Chapitre 1

Convexité

1.1 Parties convexes d'un espace vectoriel réel

1.1.1 Barycentre

Définition 1.1 *Soit n un entier non nul, soit $(x_1, \ldots, x_n)$ des éléments d'un espace vectoriel réel, soit $(\alpha_1, \ldots, \alpha_n)$ des réels tels que $\sum_{k=1}^n \alpha_k \neq 0$. Il existe un unique élément g de E tel que*

$$(\sum_{k=1}^n \alpha_k).g = \sum_{k=1}^n \alpha_k.x_k$$

c'est l'élément

$$g = \frac{1}{\sum_{k=1}^n \alpha_k}.\sum_{k=1}^n \alpha_k.x_k.$$

On l'appelle le barycentre des $(x_1, \ldots, x_n)$ affectés des poids $(\alpha_1, \ldots, \alpha_n)$.

Remarquons que les x_i ne sont pas supposés distincts.
D'autre part, la notion de barycentre étant une notion affine, on parlera plutôt de points que de vecteurs pour désigner les éléments de l'espace vectoriel.

Exemple 1.1 (L'isobarycentre) *Si $(x_1, \ldots, x_n)$ sont des éléments d'un espace vectoriel $(n \geq 1)$ leur isobarycentre est le barycentre des $(x_1, \ldots, x_n)$ affectés de poids égaux (et non nuls) c'est-à-dire l'élément*

$$\frac{1}{n}\sum_{k=1}^n x_k.$$

Remarque 1.1 *Si α est un réel non nul , le barycentre des $(x_1, \ldots, x_n)$ affectés des poids $(\lambda_1, \ldots, \lambda_n)$ est le même que celui des $(x_1, \ldots, x_n)$ affectés des poids $(\alpha\lambda_1, \ldots, \alpha\lambda_n)$. En prenant $\alpha = \dfrac{1}{\sum_{k=1}^n \lambda_k}$ on peut supposer que la somme des poids est égal à 1, sans perdre en généralité.*

Remarque 1.2 *Le programme ne propose pas de notation pour le barycentre.*

Exemple 1.2 *En Physique ou en Sciences Industrielles les éléments x_i seront interprétés comme des points matériels, les poids α_i de ces points étant leurs masses, plutôt notées m_i.*
Le barycentre du système de points pondérés $\big((x_1, m_1), \ldots, (x_n, m_n)\big)$ s'appelle de centre de masse ou centre de gravité de ce système de points. La notion de série ou d'intégrale permet d'étendre cette notion à une famille infinie de points.

Exemple 1.3 *On a vu en première année la notion d'espérance d'une variable aléatoire réelle X définie sur un espace probabilisé fini. Elle est donnée par*

$$E(X) = \sum_{\omega \in \Omega} P(\{\omega\})X(\omega).$$

Puisque $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$, on reconnaît dans l'espérance le barycentre des $(X(\omega))_{\omega \in \Omega}$ affectés des poids $(P(\{\omega\}))_{\omega \in \Omega}$.

1.1.2 Partie convexe

Définition 1.2 *Une partie A d'un espace vectoriel réel est convexe si et seulement si :*

$$\forall (x,y) \in A^2 \ \forall \lambda \in [0,1] \quad (1-\lambda)x + \lambda y \in A.$$

Dorénavant lorsqu'on parlera de la convexité/non-convexité d'une partie A il sera sous-entendu qu'il s'agit d'une partie d'un espace vectoriel réel.

Exemple 1.4 *Tout sous-espace vectoriel d'un espace vectoriel réel est convexe.*

Exercice 1.1 *Si ϕ est une forme linéaire sur l'espace vectoriel réel E et si a est un réel, les ensembles $\{x \in E; \phi(x) \geq a\}$ et $\{x \in E; \phi(x) > a\}$ sont des parties convexes. Si ϕ est non nulle, on parle de demi-espace fermé et ouvert.*

Exemple 1.5 *Les parties convexes de $\mathbb{R}$ sont les intervalles.*
En effet, si est un intervalle, si x et y sont dans I, avec $x \leq y$ sans perte de généralité, si λ est dans $[0,1]$ et si $z = (1-\lambda)x+\lambda y$, alors $x \leq z \leq y$ donc z est dans I par définition d'un intervalle.
Réciproquement soit I une parte convexe (non vide) de $\mathbb{R}$. On rappelle que I est un intervalle si et seulement si pour tout (x,y) de I^2 avec $x < y$ tout z de $\mathbb{R}$ tel que $x < z < y$ appartient I.
Or un tel z peut s'écrire

$$z = \frac{y-z}{y-x}x + \frac{z-x}{y-x}y = (1-\lambda)x + \lambda y$$

avec $\lambda = \frac{z-x}{y-x} \in [0,1]$. z appartient bien à I par convexité de I.
Cette écriture barycentrique d'un élément de I nous servira ultérieurement.

Proposition 1.1 *Une partie A est convexe si et seulement si tout barycentre poids positifs d'un système d'éléments de A est dans A.*

La condition est suffisante car si λ appartient à $[0,1]$ et x et y sont dans A alors $(1-\lambda)x+\lambda y$ est un barycentre de x et y à coefficients positifs.
Pour établir la réciproque on montre par récurrence sur n que si A est convexe, si $(x_1,\ldots,x_n)$ sont dans A, si $(\lambda_1,\ldots,\lambda_n)$ sont des réels dont la somme est non nulle alors le barycentre des $(x_1,\ldots,x_n)$ affectés des poids $(\lambda_1,\ldots,\lambda_n)$ est dans A.
Le résultat est vrai pour $n=1$ car le barycentre de (x) affecté du poids (λ) avec $\lambda \neq 0$ est x.
On le suppose vrai à l'ordre $n-1$, avec $n \geq 2$. Soit $(x_1,\ldots,x_n)$ dans A^n et $(\lambda_1,\ldots,\lambda_n)$ dans ${\mathbb{R}^+}^n$ avec $S = \sum_{k=1}^n \lambda_k \neq 0$.
Deux cas ont possibles, si $g = \frac{1}{\sum_{k=1}^n \lambda_k}.\sum_{k=1}^n \lambda_k.x_k$ est le barycentre des $(x_1,\ldots,x_n)$ affectés des poids $(\lambda_1,\ldots,\lambda_n)$:

- $\lambda_n = S$ Dans ce cas $\forall k \in [1,n-1]$ on aura $\lambda_k = 0$ (car tous les λ_i, $1 \leq i \leq n$ sont positifs). Par conséquent $g = x_n$ est bien dans A.
- $\lambda_n \neq S$ Dans ce cas on peut écrire, car $\sum_{k=1}^{n-1} \lambda_k = S - \lambda_n \neq 0$:

$$g = \frac{1}{S}\left((S-\lambda_n)\left(\frac{1}{\sum_{k=1}^{n-1}\lambda_k}\sum_{k=1}^{n-1}\lambda_k x_k\right) + \lambda_n x_n\right).$$

Or

$$g' = \frac{1}{\sum_{k=1}^{n-1}\lambda_k}\sum_{k=1}^{n-1}\lambda_k x_k$$

est dans A par hypothèse de récurrence, et ensuite

$$g = \frac{1}{S}\big((S-\lambda_n)g' + \lambda_n x_n\big)$$

est dans A car A est convexe.

1.2 Fonctions convexes d'une variable réelle

1.2.1 Définition

Définition 1.3 *Une fonction $f : I \to \mathbb{R}$ où I est un intervalle de $\mathbb{R}$ est convexe si et seulement si :*

$$\forall (x,y) \in I^2 \ \forall \lambda \in [0,1] \quad f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x) + \lambda f(y).$$

Proposition 1.2 (Inégalité de Jensen) *Si $f : I \to \mathbb{R}$ est une fonction convexe, si n est un entiers non nul, $(x_1, \ldots, x_n)$ des points de I, $(\lambda_1, \ldots, \lambda_n)$ des réels positifs de somme égale à 1 alors :*

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

La condition est nécessaire car, pour $n = 2$, elle exprime la définition de la convexité de f.
Réciproquement, si f convexe alors pour $n = 1$ le résultat est clair mais sans intérêt ($\lambda_1 = 1!$), pour $n = 2$ on retrouve la définition de la convexité.
On démontre alors le résultat par récurrence. On sait qu'il est vrai pour $n = 1$ (et $n = 2$). On le suppose vrai à l'ordre $n-1$, $n \geq 2$.
Soit $(x_1, \ldots, x_n)$ dans A^n et $(\lambda_1, \ldots, \lambda_n)$ dans $\mathbb{R}^{+n}$ avec $\sum_{k=1}^{n} \lambda_k \neq 1$.
La démonstration calque ensuite la démonstration de la sous-section précédente.
Deux cas ont possibles

- $\lambda_n = 1$ Dans ce cas $\forall k \in [1, n-1]$ on aura $\lambda_k = 0$ (car tous les λ_i, $1 \leq i \leq n$ sont positifs). Par conséquent $f(\sum_{k=1}^{n} \lambda_k . x_k) = f(x_n) = \sum_{k=1}^{n} \lambda_k f(x_k)$ et le résultat est vrai.
- $\lambda_n \neq 1$ Dans ce cas on peut écrire, car $\sum_{k=1}^{n-1} \lambda_k = 1 - \lambda_n \neq 0$:

$$\begin{aligned}
f(\sum_{k=1}^{n} \lambda_k . x_k) &= f\left((1-\lambda_n)\left(\frac{1}{\sum_{k=1}^{n-1} \lambda_k} \sum_{k=1}^{n-1} \lambda_k x_k\right) + \lambda_n x_n\right) \\
&\leq (1-\lambda_n) f\left(\frac{1}{\sum_{k=1}^{n-1} \lambda_k} \sum_{k=1}^{n-1} \lambda_k x_k\right) + \lambda_n f(x_n) \\
&\leq (1-\lambda_n)\left(\frac{1}{\sum_{k=1}^{n-1} \lambda_k} \sum_{k=1}^{n-1} \lambda_k f(x_k)\right) + \lambda_n f(x_n) \\
&\leq \sum_{k=1}^{n} \lambda_k f(x_k)
\end{aligned}$$

La deuxième ligne résulte de la convexité de f qui s'applique ici car un intervalle étant convexe $\frac{1}{\sum_{k=1}^{n-1} \lambda_k} \sum_{k=1}^{n-1} \lambda_k x_k$ est dans I d'après la proposition 1.1. La troisième ligne découle de l'hypothèse de récurrence.

Remarque 1.3 *La définition d'une fonction convexe et la démonstration de l'inégalité de Jensen resteraient valable pour une fonction définie sur une partie convexe d'un espace vectoriel réel, mais le programme a choisi de se limiter aux fonctions définies sur un intervalle réel.*

1.2.2 Caractérisation

Proposition 1.3 (Convexité de l'épigraphe) *Une fonction $f : I \to \mathbb{R}$ où I est un intervalle de $\mathbb{R}$ est convexe si et seulement si son épigraphe est convexe.*

On rappelle que l'épigraphe de f est l'ensemble

$$A = \{(x, y) \in I \times \mathbb{R}, y \geq f(x)\}.$$

Proposition 1.4 (Inégalité des pentes) *La fonction $f : I \to \mathbb{R}$ est convexe si et seulement si :*

$$(R_1) \qquad \forall (x, y, z) \in I^3 \quad x < y < z \ \Rightarrow \ \frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y}.$$

— On suppose f convexe.
Soit x, y, z tels que $x < y < z$. On a vu dans l'exemple 1.5 qu'on peut écrire $y = (1-\lambda)x + \lambda z$ avec $\lambda = \frac{y-x}{z-x} \in]0, 1[$.
Puisque f est convexe on en déduit :

$$\begin{aligned}
f(y) &\leq (1-\lambda) f(x) + \lambda f(z) \\
(1-\lambda)(f(y) - f(x)) &\leq \lambda(f(z) - f(y)) \\
\frac{z-y}{z-x}(f(y) - f(x)) &\leq \frac{y-x}{z-x}(f(z) - f(y)) \\
\frac{f(y) - f(x)}{y - x} &\leq \frac{f(z) - f(y)}{z - y}
\end{aligned}$$

Les simplifications sont compatibles avec les inégalités car $y-x$, $z-x$ et $z-y$ sont strictement positifs. Remarquons que la deuxième ligne aurait pu être

$$f(y)-f(x) \leq \lambda(f(z)-f(x))$$

dont on aurait tiré :

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

On aurait obtenu de même :

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

Tout ceci se résume en

$$(R_2) \qquad \frac{f(y)-f(x)}{y-x} \leq \frac{f(z)-f(x)}{z-x} \leq \frac{f(z)-f(y)}{z-y}.$$

— Réciproquement.
Supposons (R_1) vérifiée. Soit (x,y) dans I^2 et λ dans $[0,1]$. Si $x=y$ ou $\lambda=0$ ou $\lambda=1$, l'inégalité de convexité est bien vérifiée. Dans l'autre cas on peut supposer $x<y$ sans perdre en généralité. On prend $z=(1-\lambda)x+\lambda y$, on a $x<z<y$; la relation R_1 donne

$$\begin{array}{rcl}
\dfrac{f(z)-f(x)}{z-x} & \leq & \dfrac{f(y)-f(z)}{y-z} \\
\dfrac{f(z)-f(x)}{\lambda(y-x)} & \leq & \dfrac{f(y)-f(z)}{(1-\lambda)(y-x)} \\
(1-\lambda)(f(z)-f(x)) & \leq & \lambda(f(y)-f(z)) \\
f(z) & \leq & (1-\lambda)f(x)+\lambda f(y)
\end{array}$$

ce qui est bien la relation désirée.

Proposition 1.5 *Si $f : I \to \mathbb{R}$ est convexe et si $[x,y]$ est un segment contenu dans I alors le graphe de la restriction de f à $[x,y]$ est en dessous de la corde joignant $(x,f(x))$ à $(y,f(y))$*

On peut supposer $x<y$ car si $x=y$ le résultat est trivial. Si $x \leq z \leq y$ on peut écrire $z=(1-\lambda)x+\lambda y$ avec $\lambda = \frac{z-x}{y-x} \in [0,1]$. Le point du graphe d'abscisse z est $M=(z,f(z))=(z,f((1-\lambda)x+\lambda y))$, le point de la corde de même abscisse est $M'=(z,(1-\lambda)f(x)+\lambda f(y))$. Son ordonnée est plus grande que celle de M car f est convexe. Il est bien au dessus.

Remarque 1.4 *Cet énoncé étant essentiellement équivalent à l'inégalité des pentes sa réciproque est vraie.*

1.2.3 Fonctions concaves

Définition 1.4 *Une fonction $f : I \to \mathbb{R}$ est concave si et seulement si $-f$ est convexe, ce qui équivaut à :*

$$\forall (x,y) \in I^2 \ \forall \lambda \in [0,1] \quad f((1-\lambda)x+\lambda y) \geq (1-\lambda)f(x)+\lambda f(y).$$

La fonction $x \mapsto -x$ étant une fonction décroissante de $\mathbb{R}$ vers $\mathbb{R}$ on peut transporter tous les résultats obtenus pour les fonctions convexes aux fonctions concaves en inversant le sens des inégalités.

1.3 Fonctions convexes et dérivabilité

1.3.1 Caractérisation

Proposition 1.6 *Une fonction $f : I \to \mathbb{R}$, dérivable, est convexe si et seulement si sa dérivée est croissante.*

— Supposons f convexe et dérivable.
Soit (x,z) dans I^2 avec par exemple $x<z$. Dans la relation (R_2) faisons tendre y vers $x$dans l'inégalité de gauche et y vers z dans l'inégalité de droite , nous obtenons :

$$(R_3) \qquad f'(x) \leq \frac{f(z)-f(x)}{z-x} \leq f'(z),$$

et en particulier $f'x) \leq f'(z)$.
Nous avons bien établi la croissance de f'.

— Réciproquement supposons f dérivable et de dérivée croissante. Soit (x, y, z) dans I^3 avec $x < y < z$. D'après l'égalité des accroissements finis il existe c dans $]x, y[$ et d dans $]y, z[$ tels que

$$\frac{f(y) - f(x)}{y - x} = f'(c) \text{ et } \frac{f(z) - f(y)}{z - y} = f'(d).$$

Puisque $c < d$ et f' est croissante la relation (R_1) est vérifiée et f est convexe.

Exercice 1.2 *Réciproquement si f est une fonction de $]a, b[$ vers $\mathbb{R}$ convexe alors :*
1) f est continue sur $]a, b[$,
2) f admet en tout point une dérivée à droite et une dérivée à gauche.
3) les fonction f'_g et f'_d sont croissantes.

Proposition 1.7 *Une fonction $f : I \to \mathbb{R}$, deux fois dérivable, est convexe si sa dérivée seconde est positive.*

En effet, on sait , d'après la proposition 1.8 que f est convexe si seulement si f' est croissante. Or il a été vu en première année que f' est croissante si et seulement si $f'' = (f')'$ est positive.
Elle sera concave si et seulement si sa dérivée seconde est négative.

1.3.2 Position du graphe par rapport aux tangentes

Proposition 1.8 *Si $f : I \to \mathbb{R}$ est convexe et dérivable alors le graphe de f est au dessus de toute tangente à ce graphe.*

On part de la relation (R_3). Elle donne, i $x < z$: $f(z) \geq f(x) + f'(x)(z - x)$ et $f(x) \geq f(z) + f'(z)(x - z)$.
En étudiant les cas $t > y$, $t < y$ et $t = y$, on en déduit :

$$(R_4) \qquad \forall (y, t) \in I^2 \quad f(t) \geq (t - y) f'(y) + f(y).$$

C'est le résultat voulu.

Exercice 1.3 *Etudier la réciproque.*

1.3.3 Exemple d'inégalité de convexité

Proposition 1.9 (Inégalité de la moyenne géométrique) *La fonction* $\ln$ *est concave (et croissante). On en déduit pour toute famille $(x_1, \ldots, x_n)$ de réels strictement positifs*

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Cette inégalité s'étend trivialement au cas où l'un des x_i est nul.

Exemple 1.6 *Soit X une variable aléatoire réelle centrée (c'est-à-dire $E(X) = 0$, telle que $|X| \leq 1$ alors :*

$$\forall t \in \mathbb{R} \quad E(e^{tX}) \leq \operatorname{ch} t.$$

En effet pour tout t on peut écrire :

$$tX = \frac{1 + X}{2} t + \frac{1 - X}{2} (-t).$$

Or, $\frac{1+X}{2}$ est à valeurs dans $[0, 1]$, $\frac{1+X}{2} + \frac{1-X}{2} = 1$ et la fonction exp est convexe car sa dérivée seconde est positive. Il en résulte donc :

$$e^{tX} \leq \frac{1 + X}{2} e^t + \frac{1 - X}{2} e^{-t}.$$

On obtient le résultat demandé en passant à l'espérance qui est linéaire et conserve les inégalités (et vérifie $E(1) = 1$) :

$$E(e^{tX}) \leq \frac{1 + E(X)}{2} e^t + \frac{1 - E(X)}{2} e^{-t} = \frac{e^t + e^{-t}}{2} = \operatorname{ch} t.$$

Exemple 1.7 *Soit n un entier non nul, $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$ des réels strictement positifs. Alors*

$$\left(\sum_{i=1}^{n} x_i y_i \right)^2 \leq \left(\sum_{i=1}^{n} x_i^2 \right) \left(\sum_{i=1}^{n} y_i^2 \right).$$

On considère la fonction $f : x \mapsto x^2$. Sa dérivée seconde est positive sur $\mathbb{R}^{*+}$. Elle est donc convexe.
Posons $\lambda_i = \frac{x_i^2}{\sum_{k=1}^n x_k^2}$ et $z_i = \frac{y_i}{x_i}$, pour $1 \leq i \leq n$. Les λ_i sont positifs, de somme égale à 1. On peut donc écrire :

$$f\left(\sum_{i=1}^n \lambda_i z_i\right) \leq \sum_{i=1}^n \lambda_i f(z_i),$$

soit

$$\left(\sum_{i=1}^n \frac{x_i y_i}{\sum_{k=1}^n x_k^2}\right)^2 \leq \sum_{i=1}^n \frac{y_i^2}{\sum_{k=1}^n x_k^2}.$$

Le résultat annoncé s'obtient en multipliant par $\left(\sum_{k=1}^n x_k^2\right)^2$.

Ce résultat s'étend sans difficulté aux cas où certains des x_i ou y_i sont nuls, en l'utilisant pour un n inférieur. Comme d'autre part, si $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$ sont des nombres complexes on a

$$\left|\sum_{i=1}^n x_i y_i\right| \leq \sum_{i=1}^n |x_i||y_i|,$$

on peut affirmer :

Proposition 1.10 (Inégalité de Cauchy-Schwarz-Bounyakovskii) *Soit $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$ des nombres complexes. Alors*

$$\left|\sum_{i=1}^n x_i y_i\right|^2 \leq \left(\sum_{i=1}^n |x_i|^2\right)\left(\sum_{i=1}^n |y_i|^2\right).$$

Exercice 1.4 *En vous inspirant de la démonstration de l'exemple précédent, montrer que si p et q sont deux réels tels que $\frac{1}{p} + \frac{1}{q} = 1$, alors pour toutes familles $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$ de réels strictement positifs :*

$$\sum_{i=1}^n x_i y_i \leq \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n y_i^q\right)^{\frac{1}{q}},$$

puis écrire l'extension de l'inégalité de Cauchy-Schwarz-Bounyakovskii, qui s'appelle l'inégalité de Hölder.