A new numerical approach to calculate the eigensystem of a matrix that appear in the problem of canonical correlation *

Raimundo J.B. de Sampaio
Jorge A. Leludak
Marco A.B. Candido
Pontifical Catholic University of Paraná
Graduate Program in Poduction Engineering and Systems
(PPGEPS - PUCPR)

Abstract

In several many situations we need to calculate the eigenvalues and eigenvectors of a matrix which is given as a product of many others matrices, that is, we need to calculate the eigenvalue and eigenvectors of a matrix M which is given as $M = m_1 m_2 ... m_k$. Even without considering the complexity of the multiplication of matrices, many times M become very ill-conditioning as soon as eigenvector is the concern. Actually this happens even when M is a symmetric matrix, which will be the our only concern here.

*This work was partially supported by CNPq Brazil

1 Introduction

In many situations we need to relate some variables of one set to the variables of another set. This happens, in particular, in continuous production process, when we need to know how one of the set of variables impact the other set of variables. If we can interpret one of these set as a set of independent variables and the other set as a set of dependent variables, and if there is a linear relationship between these two sets, then we can calculate the strength of this relation, the so called canonical correlation. We must understand here that we are not looking for a relationship among the elements of one set to the elements of another set, but, in some sense, we are looking for a relationship in between all the elements of one set with all the elements of the other set. To accomplish that, we consider all the linear combinations in each of the sets, and then choose that pair that has the strongest correlation. Again we must be aware that, in general, no linear combination of the elements of one set is one variable of this set. However, practical experiments has being showing that this approach is quite reliable when the number of variables of both of the sets are reasonable.

The complexity of the canonical correlation problem accounts to by a multiplication of five matrices, and then a calculation of the pair of maximum eigenvalue and the eigenvector associated, beside the effort of obtain the matrices itself. Even if all the matrices were well-conditioned, this product could be not, that is, the eigenvector problem associated to the maximum eigenvalue of the resulting matrix may be very ill-conditioned.

Several times the most important factor to be considered in the eigensystem is exact the conditioning of the problem and not just the number of operations to get the answer, the so called complexity. Those aspects of the problem had motivated as to consider the question of calculating the eigenvector associated to the maximum eigenvalue of this five-block matrix without performing the multiplication of the matrices, even it worse a little more the complexity of the problem. The idea is to define an specific new five block matrix, where each block is one of the factors of the multiplication matrix, and then calculate it maximum eigenvalue and the eigenvector associated.

This paper is organized as follows. An Introduction Section where we give some general details about the whole work, a Preliminaries Section where we give some definitions and results that will be used later on, a Main Section , where we present the main result, and finally a section with the proposed algorithms and some conclusions.

2 Preliminaries

Let X be a random n-dimensional vector whose components are random variables. Each coordinate of X is a random variable with its own marginal probability distribution. Here it is supposed that all of the probability distributions are normal probability distributions. The marginal mean and variance of each coordinate of X is defined as $\mu_i = E(X_i)$, and the variance $\sigma_i = E(X_i - \mu_i)^2$. Thus, putting all together, we have that the expectation of X is $E(X) = \mu$, a n-dimensional vector, and the variance-covariance matrix is $\mathbf{\Sigma} = E(X-\mu)(X-\mu)^T$, a $n \times n$ symmetric matrix. Explicitly, the variance-covariance matrix is thus,

$$\mathbf{\Sigma} = Cov(X) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

where $\sigma_{ii} = \sigma_i = E(X_i - \mu_i)^2$, e $\sigma_{ik} = E(X_i - \mu_i)(X - \mu_k)$.

The correlation coefficient ρ_{ik} is calculated in terms of covariance σ_{ik} and variances σ_{ii} and σ_{kk}, as

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}.$$

This ρ_{ik} measure the strength of the linear association between the random variables X_i and X_k.

If a is a vector in R^n, then a linear combination of the variables in X is simply $a^T X$. Performing the definitions given above, we can have that,

$$E(a^T X) = a^T E(X) = a^T \mu,$$

and

$$Var(a^T X) = a^T \mathbf{\Sigma} a,$$

where $\Sigma = Cov(X)$.

Now let X and Y be random variables with $E(X) = \mu_1$ and $E(Y) = \mu_2$, and let Z the random variable,

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Then,

$$E(Z) = E \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \mu.$$

and,

$$Cov(Z) = Cov \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where Σ_{11} is the covariance matrix of X, Σ_{22} is the covariance matrix of Y and $\Sigma_{12} = (\Sigma_{21})^T$ is the covariance matrix of X and Y.

Lets call $U = a^T X$ and $V = b^T Y$, then,

$$Var(U) = a^T \Sigma_{11} a,$$

$$Var(V) = b^T \Sigma_{22} b,$$

and,

$$Cov(U,V) = a^T \Sigma_{12} b.$$

The correlation between U and V is defined as

$$Corr(U,V)= \max_{\mathbf{a},\mathbf{b}\in\mathbf{R^n}} \frac{a^T\Sigma_{12}b}{\sqrt{a^T\Sigma_{11}a}\sqrt{b^T\Sigma_{22}b}}.$$

Theorem 1 *Suppose $X \in R^m$, $Y \in R^n$ are random variables with $m \leqslant n$, $Cov(X) = \Sigma_{11}$, $Cov(Y) = \Sigma_{22}$, and $Cov(X,Y) = \Sigma_{12}$. Furthermore, suppose $Cov(Z)$=Σ has full rank. Let U= a^TX and V= b^TY, then,*

$$\max_{a,b} Corr(U,V) = \rho_1^*$$

which comes to be the biggest eigenvalue of $\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$, and a, b are respectively,

$$a^T = u_1^T\Sigma_{11}^{-\frac{1}{2}} \;\; and \;\; b^T = v_1^T\Sigma_{22}^{-\frac{1}{2}}$$

where u is the eigenvector associated to the greatest eigenvalue of

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}},$$

and v is the eigenvector associated to the greatest eigenvalue of

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}},$$

furthermore, $Var(U) = Var(V) = 1$.

Proof. See [Johnson 98] ■

Since the matrix $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ is symmetric, eigenvalue is no longer a problem, but unfortunately, eigenvector can be very ill-conditioned. It has noting we can do about conditioning acting directly on the matrix using similarity transformation, since similarity transformation preserves eigenvalues. However we can get better about conditioning if we redefine the problem so that the matrix we have to look for the eigenvalues-eigenvectors is no longer a product of matrices, and has a more favorable conditioning number.

From now on the matrix

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$$

will be immerse in a more general class of matrices $m = m_1m_2m_3m_4m_5$ where m_1, m_3, and m_5 are positive definite, $m_1 = m_5$ and $m_4^T = m_2$ are invertible. The properties used here for m_i, $i = 1, 2, ..., 5$ are exactly those one owned by their counterpart in

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}.$$

The rule of $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ is the same of that of $\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$ with the rule of Σ_{11} changed with the rule of Σ_{22}, and so, both matrices are in the same class.

3 Main Results

Let M be the matrix defined by

$$M = \begin{pmatrix} & & & & m_5 \\ m_4 & & & & \\ & m_3 & & & \\ & & m_2 & & \\ & & & m_1 & \end{pmatrix}.$$

The eigenvalues of M are closely related to those of $m = m_1m_2m_3m_4m_5$ as we are going to see.

Theorem 2 *The set of eigenvalues of M is a subset of the eigenvalues of $m = m_1m_2m_3m_4m_5$ in the following sense. If λ is an eigenvalue of M then λ^5 is an eigenvalue of m.*

Proof. Let λ be an eigenvalue of M, then there exists a non zero vector $x^T = (x_1x_2x_3x_4x_5)$ such that

$$\begin{pmatrix} & & & & m_5 \\ m_4 & & & & \\ & m_3 & & & \\ & & m_2 & & \\ & & & m_1 & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

Since $m_5x_5 = \lambda x_1$, $m_4x_1 = \lambda x_2$, $m_3x_2 = \lambda x_3$, $m_2x_3 = \lambda x_4$, $m_1x_4 = \lambda x_5$, then we have that

$$m_1m_2m_3m_4m_5x_5 = \lambda^5 x_5,$$

and the proof is complete. ■

The characteristic polynomial for M matrix has high degree than that one for m matrix, then the inclusion must stand for different eigenvalues, that is, the inclusion is in the sense of set inclusion. Therefore we have that the eigenvalues of matrix $m = m_1m_2m_3m_4m_5$ can be obtained from those of matrix M.

Now lets cast some properties of matrix M. First, matrix M is 5-cycle in the sense that if we rotate the matrix M in the clock sense to

$$M_4 = \begin{pmatrix} & & & & m_4 \\ m_3 & & & & \\ & m_2 & & & \\ & & m_1 & & \\ & & & m_5 & \end{pmatrix}$$

then we have that,

$$m_1m_2m_3m_4m_5\ x_4 = \lambda^5\ x_4,$$

that is, we have the same eigenvalue-eigenvector relation, but now the eigenvector is x_4 instead of x_5, as before. Let M_3 be the matrix,

$$M_3 = \begin{pmatrix} & & & & m_3 \\ m_2 & & & & \\ & m_1 & & & \\ & & m_5 & & \\ & & & m_4 & \end{pmatrix}$$

then we have now,

$$m_1m_2m_3m_4m_5\ x_3 = \lambda^5\ x_3.$$

If we let M_2 be the matrix

$$M_2 = \begin{pmatrix} & & & & m_2 \\ m_1 & & & & \\ & m_5 & & & \\ & & m_4 & & \\ & & & m_3 & \end{pmatrix}$$

once again we have the same kind of relation,

$$m_1 m_2 m_3 m_4 m_5 \; x_2 = \lambda^5 \; x_2.$$

Finally, if we let M_1 be the matrix,

$$M_1 = \begin{pmatrix} & & & & m_1 \\ m_5 & & & & \\ & m_4 & & & \\ & & m_3 & & \\ & & & m_2 & \end{pmatrix}$$

then, the eigenvalue-eigenvector expression comes to be,

$$m_1 m_2 m_3 m_4 m_5 \; x_1 = \lambda^5 \; x_{1.}$$

Far ahead we will see that the eigenvalue is always the same in each one of these expressions and just the eigenvectors change. This is exactly what we are going to use to have a better conditioning eigenvector problem.

Another quite striking property of matrix M is that it is invertible, up on its block components. In fact if you see M^T matrix and MM^T matrix, then you can see easily figure out how to build up M^{-1}.In fact, M^T is

$$M^T = \begin{pmatrix} & m_4^T & & & \\ & & m_3^T & & \\ & & & m_2^T & \\ & & & & m_1^T \\ m_5^T & & & & \end{pmatrix}$$

and MM^T is

$$\mathrm{MM}^T = \begin{pmatrix} m_5 m_5^T & & & & \\ & m_4 m_4^T & & & \\ & & m_3 m_3^T & & \\ & & & m_2 m_2^T & \\ & & & & m_1 m_1^T \end{pmatrix}$$

and then

$$M^{-1} = \begin{pmatrix} & m_4^{-1} & & & \\ & & m_3^{-1} & & \\ & & & m_2^{-1} & \\ & & & & m_1^{-1} \\ m_5^{-1} & & & & \end{pmatrix}.$$

The existence of this matrix is a direct consequence of invertibility of each of the components of matrix m. Note also that for each $i = 1, 2, 3, 4$, $M_i M_i^T$ is given by

$$\mathrm{M_4M_4^T} = \begin{pmatrix} m_4 m_4^T & & & & \\ & m_3 m_3^T & & & \\ & & m_2 m_2^T & & \\ & & & m_1 m_1^T & \\ & & & & m_5 m_5^T \end{pmatrix}$$

$$\mathrm{M_3M_3^T} = \begin{pmatrix} m_3 m_3^T & & & & \\ & m_2 m_2^T & & & \\ & & m_1 m_1^T & & \\ & & & m_5 m_5^T & \\ & & & & m_4 m_4^T \end{pmatrix}$$

$$\mathrm{M_3M_2^T} = \begin{pmatrix} m_2 m_2^T & & & & \\ & m_1 m_1^T & & & \\ & & m_5 m_5^T & & \\ & & & m_4 m_4^T & \\ & & & & m_3 m_3^T \end{pmatrix}$$

$$\mathrm{M_1M_1^T} = \begin{pmatrix} m_1 m_1^T & & & & \\ & m_5 m_5^T & & & \\ & & m_4 m_4^T & & \\ & & & m_3 m_3^T & \\ & & & & m_2 m_2^T \end{pmatrix}.$$

The spectrum of $M_i M_i^T$ is given by

$$\lambda(M_i M_i^T) = \bigcup_{k=1}^{5} \lambda(m_k m_k^T).$$

Now let see that all of the systems

$$M_i\ x\ = \lambda\ x,$$

has exactly the same conditioning number. In fact, to see this you only need to see that all of the matrices $M_i M_i^T$ for $i = 1, 2, ..., 5$, are diagonal matrices whose diagonal entries are always the same set Λ,

$$\Lambda = \cup_{i=1}^{5} \left\{ m_i m_i^T \right\}.$$

If we use the same reasoning for M_i^{-1} the results will be of the same nature with the rule of $m_i m_i^T$ changed respectively to $m_i^{-1} m_i^{-T}$, then the maximum of $\lambda(M_i M_i^T)$ for $i = 1, 2, ..., 5$ are the same, and $\lambda_{\max}(M_i^{-1} M_i^{-T})$ for $i = 1, 2, ..., 5$ are also the same, according to the result given above, and so cond(M_i) is always the same, no matter which matrix M_i we consider to solve the eigensystem problem.

Now we must consider the questions concerning the calculation of eigenvector. As we have seen, the calculation of eigenvalue is no longer a problem, and since the matrix m is symmetric, so it is $m - \lambda I$, then its conditioning number is just $cond(m - \lambda I) = \lambda_{\max}(m - \lambda I)/\lambda_{\min}(m - \lambda I)$. The rule to be used here is that every time the conditioning of $m - \lambda I$ is high we calculate the eigenvector using each one of the systems $(M_i - \lambda I)x = 0$ for $i = 1, 2, ..., 5$, and by the contrary, if the $m - \lambda I$ is well-conditioned we use the system $(m - \lambda I)x = 0$.

3.1 Numerical Results

Let $m = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ where Σ_{ij} is defined as before, be the matrix whose eigenvalues we are looking for. In spite of the fact that m is a product of five matrices and symmetric, since Σ_{11} and Σ_{22} are diagonal and $\Sigma_{12} = \Sigma_{21}^T$, the problem of calculating the maximum eigenvector of m can be better conditioned, since the new problem to be solved, involving M matrix, may be better conditioned, and there is not a similarity transformation between m and M.

We relat now some limited numerical experiments performed using MATLAB, where all the matrices were randomly generated.

For matrix 5x5

$v_1(M)$	$v_1(m)$
$0,1716$	$-0,3865$
$0,2326$	$-0,5241$
$0,1205$	$-0,2714$
$0,2362$	$-0,5322$
$0,2077$	$-0,4680$

$v_1(M)$	$v_1(m)$
$0,2127$	$-0,4849$
$0,2523$	$-0,5750$
$0,1404$	$-0,3201$
$0,1861$	$-0,4241$
$0,1710$	$-0,3898$

$v_1(M)$	$v_1(m)$
$-0,1961$	$0,4449$
$-0,1821$	$0,4131$
$-0,2029$	$0,4602$
$-0,1730$	$0,3924$
$-0,2272$	$0,5154$

$v_1(M)$	$v_1(m)$
$-0,0027$	$-0,0184$
$0,0263$	$0,1773$
$0,0936$	$0,6303$
$-0,0694$	$-0,4677$
$-0,0881$	$-0,5934$

$v_1(M)$	$v_1(m)$
$-0,1643$	$-0,3726$
$-0,2360$	$-0,5354$
$-0,2151$	$-0,4878$
$-0,2099$	$-0,4761$
$-0,1461$	$-0,3314$

$v_1(M)$	$v_1(m)$
$0,1618$	$-0,3685$
$0,2753$	$-0,6269$
$0,1039$	$-0,2366$
$0,1801$	$-0,4100$
$0,2183$	$-0,4971$

$v_1(M)$	$v_1(m)$
$0,2055$	$-0,4726$
$0,2039$	$-0,4690$
$0,1595$	$-0,3668$
$0,1923$	$-0,4423$
$0,2070$	$-0,4760$

$v_1(M)$	$v_1(m)$
$0,2002$	$0,4603$
$0,1969$	$0,4528$
$0,1932$	$0,4442$
$0,1650$	$0,3794$
$0,2139$	$0,4918$

$v_1(M)$	$v_1(m)$
$-0,2100$	$0,4834$
$-0,2073$	$0,4772$
$-0,1069$	$0,2462$
$-0,2048$	$0,4714$
$-0,2197$	$0,5057$

Acknowledgement 3 *We are in debt to Prof. Jin Yun Yuan for many helpful discussions we have had during the preparation of this paper.*

References

[1] RICHARD A. JOHNSON and DEAR W. WICHERN, *Applied Multivariate Statistical Analysis,* Printice Hall, 1998

[2] EMÍLIO R.F. NETO, ANSELMO C. NETO, R. J.B. DE SAMPAIO and M. ZUGE, *Canonical Correlation Analysis Applied to Verification of Important Variables at Stages of a Continuous Production Process Associated to the Quality of Final Product*, INDUSCOM 2000 - Proceeding IEEE, 2000.

[3] RAIMUNDO J. B. DE SAMPAIO and MARCO A. B. CANDIDO, *Some refletions on the matrix that appear in the problem of canonical correlation*, **Mathematics and Simulation with Biological, Economical and Musicoacoustical Applications**:, Piraeus, Grécia:, WSES Press, 2001, v. Vol. 1, p. 278-282.