SUM OF MIXING DETERMINANTS[1]

Nguyen Bac-Van
College of Natural Sciences
(Received on July 3, 1999)

ABSTRACT: *Consider some square matrices of the same order. We can withdraw a given number of rows of the first matrix, a given number of rows of the second, etc., and form a new square matrix where the rows conserve their original ordinals. The sum of determinants of all possible mixing matrices constitutes the introduced notion. The main theorem is that if the starting matrices are non-negative definite (n.n.d.) then the above sum is non-negative, a necessary and sufficient condition for its positiveness is given. Applications: measuring the steepness in multidimensional geometry, majorizing the error norm of least squares estimates.*
Keywords: kernel, cardinality, mixing determinants, least squares.
1991 Mathematical Subject Classification: 15A03

1 Introduction and summary

Investigating the matrix expression of the error of least squares estimate has led the author to the following notion.

Definition 1 *Let $\mathbf{A}_1, \ldots, \mathbf{A}_n$ be $m \times m$ matrices, $m, n \geq 1$. Consider non-negative integers $m_1, \ldots, m_n$ such that $m_1 + \cdots + m_n = m$. We define*

$$\mathbf{A}_1(m_1) \circ \cdots \circ \mathbf{A}_n(m_n) = \textit{sum of } \frac{m!}{m_1! \cdots m_n!} \textit{ determinants},$$

each determinant corresponds to a partition $\{D_1, \ldots, D_n\}$ of the set $\{1, \ldots, m\}$ with $\sharp D_i = m_i$, $i = 1, \ldots, n$, and has the form $\det \begin{pmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m \end{pmatrix}$, *where $\mathbf{r}_j$ $(j = 1, \ldots, m)$ is the jth row of $\mathbf{A}_i$ when $j \in D_i$.*

When $n = m$, $m_1 = \cdots = m_m = 1$, we shall simply write

$$\mathbf{A}_1(1) \circ \cdots \circ \mathbf{A}_m(1) = \mathbf{A}_1 \circ \cdots \circ \mathbf{A}_m \; .$$

The sum defined above is called *sum of mixing determinants.* This paper aims to present a property of such sums which, to the author's knowledge, does not figure in the existing literature, namely

Theorem 1 *Let $\mathbf{B}_1, \ldots, \mathbf{B}_n$ be $m \times m$ non-negative definite (n.n.d.) matrices and $m_1, \ldots, m_n$ non-negative integers, $m_1 + \cdots + m_n = m$, $m, n \geq 1$. Then $\mathbf{B}_1(m_1) \circ \cdots \circ \mathbf{B}_n(m_n)$ is non-negative; it vanishes if and only if*

$$(\exists \varphi \subset \{1, \ldots, n\}, \varphi \neq \emptyset) \text{ Rank } (\cdots \mathbf{B}_i \cdots)_{i \in \varphi} \leq \sum_{i \in \varphi} m_i - 1, \tag{1}$$

[1]Ce projet est réalisé avec l'appui du Fonds international de coopération universitaire–FICU (AUPELF–UREF, Agence francophone pour l'enseignement supérieur et la recherche).

where the symbol

$$(\cdots \mathbf{B}_i \cdots)_{i\in\varphi}$$

denotes a block matrix formed with blocks $\mathbf{B}_i$, *the index* i *varying over the subset* φ *from left to right.*

The proof is given in Section 2. Applications are given in Sections 3 and 4.

Notations. The following will be used in the paper.
$\mathcal{M}_{p\times q}$ = linear space of all $p \times q$ real matrices.
$\sharp\varphi$ = cardinality of the set φ.
If Q is a set in a linear space F: $\operatorname{Span} Q$ = linear hull of the set Q in F, and for F Euclidean, $Q^\perp$ is the orthogonal complement of the set Q in F.
For any $p \times p$ matrix $\mathbf{C} = (c_{ij})$ and any set $\sigma \subset \{1, \ldots, p\}$:
$\mathbf{C}(\sigma) = (c_{ij})_{i,j\in\sigma}$.

2 Proof of Theorem 1

Lemma 1 *G and H being arbitrary finite-dimensional subspaces of some linear space E, then*

i. $\dim \operatorname{Span}(G \cup H) = \dim G + \dim H - \dim(G \cap H)$.

ii. If, moreover, $\dim E = m$ *finite then* $\dim(G \cap H) = \dim G + \dim H - m$ *if and only if* $\operatorname{Span}(G \cup H) = E$.

(i) is from [2] (p. 453, formula (1)).

The "if" assertion of Theorem 1 is trivial. In the particular case $n = m$, $m_1 = \cdots = m_n = 1$. Theorem 1 is reduced to the following

Lemma 2 *Let* $\mathbf{A}_1, \ldots, \mathbf{A}_m$ *be* $m \times m$ *n.n.d. matrices. If*

$$\operatorname{Rank}(\cdots \mathbf{A}_i' \cdots)_{i\in\psi} \geq \sharp\psi \quad \forall\psi \subset \{1, \ldots, m\},\ \psi \neq \emptyset \tag{2}$$

then $\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_m > 0$.

Proof. From the requirement we have $\operatorname{Rank} \mathbf{A}_i \geq 1$, $i = 1, \ldots, m$. Consider the spectral decomposition

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{P}\Lambda\mathbf{P}' \text{ , } \mathbf{P} = (\mathbf{P}_1 \cdots \mathbf{P}_m) \text{ orthogonal,} \\ \Lambda &= \operatorname{diag}(\lambda_1, \ldots, \lambda_m) \text{, } \lambda_1 \geq \cdots \geq \lambda_\mu > 0, \end{aligned}$$

where $\mu = \operatorname{Rank} \mathbf{A}_1 \geq 1$. We can check that for $m \geq 2$

$$\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_m = \sum_{j=1}^{\mu} \lambda_j \mathbf{Q}_2(\sigma_j) \circ \cdots \circ \mathbf{Q}_m(\sigma_j), \tag{3}$$

where $\mathbf{Q}_i(\sigma_j)$ is the $(m-1) \times (m-1)$ principal submatrix of $\mathbf{Q}_i = \mathbf{P}'\mathbf{A}_i\mathbf{P}$ corresponding to $\sigma_j = \{1, \ldots, m\} - \{j\}$. In general we have

$$\mathbf{Q}_i(\sigma) = (\cdots \mathbf{P}_f \cdots)'_{f\in\sigma} \mathbf{A}_i (\cdots \mathbf{P}_f \cdots)_{f\in\sigma} \text{ for } \sigma \subset \{1, \ldots, m\}.$$

Since $\mathbf{Q}_i$ is n.n.d., so is $\mathbf{Q}_i(\sigma)$. Further let φ denote any subset of $\{2,\ldots,m\}$ and let $K_i = \ker \mathbf{A}_i$. Then

$$\begin{aligned}\dim\ker\begin{pmatrix}\vdots\\ \mathbf{Q}_i(\sigma)\\ \vdots\end{pmatrix}_{i\in\varphi} &= \dim\ker\begin{pmatrix}\vdots\\ \mathbf{A}_i\\ \vdots\end{pmatrix}_{i\in\varphi}(\cdots\mathbf{P}_f\cdots)_{f\in\sigma}\\ &= \dim\left(\cap_{i\in\varphi}K_i\right)\bigcap \operatorname{Span}\{\mathbf{P}_f, f\in\sigma\}.\end{aligned}$$

Hence from Lemma 1 (ii), the index f ranging $1,\ldots,m$, we get

$$\dim\ker\begin{pmatrix}\vdots\\ \mathbf{Q}_i(\sigma_j)\\ \vdots\end{pmatrix}_{i\in\varphi}\begin{cases} = \dim\cap_{i\in\varphi}K_i\\ \quad\text{if } \cap_\varphi K_i\subset \operatorname{Span}\{\mathbf{P}_f, f\neq j\},\\ = \dim\cap_\varphi K_i - 1 \text{ otherwise.}\end{cases} \tag{4}$$

Let us reason by induction. We make the following, equivalent to (2),

Assumption: $\mathbf{A}_1,\ldots,\mathbf{A}_m$ are $m\times m$ n.n.d. matrices satisfying the condition

$$\dim\cap_{i\in\psi}K_i\leq m-\sharp\psi\quad \forall\psi\subset\{1,\ldots,m\},\ \psi\neq\emptyset. \tag{5}$$

Let $m\geq 2$. We admit the following

Induction hypothesis: For arbitrary $(m-1)\times(m-1)$ n.n.d. matrices $\mathbf{B}_1,\ldots,\mathbf{B}_{m-1}$ there holds

$$\mathbf{B}_1\circ\cdots\circ\mathbf{B}_{m-1}>0 \text{ if } \dim\ker(\cdots\mathbf{B}_i'\cdots)_{i\in\psi}'\leq m-1-\sharp\psi$$

for every non-void set $\psi\subset\{1,\ldots,m-1\}$.

Due to formula (3), this hypothesis has three consequences:

if $\mathbf{B}_1,\ldots,\mathbf{B}_{m-1}$ are $(m-1)\times(m-1)$ n.n.d. matrices
then always $\mathbf{B}_1\circ\cdots\circ\mathbf{B}_{m-1}\geq 0,$ (6)
whereas $\mathbf{B}_1\circ\cdots\circ\mathbf{B}_{m-1}=0$ if and only if
$(\exists\psi\subset\{1,\ldots,m-1\},\psi\neq\emptyset)\ \dim\ker(\cdots\mathbf{B}_i'\cdots)_{i\in\psi}'\geq m-\sharp\psi;$ (7)
$\mathbf{A}_1\circ\cdots\circ\mathbf{A}_m\geq 0.$ (8)

Under Assumption (5) we intend to prove $\mathbf{A}_1\circ\cdots\circ\mathbf{A}_m>0$. In view of (8) we start from the converse

$$\textbf{Supposition}: \mathbf{A}_1\circ\cdots\circ\mathbf{A}_m=0. \tag{9}$$

On account of Assumption (5), Supposition (9) is equivalent to

$$\begin{aligned}&(\forall j=1,\ldots,\mu),(\exists\varphi\subset\{2,\ldots,m\},\varphi\neq\emptyset),\\ &\qquad\cap_\varphi K_i\subset\operatorname{Span}\{\mathbf{P}_f, f\neq j\} \text{ and } \dim\cap_\varphi K_i=m-\sharp\varphi.\end{aligned}\tag{10}$$

For short reference we shall say that

a subset φ of $\{2, \ldots, m\}$ suits some integer j, $1 \leq j \leq \mu$, if simultaneously

$$\varphi \neq \emptyset, \quad \cap_{\varphi} K_i \subset \text{Span}\{\mathbf{P}_f, f \neq j\} \text{ and } \dim \cap_{\varphi} K_i = m - \sharp\varphi. \tag{11}$$

Then note the following properties.

(a) If φ suits j and φ' suits $j' \neq j$ then $\varphi \cup \varphi'$ suits both j and j'.

(b) If there exists a set φ that suits a certain $j \leq \mu = \text{Rank}\,\mathbf{A}_1$, then necessarily $\mu \geq 2$ and $\cap_{\varphi} K_i$ is not included in K_1 .

(c) If φ suits j then there exists another j' such that every set φ', if any, suiting j' is not included in φ.

Now, by (10) and (11) Supposition (9) is equivalent to

$$(\forall j = 1, \ldots, \mu), (\exists \varphi \subset \{2, \ldots, m\}), \ \varphi \text{ suits } j \ . \tag{12}$$

For $\mu = 1$, from (b), (12) is impossible. Let (12) be true for some $\mu \geq 2$. Consider some $j \leq \mu$ and let φ be a set of maximal cardinality that suits j. By (12) for any $j' \leq \mu$ there exists φ' suiting j' and from (c) there exists $j' \neq j$ such that $\sharp\varphi' \cup \varphi > \sharp\varphi$, from (a) $\varphi' \cup \varphi$ suits j, which contradicts the maximality of $\sharp\varphi$. Therefore (12) is always impossible. Thus the induction hypothesis and Assumption (5) entail that $\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_m > 0$. On the other hand for $m = 2$ the induction hypothesis is trivially true. So Lemma 2 is proved.

Let us now prove Theorem 1 in the general case. Without loss of generality, we shall assume that $m_1, \ldots, m_n$ are positive. Put $l_0 = 0, \ldots, l_i = m_1 + \cdots + m_i, \ldots, l_n = m$. Consider the partition $\{1, \ldots, m\} = \cup_{i=1}^{n} \Delta_i$, where $\Delta_i = \{l_{i-1} + 1, \ldots, l_i\}$. To every set φ in $\{1, \ldots, n\}$ there corresponds a subset, called Δ-set, of $\{1, \ldots, m\}$, of the form $\Delta = \cup_{i \in \varphi} \Delta_i$, where $\Delta = \emptyset$ if and only if $\varphi = \emptyset$. Since an intersection of Δ-sets is again a Δ-set, to every set ψ in $\{1, \ldots, m\}$ there corresponds a minimal Δ-set containing ψ. From $\mathbf{B}_1, \ldots, \mathbf{B}_n$ let us generate n.n.d. matrices $\mathbf{A}_1, \ldots, \mathbf{A}_m$ by putting $\mathbf{A}_f = \mathbf{B}_i$ for all $f \in \Delta_i$, $i = 1, \ldots, n$. Then we can show that

$$m_1! \ldots m_n! \, \mathbf{B}_1(m_1) \circ \cdots \circ \mathbf{B}_n(m_n) = \mathbf{A}_1 \circ \cdots \circ \mathbf{A}_m \ ,$$

and Theorem 1 follows from Lemma 2. Q.E.D.

3 A multidimensional geometric characteristic

Theorem 1 enables us to establish a geometric fact in multidimensional vector spaces.

Theorem 2 *Let* $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$, $m \geq 1$, *be a basis for a subspace* Φ *of* R^n *and* $\{\mathbf{f}_{1K}, \ldots, \mathbf{f}_{mK}\}$ *its orthogonal projection on some subspace* K *of* R^n *according to the inner product* $\mathbf{u}'\mathbf{v}$ *for* $\mathbf{u}, \mathbf{v} \in R^n$. *Then the ratio*

$$\sigma = 1 - \frac{\det (\mathbf{f}_{1K} \cdots \mathbf{f}_{mK})' (\mathbf{f}_{1K} \cdots \mathbf{f}_{mK})}{\det (\mathbf{f}_1 \cdots \mathbf{f}_m)' (\mathbf{f}_1 \cdots \mathbf{f}_m)}$$

is independent of the choice of basis for Φ. *Moreover* $0 \leq \sigma \leq 1$, *and* $\sigma = 0$ *if and only if* $\Phi \subset K$. $\sigma = 1$ *if and only if* Φ *contains some non-null vector orthogonal to* K.

Proof. It is easily seen that σ is independent of the choice of basis.

Write $\mathbf{f}_i = \mathbf{f}_{iK} + \mathbf{t}_{iK}$, $\mathbf{f}_{iK} \in K$, $\mathbf{t}_{iK} \perp K$, $i = 1, \ldots, m$. Then

$$(\mathbf{f}_1 \cdots \mathbf{f}_m)' (\mathbf{f}_1 \cdots \mathbf{f}_m) = (\mathbf{f}'_{iK}\mathbf{f}_{jK} + \mathbf{t}'_{iK}\mathbf{t}_{jK}),\ i, j = 1, \ldots, m.$$

Consider n.n.d. matrices $\mathbf{B} = (\mathbf{f}'_{iK}\mathbf{f}_{jK})$ and $\mathbf{C} = (\mathbf{t}'_{iK}\mathbf{t}_{jK})$. We can check that

$$\det(\mathbf{f}_1 \cdots \mathbf{f}_m)' (\mathbf{f}_1 \cdots \mathbf{f}_m) = \det(\mathbf{B} + \mathbf{C}) = \det \mathbf{B} + \sum \mathbf{B}(r) \circ \mathbf{C}(s),$$

$r = 0, \ldots, m-1$, $r + s = m$. By Theorem 1, $\mathbf{B}(r) \circ \mathbf{C}(s) \geq 0$ hence $0 \leq \sigma \leq 1$, moreover, when $\det \mathbf{B} > 0$ then $\mathbf{B}(r) \circ \mathbf{C}(s) = 0$ if and only if $\operatorname{Rank} \mathbf{C} \leq s - 1$. The assertions of Theorem 2 follows. Q.E.D.

Further, $\dim K = n - \dim K^{\perp}$ entails $\dim\left(\Phi \cap K^{\perp}\right) \geq \dim \Phi - \dim K$ by Lemma 1 (ii). Hence if $\dim \Phi \leq \dim K$, σ can in fact vary from zero to one, and then σ can be used as *a measure of steepness* of the subspace Φ relatively to the subspace K in R^n.

4 Statistical application

In [1] generalized least squares (GLS) estimates (GLSE) were defined and can now be expressed more generally as follows. Consider a system of k models

$$\mathbf{Y}'_{ij} = \mathbf{b}'_{ij}\theta_i + \eta'_{ij},\ i = 1, \ldots, k,\ j = 1, \ldots, n_i,$$

the ith model contains n_i items, the $\mathbf{Y}_{ij}$ are $r \times 1$ response vectors, the $\ell(i) \times 1$ vectors $\mathbf{b}_{ij}$ are regressors, random or not, the θ_i unknown $\ell(i) \times r$ matrix parameters, the η_{ij} residuals and a prime denotes transpose. The global matrix parameter is

$$\theta = (\theta'_1 \cdots \theta'_k)'.$$

Put $\ell = \ell(1) + \cdots + \ell(k)$, then the range space of θ is $\mathcal{M}_{\ell \times r}$. Arbitrary constraints may be imposed on the parameter θ, which means its range Θ may have a quite arbitrary shape.

We call GLS value any value $\hat{\theta}$ which minimizes some norm of the global residual

$$\eta = \left(\eta'_{11} \cdots \eta'_{1n_1} \vdots \cdots \vdots \eta'_{k1} \cdots \eta'_{kn_k}\right)',$$

as in $\mathcal{M}_{\ell \times r}$ the conceptual variable θ varies over an affine manifold F containing the set Θ, if unique $\hat{\theta}$ is called the GLSE for the regression parameter θ.

The notion of sum of mixing determinants proves to be a useful tool for investigating the complex matrix expression of the error $\hat{\theta} - \theta$, namely it enables us in definite conditions to majorize the Euclidean norm $\|\hat{\theta} - \theta\|$ by $\|D\gamma\|$, where D is some positive constant independent of the data as well as of the sizes $n_1, \ldots, n_k$ and γ is a linear vector function of the global residual η. This error norm evaluation serves for proving the strong consistency of the GLSE.

TỔNG ĐỊNH THỨC HỖN HỢP

Nguyễn Bác-Văn

TÓM TẮT : Xét một số ma trận vuông cùng cấp. Rút một số cho trước các hàng của ma trận thứ nhất, một số cho trước các hàng của ma trận thứ hai, v.v., rồi tạo một ma trận vuông mới với các hàng giữ nguyên số thứ tự gốc của chúng. Chúng tôi đưa ra khái niệm tổng định thức của của mọi ma trận hỗn hợp như vậy. Định lý chính là: nếu các ma trận xuất phát xác định không âm thì tổng nói trên sẽ không âm, điều kiện cần và đủ cho tổng dương được chỉ ra. Áp dụng : đo độ dốc trong hình học nhiều chiều, đánh giá chuẩn sai của ước lượng bình phương bé nhất.

REFERENCES

[1] N: Bac-Van, *Strong consistency of least squares estimates in polygonal regression with random explanatory variables*, Acta Math. Vietnam. 19(1) (1994), 111-137.

[2] L. Chambadal and J.L. Ovaert, *Algèbre linéaire et algèbre tensorielle*, Dunod, Paris, 1968.