Matrices bistochastiques

Jean Etienne ROMBALDI

4 décembre 2012

Table des matières

1

Matrices bistochastiques

1.1 Ensembles convexes, polyèdres convexes

On désigne, pour ce paragraphe, par E un espace vectoriel réel et on désigne par E^* le dual algébrique de E, à savoir l'espace vectoriel des formes linéaires sur E.

Si a, b sont deux éléments de E, on note $[a, b]$ le segment d'extrémités a, b, à savoir la partie de E définie par :

$$[a, b] = \{(1 - \lambda) a + \lambda b \mid 0 \leq \lambda \leq 1\}.$$

Si E est un espace vectoriel normé, alors un segment dans E est fermé.

Définition 1.1 *On dit qu'une partie C de E est convexe, si pour tout couple (a, b) d'éléments de E, le segment $[a, b]$ est contenu dans E.*

Remarque 1.1 *On vérifie aisement qu'une intersection d'ensembles convexes dans E est convexe.*

Remarque 1.2 *Dans un espace vectoriel normé, l'adhérence d'un convexe est convexe. En effet si C est convexe et $a = \lim\limits_{k \to +\infty} a_k$, $b = \lim\limits_{k \to +\infty} b_k$ sont dans l'adhérence $\overline{C}$ de C, les a_k et b_k étant dans C, on a pour tout réel λ compris entre 0 et 1 :*

$$(1 - \lambda) a + \lambda b = \lim_{k \to +\infty} ((1 - \lambda) a_k + \lambda b_k) \in \overline{C}$$

Remarque 1.3 *Si F est un autre espace vectoriel et u une application linéaire de E dans F, alors pour tout convexe C dans E [resp. dans F] l'image directe [resp. l'image réciproque] de C par u est un convexe de F [resp. E]. Ces résultats se déduisent immédiatement de la linéarité de u.*

Cette remarque nous permet de donner les exemples suivants de parties convexes.

Exemple 1.1 *Si φ est une forme linéaire non nulle sur E, alors pour tout réel α l'ensemble :*

$$H = \varphi^{-1}(\alpha) = \{x \in E \mid \varphi(x) = \alpha\}$$

est convexe dans E comme image réciproque du convexe $\{\alpha\}$ de $\mathbb{R}$ par l'application linéaire φ.

On dit que H est l'hyperplan affine d'équation $\varphi(x) = \alpha$.

Exemple 1.2 *De manière analogue, pour tout $\varphi \in E^* \setminus \{0\}$, les ensembles $\varphi^{-1}([\alpha, +\infty[)$ et $\varphi^{-1}(]-\infty, \alpha])$ [resp. $\varphi^{-1}(]\alpha, +\infty[)$ et $\varphi^{-1}(]-\infty, \alpha[)$] sont convexes dans H.*

On dit que ces ensembles sont les demi espaces fermés [resp. ouverts] limités par H.

Dans ce qui suit, si H est un hyperplan affine d'équation $\varphi(x) = \alpha$, on note :

$$\left\{ \begin{array}{lllll} H^{+} & = & \varphi^{-1}([\alpha, +\infty[) & = & \{x \in E \mid \varphi(x) \geq \alpha\}, \\ H^{+,*} & = & \varphi^{-1}(]\alpha, +\infty[) & = & \{x \in E \mid \varphi(x) > \alpha\}, \\ H^{-} & = & \varphi^{-1}(]-\infty, \alpha]) & = & \{x \in E \mid \varphi(x) \leq \alpha\}, \\ H^{-,*} & = & \varphi^{-1}(]-\infty, \alpha[) & = & \{x \in E \mid \varphi(x) < \alpha\}, \end{array} \right.$$

les demi espaces fermés et ouverts limités par H.

Remarque 1.4 *En dimension finie, l'application $x \mapsto \varphi(x) - \alpha$ est continue et en conséquence les demi-espaces H^+ et H^- [resp. $H^{+,*}$ et $H^{-,*}$] sont biens bien des fermés [resp. ouverts] de E.*

Définition 1.2 *On appelle polyèdre dans un espace vectoriel réel de dimension finie E, une partie bornée de E qui peut s'écrire comme intersection d'un nombre fini de demi-espaces fermés de E.*

Dans le cas où E est un plan vectoriel, on retrouve la notion de polygone.

Un polyèdre est fermé et convexe comme intersection d'ensembles fermés et convexes. Étant fermé et borné, il est compact dans E qui est de dimension finie.

Exemple 1.3 *Dans $\mathbb{R}^n$ l'ensemble :*

$$P = \left\{ x \in \left(\mathbb{R}^+\right)^n \mid \|x\|_1 = \sum_{i=1}^{n} x_i = 1 \right\}$$

est un polyèdre convexe.

Cet ensemble est borné puisque contenu dans la boule unité fermée de $(\mathbb{R}^n, \|\cdot\|_1)$.

En désignant par $\{e_i^* \mid 1 \leq i \leq n\}$ la base duale de la base canonique de $\mathbb{R}^n$ ($e_i^*(x) = x_i$, pour tout $x \in \mathbb{R}^n$) et par φ la forme linéaire définie par $\varphi = \sum_{i=1}^{n} e_i^*$, on a :

$$x \in P \Leftrightarrow \left\{ \begin{array}{l} e_i^*(x) \geq 0 \quad (1 \leq i \leq n) \\ \varphi(x) \geq 1 \\ -\varphi(x) \geq -1 \end{array} \right.$$

ce qui prouve que P est un polyèdre convexe de $\mathbb{R}^n$.

1.2 Enveloppe convexe, théorème de Carathéodory

On désigne toujours par E un espace vectoriel réel.

L'intersection d'une famille de convexes dans E étant un convexe et, pour toute partie X de E, l'espace vectoriel E est un convexe qui contient X. Ces remarques nous permettent de donner la définition suivante.

Définition 1.3 *Si X est une partie de E, on appelle enveloppe convexe de X, l'intersection de tous les convexes de E qui contiennent X.*

On note $Cv(X)$ l'enveloppe convexe de X. C'est un convexe de E.

Cette enveloppe convexe est aussi le plus petit convexe de E contenant X.

Une définition équivalente de la notion d'enveloppe convexe est donnée par le résultat suivant.

Théorème 1.1 *Si X est une partie non vide de E, alors l'enveloppe convexe de X est l'ensemble des combinaisons linéaires convexes d'éléments de X, c'est-à-dire qu'un vecteur x est dans $Cv(X)$ si et seulement si il existe des vecteurs $x_1, \cdots, x_p$ dans X et des réels positifs ou nuls $\lambda_1, \cdots, \lambda_p$ tels que $\sum_{i=1}^{p} \lambda_i = 1$, $x = \sum_{i=1}^{p} \lambda_i x_i$, ce qui peut aussi s'exprimer en disant que x est barycentre de points de X affectés de coefficients positifs.*

Démonstration. Notons $\mathcal{B}(X)$ l'ensemble des combinaisons linéaires convexes d'éléments de X.

On montre tout d'abord que $\mathcal{B}(X)$ est convexe. C'est donc un convexe de E contenant X et en conséquence il contient $Cv(X)$.

Si $x = \sum_{i=1}^{p} \lambda_i x_i$ et $y = \sum_{i=1}^{q} \mu_i y_i$ sont dans $\mathcal{B}(X)$, avec $\lambda_i \geq 0$, $x_i \in X$ pour $1 \leq i \leq p$, $\mu_i \geq 0$, $y_i \in X$ pour $1 \leq i \leq q$ et $\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{q} \mu_i = 1$, alors pour tout réel λ compris entre 0 et 1, le vecteur $z = (1-\lambda)x + \lambda y$ s'écrit :

$$z = \sum_{i=1}^{p} (1-\lambda)\lambda_i x_i + \sum_{i=1}^{q} \lambda \mu_i y_i,$$

avec $(1-\lambda)\lambda_i > 0$, $\lambda \mu_j \geq 0$, pour $1 \leq i \leq p$, $1 \leq j \leq q$ et :

$$\sum_{i=1}^{p} (1-\lambda)\lambda_i + \sum_{i=1}^{q} \lambda \mu_i = (1-\lambda) + \lambda = 1,$$

ce qui signifie que $z \in \mathcal{B}(X)$.

On montre ensuite par récurrence sur $p \geq 1$ que toute combinaison linéaire convexe de p éléments de X est dans $Cv(X)$.

Pour $p = 1$, le résultat découle de $X \subset Cv(X)$.

Pour $p = 2$, si x_1, x_2 sont dans X et λ_1, λ_2 sont des réels positifs tels que $\lambda_1 + \lambda_2 = 1$, alors $\lambda_1 x_1 + \lambda_2 x_2$ est dans $Cv(X)$ puisque cet ensemble est convexe.

Supposons le résultat acquis pour $p \geq 2$ et soient $x_1, \cdots, x_{p+1}$ dans l'ensemble X, $\lambda_1, \cdots, \lambda_{p+1}$ dans $\mathbb{R}^+$ tels que $\sum_{i=1}^{p+1} \lambda_i = 1$. Notons $\lambda = \sum_{i=1}^{p} \lambda_i$.

Si $\lambda = 0$, alors tous les λ_i pour $1 \leq i \leq p$, sont nuls et $\lambda_{p+1} = 1$, de sorte que $\sum_{i=1}^{p+1} \lambda_i x_i = x_{p+1}$ est dans $Cv(X)$.

Si $\lambda \neq 0$, en utilisant l'hypothèse de récurrence, on a :

$$x' = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda} x_i \in Cv(X)$$

et :

$$\sum_{i=1}^{p+1} \lambda_i x_i = \lambda x' + \lambda_{p+1} x_{p+1} \in Cv(X)$$

puisque $\lambda \geq 0$, $\lambda_{p+1} \geq 0$ et $\lambda + \lambda_{p+1} = 1$.

On a donc ainsi montré que $Cv(X) = \mathcal{B}(X)$. ■

Dans le cas particulier où l'ensemble X est contenu dans un espace vectoriel E de dimension finie, le théorème de Carathéodory nous permet de préciser que dans le résultat qui précède on peut toujours avoir $p \leq n+1$, où n est la dimension de E.

Théorème 1.2 (Carathéodory) *Si X est une partie non vide dans un espace vectoriel réel E de dimension $n \geq 1$, alors tout élément de l'enveloppe convexe de X est combinaison linéaire convexe de p éléments de X avec $p \leq n+1$.*

Démonstration. On sait déjà que tout élément de $Cv\,(X)$ s'écrit $x = \sum_{i=1}^{p} \lambda_i x_i$ avec $x_i \in X$, $\lambda_i \in \mathbb{R}^+$ et $\sum_{i=1}^{p} \lambda_i = 1$.

Si $p \leq n+1$, il n'y a rien à montrer. Supposons donc que $p > n+1$. Le système $\{x_i - x_1 \mid 2 \leq i \leq p\}$ formé de $p-1 > n$ vecteurs est alors lié et il existe des réels $\mu_2, \cdots, \mu_p$ non tous nuls tels que :

$$\sum_{i=2}^{p} \mu_i (x_i - x_1) = 0,$$

ce qui peut aussi s'écrire en posant $\mu_1 = -\sum_{i=2}^{p} \mu_i$:

$$\sum_{i=1}^{p} \mu_i x_i = 0,$$

avec $\sum_{i=1}^{p} \mu_i = 0$.

On peut alors écrire pour tout réel positif t :

$$x = \sum_{i=1}^{p} (\lambda_i - t\mu_i)\, x_i,$$

avec $\sum_{i=1}^{p} (\lambda_i - t\mu_i) = 1$.

Comme les coefficients μ_i sont tous non nuls de somme nulle, il en existe au moins un qui est strictement positif et on peut poser :

$$t_0 = \inf \left\{ \frac{\lambda_i}{\mu_i} \mid 1 \leq i \leq p,\ \mu_i > 0 \right\} = \frac{\lambda_k}{\mu_k}$$

pour un indice k compris entre 1 et p.

En posant $\delta_i = \lambda_i - t_0\mu_i$, pour $1 \leq i \leq p$, on a $\delta_k = 0$, $\delta_i \geq 0$ pour $1 \leq i \leq p$ (pour $\mu_i > 0$, on a $\dfrac{\lambda_i}{\mu_i} \geq t_0$, soit $\delta_i \geq 0$ et pour $\mu_i \leq 0$, $\delta_i \geq \lambda_i \geq 0$) et $\sum_{i=1}^{p} \delta_i = 1$. On a donc $x = \sum_{i=1,\, i\neq k}^{p} \delta_i x_i$, avec $\delta_i \geq 0$, $\sum_{i=1,\, i\neq k}^{p} \delta_i = 1$, c'est-à-dire que x est combinaison linéaire convexe de $p-1$ éléments de X.

Une récurrence descendante nous permet alors de conclure. ∎

Une première application de ce théorème, dans le cas où E est un espace vectoriel normé de dimension finie, est la suivante.

Corollaire 1.1 *Dans un espace vectoriel normé de dimension finie l'enveloppe convexe d'un compact est compacte.*

Démonstration. On désigne par Δ le compact de $\mathbb{R}^{n+1}$ défini par :

$$\Delta = \left\{ \lambda \in \left(\mathbb{R}^{+}\right)^{n+1} \mid \|\lambda\|_1 = \sum_{i=1}^{n+1} \lambda_i = 1 \right\}$$

et si X est un compact non vide de l'espace vectoriel normé E de dimension n, par φ l'application définie par :

$$\forall (\lambda, x) \in \Delta \times X^{n+1}, \quad \varphi(\lambda, x) = \sum_{i=1}^{n+1} \lambda_i x_i.$$

Le théorème de Carathéodory nous dit que l'image de φ est exactement l'enveloppe convexe de X dans E. On en déduit alors que $Cv(X) = \varphi(\Delta \times X^{n+1})$ est compacte dans E comme image du compact $\Delta \times X^{n+1}$ (produit de compacts) par l'application continue φ. ■

Corollaire 1.2 *Dans un espace vectoriel normé de dimension finie l'enveloppe convexe d'une partie bornée est bornée.*

Démonstration. Si X est une partie bornée dans l'espace vectoriel normé E de dimension n, elle est alors contenue dans une partie compacte Y et $Cv(X)$ est bornée car contenue dans le compact $Cv(Y)$. ■

1.3 Le théorème de projection sur un convexe fermé dans un espace de Hilbert

Pour ce paragraphe, E désigne un espace de Hilbert. On note respectivement $\langle \cdot \mid \cdot \rangle$ et $\|\cdot\|$ le produit scalaire et la norme associée sur E.

Théorème 1.3 *Soit C une partie non vide de E convexe et fermée. Pour tout x dans E, il existe un unique y dans C tel que :*

$$\|x - y\| = \inf_{z \in C} \|x - z\|. \tag{1.1}$$

Ce vecteur $y \in C$ est également caractérisé par :

$$\forall z \in C, \quad \langle x - y \mid z - y \rangle \leq 0. \tag{1.2}$$

Démonstration. L'ensemble $\{\|x - z\| \mid z \in C\}$ est une partie non vide minorée de $\mathbb{R}$, elle admet donc une borne inférieure :

$$\delta = \inf_{z \in C} \|x - z\|.$$

Par définition de cette borne inférieure, on peut construire une suite $(y_k)_{k \geq 1}$ d'éléments de C telle que :

$$\forall k \geq 1, \quad \delta^2 \leq \|x - y_k\|^2 \leq \delta^2 + \frac{1}{k} \tag{1.3}$$

En utilisant l'identité de la médiane, on peut écrire, pour $q > p \geq 1$:

$$\begin{aligned} \|y_q - y_p\|^2 &= \|(y_q - x) + (x - y_p)\|^2 \\ &= 2\left(\|y_q - x\|^2 + \|x - y_p\|^2\right) - \|(y_q - x) - (x - y_p)\|^2, \end{aligned}$$

avec :

$$\|(y_q - x) - (x - y_p)\|^2 = 4\left\|x - \frac{1}{2}(y_p + y_q)\right\|^2 \geq 4\delta^2$$

puisque $\frac{1}{2}(y_p + y_q) \in C$ (qui est convexe).

On a donc, pour $q > p \geq 1$:

$$\begin{aligned}\|y_q - y_p\|^2 &\leq 2\left(\|y_q - x\|^2 + \|x - y_p\|^2\right) - 4\delta^2 \\ &\leq 2\left(\frac{1}{q} + \frac{1}{p}\right) - 2\delta^2 \leq \frac{4}{p}\end{aligned}$$

et il en résulte que la suite $(y_k)_{k\geq 1}$ est de Cauchy dans l'espace complet E, elle est donc convergente et sa limite y est dans C qui est fermé.

Et en faisant tendre k vers l'infini dans (1.3), on obtient $\|x - y\| = \delta$.

On a donc l'existence de $y \in C$ vérifiant (1.1) et il reste à montrer l'unicité.

Si z est un autre élément de C vérifiant (1.1), l'identité de la médiane nous donne alors :

$$\begin{aligned}\|y - z\|^2 &= \|(y - x) + (x - z)\|^2 \\ &= 2\left(\|y - x\|^2 + \|x - z\|^2\right) - 4\left\|x - \frac{1}{2}(y + z)\right\|^2 \\ &= 4\delta^2 - 4\left\|x - \frac{1}{2}(y + z)\right\|^2 \leq 4\delta^2 - 4\delta^2 = 0,\end{aligned}$$

puisque $\frac{1}{2}(y + z) \in C$ (qui est convexe) et nécessairement $z = y$.

Soit y l'élément de C vérifiant (1.1). Cet ensemble étant convexe, pour tout z dans C et tout λ dans $]0, 1]$, le vecteur $v = (1 - \lambda)y + \lambda z$ est dans C et :

$$\|x - y\|^2 \leq \|x - v\|^2 = \|x - y\|^2 - 2\langle x - y \mid z - y\rangle\lambda + \|z - y\|^2\lambda^2,$$

ce qui équivaut à :

$$\forall z \in C, \quad \forall \lambda \in]0, 1], \quad 2\langle x - y \mid z - y\rangle \leq \|z - y\|^2\lambda.$$

En faisant tendre λ vers 0, on aboutit à :

$$\forall z \in C, \quad \langle x - y \mid z - y\rangle \leq 0$$

Réciproquement supposons que $t \in C$ vérifie (1.2). Pour tout $z \in C$, on peut écrire :

$$\begin{aligned}\|x - z\|^2 &= \|(x - t) + (t - z)\|^2 \\ &= \|x - t\|^2 - 2\langle x - t \mid z - t\rangle + \|t - y\|^2 \geq \|x - t\|^2\end{aligned}$$

ce qui équivaut à dire que $\|x - t\| = \inf_{z\in C}\|x - z\|$ et nécessairement $t = y$. ■

Avec les hypothèses et notations du théorème précédent, la borne inférieure $\delta = \inf_{z\in C}\|x - z\|$ est la distance de x à C. Elle est notée $d(x, C)$. Le vecteur $y \in C$ réalisant cette distance est la meilleure approximation de $x \in E$ par des éléments du convexe C. Ce vecteur y étant également caractérisé par (1.2) est de ce fait aussi appelé la projection de x sur C et noté $y = p_C(x)$. L'application p_C ainsi définie de E sur C est la projection de E sur C.

Corollaire 1.3 *Soit C une partie convexe fermée non vide de E, distincte de E. Si pour x dans $E \setminus C$, on désigne par D la droite vectorielle dirigée par $x - p_C(x)$ et par H l'hyperplan affine passant par x et orthogonal à D, soit :*

$$H = x + D^{\perp} = \{z \in E \mid \langle x - p_C(x) \mid x - z \rangle = 0\},$$

alors cet hyperplan contient x et C est contenu dans le demi-espace ouvert :

$$H^{+,*} = \{z \in E \mid \langle x - p_C(x) \mid x - z \rangle > 0\}.$$

Démonstration. On a bien $x \in H$ et pour tout $z \in C$, on a :

$$\begin{aligned} \langle x - p_C(x) \mid x - z \rangle &= \langle x - p_C(x) \mid (x - p_C(x)) + (p_C(x) - z) \rangle \\ &= \|x - p_C(x)\|^2 - \langle x - p_C(x) \mid z - p_C(x) \rangle \\ &\geq \|x - p_C(x)\|^2 > 0 \end{aligned}$$

puisque $x \notin C$. ■

1.4 Le théorème de Krein-Milman

Pour ce paragraphe, E est un espace vectoriel euclidien de dimension $n \geq 2$. On note $(e_1, \cdots, e_n)$ une base orthonormée de E et $(e_1^*, \cdots, e_n^*)$ la base duale ($e_i^* \left(\sum_{k=1}^{n} x_k e_k \right) = x_i$ pour $1 \leq i \leq n$).

Si X est une partie non vide de E, on définit sa frontière par $\mathrm{Fr}(X) = \overline{X} \setminus \overset{\circ}{X}$. Dans le cas où X est fermé, cette frontière est $\mathrm{Fr}(X) = X \setminus \overset{\circ}{X}$.

Définition 1.4 *Soit C un convexe dans E non vide et distinct de E. On dit qu'un hyperplan affine H est un hyperplan d'appui de C si $H \cap C$ est non vide et C est contenu dans l'un des demi-espaces fermés limités par H.*

Lemme 1.1 *Soit C un convexe dans E non vide et distinct de E. Si H est un hyperplan d'appui de C, alors tout point de $H \cap C$ est un point frontière de C.*

Démonstration. Soit H un hyperplan d'appui de C d'équation $\varphi(x) = \alpha$.

On rappelle que pour toute forme linéaire φ sur E, on peut trouver un unique vecteur v dans E tel que pour tout x dans E, on ait $\varphi(x) = \langle x \mid v \rangle$ ($v = \sum_{k=1}^{n} \varphi(e_k) e_k$) et ce vecteur est non nul si φ est non nulle.

Soit $a \in H \cap C$ et supposons que $\varphi(x) = \langle x \mid v \rangle \geq \alpha$ pour tout x dans C.

Si a n'est pas dans la frontière de C, il existe un réel $\varepsilon > 0$ tel que la boule ouverte $B(a, \varepsilon)$ de centre a et de rayon ε soit contenue dans C. Pour tout réel $\lambda > 0$ assez petit, on a $a - \lambda v \in C$ et :

$$\varphi(a - \lambda v) = \langle a \mid v \rangle - \lambda \|v\|^2 < \langle a \mid v \rangle = \alpha$$

en contradiction avec $\varphi(a - \lambda v) \geq \alpha$. On a donc $a \in \mathrm{Fr}(C)$. ■

Exemple 1.4 *Soit $C = \bigcap_{i=1}^{p} H_i^+$ un polyèdre convexe dans E, avec :*

$$H_i^+ = \{x \in E \mid \varphi_i(x) \geq \alpha_i\} \quad (1 \leq i \leq p),$$

où les φ_i sont des formes linéaires non nulles sur E et les α_i des réels.
Si $x \in C$ est tel que $\varphi_i(x) > \alpha_i$ pour tout i compris entre 1 et p, avec la continuité des applications φ_i, on déduit alors qu'il existe un réel $\varepsilon > 0$ tel que la boule ouverte $B(x, \varepsilon)$ de centre x et de rayon ε soit contenue dans $C = \bigcap_{i=1}^{p} H_i^+$ et en conséquence x est dans l'intérieur de C, donc $x \notin \mathrm{Fr}(C)$.
On a donc ainsi montré que pour tout $x \in \mathrm{Fr}(C)$, il existe un indice i compris entre 1 et p tel que $\varphi_i(x) = \alpha_i$ et H_i est un hyperplan d'appui de C qui contient x. C'est-à-dire que tout point de la frontière de C est contenu dans un hyperplan d'appui.

En fait ce résultat est valable pour tout convexe fermé dans l'espace euclidien E.

Lemme 1.2 *Si C est un convexe fermé dans E non vide et distinct de E, alors tout point de la frontière de C est contenu dans un hyperplan d'appui de C.*

Démonstration. Soit a dans la frontière de C. Pour tout entier naturel non nul k on peut trouver un élément x_k dans la boule ouverte $B\left(a, \frac{1}{k}\right)$ de centre a et de rayon $\frac{1}{k}$ qui n'appartient pas à C. On note alors $y_k = p_C(x_k)$ la projection de x_k sur C, $z_k = \frac{1}{\|x_k - y_k\|}(x_k - y_k)$ et par H_k l'hyperplan affine passant par x_k et orthogonal à z_k, soit :

$$H_k = \{z \in E \mid \langle z_k \mid x_k - z\rangle = 0\}.$$

Le corollaire 1.3 nous dit alors que C est contenu dans le demi-espace ouvert :

$$H_k^{+,*} = \{z \in E \mid \langle z_k \mid x_k - z\rangle > 0\}.$$

Chaque vecteur z_k étant dans la sphère unité de E qui est compacte puisque E est de dimension finie, on peut extraire de la suite $(z_k)_{k\geq 1}$ une sous-suite $\left(z_{\varphi(k)}\right)_{k\geq 1}$ qui converge vers un vecteur v de norme 1. En considérant que la suite $(x_k)_{k\geq 1}$ converge vers a et en utilisant la continuité du produit scalaire, on déduit que pour tout z dans C on a :

$$\langle v \mid a - z\rangle = \lim_{k\to+\infty} \left\langle z_{\varphi(k)} \mid x_{\varphi(k)} - z\right\rangle \geq 0,$$

c'est-àdire que C est contenu dans le demi-espace :

$$H^+ = \{z \in E \mid \langle v \mid a - z\rangle \geq 0\},$$

le vecteur a étant dans l'hyperplan H d'équation $\langle v \mid a - z\rangle = 0$. Cet hyperplan H est donc un hyperplan d'appui de C. ■

Par analogie à la notion de sommet d'un polygone dans $\mathbb{R}^2$, on définit de manière plus générale les sommets, ou points extrémaux, d'un convexe de la manière suivante.

Définition 1.5 *Soit C un convexe non vide de E. On dit qu'un point a de C est un point extrémal si tout segment dans C qui contient a admet ce point pour extrémité.*

Dire que a dans le convexe C est extrémal équivaut à dire que si $a \in [x, y]$ avec x, y dans C, alors $a = x$ ou $a = y$, encore équivalent à dire que si $a = (1-\lambda)x + \lambda y$ avec x, y dans C et $0 < \lambda < 1$, alors $a = x = y$.

Une définition équivalente de point extrémal d'un convexe est donnée par le résultat suivant.

Lemme 1.3 *Soit C un convexe non vide de E. Un point a de C est extrémal si et seulement si $C \setminus \{a\}$ est convexe.*

Démonstration. Soit $a \in C$ extrémal et x, y dans $C \setminus \{a\}$.

L'ensemble C étant convexe, on a $[x, y] \subset C$ et $a \in [x, y]$ entraîne $x = a$ ou $y = a$, ce qui est exclu. On a donc $[x, y] \subset C \setminus \{a\}$. On a donc ainsi montré que $C \setminus \{a\}$ est convexe.

Réciproquement supposons $C \setminus \{a\}$ convexe et soit $[x, y]$ un segment dans C qui contient a. Si x, y sont tous deux dans $C \setminus \{a\}$ qui est convexe, on a alors $a \in [x, y] \subset C \setminus \{a\}$, ce qui est impossible. On a donc $x = a$ ou $y = a$, ce qui prouve que a est extrémal dans C. ■

Exemple 1.5 *Les points extrémaux du convexe de $\mathbb{R}^n$:*

$$P = \left\{ x \in \left(\mathbb{R}^+\right)^n \mid \|x\|_1 = \sum_{i=1}^{n} x_i = 1 \right\}$$

sont les vecteurs $e_1, \cdots, e_n$ de la base canonique.
En effet si $e_i = (1-\lambda)x + \lambda y$ avec x, y dans P et $0 \leq \lambda \leq 1$, alors :

$$(1-\lambda)x_j + \lambda y_j = \begin{cases} 0 \text{ si } j \neq i, \\ 1 \text{ si } j = i, \end{cases}$$

avec x_j, y_j positifs pour $1 \leq j \leq n$. Si $0 < \lambda < 1$, alors $x_j = y_j = 0$ pour $j \neq i$ et $x_i = y_i = 1$ puisque $\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j = 1$, c'est-à-dire que $x = y = e_i$. Chaque vecteur e_i est donc extrémal dans P.
Réciproquement si a est un élément extrémal de P et a n'est égal à aucun des e_i, alors ce vecteur a au moins deux composantes a_i et a_j strictement positives avec $1 \leq i < j \leq n$. Si $t = \min(a_i, a_j)$, alors $0 < t < 1$ et en posant $x = a + t(e_i - e_j)$ et $y = a + t(-e_i + e_j)$, on a :

$$\begin{cases} x = (a_1, \cdots, a_{i-1}, a_i + t, a_{i+1}, \cdots, a_{j-1}, a_j - t, a_{j+1}, \cdots, a_n) \in P, \\ y = (a_1, \cdots, a_{i-1}, a_i - t, a_{i+1}, \cdots, a_{j-1}, a_j + t, a_{j+1}, \cdots, a_n) \in P \end{cases}$$

avec $a = \frac{1}{2}(x + y)$, c'est-à-dire que a est le milieu du segment $[x, y] \subset P$ avec $a \neq x$, $a \neq y$, en contradiction avec a extrémal. On a donc ainsi montré que les e_i sont les seuls points extrémaux de P.

De manière plus générale un convexe compact admet des points extrémaux.

Lemme 1.4 *Un convexe compact non vide dans E a des points extrémaux.*

Démonstration. Soit C un convexe compact non vide dans E. L'application e_1^* étant continue sur le compact C, elle y est bornée et atteint sa borne inférieure, on peut donc poser :

$$t_1 = \inf_{x \in C} e_1^*(x)$$

(t_1 est la plus petite des premières composantes d'éléments de C). L'ensemble :

$$C_1 = \left\{ x = t_1 e_1 + \sum_{i=2}^{n} x_i e_i \mid x \in C \right\}$$

est alors un compact non vide de E (il est fermé et borné) et on peut poser :

$$t_2 = \inf_{x \in C_1} e_2^* (x) .$$

En continuant ainsi de suite on construit un vecteur $t = \sum_{i=1}^{n} t_i e_i$ dans C et on vérifie qu'il extrémal.

Si $t = (1 - \lambda) x + \lambda y$ avec x, y dans C et λ dans $]0, 1[$, alors de $t_1 = (1 - \lambda) x_1 + \lambda y_1$ avec $t_1 \leq x_1$, $t_1 \leq y_1$ et $0 < \lambda < 1$, on déduit que nécessairement $t_1 = x_1 = y_1$. Puis par récurrence, vue la construction des t_k, on déduit que $t_k = x_k = y_k$ pour tout k compris entre 1 et n. On a donc $t = x = y$, ce qui prouve que t est extrémal dans C. ■

Lemme 1.5 *Si C est un convexe compact non vide dans E alors pour tout hyperplan d'appui H de C, $C \cap H$ (qui est convexe compact et non vide) admet des points extrémaux qui sont aussi des points extrémaux de C.*

Démonstration. Notons $H = \{x \in E \mid \varphi (x) = \alpha\}$ un hyperplan d'appui de C. On a $\varphi (x) \geq \alpha$ pour tout $x \in C$ et H contient un point frontière de C. L'intersection $C \cap H$ est alors un convexe compact non vide et il admet des points extrémaux (lemme 1.4).

Soit x un point extrémal de $C \cap H$. Si il existe y, z dans C et λ dans $]0, 1[$ tels que $x = (1 - \lambda) y + \lambda z$, alors :

$$\alpha = \varphi (x) = (1 - \lambda) \varphi (y) + \lambda \varphi (z) > (1 - \lambda) \alpha + \lambda \alpha = \alpha$$

si $\varphi (y) > \alpha$ ou $\varphi (z) > \alpha$, ce qui est impossible. On a donc $\varphi (y) = \alpha$ et $\varphi (z) = \alpha$, c'est-à-dire que y et z sont dans $C \cap H$ et $y = z = x$ puisque x est un point extrémal de $C \cap H$. En conclusion x est un point extrémal de C. ■

Théorème 1.4 (Krein-Milman) *Tout compact convexe dans l'espace euclidien E est l'enveloppe convexe de ses points extrémaux.*

Démonstration. Soit C un convexe compact non vide dans E. On note $S (C)$ l'enveloppe convexe de l'ensemble des points extrémaux de C. On a $S (C) \subset C$.

Supposons qu'il existe a dans C qui n'est pas dans $S (C)$. On a alors $a \notin \overline{S (C)}$, ce dernier ensemble étant convexe (l'adhérence d'un convexe est convexe) et fermé dans E. On peut donc utiliser le corollaire 1.3 pour dire qu'il existe un hyperplan affine d'équation $\varphi (x) = \alpha$ contenant a et tel que $\varphi (x) > \alpha$ pour tout $x \in \overline{S (C)}$. L'image de C par φ est convexe (image d'un convexe par l'application linéaire φ) et compacte (image du compact C par l'application continue φ) dans $\mathbb{R}$, c'est donc un intervalle réel $[u, v]$ qui contient α (puisque $\varphi (a) = \alpha$). Désignons par K l'hyperplan affine d'équation $\varphi (x) = u$. On a $K \cap C \neq \emptyset$ et $\varphi (x) \geq u$ pour tout $x \in C$ car $[u, v] = \varphi (C)$, c'est-à-dire que K est un hyperplan d'appui de C et le lemme 1.5 nous dit que K contient des points extrémaux de C, si x est l'un de ces points il est alors dans $S (C)$ et $\varphi (x) > \alpha \geq u$ en contradiction avec $\varphi (x) = u$ (x est dans K). On a donc $C \subset S (C)$ et $C = S (C)$. ■

Pour des énoncés dans un cadre plus général, on peut consulter [1] ou [2].

1.5 Matrices bistochastiques

Si n est un entier naturel supérieur ou égal à 2, on désigne par $\mathfrak{S}_n$ le groupe des permutations de l'ensemble $\{1, \cdots, n\}$ et par $\mathcal{B} = (e_1, \cdots, e_n)$ la base canonique de $\mathbb{R}^n$.

Pour tout couple (i, j) d'entiers naturels, on note $\delta_{i,j}$ le symbole de Kronecker ($\delta_{ii} = 1$ et $\delta_{i,j} = 0$ pour $i \neq j$).

Définition 1.6 *Si $\sigma \in \mathfrak{S}_n$, on appelle matrice de permutation associée à σ, la matrice de passage P_σ de la base canonique de $\mathbb{R}^n$ à la base $\mathcal{B}_\sigma = \left(e_{\sigma(1)}, \cdots, e_{\sigma(n)}\right)$.*

On a donc, si P_σ est une matrice de permutation, $P_\sigma e_j = e_{\sigma(j)}$ pour tout entier j compris entre 1 et n, ce qui revient à dire que :

$$P_\sigma = \left(\left(\delta_{i,\sigma(j)}\right)\right)_{1 \leq i,j \leq n}.$$

Définition 1.7 *On dit qu'une matrice $A = ((a_{i,j}))_{1 \leq i,j \leq n}$ dans $\mathcal{M}_n(\mathbb{R})$ est stochastique si elle est positive et :*

$$\forall i \in \{1, \cdots, n\}, \ \sum_{j=1}^{n} a_{ij} = 1$$

Exemple 1.6 *Une matrice de permutation est stochastique.*

Il est facile de vérifier que l'ensemble $\mathcal{P}_n(\mathbb{R})$ des matrices stochastiques est convexe et compact.

Définition 1.8 *On appelle matrice doublement stochastique une matrice stochastique A telle que ${}^tA$ soit aussi stochastique.*

Les matrices de permutation sont des matrices doublement stochastiques.

On note $\mathcal{B}_n(\mathbb{R})$ l'ensemble des matrices bistochastiques dans $\mathcal{M}_n(\mathbb{R})$ et $\mathcal{E}_n(\mathbb{R})$ le sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$ constitué des matrices A vérifiant $\sum\limits_{k=1}^{n} a_{ik} = \sum\limits_{j=1}^{n} a_{kj} = 0$ pour tous i, j compris entre 1 et n.

Lemme 1.6 *L'espace vectoriel $\mathcal{E}_n(\mathbb{R})$ est de dimension égale à $(n-1)^2$.*

Démonstration. On désigne par φ l'application linéaire qui à toute matrice $X = ((x_{i,j}))_{1 \leq i,j \leq n}$ dans $\mathcal{E}_n(\mathbb{R})$ associe la matrice $Y = ((x_{i,j}))_{1 \leq i,j \leq n-1}$.
Si $X \in \mathcal{E}_n(\mathbb{R})$ est telle que $\varphi(X) = 0$, alors $x_{ij} = 0$ pour tous i, j compris entre 1 et $n-1$, $x_{in} = -\sum\limits_{k=1}^{n-1} x_{ik} = 0$ pour i compris entre 1 et $n-1$, et $x_{nn} = -\sum\limits_{k=1}^{n-1} x_{kn} = 0$. C'est-à-dire que $X = 0$ et φ est injective.
Pour toute matrice $Y \in \mathcal{M}_{n-1}(\mathbb{R})$, en posant :

$$x_{in} = -\sum_{k=1}^{n-1} y_{ik}, \ x_{nj} = -\sum_{k=1}^{n-1} y_{kj}, \ (1 \leq i, j \leq n-1)$$

on a :

$$\sum_{i=1}^{n-1} x_{i,n} = -\sum_{i=1}^{n-1} \sum_{k=1}^{n-1} y_{ik} = -\sum_{k=1}^{n-1} \sum_{i=1}^{n-1} y_{ik} = \sum_{k=1}^{n-1} x_{nk}$$

de sorte que la matrice :

$$X = \begin{pmatrix} y_{11} & \cdots & y_{1,n-1} & x_{1n} \\ \vdots & \ddots & \vdots & \vdots \\ y_{n-1,1} & \cdots & y_{n-1,n-1} & x_{n-1,n} \\ x_{n1} & \cdots & x_{n,n-1} & -\sum\limits_{i=1}^{n-1} x_{i,n} \end{pmatrix}$$

est dans $\mathcal{E}_n(\mathbb{R})$ avec $\varphi(X) = Y$. L'application φ est donc surjective. Il en résulte que c'est un isomorphisme et :

$$\dim(\mathcal{E}_n(\mathbb{R})) = \dim(\mathcal{M}_{n-1}(\mathbb{R})) = (n-1)^2$$

■

Lemme 1.7 *$\mathcal{B}_n(\mathbb{R})$ est un polyèdre convexe dans $\mathcal{M}_n(\mathbb{R})$ et les points extrémaux de ce polyèdre sont les matrices de permutations.*

Démonstration. Il est facile de vérifier que $\mathcal{B}_n(\mathbb{R})$ est fermé et borné, donc compact.
Dire que $A = ((a_{i,j}))_{1\leq i,j\leq n}$ est dans $\mathcal{B}_n(\mathbb{R})$ équivaut à :

$$a_{ij} \geq 0,\ \sum_{k=1}^{n} a_{ik} = 1,\ \sum_{j=1}^{n} a_{kj} = 1 \quad (1 \leq i,j \leq n).$$

En notant $\left(E_{ij}^* \mid 1 \leq i,j \leq n\right)$ la base duale de la base canonique de $\mathcal{M}_n(\mathbb{R})$ ($E_{ij}^*(A) = a_{ij}$), et pour $1 \leq i,j \leq n$, L_i^* et C_j^* les formes linéaires définies sur $\mathcal{M}_n(\mathbb{R})$ par :

$$L_i^* = \sum_{k=1}^{n} E_{ik}^*,\ C_j^* = \sum_{k=1}^{n} E_{kj}^*,$$

on déduit que $A \in \mathcal{B}_n(\mathbb{R})$ équivaut à :

$$E_{ij}^*(A) \geq 0,\ L_i^*(A) = 1,\ C_j^*(A) = 1 \quad (1 \leq i,j \leq n).$$

Il en résulte que $\mathcal{B}_n(\mathbb{R})$ est un polyèdre convexe dans $\mathcal{M}_n(\mathbb{R})$.

Soit A une matrice de permutation et supposons que $A = (1-\lambda)X + \lambda Y$ avec X, Y dans $\mathcal{B}_n(\mathbb{R})$ et $0 < \lambda < 1$. Sur chaque ligne i de la matrice A il y a un seul coefficient non nul $a_{ij} = 1$ et pour $k \neq j$, on a :

$$0 = a_{ik} = (1-\lambda)x_{ik} + \lambda y_{ik}$$

qui entraîne $x_{ik} = y_{ik} = 0$ et $x_{ij} = y_{ij} = 1$ puisque tous les coefficients d'une matrice bistochastique sont positifs ou nuls et la somme des termes d'une même ligne vaut 1. On a donc $X = Y = A$.

On a donc ainsi montré que les matrices de permutation sont des points extrémaux de $\mathcal{B}_n(\mathbb{R})$.

Pour la réciproque, on procède par récurrence sur $n \geq 2$.

Pour $n = 2$, une matrice stochastique est de la forme :

$$A = \begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix},$$

avec $0 \leq a \leq 1$. Si cette matrice est de permutation c'est alors un point extrémal de $\mathcal{B}_2(\mathbb{R})$, sinon on a $0 < a < 1$. Supposons, ce qui n'est pas restrictif, que $0 < a \leq 1 - a < 1$, soit $0 < a \leq \frac{1}{2}$. En posant :

$$\begin{aligned} X &= A + a \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ Y &= A - a \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2a & 1-2a \\ 1-2a & 2a \end{pmatrix}, \end{aligned}$$

on a $X \in \mathcal{B}_2(\mathbb{R})$, $Y \in \mathcal{B}_2(\mathbb{R})$, $A = \frac{1}{2}(X + Y)$, avec $A \neq X$, $A \neq Y$, ce qui signifie que A n'est pas extrémal. Le résultat est donc montré pour $n = 2$.

Supposons le acquis pour $n - 1 \geq 2$ et soit $A \in \mathcal{B}_n(\mathbb{R})$ un élément extrémal.
On montre tout d'abord que la matrice A a au plus $2n - 1$ coefficients non nuls.
Supposons que A ait au moins $2n$ coefficients non nuls que nous notons a_{i_k,j_k} avec $1 \leq k \leq 2n$ et les couples (i_k, j_k) deux à deux distincts. On désigne par H le sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$ de dimension $2n$ engendré par les matrices E_{i_k,j_k}, pour $1 \leq k \leq 2n$, où $(E_{ij} \mid 1 \leq i, j \leq n)$ désigne la base canonique de $\mathcal{M}_n(\mathbb{R})$. On a $H \cap \mathcal{E}_n(\mathbb{R}) \neq \{0\}$ à cause des dimensions. Il existe donc une matrice B dans $H \cap \mathcal{E}_n(\mathbb{R}) \setminus \{0\}$. Si pour tout réel t, on note $C_t = A + tB$, on a alors $c_{ij} = a_{ij}$ pour $(i,j) \neq (i_k, j_k)$ et $c_{i_k,j_k} = a_{i_k,j_k} + tb_{i_k,j_k} > 0$ pour t voisin de 0 (on a $a_{i_k,j_k} > 0$ pour tout k), puis avec $B \in \mathcal{E}_n(\mathbb{R})$, on déduit que $\sum\limits_{k=1}^{n} c_{ik} = \sum\limits_{k=1}^{n} a_{ik} = 1$, $\sum\limits_{k=1}^{n} c_{kj} = \sum\limits_{k=1}^{n} a_{kj} = 1$ pour tous i, j. Pour $t \in \left]-\varepsilon, \varepsilon\right[\setminus \{0\}$ avec $\varepsilon > 0$ assez petit, les matrice C_t et C_{-t} sont donc dans $\mathcal{B}_n(\mathbb{R})$ et $A = \frac{1}{2}(C_t + C_{-t})$ avec $A \neq C_t$, $A \neq C_{-t}$, en contradiction avec A extrémal.

La matrice A a donc au plus $2n - 1$ termes non nuls et il existe nécessairement une ligne d'indice i de cette matrice avec un seul coefficient a_{ij} non nul, ce coefficient valant 1. La matrice A étant bistochastique, tous les coefficients de la colonne j, excepté celui en ligne i, sont nuls. La matrice A' extraite de A en supprimant la ligne i et la colonne j est alors dans $\mathcal{B}_{n-1}(\mathbb{R})$ et extrémale. En effet si $A' \in [B', C']$ avec B', C' dans $\mathcal{B}_{n-1}(\mathbb{R})$ alors $A \in [B, C]$, où B, C sont dans $\mathcal{B}_n(\mathbb{R})$ telles que $b_{ij} = c_{ij} = 1$ et B', C' sont extraites de B, C en supprimant la ligne i et la colonne j et $A = B$ ou $A = C$, ce qui entraîne $A' = B'$ ou $A' = B'$. Avec l'hypothèse de récurrence, on déduit alors que A' est une matrice de permutation dans $\mathcal{M}_{n-1}(\mathbb{R})$ et A est une matrice de permutation dans $\mathcal{M}_n(\mathbb{R})$. ■

De ce résultat et des théorèmes de Krein-Milman et de Carathéodory (théorèmes 1.4 et 1.2), on déduit alors le résultat suivant.

Théorème 1.5 (Birkhoff) *Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est doublement stochastique si et seulement si elle s'écrit $A = \sum\limits_{k=1}^{p} \lambda_k P_k$, où $p \leq (n-1)^2 + 1$, les P_k sont des matrices de permutation et les λ_k des réels positifs tels que $\sum\limits_{k=1}^{p} \lambda_k = 1$.*

Démonstration. L'ensemble des matrices bistochastiques étant un polyèdre convexe dans l'espace vectoriel $\mathcal{M}_n(\mathbb{R})$, le théorème de Krein-Milman nous dit que c'est l'enveloppe convexe de ses points extrémaux, donc de l'ensemble des matrices de permutations.
Toute matrice $A \in \mathcal{B}_n(\mathbb{R})$ s'écrit donc $A = \sum\limits_{k=1}^{q} \mu_k P_k$, où les P_k sont des matrices de permutation et les μ_k des réels positifs tels que $\sum\limits_{k=1}^{q} \mu_k = 1$. En écrivant $A - I_n = \sum\limits_{k=1}^{q} \mu_k (P_k - I_n)$, on

déduit que $A - I_n$ est dans l'enveloppe convexe de l'ensemble X, contenu dans $\mathcal{E}_n(\mathbb{R})$, formé des matrices $P_\sigma - I_n$, où σ parcourt l'ensemble des permutations de $\{1, \cdots, n\}$. Le théorème de Carathéodory nous dit alors que $A - I_n = \sum_{k=1}^{p} \lambda_k (P_k - I_n)$, où $p \leq \dim(\mathcal{E}_n(\mathbb{R})) + 1 = (n-1)^2 + 1$, les P_k sont des matrices de permutation et les λ_k des réels positifs tels que $\sum_{k=1}^{p} \lambda_k = 1$, ce qui équivaut à $A = \sum_{k=1}^{p} \lambda_k P_k$.

■

1.6 Exercices

Exercice 1.1 *Montrer que si C est un convexe compact dans un espace euclidien E, il est alors l'enveloppe convexe de sa frontière.*

Solution 1.1 *L'ensemble C étant fermé, on a $\mathrm{Fr}\,(C) \subset C$ et $Cv\,(\mathrm{Fr}\,(C))$ est contenu dans C puisque C est convexe.*
Soit x un élément de C. Pour toute droite D passant par x, $D \cap C$ est convexe comme intersection de convexes, fermé comme intersection de fermés et borné car contenu dans C qui est compact, c'est donc un convexe compact de D, c'est-à-dire un segment de D (qui peut être identifiée à $\mathbb{R}$ par le choix d'une origine). On a donc $x \in D \cap C = [y, z]$ avec y, z dans la frontière de C, ce qui entraîne $x \in Cv\,(\mathrm{Fr}\,(C))$. On a donc bien $C = Cv\,(\mathrm{Fr}\,(C))$.

Bibliographie

[1] S. Lang — *Real analysis*. Addison-Wesley (1969).

[2] W. Rudin — *Analyse fonctionnelle*. Ediscience (1995).

[3] P. Tauvel — *Mathématiques générales pour l'agrégation*. Masson (1992).