonnes sont nuls sauf le vecteur C_i qui est égal à $\bar{C}_j$, on montre que $\lambda_{i,j}$ est nul.
On peut donc appliquer le théorème des extrema liés qui donne une condition nécessaire sur les points où le maximum est atteint. Ainsi, si le maximum de φ_A sur Ω_m est atteint en $\bar{Q}$, il existe des réels $(\lambda_{i,j})_{1 \leq i,j \leq m}$ tels que

$$\forall Q \in M_{n,m}(\mathbb{R}), \quad D\varphi_A(\bar{Q}) \cdot Q = \sum_{i,j=1}^{m} \lambda_{i,j} Dh_{i,j}(\bar{Q}) \cdot Q.$$

Il en découle que

$$\forall C_1, \cdots C_m \in \mathbb{R}^n, \quad \sum_{j=1}^{m} \langle A\bar{C}_j, C_j \rangle = \sum_{i,j=1}^{m} \lambda_{i,j} \langle \bar{C}_i, C_j \rangle,$$

donc

$$A\bar{C}_j = \sum_{i=1}^{m} \lambda_{i,j} \bar{C}_i, \quad 1 \leq j \leq m.$$

Ainsi, l'endomorphisme canoniquement associé à la matrice A laisse stable le sous-espace de $\mathbb{R}^n$ engendré par les vecteurs $\bar{C}_1, \cdots \bar{C}_m$, et la matrice $\Lambda = (\lambda_{i,j})_{1 \leq i,j \leq m}$ n'est autre que la matrice de la restriction de cet endomorphisme dans la base $(\bar{C}_1, \cdots \bar{C}_m)$. On a alors : $\varphi_A(\bar{Q}) = \sum_{j=1}^{m} \langle A\bar{C}_j, \bar{C}_j \rangle = tr(\Lambda)$. Or les valeurs propres de Λ sont m valeurs propres de A, on a donc : $\varphi_A(\bar{Q}) = tr(\Lambda) \leq f_m(A)$, et donc :

$$\forall Q \in \Omega_m, \quad \varphi_A(Q) \leq \varphi_A(\bar{Q}) \leq f_m(A).$$

3ème étape : Il est facile de vérifier que la valeur $f_m(A)$ est effectivement atteinte par φ_A sur Ω_m. Il suffit pour cela de choisir une matrice $\bar{Q}$ dont les colonnes sont les vecteurs propres $(v_1, \cdots v_m)$ associés aux valeurs propres $\lambda_1(A), \cdots \lambda_m(A)$, et d'appliquer la formule (2). On a donc démontré l'égalité (1).

4ème étape : La convexité de f_m découle directement de cette égalité car on a vu qu'un sup de fonctions convexes est toujours convexe (proposition 1.1).
En prenant $m = 1$, on obtient en particulier la convexité de la première valeur propre λ_1. En prenant $m = n-1$, on obtient que $f_{n-1} = tr - \lambda_n$ est convexe. Or la trace est linéaire, donc concave ; on en déduit donc que la plus petite valeur propre $\lambda_n = tr - f_{n-1}$ est concave comme somme de fonctions concaves.

3 Quelques résultats importants sur les fonctions convexes

3.1 Inégalités de convexité

Théorème 3.1 – *Soit $f : C \to \mathbb{R}$, convexe. Alors, pour tout entier $k \geq 2$,*

$$\forall x_1, \cdots x_k \in C,\ \forall \alpha = (\alpha_1, \cdots \alpha_k) \in [0,1]^k,\ \sum_{i=1}^{k} \alpha_i = 1,$$

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

Preuve. – Pour $k = 2$, c'est la définition ; on passe à des entiers $k \geq 3$ par une récurrence facile. Voir par exemple [2].

Applications : voir par exemple [2]

$$-\ln \text{ convexe} \quad \Rightarrow \quad \prod_{i=1}^{k} x_i^{\alpha_i} \leq \sum_{i=1}^{k} \alpha_i x_i.$$

$$x \mapsto x \ln x \text{ convexe} \quad \Rightarrow \quad \left(\sum_{i=1}^{k} \alpha_i x_i\right)^{\sum_{i=1}^{k} \alpha_i x_i} \leq \prod_{i=1}^{k} x_i^{\alpha_i x_i}.$$

$$x \mapsto \frac{-1}{1+e^{-x}} \text{ convexe} \quad \Rightarrow \quad \sum_{i=1}^{k} \frac{\alpha_i}{1+x_i} \leq \left(1 + \prod_{i=1}^{k} x_i^{\alpha_i}\right)^{-1}$$

3.2 Continuité, dérivabilité

Théorème 3.2 – *Soient Ω ouvert convexe de $\mathbb{R}^N$, et f convexe bornée sur Ω. Alors pour toute partie C incluse dans Ω, telle qu'il existe $\epsilon > 0$ avec $C + B_\epsilon \subset \Omega$, f est lipschitzienne sur C.*

Preuve. – voir par exemple [2] pour la preuve dans $\mathbb{R}^N$, ou [3] pour le cas $N = 1$ (c'est exactement la même preuve). Le théorème et sa preuve sont encore valables en dimension infinie.

Corollaire 3.3 – *Soit f une fonction convexe définie sur un ouvert convexe Ω de $\mathbb{R}^N$. Alors f est continue sur Ω, et lipschitzienne sur tout compact de Ω.*

Preuve. – voir par exemple [2]. Ici, le résultat ne se généralise pas à la dimension infinie (contre-exemple simple : il existe des fonctions linéaires non continues). Cela tient au fait qu'en dimension infinie, les fonctions continues ne sont plus nécessairement bornées sur les bornés. On ne peut donc pas leur appliquer le théorème précédent.

Remarque 3.1 – *De la propriété de Lipschitz découle, en utilisant le théorème de Rademacher (dur) que toute fonction convexe est différentiable presque partout sur son domaine (presque partout au sens de la mesure de Lebesgue). Voir par exemple [1] pour une preuve du théorème de Rademacher. (Attention, la démonstration proposée par [2] de ce résultat me semble fausse.)*

En dimension $N = 1$, on montre facilement que toute fonction convexe est dérivable sauf en un nombre au plus dénombrable de points, voir [2].

3.3 Minimisation "sans contrainte"

On parle de minimisation sans contrainte lorsqu'on minimise une fonction sur l'espace entier. (Si on minimise sur une partie C stricte de $\mathbb{R}^N$, on impose sur le minimiseur une certaine contrainte, qui est d'appartenir à C.)

Théorème 3.4 – *On suppose ici que $C = \mathbb{R}^N$. Soit f convexe sur $\mathbb{R}^N$, et $x_0 \in \mathbb{R}^N$. Alors*

1. *x_0 minimum local de f $\iff$ x_0 minimum global de f*
2. *l'ensemble des points de minimum est un convexe fermé (éventuellement vide);*
3. *si de plus, f est différentiable en x_0, alors x_0 minimum local de f $\iff$ $Df(x_0) = 0$.*
4. *si f est strictement convexe, alors il existe au plus un point de minimum ;*
5. *si f est fortement convexe, alors il existe exactement un point de minimum ;*

Preuve. – La démonstration est facile. Voir par exemple [2].

Références

[1] Azé, Eléments d'analyse convexe et variationnelle

[2] Hiriart-Urruty, Lemaréchal, Convex Analysis and Minimization Algorithms I, Springer-Verlag, 1996

[3] Zuily-Queffélec, analyse pour l'agrégation.