ACTA MATHEMATICA VIETNAMICA
Volume 31, Number 3, 2006, pp. 205-217

AN OPTIMIZATION PROBLEM IN MATRIX DATA DEPENDENCE ANALYSIS

NGUYEN BAC VAN

Abstract. The paper treats the problem of stepwise maximizing a sum of squared bilinear forms which stems from the dependence analysis of several data matrices. A constructive approach is presented which leads to clear-cut properties of the maximization process.

1. Introduction

The problem originates from the works of Lafosse and Hanafi (1997, 2001). Let us consider some real data matrix X of order $n \times p$. Its rows are viewed as multidimensional observations on n individuals, and each column is formed by observed values of some variable. The set of individuals is given, so every considered data matrix will be of n rows.

Let M be a $p \times p$ positive definite (p.d.) constant matrix. For every unit vector u in (R^p, M), i.e. $u'Mu = 1$, we shall call XMu the compoment of X on the axis u. XMu is also a linear combination of columns of X.
Let the individuals have positive weights $d_1, \dots, d_n$ with $\sum_{i=1}^n d_i = 1$. Assume the weighted mean of every column is zero, that will be expressed below by saying that every data matrix is centered.

Put $D = \operatorname{diag}(d_1, \dots, d_n)$. Then two variables whose n observed values are respectively $x_1, \dots, x_n$ and $y_1, \dots, y_n$ have the empirical covariance

$$(x_1 \cdots x_n) D (y_1 \cdots y_n)'.$$

Viewing the $n \times p$ data matrix X as the empirical realization of p variables, the dispersion matrix of this set X of p columns is then $X'DX$. The covariance matrix between two data matrices $X_{n\times p}$ and $Y_{n\times q}$ is $X'DY$. The variance of the above-mentioned component XMu is $u'MX'DXMu$. In the case rank $X = p$, if the Mahalanobis (1936) distance is used, i.e. if one chooses $M = (X'DX)^{-1}$, then $\operatorname{Var}(XMu) = 1$ for all axes u.

The classical Principal Component Analysis (PCA) starts from stepwise maximizing the variance $u'MX'DXMu$ when investigating a single data matrix X,

Received October 27, 2003; in revised form June 30, 2006.
1991 *Mathematics Subject Classification.* Primary 62H 25, Secondary 15A18.
Key words and phrases. Dependence, covariance, projector, singular value.
This work was supported by the Natural Science Council of Viet Nam.

where p.d. M is given. The solution u at any maximization step is constrained by the condition that it should be M-orthogonal to the solutions at all previous steps.

Generalizing the idea of PCA, Lafosse and Hanafi (1997), Hanafi and Lafosse (2001) considered real data matrices X_i $(i = 1, \ldots, l)$ and Y of order $n \times p_i$ and $n \times q$, respectively. They set the problem of discovering the dependence between Y and the collection $\{X_1, \ldots, X_l\}$ by stepwise maximizing the sum of squared covariances between components $X_i M_i u_i$ and YNb, where $u_i \in (R^{p_i}, M_i)$ and $b \in (R^q, N)$ are unit vectors, M_i and N being given p.d. matrices. At every maximization step, the vector u_i $(i = 1, \ldots, l)$ and b are to be sought so as to maximize the sum

$$\sum_{i=1}^{l} (u_i' M_i X_i' D Y N b)^2 = \sum_{i=1}^{l} \mathrm{cov}^2(YNb, X_i M_i u_i) \tag{1.1}$$

subject to the constraint that u_i and b respectively have to be M_i-orthogonal and N-orthogonal to u_i-solutions and b-solutions at all previous steps.

The arbitrary choice of p.d. matrices M_i and N gives full generality to the expression of dependence. However, by assuming linear independence of the columns of Y as well as of those of X_i, if the Mahalanobis distance is to be used, i.e. if one takes $N = (Y'DY)^{-1}$ and $M_i = (X_i' D X_i)^{-1}$, then $u_i' M_i X_i' D Y N b$ is just the correlation coefficient between the components $X_i M_i u_i$ and YNb. In this case the expression to be maximized (1.1) is the sum of squared correlation coefficients.

With a view to comparison, we note that when defining a measure of fit between two centered data matrices $X_{n\times p}$ and $Y_{n \times p}$ Lingoes and Schönemann (1974) tried to maximize the trace of $T'Y'X$ over the set of all $p \times p$ orthogonal matrices T. By choosing $D = \mathrm{diag}(1/n, \ldots, 1/n)$, computation gives

$$n^{-1} \mathrm{Tr}(T'Y'X) = \sum_{k=1}^{p} \mathrm{cov}(Xu_k, (YT)u_k),$$

where Xu_k and YTu_k are respectively the components of X and YT along the kth axis u_k in R^p. Thus an important step in the construction of the measure of fit by Lingoes and Schönemann is *to maximize a sum of covariances* of components of the data matrices along different axes.

We also note that in generalized canonical analysis, see Carroll (1968), Kiers et al. (1994), X_k denoting the $n \times p_k$ data matrix for set k, $k = 1, \ldots, K$, and a_k a vector of weights to form a canonical variate $X_k a_k$ which is the component of X_k on the axis a_k, a consensus variable z is to be found so as *to maximize the weighted sum of squared correlations*

$$\sum_{k=1}^{K} w_k r^2(z, X_k a_k)$$

over $a_1, \dots, a_k$ and z, where $r(\cdot, \cdot)$ denotes the correlation between the variables in parentheses, and w_k denotes a fixed (nonnegative) weight for set k.

The method of dependence analysis, called concordance analysis, introduced by Lafosse (1997) and developped by Lafosse and Hanafi (1997), then by Hanafi and Lafosse (2001), has practical applications. It is time to review the mathematical context.

The above consideration concerns empirical distributions. We shall now restate the same consideration in a general form. Consider $l+1$ collections $\xi_1, \dots, \xi_l, \eta$ consisting respectively of $p_1, \dots, p_l, q$ real-valued random variables (r.v.). We shall write $\xi_1, \dots, \xi_l, \eta$ as column vectors.

When $l = 0$, let B be the dispersion matrix of η. It is known that given a $q \times q$ p.d. matrix N, by stepwise maximizing the function of a q-vector u

$$u'NBNu = \operatorname{var}(u'N\eta)$$

over u, $u'Nu = 1$, under the constraints that at the kth step u has to be N-orthogonal to the solutions $u_1, \dots, u_{k-1}$ at $k-1$ previous steps, we arrive at an N-orthonormal (o.n.) system $\{u_1, \dots, u_m\}$, where $m = \operatorname{rank} B$. The r.v'.s $u_1'N\eta, \dots, u_m'N\eta$ are just the principal components of η.

When $l \geqslant 1$, let A_i be the covariance matrix between ξ_i and η, $i = 1, \dots, l$. Let M_i and N be respectively $p_i \times p_i$ and $q \times q$ p.d. matrices. Then, generalizing the idea of principal components, we set the problem of stepwise maximizing the sum of squared covariances

$$\sum_{i=1}^{l} (e_i' M_i A_i N b)^2 = \sum_{i=1}^{l} \operatorname{cov}^2(b'N\eta, e_i' M_i \xi_i)$$

over e_i and b, $e_i' M_i e_i = 1$, $b'Nb = 1$. With no constraint at the first step, the constraint at the kth ($k > 1$) maximizing step is that the vectors e_i and b respectively have to be M_i-orthogonal and N-orthogonal to the solutions $e_i(j)$ and $b(j)$ at every previous jth step. In mathematical form, the problem to which the paper is devoted is as follows, the corresponding notations are to be kept in mind.

Given l real matrices A_i of order $p_i \times q$ and p.d. M_i and N, we shall construct unit vectors e_i and b, i.e. $e_i' M_i e_i = 1$ and $b'Nb = 1$, so that the function

$$L(e_1, \dots, e_l, b) = \sum_{i=1}^{l} (e_i' M_i A_i N b)^2 \tag{1.2}$$

will be maximized stepwise. At each maximizing step additional constraints mean the solutions e_i and b have to be respectively M_i-orthogonal and N-orthogonal to the corresponding solutions at all previous steps.

The aim of this paper is to present a general constructive approach to the above problem, which enables us to give a thorough discussion about the solution, to establish clear-cut properties of the maximization process, and to clarify the geometrical meaning of the assertions.

The main results of the paper are contained in Sections 3 and 4. In Section 3 we introduce a consistent construction which gives the first o.n. systems in all cases, whereas the construction given by Hanafi and Lafosse (2001) did not. Section 4 is devoted to establish an accurate formula for the cardinality of these o.n. systems which, together with others in Section 5, constitute the crucial tool in dependence analysis. Notice that Hanafi and Lafosse (2001) gave no precise indication about this cardinality.

2. Two lemmas on maximization

We shall make use of the Cauchy-Schwarz inequality: for real column vectors X, Y and a p.d. matrix M

$$(X'MY)^2 \leqslant (X'MX)(Y'MY),$$

the equality is attained if and only if X and Y are colinear.
Let us start from the matrices A_i of size $p_i \times q$, p.d. M_i of size $p_i \times p_i$, and N of size $q \times q$, $i = 1, \dots, l$. Put

$$p = \sum_{i=1}^{l} p_i,\ A = (A'_1 \cdots A'_l)',\ M = \operatorname{diag}(M_1, \dots, M_l).$$

The null matrix of any order will be denoted by the symbol 0. The notation (R^p, M) means the space R^p endowed with the inner product $u'Mv$, $u, v \in R^p$. The induced norm is denoted by $\|\cdot\|_M$. Note the formula:

$$\|x\|_M^2 = \sum_{i=1}^{l} \|x_i\|_{M_i}^2 \tag{2.1}$$

for $x = (x'_1 \cdots x'_l)' \in R^p,\ x_i \in R^{p_i}$.

Lemma 2.1. *For $e_i \in R^{p_i},\ e'_i M_i e_i = 1,\ i = 1, \dots, l$, and $b \in R^q$, we have*

$$\sum_{i=1}^{l} (e'_i M_i A_i N b)^2 \leqslant b' N A' M A N b, \tag{2.2}$$

where the right hand side is just $\|ANb\|_M^2$. For $ANb \neq 0$, the equality is attained if and only if

$$\frac{ANb}{\|ANb\|_M} = \begin{pmatrix} \mu_{11} e_1 \\ \cdots \\ \mu_{l1} e_l \end{pmatrix},$$

where $\mu_{11}, \dots, \mu_{l1}$ are norming coefficients, i.e. $\sum_{i=1}^{l} \mu_{i1}^2 = 1$.

Proof. Put $\lambda = \|ANb\|_M \geqslant 0$, $ANb = \lambda c$, with $c \in R^p$, $c'Mc = 1$. Then, writing $c = (c'_1 \cdots c'_l)'$ with $\lambda c_i = A_i N b$, we have

$$\sum_{i=1}^{l} (e'_i M_i A_i N b)^2 = \lambda^2 \sum_{1}^{l} (e'_i M_i c_i)^2.$$

By Cauchy-Schwarz inequality $(e_i'M_ic_i)^2 \leqslant (e_i'M_ie_i)(c_i'M_ic_i)$ with equality if and only if $c_i = \mu_{i1}e_i$, where μ_{i1} is some number. From (2.1) we get

$$\sum_{i=1}^{l} c_i'M_ic_i = (c'Mc)^2 = 1,$$

hence

$$\sum_{i}^{l}(e_i'M_iA_iNb)^2 \leqslant \lambda^2,$$

the equality is attained if and only if

$$c = \begin{pmatrix} \mu_{11}e_1 \\ \cdots \\ \mu_{l1}e_l \end{pmatrix},$$

where $\sum_{i=1}^{l}\mu_{i1}^2 = 1$ by (2.1). □

The following definition is very important for our subject. Let B be any $p\times q$ real matrix, S and T be respectively $p\times p$ and $q\times q$ p.d. ones, S and T being given once for all to specify norms in R^p and R^q respectively.

Definition 2.1. A positive number λ is called a singular value of a nonnull matrix B if and only if λ^2 is an eigenvalue of BTB' (resp. $B'SB$). The latent subspace of BTB' (resp. $B'SB$) induced by λ^2 is called the left (resp. right) singular subspace of B corresponding to the singular value λ. Any unit vector in the left (resp. right) singular subspace endowed with the norm $\|\cdot\|_S$ (resp. $\|\cdot\|_T$) is called a left (resp. right) singular vector corresponding to the singular value λ.

If $B = 0$, any singular value is zero. In this case, the right singular subspace corresponding to the singular value zero is R^q.

Lemma 2.2. *Let λ_1 be the greatest singular value of the $p\times q$ matrix B. Then*

$$\|BTb\|_S^2 \leqslant \lambda_1^2 \quad (\forall b\in R^q,\ b'Tb = 1).$$

If $B\neq 0$ the equality is attained if and only if b is a right singular vector of B corresponding to λ_1.

Proof. Let $C = B'SB$ and $C = P\Phi P'$ be the spectral decomposition, $\Phi = \operatorname{diag}(\varphi_1,\ldots,\varphi_q)$, $\varphi_1 \geqslant \cdots \geqslant \varphi_q \geqslant 0$, $P'TP = TPP' = PP'T = I_q$.
Put $P'Tb = c = (c_1\cdots c_q)'$. Then $c'c = b'Tb = 1$, and

$$\|BTb\|_S^2 = b'TCTb = c'\Phi c = \sum_{i=1}^{q}\varphi_i c_i^2 \leqslant \varphi_1,$$

which proves the inequality. The equality occurs if and only if the following equivalent conditions are fulfilled:

$$(\|BTb\|_S^2 = \varphi_1) \Longleftrightarrow (\Phi c = \varphi_1 c) \Longleftrightarrow$$
$$\Longleftrightarrow (P\Phi c = \varphi_1 Pc = \varphi_1 PP'Tb = \varphi_1 b) \Longleftrightarrow (P\Phi P'Tb = \varphi_1 b) \Longleftrightarrow$$

$\Longleftrightarrow (CTb = \varphi_1 b) \Longleftrightarrow b$ belongs to the right singular subspace of B corresponding to the greatest singular value $\lambda_1 = \varphi_1^{1/2}$. □

3. Construction of the first orthonormal systems

This section presents an iterative construction to generate an o.n. system in (R^{p_i}, M_i), $i = 1, \dots, l$. From now on, the notation $\mathcal{M}(T)$ will denote the linear hull of a set T of vectors as well as that of the set of column vectors of a matrix T.

In R^{p_i} let $\{a_{i1}, \dots, a_{ij}\}$, $j \leqslant p_i$, be an M_i-o.n. system. We shall put

$P_i^{(j)} = M_i$-orthogonal projector of R^{p_i} onto $\mathcal{M}(a_{i1}, \dots, a_{ij})$, $j = 1, \dots, p_i$,

$P_i^{(0)} =$ null projector,

$P_i^{(p_i)} =$ unit matrix I_{p_i} .

Then $I_{p_i} - P_i^{(j)}$ is also an M_i-orthogonal projector. We have in (R^{p_i}, M_i)

$$\begin{aligned} \mathcal{M}(P_i^{(j)}) &= \mathcal{M}(a_{i1}, \dots, a_{ij}), \\ \mathcal{M}(I_{p_i} - P_i^{(j)}) &= \mathcal{M}(a_{i1}, \dots, a_{ij})^{\perp}. \end{aligned} \tag{3.1}$$

Thus

$$a'_{ik} M_i (I_{p_i} - P_i^{(j)}) = 0 \ \forall k = 1, \dots, j. \tag{3.2}$$

The following matrices will be of constant use :

$$A_i^{(j)} = (I_{p_i} - P_i^{(j-1)}) A_i, \ j = 1, \dots, p_i + 1, \tag{3.3}$$

$$A_i^{(p_i+1)} = 0, \tag{3.4}$$

$$A_i^{(1)} = A_i.$$

Put $A^{(j)} = ((A_1^{(j)})' \cdots (A_l^{(j)})')'$. In particular $A^{(1)} = A$. For the existence of $A^{(j)}$, all the $A_i^{(j)}$ have to exist. Hence $j \leqslant \ \min_{1 \leqslant i \leqslant l} p_i + 1$. By (3.1) and (3.3) we have

$$\mathcal{M}(A_i^{(j)}) \subset \mathcal{M}(a_{i1}, \dots, a_{i,j-1})^{\perp}. \tag{3.5}$$

An o.n. system will be generated as follows.

Construction process

If $A^{(j)} \neq 0$, choose b_j arbitrarily in R^q so that $A^{(j)} N b_j \neq 0$.

Define a_{ij} in R^{p_i}, $i = 1, \dots, l$ so that $\|a_{ij}\|_{M_i} = 1$ and

$$\begin{pmatrix} \mu_{1j} a_{1j} \\ \cdots \\ \mu_{lj} a_{lj} \end{pmatrix} = \frac{A^{(j)} N b_j}{\|A^{(j)} N b_j\|_M}, \tag{3.6}$$

where

$$\mu_{ij} = \frac{\|A_i^{(j)} N b_j\|_{M_i}}{\|A^{(j)} N b_j\|_M} \text{ and } \sum_{i=1}^{l} \mu_{ij}^2 = 1.$$

(i) If $A_i^{(j)} Nb_j \neq 0$, it is colinear with a_{ij}, namely $a_{ij} = \frac{A_i^{(j)} Nb_j}{\|A_i^{(j)} Nb_j\|_{M_i}}$.

(ii) If $A_i^{(j)} Nb_j = 0$ and $A_i^{(j)} \neq 0$, we shall choose a_{ij} arbitrarily in $\mathcal{M}(A_i^{(j)})$.

(iii) If $A_i^{(j)} = 0$ choose a_{ij} in R^{p_i} so that $a_{ij} \perp_{M_i} \mathcal{M}(a_{i1}, \dots, a_{i,j-1})$ (choose a_{i1} arbitrarily if $j = 1$).

From (i), (ii), (iii) one can construct a_{ij} for each given i providing $A^{(j)} \neq 0$. To ensure that $A^{(j)}$ exists and may be non-null, the construction of a_{ij} for given i assume that $j \leqslant \min_{1 \leqslant k \leqslant l} p_k + 1$ and that $j \leqslant p_i$. Furthermore, the condition $j \leqslant p_i$ is necessary for (ii) and (iii) to be effective.

Remark. Hanafi and Lafosse (2001) solved the problem of seeking o.n. systems $\{a_{ij},\ j = 1, \dots\}$ that stepwise maximize the expression (1.1) which can be rewritten as $\sum_{i=1}^{l} (e_i' M_i A_i N b)^2$ with $A_i = X_i' D Y$, but they allowed the existence of null vectors a_{ij}. The present construction always leads to an o.n. system $\{a_{ij}\}$ for each i, moreover, on the very basis of such a consistent construction we can prove the accurate formula (4.10) for the cardinality of these o.n. systems, which formula is lacking in the above-mentioned paper.

We now highlight some properties of the vectors a_{ij}.

Lemma 3.1. *Assume $j \leqslant \min_k p_k + 1$ and $j \leqslant p_i$ for given i. If $A^{(j)} \neq 0$ then $\{a_{i1}, \dots, a_{ij}\}$ is an M_i-o.n. system in R^{p_i}.*

Proof. From (3.2), for $j \geqslant 2$, $a_{ik}' M_i A_i^{(j)} N b_j = 0$ ($k = 1, \dots, j-1$). From the construction (i), (iii), (ii) and (3.5) the property follows. □

Lemma 3.2. *Under the constraints $e_i \perp_{M_i} \mathcal{M}(a_{i1}, \dots, a_{i,j-1})$ $(i = 1, \dots, l)$, we have*

$$(3.7) \qquad (\forall b \in R^q) \quad \sum_{i=1}^{l} (e_i' M_i A_i N b)^2 = \sum_{i=1}^{l} (e_i' M_i A_i^{(j)} N b)^2, \ 1 \leqslant j \leqslant \min_i p_i.$$

Proof. Because

$$e_i \perp_{M_i} \mathcal{M}(a_{i1}, \dots, a_{i,j-1}) \Longleftrightarrow e_i' M_i P_i^{(j-1)} = 0,$$

which entails $e_i' M_i A_i^{(j)} = e_i' M_i A_i$ on account of (3.3). □

Lemma 3.3. *Given b_j such that $A^{(j)} N b_j \neq 0$. The vectors $\pm a_{ij}$ $(i = 1, \dots, l)$ are the only vectors e_i that maximize $\sum_{i=1}^{l} (e_i' M_i A_i^{(j)} N b_j)^2$ unconditionally, and also the only vectors e_i that maximize $\sum_{i}^{l} (e_i' M_i A_i N b_j)^2$ under the constraints $e_i \perp_{M_i} \mathcal{M}(a_{i1}, \dots, a_{i,j-1}), i = 1, \dots, l$ (no constraints when $j = 1$).*

Proof. The first assertion follows from Lemma 2.1 and (3.6), the second from Lemmas 3.2 and 3.1. □

4. The stopping step

This section aims at establishing an accurate formula for the cardinality of the o.n. systems constructed in Section 3.

The jth step produces $\{a_{1j},\ldots,a_{lj}\}$, $j \leqslant \min_i p_i$. By the construction j is the stopping step, i.e. the process stops just after producing a_{ij}, if and only if either $j=\min_i p_i$ or $A^{(j)} \neq 0$ and $A^{(j+1)}=0$.
We shall first study the successive matrices $A_i^{(j)}$. The symbol P_F will denote the orthogonal projector onto a subspace F of some vector space.
Consider the vector space (R^{p_i}, M_i). Put

$$\begin{aligned} F &= \mathcal{M}(a_{i1},\ldots,a_{i,j-1})^{\perp}, \\ F &= R^{p_i} \text{ when } j=1. \end{aligned}$$

Consider some given i. For $p_i+1 \geqslant h > j \geqslant 1$ we have

$$F=\mathcal{M}(a_{ij},\ldots,a_{i,h-1}) \oplus H, \text{ with } H \subset F. \tag{4.1}$$

Then $H=\mathcal{M}(a_{i1},\ldots,a_{i,h-1})^{\perp}$. When $h=p_i+1$, $H=\{0\}$. By (3.1) we have $I_{p_i}-P_i^{(j-1)}=P_F$, then by (3.3)

$$A_i^{(j)}=P_FA_i,\ A_i^{(h)}=P_HA_i.$$

Since H is a subspace of F we have $P_H=P_HP_F$. Thus

$$A_i^{(h)}=P_HP_FA_i=P_HA_i^{(j)}.$$

Thus from (4.1) we have the orthogonal decomposition in the vector space F

$$A_i^{(j)}=A_i^{(h)}+P_{\mathcal{M}(a_{ij},\cdots,a_{i,h-1})}A_i^{(j)},\ 1 \leqslant j < h \leqslant p_i+1. \tag{4.2}$$

This formula leads directly to the following

Proposition 1. *For given i and h, $p_i+1 \geqslant h > j \geqslant 1$, if $\mathcal{M}(a_{ij},\ldots,a_{i,h-1}) \subset \mathcal{M}(A_i^{(j)})$ then $\mathcal{M}(A_i^{(h)}) \subset \mathcal{M}(A_i^{(j)})$.*

By Lemma 2.1, given any $b \in R^q$ the maximum of $\sum_{i=1}^{l}(e_i'M_iA_i^{(j)}Nb)^2$ equals $\|A^{(j)}Nb\|_M^2$ which enjoys the following lemma

Lemma 4.1. *As j increases we have*

(i) *$\|A_i^{(j)}Nb\|_{M_i}$ and rank$A_i^{(j)}$ are non-increasing for $1 \leqslant j \leqslant p_i+1$,*
(ii) *so are $\|A^{(j)}Nb\|_M$ and rank$A^{(j)}$ for $1 \leqslant j \leqslant \min_{1\leqslant i\leqslant l} p_i+1$.*

Proof. From (4.2) for any $b \in R^q$ we have the orthogonal decomposition in R^{p_i}

$$A_i^{(j)}Nb=A_i^{(h)}Nb+P_{\mathcal{M}(a_{ij},\cdots,a_{i,h-1})}A_i^{(j)}Nb,\ 1 \leqslant j < h \leqslant p_i+1,$$

hence the M_i-norm of the left-hand side equals the sum of norms of the two summands on the right-hand side. The first assertion (i) follows. Then $\ker A_i^{(j)} \subset \ker A_i^{(h)}$, hence rank$A_i^{(j)} \geqslant$ rank$A_i^{(h)}$ by the dimension formula. From (2.1) we have $\|A^{(j)}Nb\|_M^2 = \sum_{i=1}^{l} \|A_i^{(j)}Nb\|_{M_i}^2$, then $\|A^{(j)}Nb\|_M \geqslant \|A^{(h)}Nb\|_M$ when $1 \leqslant j < h \leqslant \min_i p_i + 1$. □

Lemma 4.2. *For given i and $1 \leqslant j \leqslant h \leqslant p_i$, the following hold.*

$$(A_i^{(h)} \neq 0) \iff (\mathcal{M}(A_i^{(j)}) \supset \mathcal{M}(a_{ij}, \dots, a_{ih})), \tag{4.3}$$

$$(A_i^{(h)} \neq 0) \implies (\text{rank} A_i^{(j)} \geqslant h - j + 1), \tag{4.4}$$

$$(A_i^{(h+1)} = 0) \iff (\mathcal{M}(A_i^{(j)} \subset \mathcal{M}(a_{ij}, \dots, a_{i,h})), \tag{4.5}$$

$$(A_i^{(h+1)} = 0) \implies (\text{rank} A_i^{(j)} \leqslant h - j + 1). \tag{4.6}$$

Proof. Let i be given. First note that by (4.2) for $p_i + 1 \geqslant h + 1 > j \geqslant 1$

$$\begin{aligned}(A_i^{(h+1)} = 0) &\iff (\forall b \in R^q,\ A_i^{(j)}Nb = P_{\mathcal{M}(a_{ij},\cdots,a_{i,h})}A_i^{(j)}Nb) \\ &\iff (\mathcal{M}(A_i^{(j)}) \subset \mathcal{M}(a_{ij}, \dots, a_{i,h})).\end{aligned}$$

Thus we get (4.5) and (4.6). Let $1 \leqslant j \leqslant k \leqslant h \leqslant p_i$. Then by Lemma 4.1(i)

$$\forall h \leqslant p_i,\quad (A_i^{(h)} \neq 0) \implies (A_i^{(k)} \neq 0,\ k = j, \dots, h).$$

Then by the construction (i), (ii) in Section 3 we see that

$$(A_i^{(h)} \neq 0) \implies (a_{ik} \in \mathcal{M}(A_i^{(k)}),\ k = j, \dots, h).$$

Assume $A_i^{(h)} \neq 0,\ h \geqslant j$, then $a_{ij} \in \mathcal{M}(A_i^{(j)})$. Using Proposition 4.1 it follows that

$$(a_{ij} \in \mathcal{M}(A_i^{(j)})) \implies (\mathcal{M}(A_i^{(j+1)}) \subset \mathcal{M}(A_i^{(j)})) \implies (a_{ij}, a_{i,j+1} \in \mathcal{M}(A_i^{(j)}))$$

when $h \geqslant j + 1$, by Proposition 4.1 the last inclusion in turn entails

$$(\mathcal{M}(A_i^{(j+2)}) \subset \mathcal{M}(A_i^{(j)})) \implies (a_{ij}, a_{i,j+1}, a_{i,j+2} \in \mathcal{M}(A_i^{(j)}))$$

when $h \geqslant j + 2$, and so on. Finally we get

$$(A_i^{(h)} \neq 0) \implies (\mathcal{M}(a_{ij}, \dots, a_{ih}) \subset \mathcal{M}(A_i^{(j)})),\ 1 \leqslant j \leqslant h \leqslant p_i,$$

and by the way we get (4.4).
To prove the converse note that from (4.5)

$$(A_i^{(h)} \neq 0) \iff (\mathcal{M}(A_i^{(j)}) \not\subset \mathcal{M}(a_{ij}, \dots, a_{i,h-1})),\ 1 \leqslant j < h \leqslant p_i.$$

Let $\mathcal{M}(a_{ij}, \dots, a_{ih}) \subset \mathcal{M}(A_i^{(j)})$, then $\mathcal{M}(A_i^{(j)}) \not\subset \mathcal{M}(a_{ij}, \dots, a_{i,h-1})$, $j < h$, thus $A_i^{(h)} \neq 0$. For $j = h$, if $\mathcal{M}(a_{ij}) \subset \mathcal{M}(A_i^{(j)})$ then $A_i^{(h)} \neq 0$ too. Thus

$$(\mathcal{M}(a_{ij}, \dots, a_{ih}) \subset \mathcal{M}(A_i^{(j)})) \implies (A_i^{(h)} \neq 0),\ 1 \leqslant j \leqslant h \leqslant p_i,$$

and (4.3) is proved. □

Lemma 4.3. *For given i and $p_i \geqslant h \geqslant j \geqslant 1$,*

$$(A_i^{(h)} \neq 0,\ A_i^{h+1} = 0) \iff (\mathcal{M}(A_i^{(j)}) = \mathcal{M}(a_{ij}, \dots, a_{ih})), \tag{4.7}$$

$$(A_i^{(h)} \neq 0,\ A_i^{(h+1)} = 0) \iff (\text{rank} A_i^{(j)} = h - j + 1,\ j \leqslant h \leqslant p_i). \tag{4.8}$$

The relation

$$\text{rank}A_i^{(j)} = \text{rank}A_i - j + 1 \tag{4.9}$$

holds for $1 \leqslant j \leqslant \min(p_i, \text{rank}A_i + 1)$.

Proof. (4.7) follows immediately from (4.3) and (4.5). From (4.7), for $p_i \geqslant h \geqslant j \geqslant 1$ we have

$$(A_i^{(h)} \neq 0,\ A_i^{(h+1)} = 0) \Longrightarrow (\text{rank}A_i^{(j)} = h - j + 1).$$

Let us prove the converse. For $h = p_i$, from (3.4) $A_i^{(h+1)} = 0$ is trivially true. For $h < p_i$, i.e. $h + 1 \leqslant p_i$, by (4.4)

$$(A_i^{(h+1)} \neq 0) \Longrightarrow (\text{rank}A_i^{(j)} \geqslant h - j + 2).$$

On the other hand, by (4.6), for $h \leqslant p_i$

$$(A_i^{(h)} = 0) \Longrightarrow (\text{rank}A_i^{(j)} \leqslant h - j).$$

Therefore,

$$(\text{rank}A_i^{(j)} = h - j + 1,\ j \leqslant h \leqslant p_i) \Longrightarrow (A_i^{(h)} \neq 0,\ A_i^{(h+1)} = 0).$$

Thus (4.8) is proved. Letting $j = 1$ in (4.8) we get (4.9) which holds only for $\text{rank}A_i - j + 1 \geqslant 0$, i.e. for $j \leqslant \text{rank}A_i + 1$. □

Lemma 4.4. $(\forall j:\ 1 \leqslant j \leqslant h \leqslant \min_{1\leqslant i \leqslant l} p_i)$

$$\begin{aligned}(A^{(h)} \neq 0,\ A^{(h+1)} = 0) \quad &\Leftrightarrow \quad (\max_{i=1,\dots,l} \text{rank}A_i^{(j)} = h - j + 1,\ j \leqslant h \leqslant \min_i p_i) \\ &\Leftrightarrow \quad (\max_{i=1,\dots,l} \text{rank}A_i = h,\ h \leqslant \min_i p_i).\end{aligned}$$

Proof. Using (4.4) and (4.6) for $1 \leqslant j \leqslant h \leqslant \min_i p_i$ we get

$$\begin{aligned}(A^{(h)} \neq 0) \quad &\Longrightarrow \quad ((\exists i)\, \text{rank}\, A_i^{(j)} \geqslant h - j + 1), \\ (A^{(h+1)} = 0) \quad &\Longrightarrow \quad ((\forall i)\, \text{rank}\, A_i^{(j)} \leqslant h - j + 1),\end{aligned}$$

hence

$$(A^{(h)} \neq 0,\ A^{(h+1)} = 0) \Longrightarrow (\max_i \text{rank}A_i^{(j)} = h - j + 1).$$

Let us prove the converse. From above it follows that

$$(\max_i \text{rank}A_i^{(j)} = h - j + 1,\ j \leqslant h \leqslant \min_i p_i) \Longrightarrow (A^{(h)} \neq 0).$$

Now assume that $\max_i \text{rank}A_i^{(j)} = h - j + 1,\ j \leqslant h \leqslant \min_i p_i$. Then $A^{(h)} \neq 0$. We shall show that

$$(\max_i \text{rank}A_i^{(j)} = h - j + 1,\ j \leqslant h \leqslant \min_i p_i) \Longrightarrow (A^{(h+1)} = 0).$$

Indeed, $A^{(h)}$ being nonnull, $A^{(h+1)}$ exists from the construction in Section 3. If $A^{(h+1)} \neq 0$ then $\exists i' :\ A_{i'}^{(h+1)} \neq 0$, hence by (3.4) $h+1 \leqslant\ p_{i'}$. Therefore, by (4.4) it would follow that rank $A_{i'}^{(j)} \geqslant\ h-j+2$. Thus, for $1 \leqslant\ j \leqslant\ h \leqslant\ \min_i p_i$

$$(\max_i \operatorname{rank} A_i^{(j)} = h-j+1,\ j \leqslant\ h \leqslant\ \min_i p_i) \Longleftrightarrow (A^{(h)} \neq 0,\ A^{(h+1)} = 0). \quad \square$$

From the preceding consideration we shall now get a clear-cut formula for the stopping step m of the construction process. Since $\{a_{i1}, \ldots, a_{im}\}$ is an o.n. system in R^{p_i} ($\forall i = 1, \ldots, l$), it is required that $m \leqslant\ \min_i p_i$.

Theorem 4.1. *We have*

$$m \;=\; \min\{p_1, \ldots, p_l, \max_i \operatorname{rank} A_i^{(j)} + j - 1\},$$

$$m \;=\; \min\{p_1, \ldots, p_l, \max_i \operatorname{rank} A_i\}. \tag{4.10}$$

Proof. It suffices to prove (4.10). For the stopping step m, always $A^{(m)} \neq 0$. Assume $h = \max_i \operatorname{rank} A_i \leqslant\ \min_i p_i$. Then from Lemma 4.4 it follows that $A^{(h)} \neq 0,\ A^{(h+1)} = 0$. Thus the stopping step $m = h$. If $h = \max \operatorname{rank} A_i > \min p_i$, by Lemma 4.4 we cannot have $A^{(k)} \neq 0,\ A^{(k+1)} = 0$ for any $k \leqslant\ \min p_i$. Then the stopping step m must be $\min p_i$. $\square$

5. Generating an o.n. system in (R^q, N)

We shall see that the construction in Section 3 induces two other o.n. systems. The jth step of the construction in Section 3 is based on $A^{(j)} \neq 0,\ j \leqslant\ m$. b_j is arbitrarily chosen subject to $A^{(j)} N b_j \neq 0$. Let us put

$$\lambda_j = \|A^{(j)} N b_j\|_M > 0,\ a_j = \frac{A^{(j)} N b_j}{\|A^{(j)} N b_j\|_M},\ j \leqslant\ m.$$

Then $A^{(j)} N b_j = \lambda_j a_j,\ \|a_j\|_M = 1,\ \lambda_j > 0$.
Assume that the system $\{a_{i1}, \ldots\}$ is o.n.

Lemma 5.1. *For any (j,k), $1 \leqslant\ j \leqslant\ k \leqslant\ m$, we have*

$$A_i' M_i a_{ik} \;=\; (A_i^{(j)})' M_i a_{ik}, \tag{5.1}$$

$$A' M a_k \;=\; (A^{(j)})' M a_k. \tag{5.2}$$

Moreover, $\{a_1, \ldots, a_m\}$ is an o.n. system in (R^p, M).

Proof. Since $a_{ik} \perp \mathcal{M}(a_{i1}, \ldots, a_{i,j-1})$ we get $a_{ik}' M_i P_i^{(j-1)} = 0$. Then (5.1) follows from (3.3), whereas (5.2) from (5.1) and (3.6). Because of (3.6) $a_h' M a_j = \sum_{i=1}^{l} \mu_{ih} \mu_{ij} a_{ih}' M_i a_{ij} = 0$ for $h \neq j$, the last assertion follows. $\square$

Let us now restrict the choice of b_j. We shall choose b_j so that, $1 \leqslant\ j \leqslant\ m$,

$$A^{(j)} N b_j \;=\; \lambda_j a_j,\ (\lambda_j > 0,\ \|a_j\|_M = 1), \tag{5.3}$$

$$(A^{(j)})' M a_j \;=\; \lambda_j b_j. \tag{5.4}$$

From this choice we get the

Lemma 5.2. *The b_j's form an N-o.n. system in R^q.*

Proof. For $h \geqslant j$, using (5.3) and (5.4) we get

$$b'_h N b_j = \lambda_h^{-1} a'_h M A^{(h)} N bj = \lambda_h^{-1}(a'_h M A^{(j)}) N b_j = \lambda_h^{-1}\lambda_j a'_h M a_j = \begin{cases} 0 & \text{if} \quad j < h \\ 1 & \text{if} \quad j = h. \end{cases}$$ □

(5.3) and (5.4) mean that (a_j, b_j, λ_j) is a trio, always existing, consisting respectively of a left singular vector, the associated right one and the corresponding positive singular value of the matrix $A^{(j)} \neq 0,\ 1 \leqslant j \leqslant m$.

6. Maximum of a sum of squared bilinear forms

We are now in a position to solve the maximization problem.
From Lemma 3.3 it is known that the system $\{\epsilon_1 a_{1j}, \dots, \epsilon_l a_{lj}\}$, $\epsilon_i = \pm 1$, is the unique one that maximizes $\sum_{i=1}^{l}(e'_i M_i A_i^{(j)} N b_j)^2$, b_j being given so that $A^{(j)} N b_j \neq 0$. The maximum equals $\|A^{(j)} N b_j\|_M^2$. By Lemma 2.2 the maximum value of $\|A^{(j)} N b_j\|_M^2$ equals the squared greatest singular value $\lambda_{(j)}^2$ of $A^{(j)}$. Since $A^{(j)} \neq 0,\ 1 \leqslant j \leqslant m$, this maximum is attained if and only if b_j is a right singular vector $b^{(j)}$ of $A^{(j)}$ corresponding to the greatest singular value $\lambda_{(j)}$.
By Lemma 3.2 and 3.3 to maximize $\sum_i (e'_i M_i A_i^{(j)} N b_j)^2$ unconditionally is equivalent to maximize $\sum_i (e'_i M_i A_i N b_j)^2$ under the constraints

$$e_i \perp_{M_i} \mathcal{M}(a_{i1}, \dots, a_{i,j-1}),\ i = 1, \dots, l. \tag{6.1}$$

Therefore, by (3.7), the system $\{a_{1j}, \dots, a_{lj}, b^{(j)}\}$ not only unconditionally maximizes $\sum_i (e'_i M_i A_i^{(j)} N b_j)^2$, but also maximizes $\sum_i (e'_i M_i A_i N b_j)^2$ under the constraints (6.1) and under even the constraints

$$e_i \perp_{M_i} \mathcal{M}(a_{i1}, \dots, a_{i,j-1}),\ b_j \perp_N \mathcal{M}(b^{(1)}, \dots, b^{(j-1)}).$$

Indeed, the system $\{a_{1j}, \dots, a_{lj}, b^{(j)}\}$ satisfies the last constraints because of Lemma 5.2. Thus, we have solved the stepwise maximization problem in Section 1. The maximum at every step enjoys the obvious following lemma:

The maximum $\lambda_{(j)}^2$ at the jth maximization step is non-increasing as j increases.

7. Conclusion

The number m, given by (4.10), being the stopping step of the construction in Section 3, the solution to the stepwise maximization problem for the sum (1.2) of squared bilinear forms highlights the following o.n. systems:
$\{a_{i1}, \dots, a_{im}\}$ in $(R^{p_i}, M_i),\ i = 1, \dots, l$,
$\{a^{(1)}, \dots, a^{(m)}\}$ in (R^p, M), where $a^{(j)}$ is some left singular vector corresponding to the greatest singular value $\lambda_{(j)}$ of $A^{(j)}$,

$\{b^{(1)}, \ldots, b^{(m)}\}$ in (R^q, N), where $b^{(j)}$ is the right singular vector associated to the left one $a^{(j)}$.
If $l = 1,\ p_1 = p = q$ these systems all coincide to become the o.n. system $\{u_1, \ldots, u_m\}$ in PCA, $m = \text{rank}A$. Hanafi and Lafosse (2001) discussed the use of these o.n. systems for analyzing the dependence between several collections of variables.

Acknowledgment

The author would like to thank the referee for helpful remarks and suggestions.

References

[1] J. D. Carroll, *Generalization of canonical correlation analysis to three or more sets of variables*, Proc. 76th Annual Convention APA (1968), 227-228.

[2] M. Hanafi and R. Lafosse, *Généralisation de la régression simple pour analyser la dépendance de K ensembles de variables avec un (K + 1) ème*, Revue de Statistique Appliquée **49** (1) (2001), 5-30.

[3] H. A. L. Kiers, R. Cléroux and J. M. F Ten Berge, *Generalized canonical analysis based on optimizing matrix correlations and a relation with IDIOSCAL*, Computational Statistics and Data Analysis **18** (1994), 331-340.

[4] R. Lafosse, *Analyse de concordance de deux tableaux: monogamies, simultanéités et découpages*, Revue de Statistique Appliquée **45** (3) (1997), 45-72.

[5] R. Lafosse and M. Hanafi, *Concordance d'un tableau avec K tableaux: définition de K + 1 uples synthétiques*, Revue de Statistique Appliquée **45** (4) (1997), 111-126.

[6] J. C. Lingoes and P.H. Schönemann, *Alternative measures of fit for the Schönemann-Carroll matrix fitting algorithm*, Psychometrika **39** (1974), 423-427.

[7] P. C. Mahalanobis, *On the generalized distance in statistics*, Proc. Nat. Inst. Sci. India **12** (1936), 49-55.

Department of Mathematical Statistics
University of Natural Sciences
227 Nguyen Van Cu, Dist. 5
Ho Chi Minh City, Vietnam

E-mail address: nbvan@hcmuns.edu.vn