Développement: Déterminants de Gram et de Cauchy

Arthur Maritch-Roy

Dans tout ce développement, E désignera un espace préhilbertien (éventuellement hermitien quitte à remplacer la transposée par l'adjoint et à changer le produit scalaire).

Définition 1. Soit $(x_1, \dots, x_n)$ une famille de vecteurs de E. La *matrice de Gram* associé à cette famille est la matrice :

$$M = (\langle x_i, x_j \rangle)_{1 \leq i,j \leq n} \in \mathcal{M}_n(\mathbb{R}).$$

Son déterminant est appelé *déterminant de Gram* de la famille $(x_1, \dots, x_n)$.

Proposition 1. Les matrices de Gram de taille n sont exactement les matrices symétriques positives de taille n.

Démonstration. Si M est la matrice de Gram de $(x_1, \dots, x_n)$, alors $M = X^\top X$, où X est la matrice dont les colonnes sont les $(x_j)_j$, donc M est symétrique positive. Réciproquement, si A est symétrique positive, c'est la matrice de Gram de la famille des colonnes de la racine carrée de A. □

Proposition 2. Soit $(x_1, \dots, x_n)$ une famille de vecteurs de E. Elle est liée si et seulement si son déterminant de Gram est nul.

Démonstration. Si la famille est liée, on peut se ramener à une colonne de zéros en faisant des opérations élémentaires sur les colonnes par bilinéarité du produit scalaire. Réciproquement, si la matrice de Gram est de déterminant nul, alors ses colonnes sont liées, donc il existe des scalaires $(\lambda_2, \dots, \lambda_n)$ tels que :

$$\forall i \in [\![1, n]\!], \langle x_i, x_1 \rangle = \sum_{k=2}^{n} \lambda_k \langle x_i, x_k \rangle$$

soit par bilinéarité :

$$\forall i \in [\![1, n]\!], \left\langle x_i, x_1 - \sum_{k=2}^{n} \lambda_k x_k \right\rangle = 0.$$

Ainsi, $x_1 - \sum_{k=2}^{n} \lambda_k x_k$ appartient à la fois à $\mathrm{Vect}(x_1, \dots, x_n)$ et à son orthogonale, *i.e.* est nul, donc la famille est liée. □

Exemple 1. Si X est un vecteur aléatoire dans $\mathbb{R}^n$ dont chaque composante admet une variance, sa matrice de covariance correspond à la matrice de Gram de la famille $(X_1 - \mathbb{E}(X_1), \dots, X_n - \mathbb{E}(X_n))$ pour le produit scalaire de l'espérance. Le déterminant de Gram associé est appelé *variance généralisée* de X.

Théorème 3. Soit V un sous-espace vectoriel de E admettant une base $(x_1, \dots, x_n)$. Alors pour tout $u \in E$:

$$d(u, V)^2 = \frac{G(u, x_1, \dots, x_n)}{G(x_1, \dots, x_n)}$$

Démonstration. On sait que $d(u,V)^2 = \|x\|^2 - \|p_V(x)\|^2 = \|x - p_V(x)\|^2$ par Pythagore. Notons $y = p_V(u)$ et $z = x - y$. Pour tout i, on a $\langle x_i, y\rangle = \langle x_i, u\rangle$, ainsi, si on note M la matrice de Gram de la famille $(x_1, \ldots, x_n, u)$, on a :

$$M = \begin{pmatrix} \langle x_1, x_1\rangle & \ldots & \langle x_1, x_n\rangle & \langle x_1, y\rangle \\ \ldots & \ldots & \ldots & \ldots \\ \langle x_n, x_1\rangle & \ldots & \langle x_n, x_n\rangle & \langle x_n, y\rangle \\ \langle y, x_1\rangle & \ldots & \langle y, x_n\rangle & \|y\|^2 + \|z\|^2 \end{pmatrix}$$

Puis, par multlinéarité du déterminant :

$$\det(M) = \begin{vmatrix} \langle x_1, x_1\rangle & \ldots & \langle x_1, x_n\rangle & \langle x_1, y\rangle \\ \ldots & \ldots & \ldots & \ldots \\ \langle x_n, x_1\rangle & \ldots & \langle x_n, x_n\rangle & \langle x_n, y\rangle \\ \langle y, x_1\rangle & \ldots & \langle y, x_n\rangle & \|y\|^2 \end{vmatrix} + \begin{vmatrix} \langle x_1, x_1\rangle & \ldots & \langle x_1, x_n\rangle & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \langle x_n, x_1\rangle & \ldots & \langle x_n, x_n\rangle & 0 \\ \langle y, x_1\rangle & \ldots & \langle y, x_n\rangle & \|z\|^2 \end{vmatrix}$$

Or le premier déterminant vaut $G(x_1, \ldots, x_n, y) = 0$ car la famille en question est liée, et le deuxième vaut $\|z\|^2 G(x_1, \ldots, x_n)$, ce qui fournit le résultat. □

Exemple 2. Dans le cas d'un vecteur gaussien centré $(X = X_1, \ldots, X_n, Y) \in \mathrm{L}^2$, la projection correspondant à l'espérance conditionnelle (et celle-ci étant une combinaison linéaire des (X_i) car on est dans le monde gaussien), on a l'identité suivante :

$$\mathbb{E}(\mathrm{Var}(Y|X)) = \mathbb{E}((Y - \mathbb{E}(Y|X))^2) = \mathbb{E}((Y - \mathbb{E}(Y|X_1, \ldots, X_n))^2) = \frac{\mathrm{Var}(Y, X)}{\mathrm{Var}(X)}.$$

Théorème 4. (Inégalité de Hadamard)
Soit $(x_1, \ldots, x_n)$ une famille de vecteurs non nuls de E. Alors

$$G(x_1, \ldots, x_n) \leq \prod_{i=1}^{n} \|x_i\|^2,$$

avec égalité si et seulement si la famille en question est orthogonale.
Plus généralement,

$$|\det(x_1, \ldots, x_n)| \leq \prod_{i=1}^{n} \|x_i\|.$$

Démonstration. On peut toujours supposer que la famille est libre. Dans ce cas, on procède par récurrence. Si $n = 1$, le résultat est immédiat, on le suppose vrai pour un entier n quelconque. On considère alors une famille libre $(x_1, \ldots, x_n)$ qui engendre V et un vecteur x_{n+1} libre avec cette famille. Notons y la projection orthogonale de x_{n+1} sur V. D'après le résultat précédent, on a

$$G(x_1, \ldots, x_{n+1}) = G(x_1, \ldots, x_n)\|x - y\|^2 \leq G(x_1, \ldots, x_n)\|x_{n+1}\|^2,$$

d'où le résultat par hypothèse de récurrence. Le cas d'égalité à l'étape n nécessite d'avoir eu à chaque étape $\|x_{k+1} - y\|^2 = \|x_{k+1}\|^2$ *i.e.* x_{k+1} est orthogonal à tous les vecteurs d'avant *i.e.* la famille est orthogonale. Pour la généralisation, il suffit de se rappeler que l'on peut écrire la matrice de Gram comme $X^\top X$, où X est de déterminant $\det(x_1, \ldots, x_n)$. □

Exemple 3. Si on reprend l'exemple de notre vecteur aléatoire X, on obtient :

$$\mathrm{Var}(X) \leq \prod_{i=1}^{n} \mathrm{Var}(X_i),$$

avec égalité si et seulement si le vecteur est décorrélé, ou dans le cas gaussien si et seulement si le vecteur est indépendant.

Application 1. Montrer que l'expression $\int_0^1 (1 + a_1 + a_2x^2 + \ldots + a_nx^n)^2 \,\mathrm{d}x$ atteint son minimum μ en un unique n-uplet.

Démonstration. On considère donc E l'ensemble des fonctions continues sur $[0, 1]$ muni du produit scalaire canonique. Le problème revient donc à calculer la distance du vecteur 1 à la famille $(x, x^2, \ldots, x^n)$. On sait qu'elle est atteinte en un unique n-uplet et d'après le premier théorème,

$$\mu = \frac{G(1, x, \ldots, x^n)}{G(x, \ldots, x^n)}$$

Pour $1 \leq i, j \leq n$, $\langle x^i, x^j \rangle = \int_0^1 x^{i+j} \,\mathrm{d}x = \frac{1}{i+j+1}$. Les déterminants à calculer sont donc des déterminants de Cauchy (*cf.* fait suivant) :

$$\mu = \frac{\prod_{1 \leq i<j \leq n+1} (i-j)^2}{\prod_{1 \leq i,j \leq n+1} (i+j-1)} \frac{\prod_{1 \leq i,j \leq n} (i+j+1)}{\prod_{1 \leq i<j \leq n} (i-j)^2},$$

qu'on peut simplifier pour obtenir $\mu = \frac{1}{(n+1)^2}$. □

Fait 1. Pour des complexes $(a_j)_j, (b_j)_j$ tels que $a_i + b_j \neq 0$ pour tout (i, j),

$$\Delta_n := \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_n} \\ \vdots & \vdots & \vdots \\ \frac{1}{a_1+b_n} & \cdots & \frac{1}{a_n+b_n} \end{vmatrix} = \frac{\prod_{i<j} (a_j - a_i) \prod_{i<j} (b_j - b_i)}{\prod_{1 \leq i,j \leq n} (a_i + b_j)}.$$

Démonstration. On suppose que les $(a_i)_i$ sont distincts deux à deux (sinon le déterminant est nul et la formule est vraie). On considère R la fraction rationnelle

$$R(X) = \frac{(b_1 - X) \ldots (b_{n-1} - X)}{(a_1 + X) \ldots (a_n + X)},$$

à laquelle on peut appliquer le théorème de décomposition en éléments simples. Il existe donc des scalaires $\lambda_1, \ldots, \lambda_n$ tels que

$$R(X) = \frac{\lambda_1}{X + a_1} + \ldots + \frac{\lambda_n}{X + a_n}.$$

Par la méthode des résidus, on obtient :

$$\lambda_k = \frac{\prod_i^{n-1} (b_i + a_k)}{\prod_{i \neq k} (a_i - a_k)} \neq 0.$$

Ainsi, si on note $(L_i)_i$ les lignes de la matrice dont on veut le déterminant, on a :

$$\Delta_n = \frac{1}{\lambda_n} \begin{vmatrix} L_1 \\ \vdots \\ \sum_{i=1}^n \lambda_i L_i \end{vmatrix} = \frac{1}{\lambda_n} \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_{n-1}} & \frac{1}{a_1+b_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{a_1+b_n} & \cdots & \frac{1}{a_{n-1}+b_{n-1}} & \frac{1}{a_{n-1}+b_n} \\ R(b_1) & \ldots & R(b_{n-1}) & R(b_n) \end{vmatrix}$$

Puisque $R(b_k) = 0$ dès que $k \neq n$, on peut simplement développer par rapport à la dernière ligne et avoir :

$$\Delta_n = \frac{R(b_n)}{\lambda_n} \Delta_{n-1}$$

soit :

$$\Delta_n = \frac{\prod_{i=1}^{n-1} (b_i - b_n)}{\prod_{i=1}^{n} (b_n + a_i)} \frac{\prod_{i=1}^{n-1} (a_i - a_n)}{\prod_{i=1}^{n-1} (b_i + a_n)} \Delta_{n-1}.$$

Cela donne le résultat par récurrence. □

Référence. X.Gourdon, *Analyse*. Pour la partie probabilités, j'ai écrit ça rapidement et c'est donc peut-être imprécis.