Le théorème déterminantal en bref

par Jean-Baptiste HIRIART-URRUTY

Institut de mathématiques, Université Paul Sabatier, TOULOUSE

RÉSUMÉ. *Dans cette note à caractère essentiellement pédagogique, nous présentons le théorème déterminantal de* CRAIG-SAKAMOTO, *accompagné de notre démonstration favorite. Son histoire et une de ses applications sont brossées à grands traits. Nous concluons en suggérant une nouvelle piste de démonstration, que nous conjecturons être possible mais que nous n'avons pu mener à bien.*

MOTS-CLÉS : *Matrice symétrique rélle. Déterminant. Développement en série entière.*

Introduction.

Commençons par le commencement, à savoir l'énoncé du théorème auquel on va s'intéresser.

Théorème 1 *Le théorème déterminantal.*

Soit A et B deux matrices symétriques de taille p à coefficients réels. Alors, les deux assertions suivantes sont équivalentes :

$(\mathcal{H})$ $\quad \det(I_p + xA + yB) = \det(I_p + xA)\det(I_p + yB)$ *pour tous x et y réels ;*

$(\mathcal{C})$ $\quad AB = 0.$

Ce théorème, dont on pourrait penser qu'il est un simple exercice d'algèbre linéaire, gratuit et sans applications, a toute une histoire et (au moins) une application en Probabilités-Statistiques (domaine de son origine). Toute une histoire car les premiers énoncés étaient faux, les premières démonstrations incomplètes (faites seulement sous des hypothèses additionnelles comme la commutation de A et B), puis découvertes ou redécouvertes indépendamment par plusieurs auteurs. L'histoire est bien racontée dans [2], [10] et [6], cette dernière référence ayant été notre principale source d'inspiration. Historiquement, c'est A. T. CRAIG (1943) qui énonce le premier le résultat dans un contexte de Probabilités-Statistiques ; en version contextualisée, cela donne : si X est un vecteur aléatoire de $\mathbb{R}^p$ distribué suivant la loi normale $\mathcal{N}(\mu, I_p)$, alors les formes quadratiques $X^T AX$ et $X^T BX$ sont des variables aléatoires réelles indépendantes si, et seulement si, $AB = 0$; apparemment, sa démonstration n'était pas convaincante. Un peu plus tard, en 1949, H. SAKAMOTO conjecture également le résultat mais ne donne pas de démonstration. Les premières démonstrations complètes et correctes sont venues vite après, notamment avec l'article de K. MATUSITA en 1949. Cela explique que le théorème au-dessus est souvent référencé comme celui de CRAIG-SAKAMOTO ou même de CRAIG-SAKAMOTO-MATUSITA. Devant le nombre de contributeurs au sujet, nous avons décidé de lui donner un nom neutre, quitte à pécher par néologisme : *le théorème déterminantal.*

Ceci est une illustration de ce qui peut arriver en sciences, en mathématiques en particulier ; en effet, le processus de création dans ces domaines n'est pas linéaire : il peut arriver que des résultats soient annoncés mais pas vraiment démontrés, puis les énoncés sont démontrés et même modifiés, enfin ils sont simplifiés et généralisés. Ce n'est pas sans rappeler ce que disait S. BANACH : *"La première version d'un théorème est toujours dégueulasse..."*.

1. Les premières idées et commentaires.

- Si on devait se servir de ce théorème déterminantal pour faire un sujet d'examen, on commencerait bien sûr par une question simple : montrer que $(\mathcal{C})$ implique $(\mathcal{H})$. Le déterminant du produit de deux matrices étant le produit de leurs déterminants, ça va tout seul....C'est donc l'implication $[(\mathcal{H}) \Longrightarrow (\mathcal{C})]$ qui pose problème.

- Le résultat ne saurait être vrai si on enlève l'hypothèse de symétrie des matrices A et B. Par exemple (*cf.* [12]), si

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \text{ et } B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

  on a $\det(I_p + xA + yB) = \det(I_p + xA) = \det(I_p + yB) = 1$ pour tous les x et y réels, et pourtant $AB \neq 0$.

- Que A et B commutent est une *conséquence* du théorème, on ne voit pas comment cette propriété pourrait être directement déduite de $(\mathcal{H})$.

- Contrairement à ce qu'on pourrait penser au premier abord, on ne peut réduire le problème à la situation où l'une des matrices, disons A, est semidéfinie positive... ce qui est fort dommage.

- Étant donné que c'est l'égalité de deux fonctions (très régulières) de deux variables réelles x et y qui est en jeu dans $(\mathcal{H})$, $f(x, y) = g(x)h(y)$, il est normal de penser à utiliser les outils et résultats du calcul différentiel : les deux fonctions f et $g.h$ coïncidant au point $(0, 0)$, l'égalité fonctionnelle de départ équivaut à l'égalité des vecteurs gradients : $\nabla f(x, y) = \begin{pmatrix} g'(x)h(y) \\ g(x)h'(y) \end{pmatrix}$ pour tout $(x, y) \in \mathbb{R}^2$. Cela donne quelques résultats intéressants sur A et B, mais on reste loin du compte, qui est : $AB = 0$. De même, des développements comme $(I_p - uA)^{-1} = \sum_{k=0}^{+\infty} u^k A^k$ (pour u réel petit), alliés au calcul différentiel du premier ordre de la fonction déterminant, permettent d'accéder à des résultats sur des traces de matrices construites à partir de A et B, par exemple $\mathrm{tr}(A^k B) = \mathrm{tr}(AB^k) = 0$ pour tout entier $k \geqslant 1$. Il manque toutefois un maillon dans la chaîne de la démonstration pour conclure.

C'est néanmoins un point de vue que nous adoptons, et notre démonstration favorite repose sur un développement en série entière de la fonction $u \longmapsto \ln\det(I_p + uM)$. Elle est due à T. OGAZAWARA et M. TAKAHASHI en 1951 ([8]). Nous la présentons dans le § 2 qui suit.

- Le théorème déterminantal en question étant un résultat d'algèbre linéaire, c'est en puisant dans les outils et résultats de ce domaine des mathématiques que sont faites les autres démonstrations que nous connaissons : tout d'abord K. MATUSITA (1949) et H.O. LANCASTER (1954), leur démonstration est présentée dans ([6], pages 794-795) ; également O. TAUSSKY [12] ; plus récemment, I. OLKIN [11], CHI-KWONG LI [1]. C'est assez naturellement que l'on pense à des démonstrations par récurrence sur la taille p de A ; des exemples en sont [1] et [5].

2. Démonstration du théorème déterminantal à l'aide de développements en séries entières.

2.1. *Le développement de la fonction* $z \mapsto \ln\det(I_p + zA)$ *au voisinage de* 0.

Soit $A \in \mathcal{M}_p(\mathbb{C})$. Nous allons démontrer que, pour tout $z \in \mathbb{C}$ dans un voisinage de 0, on a le développement suivant :

$$\ln\det(I_p + zA) = \sum_{n=1}^{+\infty} \frac{(-1)^{n-1}}{n}\,\mathrm{tr}(A^n) z^n. \tag{1}$$

Cette généralité (sur A) n'est pas vraiment nécessaire ici, puisque nous n'avons affaire qu'à des matrices symétriques réelles, mais la démonstration de (1) est intéressante en soi ; faisons-la en plusieurs étapes.

Étape 1. Expression de $\det(A^n)$ et de $\mathrm{tr}(A^n)$ en fonction des éléments diagonaux d'une matrice triangulaire T semblable à A.

Le polynôme caractéristique de A étant scindé (puisque le corps de référence est $\mathbb{C}$ ici), A est semblable à une matrice triangulaire supérieure $T = [t_{ij}] \in \mathcal{M}_p(\mathbb{C})$. Par suite, pour tout entier $n \geqslant 1$,

$$\det(A^n) = \det(T^n) = \prod_{i=1}^{p} t_{ii}^n \quad , \quad \mathrm{tr}(A^n) = \mathrm{tr}(T^n) = \sum_{i=1}^{p} t_{ii}^n. \tag{2}$$

Étape 2. Convergence absolue de la série entière de terme général $\dfrac{(-1)^{n-1}}{n}\,\mathrm{tr}(A^n)z^n$.

Si A est nilpotente, $A^{n_0} = 0$ pour un certain entier n_0. Dans ce cas, A est semblable à une matrice triangulaire supérieure à diagonale nulle. En fait, $\mathrm{tr}(A^n) = 0$ pour tout $n \geqslant 1$, ce qui assure clairement l'égalité (1), qui n'est autre que la relation $\ln 1 = 0$.

Considérons le cas général, c'est-à-dire lorsque A n'est pas nilpotente. Au vu de l'expression (2) de $\mathrm{tr}(A^n)$, le terme général de la série qui nous préoccupe est $\displaystyle\sum_{i=1}^{p} \frac{(-1)^{n-1}}{n}(t_{ii}z)^n$. Ainsi, si $|t_{ii}z| < 1$ pour tout $i = 1, ..., p$, les p séries de termes généraux $\dfrac{(-1)^{n-1}}{n}(t_{ii}z)^n$ sont absolument convergentes et

$$\sum_{n=1}^{+\infty}\left(\sum_{i=1}^{p}\frac{(-1)^{n-1}}{n}(t_{ii}z)^n\right) = \sum_{i=1}^{p}\left(\sum_{n=1}^{+\infty}\frac{(-1)^{n-1}}{n}(t_{ii}z)^n\right).$$

En résumé, la série entière de terme général $\dfrac{(-1)^{n-1}}{n}\,\mathrm{tr}(A^n)z^n$ est absolument convergente pour $|z| < \dfrac{1}{\max_i |t_{ii}|}$.

Étape 3. Développement de $\det(I_p + zA)$ au voisinage de 0.

Pour $|z| \max_i |t_{ii}| < 1$, on a :

$$\begin{aligned}\exp\left[\sum_{i=1}^{p}\left(\sum_{n=1}^{+\infty}\frac{(-1)^{n-1}}{n}(t_{ii}z)^n\right)\right] &= \prod_{i=1}^{p}\exp\left[\sum_{n=1}^{+\infty}\frac{(-1)^{n-1}}{n}(t_{ii}z)^n\right]\\ &= \prod_{i=1}^{p}\exp\left[\ln(1+t_{ii}z)\right],\end{aligned}$$

car on a reconnu, entre les crochets, le développement de la fonction logarithme (principal) $\ln(1 + u)$. Or, $\exp\left[\ln(1 + u)\right] = 1 + u$ pour tout $u \in \mathbb{C}$ tel que $|u| < 1$. Par suite, le produit ci-dessus devient $\displaystyle\prod_{i=1}^{p}(1 + t_{ii}z) = \det(I_p + zT) = \det(I_p + zA)$.

Étape 4. Synthèse des résultats : on a démontré que pour tout $z \in \mathbb{C}$ dans un voisinage de 0

$$\det(I_p + zA) = \exp\left[\sum_{n=1}^{+\infty}\frac{(-1)^{n-1}}{n}\,\mathrm{tr}(A^n)z^n\right].$$

La conséquence, pour ce qui nous concerne, est la belle formule suivante : pour tout $z \in \mathbb{C}$ dans un voisinage de 0,

$$-\ln\det(I_p - zA) = \sum_{n=1}^{+\infty}\frac{\mathrm{tr}(A^n)}{n}z^n. \tag{3}$$

Commentaires.

- On reconnaît en (3) la « cousine matricielle » de la formule plus familière que voici : $-\ln(1-az)=\sum_{n=1}^{+\infty}\frac{a^n}{n}z^n$, valable pour tout $z\in\mathbb{C}$ dans un voisinage de 0.
- Si $f:z\longmapsto f(z):=\det(I_p+zA)$, la $\mathbb{C}$-dérivée de f en 0, c'est-à-dire $f'(0):=\lim_{z\to 0}\frac{f(z)-f(0)}{z}$ est le coefficient a_1 du terme a_1z dans le développement en série entière (à l'origine) de f ; c'est donc $\operatorname{tr}A$. Ce résultat était déjà acquis dès l'étude du polynôme caractéristique de A :

$$\lambda\in\mathbb{C}\mapsto P_A(\lambda):=\det(A-\lambda I_p)=(-1)^p\lambda^p+(-1)^{p-1}(\operatorname{tr}A)\lambda^{p-1}+...\det A.$$

- A l'aide des développements au-dessus, notamment de l'évaluation du rayon de convergence de la série entière de (3), on peut démontrer l'expression suivante du rayon spectral de A :

$$\rho(A):=\max\left(|\lambda_i|\ ;\ \lambda_i \text{ valeur propre de } A\right)=\limsup_{n\to+\infty}|\operatorname{tr}(A^n)|^{\frac{1}{n}}\ .$$

2.2. ***Démonstration du théorème déterminantal à l'aide des développements en séries entières ci-dessus.***

Dans le contexte qui est le nôtre, on se limite à $A\in\mathcal{S}_p(\mathbb{R}),B\in\mathcal{S}_p(\mathbb{R})$. À partir de l'hypothèse de départ, écrite en changeant les signes devant x et y, nous disposons de la relation suivante : pour tout $(x,y)\in\mathbb{R}^2$ dans un voisinage de $(0,0)$,

$$-\ln\det(I_p-xA-yB)=-\ln\det(I_p-xA)-\ln\det(I_p-yB). \tag{4}$$

Nous allons exploiter cette relation en voyant ce que donnent les égalités entre les coefficients des polynômes homogènes de degré $k=1,2,...$apparaissant à l'identique dans chaque membre, après les développements comme en (3). En faisant cela, nous utilisons constamment le résultat suivant sur les traces de matrices : $\operatorname{tr}(UV)=\operatorname{tr}(VU)$.

Conséquence de l'égalité des coefficients des polynômes homogènes :

- du 1$^{\text{er}}$ degré : $\operatorname{tr}(A+B)=\operatorname{tr}A+\operatorname{tr}B$, ce que nous savions déjà ;
- du 2$^{\text{e}}$ degré : $\operatorname{tr}(AB)=0$, intéressant mais insuffisant ;
- du 3$^{\text{e}}$ degré : $\operatorname{tr}(A^2B)=\operatorname{tr}(AB^2)=0$.

On conçoit aisément qu'une égalité des coefficients des polynômes du k^{e} degré conduira à $\operatorname{tr}(A^kB)=\operatorname{tr}(AB^k)=0$. Mais ceci n'est pas suffisant, et c'est l'égalité des coefficients *du*

4$^{\text{e}}$ degré qui va déclencher le résultat annoncé. Ces égalités de coefficients se traduisent par : $\mathrm{tr}(A^3B) = \mathrm{tr}(AB^3) = 0$ (comme déjà dit), mais surtout, en considérant les coefficients de x^2y^2 :

$$\mathrm{tr}(ABAB) + 2\,\mathrm{tr}(A^2B^2) = 0.$$

Considérons à présent la matrice $X := AB$, de terme général noté x_{ij}, et définissons la matrice $M := (X + X^T)^2 + 2XX^T$. Par construction, les matrices $(X + X^T)^2$, XX^T, et donc M, sont symétriques semidéfinies positives. De plus, avec ce qui a été vu au-dessus, $\mathrm{tr}(M) = 2\,\mathrm{tr}(ABAB) + 4\,\mathrm{tr}(A^2B^2) = 0$. Par suite, $\mathrm{tr}(A^2B^2) = \mathrm{tr}(XX^T) = 0$. Comme $\mathrm{tr}(XX^T) = \sum_{i,j=1}^{n} x_{ij}^2$, il s'ensuit que $x_{ij} = 0$ pour tout i, j, c'est-à-dire $X = AB = 0$.

3. Piste pour une autre démonstration du théorème déterminantal via l'analyse convexe.

Une fonction centrale en analyse convexe, dans son développement théorique comme en applications à l'optimisation dite SDP (*i.e.,* avec des contraintes de semidéfinie positivité sur les variables-matrices en jeu), est la fonction $M \mapsto \theta(M) := -\ln\det(M)$, définie sur le cône convexe ouvert des matrices symétriques définies positives (de taille p, disons) ; on l'étend à tout $\mathcal{S}_p(\mathbb{R})$ en lui assignant la valeur $+\infty$ dès que M n'est pas définie positive. Outre la propriété de stricte convexité de θ, les règles de calcul différentiel (du premier et deuxième ordre) qui sont simples, c'est l'expression de sa transformée de LEGENDRE-FENCHEL θ^* qui est intéressante.

$$\theta^*(N) := \sup_{M\in\mathcal{S}_p(R)} \left[\ll N, M \gg -\theta(M)\right].$$

Ici $\ll \cdot,\cdot \gg$ désigne le produit scalaire usuel dans $\mathcal{S}_p(\mathbb{R})$, à savoir celui défini par

$\ll N, M \gg := \mathrm{tr}(MN)$.

En fait ([3], p. 80-81) :

$$\theta^*(N) = -\ln\det(-N) - p \text{ si } N \text{ est définie négative, et } +\infty \text{ sinon.} \tag{5}$$

Que dit l'hypothèse $(\mathcal{H})$ du théorème déterminantal en termes de convexité ? Si on définit

$$(x, y) \longmapsto \varphi(x, y) := -\ln\det(I_p + xA + yB),$$

la fonction φ n'est autre que la composée de θ avec la fonction affine $(x, y) \longmapsto I_p + xA + yB$; elle est donc convexe, à valeurs finies dans un voisinage convexe de $(0, 0)$. Ce qu'exprime $(\mathcal{H})$ n'est autre que

$$\varphi(x, y) = \varphi(x, 0) + \varphi(0, y) \text{ pour tout } (x, y) \in \mathbb{R}^2, \tag{6}$$

une égalité entre deux fonctions convexes de deux variables, dont une (celle de droite) est la somme de deux fonctions d'une seule variable à la fois. Il est alors tentant de prendre la conjuguée de LEGENDRE-FENCHEL dans (6), d'autant que des règles de calcul existent pour ce format de fonctions composées ([3], chapitre X). Nous l'avons tenté, avons obtenu quelques résultats partiels, mais n'avons pu conclure... Il nous paraît inimaginable qu'avec toute l'information que concentre la fonction convexe φ, et la puissance de résultats comme celle de biconjugaison $(\varphi^*)^* = \varphi$, la relation duale (et équivalente) de (6) obtenue par conjugaison ne conduise pas à la conclusion... En attendant, nous laissons cela à la sagacité du lecteur.

4. Extensions

Pour le théorème déterminantal, il y a deux extensions (*i.e.*, affaiblissements) de l'hypothèse $(\mathcal{H})$ qui impliquent encore la même conclusion $AB = 0$, ce sont :

$$(\mathcal{H}_1) : \quad \begin{array}{c} \det(I_p + xA - xB) = \det(I_p + xA)\det(I_p - xB) \text{ et} \\ \det(I_p + xA + xB) = \det(I_p + xA)\det(I_p + xB) \text{ pour tout } x \in \mathbb{R} \end{array}$$

et

$$(\mathcal{H}_2) : \quad \det(I_p + xA + xB) = \det(I_p + xA)\det(I_p + xB) \text{ pour tout } x \in \mathbb{R}$$

Si l'on sait faire avec l'hypothèse $(\mathcal{H})$, on se débrouille aussi avec l'hypothèse $(\mathcal{H}_1)$: la démonstration est une adaptation de ce qui a été fait au §2.2, toujours en égalant les coefficients du 4^e degré dans les développements comme (3). En revanche, sous l'hypothèse $(\mathcal{H}_2)$, cela est plus difficile et long (J. OGAWA, 1950, *cf.* [9, 10]), nous n'en connaissons pas de démonstration simple... En effet, en suivant le cheminement de la démonstration du § 2, les variables x et y se « coagulent » et il devient difficile de conclure.

Références

[1] CHI KWONG-LI, A simple proof of the Craig-Sakamoto theorem, Linear Algebra and its Applications 321 (2000), 281-283

[2] M. F. DRISCOLL & X. R. GUNDBERG, Jr, A history of the development of Craig's theorem, The American Statistician, Vol. 40, N1 (1986), 65-70

[3] J.-B. HIRIART-URRUTY & C. LEMARECHAL, Convex Analysis and Minimization Algorithms II, Grundlehren der mathematischen Wissenschaften 306, Springer-Verlag (1993)

[4] H. O. LANCASTER, Traces and cumulants of quadratic forms in the normal variables, J. Roy. Statist. Soc. ser. B, 16 (1954), 247-254

[5] P. LASSÈRE & H. CARRIEU, One simple proof of the Craig-Sakamoto theorem, à paraître, Linear Algebra and its Applications

[6] G. LETAC & H. MASSAM, Craig-Sakamoto's theorem for the Wishart distributions on symmetric cones, Ann. Inst. Statist. Math. 47(1995), 785-799

[7] K. MATUSITA, Note on the independence of certain statistics, Ann. Inst. Statist. Math. 1 (1949), 79-82

[8] T. OGASAWARA & M. TAKAKASHI, Independence of quadratic quantities in a normal system, Journal of Sciences of the Hiroshima University, Vol. 15 (1951), 1-9

[9] J. OGAWA, On the independence of quadratic forms in a non central normal system, Osaka Math. Journal 2 (1950), 151-159

[10] J. OGAWA, A history of the development of Craig-Sakamoto's theorem viewed from Japanese viewpoint, Proc. Inst. Statist. Math. 21 (1993), 47-59 (en japonais)

[11] I. OLKIN, A determinantal proof of the Craig-Sakamoto theorem, Linear Algebra and its Applications 264 (1997), 217-223

[12] O. TAUSSKY, On a matrix theorem of A.T. Craig and H. Hotelling, Indagationes Mathematicae 20 (1958), 139-141