Cadre : I est un intervalle de $\mathbb{R}$ non réduit à un singleton, et C une partie d'un espace vectoriel E.

I Généralités autour de la convexité

1) Définition et premières propriétés

Définition 1. C est convexe si : $\forall\, a, b \in C,\ \forall\, t \in [0,1],\ (1-t)a + tb \in C$.

Exemple 2. *Un intervalle de $\mathbb{R}$ est convexe, une boule de E est convexe.*

Définition 3. On dit que la fonction $f : C \to \mathbb{R}$ est convexe lorsque, pour tous $a, b \in C$ et tout $\lambda \in [0,1]$, on a :

$$f((1-\lambda)a + \lambda b) \leqslant (1-\lambda) f(a) + \lambda f(b)$$

On dit que f concave si $-f$ est convexe. Lorsque l'inégalité est stricte pour $a \neq b$ et $0 < \lambda < 1$, f est strictement convexe. Pour $\alpha > 0$, on dit que f est α-convexe si pour tous $a, b \in C$ distincts et tout $\lambda \in]0,1[$, on a :

$$f((1-\lambda)a + \lambda b) \leqslant (1-\lambda) f(a) + \lambda f(b) - \frac{\alpha}{2} \|a - b\|^2 \lambda(1-\lambda)$$

Remarque 4. *L'α-convexité implique la stricte convexité, qui implique la convexité.*

Remarque 5. *Une fonction $f : C \to \mathbb{R}$ est convexe si l'ensemble $\{(x,y) \in C \times \mathbb{R} \,|\, y \geqslant f(x)\}$ est convexe.*

Exemple 6. *L'application $x \mapsto \|x\|$ est convexe*

Théorème 7. *Une fonction $f : C \to \mathbb{R}$ est convexe si, et seulement si, pour tous $x, y \in C$, $t \mapsto f((1-t)x + ty)$ est convexe sur $[0,1]$.*

Proposition 8.
- *(i) Une combinaison linéaire à coefficients réels positifs de fonctions convexes est convexe.*
- *(ii) La composée d'une fonction convexe croissante avec une fonction convexe est convexe.*
- *(iii) Une limite simple de fonctions convexes est convexe.*
- *(iv) Le maximum de deux fonctions convexes est convexe.*

Remarque 9. *Le produit de deux fonctions convexes n'est pas nécessairement convexe ($-x \cdot x^2 = x^3$), et leur composition non plus ($(x \mapsto -x) \circ (x \mapsto x^2) = (x \mapsto -x^2)$).*

Proposition 10. *Une fonction convexe sur I est continue sur $\mathring{I}$.*

2) Caractérisation des fonctions convexes

En dimension 1

Théorème 11. *Pour $f : I \to \mathbb{R}$, il y a équivalence entre :*
- *(i) f est convexe sur I.*
- *(ii) Pour $a < b < c$ dans I, on a : $\frac{f(b)-f(a)}{b-a} \leqslant \frac{f(c)-f(a)}{c-a} \leqslant \frac{f(c)-f(b)}{c-b}$*
- *(iii) Pour $a \in I$, la fonction $x \mapsto \frac{f(x)-f(a)}{x-a}$ est croissante sur $I \setminus \{a\}$.*

Corollaire 12. *Une fonction de $\mathbb{R}$ dans $\mathbb{R}$ est affine si, et seulement si, elle est convexe et concave.*

Théorème 13. *Soit $f : I \to \mathbb{R}$ dérivable. Il y a équivalence entre :*
- *(i) f est (strictement) convexe sur I.*
- *(ii) La fonction dérivée f' est (strictement) croissante.*
- *(iii) La courbe représentative de f est située (strictement) au-dessus de sa tangente en tout point de I.*

Proposition 14. *Si f est deux fois dérivable sur I, elle est alors convexe si, et seulement si, $f'' \geqslant 0$.*

En dimension $n \geqslant 1$

Théorème 15. *Soit $J : C \to \mathbb{R}$ différentiable. Il y a équivalence entre :*
- *(i) J est convexe sur C.*
- *(ii) $\forall\, x, y \in C,\ \langle \nabla J(x) - \nabla J(y), x - y \rangle \geqslant 0$.*
- *(iii) $\forall\, x, y \in C,\ J(x) \geqslant J(y) + \langle \nabla J(y), x - y \rangle$.*

Si J est deux fois différentiable, on a aussi : $\left\langle d^2 J(x) \cdot y, y \right\rangle \geqslant 0$.

Théorème 16. *Soit $J : C \to \mathbb{R}$ différentiable. Il y a équivalence entre :*
- *(i) J est α-convexe sur C.*
- *(ii) $\forall\, x, y \in C,\ \langle \nabla J(x) - \nabla J(y), x - y \rangle \geqslant \alpha \|x - y\|^2$.*
- *(iii) $\forall\, x, y \in C,\ J(x) \geqslant J(y) + \langle \nabla J(y), x - y \rangle + \frac{\alpha}{2} \|x - y\|^2$.*

Si J est deux fois différentiable, on a aussi : $\left\langle d^2 J(x) \cdot y, y \right\rangle \geqslant \alpha \|y\|^2$.

Exemple 17. *Si A est une matrice symétrique définie positive, alors la fonctionnelle quadratique $J : X \mapsto \langle AX, X \rangle - \langle B, X \rangle$ est λ_1-convexe, où λ_1 est la plus petite valeur propre de A.*

3) Inégalités de convexité

Proposition 18. *Soient $n \in \mathbb{N}^\star$ et $a_1, \ldots, a_n \in \mathbb{R}^+$, alors :*

$$\sqrt[n]{\prod_{i=1}^{n} a_i} \leqslant \frac{1}{n} \sum_{i=1}^{n} a_i$$

Il y a égalité si, et seulement si, tous les a_i sont égaux.

Proposition 19 (Young)**.** *Soient $p, q > 0$ tels que $\frac{1}{p} + \frac{1}{q} = 1$ et $a, b \in \mathbb{R}^+$:*

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

Il y a égalité si, et seulement si, $a^p = b^q$.

II Applications dans certains espaces

1) Espaces de Hilbert

Soit $(H, \langle ., . \rangle)$ un espace de Hilbert.

Théorème 20. *Pour tout $f \in H$, il existe un unique élément de K, noté $P_K(f)$, et appelé projection de f sur K, tel que :*

$$\|P_K(f) - f\| = \inf_{v \in K} \|v - f\|$$

De plus, $P_K(f)$ est caractérisée par :

$$\forall v \in K,\ \mathrm{Re}\left(\langle f - P_K(f), v - P_K(f) \rangle\right) \leqslant 0$$

Remarque 21. *L'application $x \mapsto P_K(x)$ est 1-lipschitzienne.*

Corollaire 22. *Soient M un sous-espace vectoriel fermé de H et $f \in H$. Alors $P_M(f)$ est caractérisé par :*

$$P_M(f) \in M \qquad \text{et} \qquad \forall\, v \in M,\ \mathrm{Re}\left(\langle f - P_M(f), v \rangle\right) = 0$$

De plus, P_M est un opérateur linéaire.

Théorème 23 (Riesz-Fréchet)**.** *Soit $\varphi \in H'$. Alors :*

$$\exists!\, f \in H,\ \forall\, v \in H,\ \langle \varphi, v \rangle = \langle f, v \rangle$$

Théorème 24 (Lax-Milgram)**.** *Soient H un espace de Hilbert, a une forme bilinéaire continue et coercive sur H, et $\ell \in H'$. Alors il existe un unique $u \in H$ tel que, pour tout $v \in H$, $a(u, v) = \ell(v)$. Si de plus a est symétrique, u réalise le minimum sur H de $v \mapsto \frac{1}{2} a(v, v) - \ell(v)$.*

Application 25 (Dirichlet)**.** *Pour $f \in L^2$, on considère le problème :*

$$\begin{cases} -u'' + u = f & \textit{sur }]0, 1[\\ u(0) = u(1) = 0 \end{cases}$$

Il existe une unique solution faible $u \in H_0^1([0, 1])$ à ce problème.

2) Espaces L^p

Soient $(E, \mathcal{A}, \mu)$ un espace mesuré, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, et $p, q > 0$ conjugués.

Définition 26. Pour tout réel $p > 0$, on définit le $\mathbb{K}$-espace vectoriel :

$$\mathscr{L}^p_{\mathbb{K}}(X, \mathcal{A}, \mu) = \left\{ f : X \to \mathbb{K} \text{ mesurable} \,\middle|\, \int_X |f|^p \, d\mu < +\infty \right\}$$

Sauf situation ambiguë, on privilégiera la notation plus concise $\mathscr{L}^p_{\mathbb{K}}(\mu)$.

Définition 27. Pour toute fonction $f : X \to \mathbb{K}$ et tout $p > 0$, on définit :

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{\frac{1}{p}} \quad \left(\text{convention : } \infty^{\frac{1}{p}} = \infty\right)$$

Proposition 28 (Hölder et Minkowski)**.** *Soient $a, b \in \mathbb{R}^+$ et $f, g : E \to \mathbb{K}$ mesurables, alors :*

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q \quad \text{et} \quad \|f + g\|_p \leqslant \|f\|_p + \|g\|_p$$

Définition 29. Pour $1 \leqslant p < +\infty$, on définit $L^p_{\mathbb{K}}(\mu)$ comme l'espace vectoriel normé quotient de $\mathscr{L}^p_{\mathbb{K}}(\mu)$ par les fonctions presque nulles. On associera par abus de langage un élément de $\mathscr{L}^p_{\mathbb{K}}(\mu)$ à sa classe dans $L^p_{\mathbb{K}}(\mu)$.

Théorème 30. $(L^p, \|.\|_p)$ *est un espace vectoriel normé.*

3) Espaces de probabilité

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire.

Proposition 31 (Jensen)**.** *Si $X \in L^1$ et ϕ est convexe, alors $\phi(\mathbb{E}[X]) \leqslant \mathbb{E}[\phi(X)]$. De plus, si ϕ est strictement convexe, on a égalité si, et seulement si, X est constante presque sûrement.*

Remarque 32. *En particulier, $|\mathbb{E}[X]| \leqslant \mathbb{E}[|X|]$ et $\mathbb{E}[X]^2 \leqslant \mathbb{E}[X^2]$.*

III Recherche d'extrema et de points fixes

1) Optimisation

Théorème 33. *Si $J : C \to \mathbb{R}$ est différentiable en $u \in C$ et admet un minimum local en u, alors :*

$$\forall\, v \in C,\ \langle \nabla J(u), v - u \rangle \geqslant 0$$

Théorème 34. *On considère $J : C \to \mathbb{R}$.*

(i) Si J est convexe, tout minimum local est global.

(ii) Si J est strictement convexe, J admet au plus un minimum global.

(iii) Si J est α-convexe, J admet un unique minimum global.

(iv) Si J est définie sur un ouvert contenant C et différentiable en $u \in C$, alors le théorème précédent donne en fait une équivalence.

(v) Si C est ouvert, le théorème précédent équivaut à $\nabla J(u) = 0$.

2) Méthodes de gradient

Soit $J : \mathbb{R}^n \to \mathbb{R}$. On suppose J différentiable. On cherche, s'il existe, un élément $u \in \mathbb{R}^n$ tel que :

$$J(u) = \inf_{v \in \mathbb{R}^n} J(v)$$

Pour cela, on utilise les méthodes de gradient. On considère la suite :

$$u_0 \in \mathbb{R}^n \quad \text{et} \quad \forall\, k \in \mathbb{N},\ u^{k+1} = u^k - \rho^k \nabla J(u^k)$$

Il existe plusieurs possibilités pour choisir les ρ^k, par exemple :

(i) Gradient à pas fixe : $\rho^k = \rho$ une constante positive fixée.

(ii) Gradient à pas optimal : ρ^k minimise $\rho \mapsto J\left(u^k - \rho \nabla J(u^k)\right)$.

Théorème 35. *Si J est α-convexe et différentiable, et que ∇J est L-lipschitzienne, alors la méthode de gradient à pas optimal converge vers l'unique minimum de J.*

Application 36. *Soient $A \in \mathcal{S}_n^{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. On considère la fonctionnelle quadratique $J : \mathbb{R}^n \to \mathbb{R}$ définie par :*

$$J(X) = \langle AX, X \rangle - \langle b, X \rangle + c$$

Cette fonctionnelle satisfait les conditions du théorème précédent. De plus, son minimum est atteint en $X_0 \in \mathbb{R}^n$ qui vérifie $\nabla J(X_0) = AX - b = 0$. On a donc une méthode itérative pour approcher la solution de $AX = b$.

3) Méthode de Newton

La méthode de Newton consiste à approcher une solution d'une équation $f(x) = 0$ en partant d'une approximation plus grossière. L'idée est de remplacer la courbe de f par sa tangente.

Théorème 37 (Méthode de Newton)**.** *Soient $a, b \in \mathbb{R}$ tels que $a < b$, et soit $f : [a, b] \to \mathbb{R}$ une fonction de classe $\mathcal{C}^2$ telle que $f(a) < 0 < f(b)$ et $f' > 0$ sur $[a, b]$. On considère la suite $(x_n)_{n \in \mathbb{N}}$ définie par :*

$$x_0 \in [a, b] \quad \text{et} \quad \forall\, n \in \mathbb{N},\ x_{n+1} = \phi(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}$$

La fonction f admet un unique zéro $\alpha \in]a, b[$, et on a :

(i) Il existe $\varepsilon > 0$ tel que, pour $x_0 \in I =]\alpha - \varepsilon, \alpha + \varepsilon[$, la suite $(x_n)_{n \in \mathbb{N}}$ converge quadratiquement vers α, et il existe $C > 0$ tel que :

$$\forall\, n \in \mathbb{N},\ |x_{n+1} - \alpha| \leqslant C |x_n - \alpha|^2$$

(ii) Si de plus $f'' > 0$ sur $[\alpha, b]$, alors, pour $x \in]\alpha, b]$, la suite $(x_n)_{n \in \mathbb{N}}$ est strictement décroissante, et pour tout $n \in \mathbb{N}$ on a :

$$0 \leqslant x_{n+1} - \alpha \leqslant C(x_n - \alpha)^2 \quad \text{et} \quad x_{n+1} - \alpha \sim \frac{f''(\alpha)}{2f'(\alpha)}(x_n - \alpha)^2$$

Développements

— Projection sur un convexe fermé et théorème de Riesz (20,22,23) [Bre87]

— Algorithme de gradient à pas optimal (35) [Cia88]

Références

[RDO91] E. Ramis, C. Deschamps, et J. Odoux. *Cours de Mathématiques, Topologie et éléments d'analyse.* Masson

[Rom19] J.-E. Rombaldi. *Éléments d'analyse réelle.* EDP Sciences

[BP12] M. Briane et G. Pagès. *Théorie de l'intégration.* Vuilbert

[Cia88] P. Ciarlet. *Introduction à l'analyse numérique et à l'optimisation.* Masson

[Bre87] H. Brezis. *Analyse fonctionelle.* Masson

Annexes

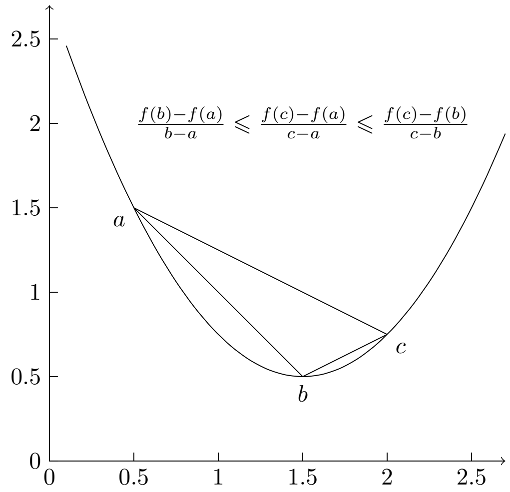

FIGURE 1 – Inégalité des trois pentes

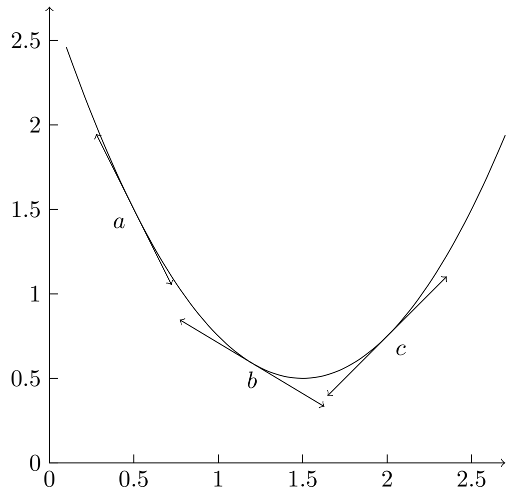

FIGURE 2 – Tangentes d'une fonction convexe

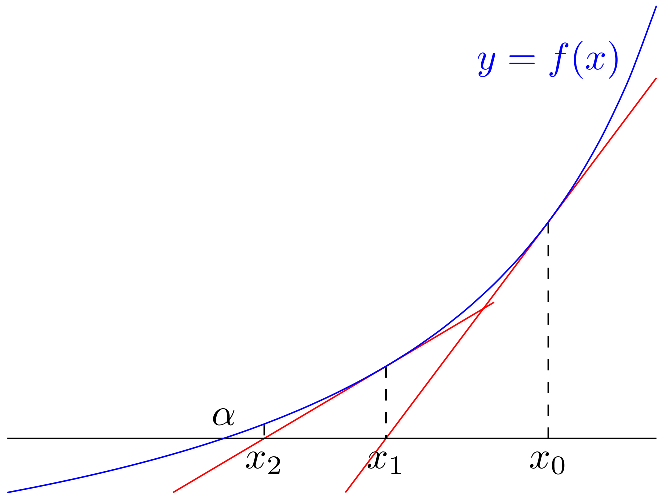

FIGURE 3 – Méthode de Newton