 Vol. 20: 361–367, 1993

Eigenanalysis of Symmetrizable Matrix Products: a Result with Statistical Applications

NANNY WERMUTH and HELMUT RÜSSMANN

Universität Mainz, Germany

ABSTRACT. A theorem is proven which relates the matrices of eigenvalues and eigenvectors of matrix products AB^T, A^TB, BA^T, B^TA if they are symmetrizable, that is if each product itself is expressible as the product of a symmetric and of a positive definite matrix. The result is used to derive properties of a number of different multivariate statistical techniques.

Key words: canonical correlations, correspondence analysis, derived responses, eigenvalues, eigenvectors, principal components

1. Introduction

In several multivariate statistical techniques eigenanalyses are performed, i.e. eigenvalues and eigenvectors of real matrices are computed, to obtain representations which have the attractive feature of being unique up to scalar multiplications or up to orthogonal transformations. These eigenanalyses often involve matrix products of two rectangular matrices A and B, for which the number of columns may differ substantially from the number of rows. As a consequence the dimensions of the two quadratic matrix products AB^T and BA^T agree but may be much smaller say than the dimensions of the two quadratic matrix products A^TB and B^TA. Relations between the products may be exploited to simplify calculations and interpretations. Such relations are derived in this paper in the case the products are *symmetrizable*, that is if the products can be written as products of a symmetric and a positive definite matrix. This notion has been used previously in the theory of compact linear operators in Hilbert spaces (Zaanen, 1953, ch. 12), while we are concerned here with real matrices.

Our result on eigenanalyses of symmetrizable matrix products (section 2, theorem, equations (4)–(12)) is based on the singular value decomposition (Eckart & Young, 1931; Lanczos, 1958; Schwerdtfeger, 1960; Searle, 1982, pp. 316–317) and on the generalized singular value decomposition of a rectangular matrix (Paige, 1985, lem., sect. 2). A short proof of the latter is possible if a formulation and proof are used (appendix) for the singular value decomposition which mimics early arguments given by Schmidt (1907) for an analogous result that has been called the canonical expansion of compact linear operators in a Hilbert space (Kato, 1966, p. 261).

2. Matrices of eigenvalues and of eigenvectors of symmetrizable matrix products

We restate the generalized value decomposition of a rectangular matrix as

Lemma

(Decomposition of a rectangular matrix relative to two given matrices.) For every $s \times t$ matrix M of rank $r > 0$, a positive definite $s \times s$ matrix F and a positive definite $t \times t$ matrix G it is possible to find an $s \times r$ matrix U_+ of rank r, an $t \times r$ matrix V_+ of rank r, and an $r \times r$ positive definite diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ such that

$$M = U_+ \Lambda V_+^T, \quad U_+^T F U_+ = I_r, \quad V_+^T G V_+ = I_r, \tag{1}$$

$$V_+ \Lambda = M^T F U_+, \quad U_+ \Lambda = M G V_+. \tag{2}$$

The proof is direct with (A1) *and* (A2) *after denoting by* U'_+ *and* V'_+ *the column-orthogonal matrices which reduce an* $s \times t$ *matrix* M' *to diagonal form and by defining*

$$M = F^{1/2}M'G^{1/2}, \quad U_+ = F^{1/2}U'_+, \quad V_+ = G^{1/2}V'_+.$$

We are now ready to state the main result concerning matrices of eigenvectors of symmetrizable matrix products. Two rectangular $s \times t$ *matrices* A *and* B *have symmetrizable matrix products* AB^T, A^TB, BA^T, *and* B^TA *if they are defined as*

$$A = FC, \quad B = CG, \tag{3}$$

where C *is a rectangular* $s \times t$ *matrix of rank* r, F *is a positive definite* $s \times s$ *matrix, and* G *is a positive definite* $t \times t$ *matrix*; $0 < r \leqslant s \leqslant t$.

Theorem

(Eigenvalues and eigenvectors of symmetrizable matrix products.) For every two rectangular $s \times t$ *matrices* $A = FC$ *and* $B = CG$ *specified as in* (3) *there exists an* $s \times r$ *matrix* U_+ *of rank* r, *an* $t \times r$ *matrix* V_+ *of rank* r, *an* $r \times r$ *positive definite diagonal matrix* $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ *such that we get after defining*

$$W_+ = FU_+, \quad Z_+ = GV_+ \tag{4}$$

(*i*) *a decomposition and diagonal form of the rectangular matrix* C:

$$C = U_+\Lambda V_+^T, \quad \Lambda = W_+^T C Z_+; \tag{5}$$

(*ii*) *a diagonal form of each of the positive definite matrices* F, F^{-1}, *and* G, G^{-1}:

$$I_r = U_+^T F U_+ = V_+^T G V_+ = W_+^T F^{-1} W_+ = Z_+^T G^{-1} Z_+; \tag{6}$$

(*iii*) *the same matrix* K *of positive eigenvalues for each of the symmetrizable matrix products* BA^T, A^TB, AB^T, *and* B^TA:

$$BA^TU_+ = U_+K, \quad A^TBV_+ = V_+K, \tag{7}$$

$$AB^TW_+ = W_+K, \quad B^TAZ_+ = Z_+K, \tag{8}$$

where $K = \mathrm{diag}(\kappa_1, \ldots, \kappa_r) = \Lambda^2$;

(*iv*) *a decomposition and diagonal form of each of the two symmetric matrices* CGC^T, *and* C^TFC:

$$CGC^T = U_+KU_+^T, \quad C^TFC = V_+KV_+^T, \tag{9}$$

$$K = Z_+^T C^T F C Z_+ = W_+^T C G C^T W_+, \tag{10}$$

(*v*) *one of the matrices of eigenvectors* U_+, V_+, W_+ *and* Z_+ *together with the positive eigenvalues of the symmetrizable products determine all the other three matrices*:

$$V_+\Lambda = A^TU_+, \quad U_+\Lambda = BV_+, \tag{11}$$

$$Z_+\Lambda = B^TW_+, \quad W_+\Lambda = AZ_+. \tag{12}$$

Proof. The application of the lemma to the matrix C yields the existence of U_+ and V_+ satisfying (6) such that $C = U_+\Lambda V_+^T$. The relations in (11) are then a direct consequence of (2); those in (12) follow from (11) with (4) and (3). The definition of W_+ and Z_+ in (4) provides the link between equations (11) and (12). By using (4) the diagonalization of F^{-1} in terms of W_+ and of G^{-1} in terms of Z_+ follows from that of F in terms of U_+ and of G

in terms of V_+. The diagonal form of C in (5) is a consequence of the decomposition of C and of (4) and (6):

$$W_+^T CZ_+ = W_+^T(U_+ \Lambda V_+^T)Z_+ = (U_+^T FU_+)\Lambda(Z_+^T GZ_+) = \Lambda.$$

The relations (7) and (8) can be verified by using (3), (4), (5) and (6). Finally, (10) and (9) result with (5), (6), (4), and $K = \Lambda^2$ since e.g.

$$C^T FC = V_+ \Lambda(U_+^T FU_+)\Lambda V_+^T = V_+ KV_+^T,$$

$$Z_+^T C^T FCZ_+ = (Z_+^T V_+)K(V_+^T Z_+) = K,$$

so that the proof is complete. □

The matrices U_+, V_+, W_+ and Z_+ are not uniquely determined by equations (7) and (8); in particular the solutions need not satisfy (6). However, for every solution U_+ to $BA^T U_+ = U_+ K$, say, the matrix $L = U_+^T FU_+$ is positive definite, so that the square root can be obtained and $\tilde{U}_+ = U_+ L^{-1/2}$ is a solution which satisfies $BA^T \tilde{U}_+ = \tilde{U}_+ K$ and $\tilde{U}_+^T F\tilde{U}_+ = I_r$ as well. If all eigenvalues of BA^T, i.e. all diagonal elements of K, are distinct then the matrix L is diagonal and hence computing $L^{1/2}$ just means taking square roots of the r positive diagonal elements. In the general case L has block-diagonal form with non-zero elements at most in submatrices along the main diagonal of a size equal to the multiplicity of the corresponding eigenvalue.

A major numerical gain of the theorem is in situations with widely different dimensions of rows and columns: if e.g. $r = s = 4$ and $t = 100$ then the eigenanalysis of $B^T A$ to determine Z_+ and K would involve computations with a 100×100 matrix, while (8) and (12) show that only an eigenanalysis of the 4×4 matrix AB^T is needed.

Even if $r = s = t$ use of the theorem may yield considerable simplifications. For instance, the eigenvalues and the matrix of eigenvectors Z_+ of a matrix $B^T A$ can be given in closed form if AB^T is a diagonal matrix K having the distinct positive eigenvalues of $B^T A$ as diagonal elements.

3. Properties of some statistical techniques

With the result in the previous section properties of several statistical techniques may be derived, so that these properties need no longer be justified within the specific context.

3.1. *The dual of principal component analysis* (Hotelling, 1933; Gower, 1966)

Let X be an $n \times q$ data matrix composed entirely of variates with zero sample means, i.e. the i-th element in column j of X is the observation for the i-th individual in the sample on the j-th variable, recorded as deviation from the variable's sample mean: $x_{ij} - \bar{x}_j$. A principal component analysis of the variables involves then the eigenanalysis of the matrix $X^T X$ and yields the diagonal matrix K_q of eigenvalues and a $q \times q$ matrix U of corresponding eigenvectors. The eigenanalysis of the typically much larger matrix XX^T has been called the dual to the principal component analysis of the variables or the principal component analysis of the individuals. After defining $U = (U_+, U_0)$ as the eigenvectors corresponding to the $r \leqslant q$ non-zero eigenvalues of K_q in K say and to the $q - r$ zero eigenvalues, respectively, we obtain from (11) with $s = q$, $t = n$, $F = I_s$, $G = I_t$, and $C = X^T$ that the matrix of eigenvectors V_+ of the non-zero eigenvalues of XX^T is determined by K and U_+ as $V_+ = XU_+ K^{-1}$. This specializes to the result by Gower if $X^T X$ has full rank q and column j in V_+ is normalized to have as length the j-th eigenvalue κ_j. An interpretation is that principal components of the individuals is an inflated summary of the information provided by principal components of the variables.

3.2. Hirschfeld's theorem for correspondence analysis (Hirschfeld, 1935; de Leeuw, 1988)

Let C be an $s \times t$ matrix of counts of two discrete variables with s and t categories, F^{-1} be a diagonal matrix with the s marginal counts of the first variable along the diagonal, and G^{-1} a diagonal matrix with the t marginal counts of the second variable. Then Hirschfeld's (1935) theorem restated in this notation says that there exists an $s \times r$ matrix W_+ and an $t \times r$ matrix Z_+ such that the following two equations hold simultaneously for an $r \times r$ diagonal matrix Λ:

$$FCZ_+ = W_+\Lambda, \quad GC^TW_+ = Z_+\Lambda.$$

These are just the equations in (12) with $A = FC$ and $B = CG$. An interpretation is that the solutions W_+ and Z_+ permit a rescaling of the two discrete variables so that the association between them becomes linear.

3.3. Properties of Hotelling's canonical variables (Hotelling, 1936; Dempster, 1969, pp. 98–100; Rao, 1973, pp. 582–585; Chambers, 1977, p. 126)

Let Y be an $p \times 1$ vector of variables and X be an $q \times 1$ vector of variables both measured in deviations from their means so that the covariance matrices (Σ_{yy} and Σ_{xx}) and covariances (Σ_{yx}) of Y and X are given by

$$\Sigma_{yy} = E(YY^T), \quad \Sigma_{xx} = E(XX^T), \quad \Sigma_{yx} = E(YX^T),$$

then Hotelling's canonical variables $\tilde{Y} = Z^TY$ and $\tilde{X} = W^TX$ are the linearly transformed variables obtained from the solutions Z and W of

$$\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}Z = Z\Lambda_p, \quad \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}W = W\Lambda_q.$$

Some of the properties of these canonical variables are obtainable from (4) to (12) since W and Z are the matrices of eigenvectors of symmetrizable matrix products. By taking $F = \Sigma_{xx}^{-1}$, $C = \Sigma_{xy} = \Sigma_{yx}^T$, and $G = \Sigma_{yy}^{-1}$ it follows in particular from (5) and (6) that the covariance matrices ($\Sigma_{\tilde{y}\tilde{y}}$ and $\Sigma_{\tilde{x}\tilde{x}}$) and covariances ($\Sigma_{\tilde{y}\tilde{x}}$) of the canonical variables corresponding to non-zero eigenvalues are given by

$$\begin{pmatrix} \Sigma_{\tilde{y}\tilde{y}} & \Sigma_{\tilde{y}\tilde{x}} \\ \cdot & \Sigma_{\tilde{x}\tilde{x}} \end{pmatrix} = \begin{pmatrix} Z_+^T\Sigma_{yy}Z_+ & Z_+^T\Sigma_{yx}W_+ \\ \cdot & W_+^T\Sigma_{xx}W_+ \end{pmatrix} = \begin{pmatrix} I_r & \Lambda \\ \cdot & I_r \end{pmatrix}, \tag{13}$$

where as before we have $Z = (Z_+, Z_0)$ and $W = (W_+, W_0)$, i.e. Z_0 and W_0 correspond to zero eigenvalues of multiplicities $p - r$ and $q - r$. The interpretation is here that Z_0^TY and W_0^TX would just give those components of the canonical variables which are unimportant since they are uncorrelated, while the $r \times r$ matrix Λ contains the r non-zero canonical correlations along the diagonal. In the case the X-variables are uncorrelated, i.e. if Σ_{xx} is diagonal and W_+ in (6) is the identity matrix, it follows from (12) that the squared canonical correlations are the diagonal elements of AB^T, i.e. $\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} = AB^T = \Lambda^2$. Conversely, if AB^T is a diagonal matrix of distinct diagonal elements, then (8) implies that they are the squared canonical correlations and that $W_+ = I_q$, and (12) implies that each column of Z_+ is proportional to the regression coefficients obtained when regressing X_i on Y so that the corresponding squared canonical correlation is the coefficient of determination associated with this regression.

3.4. *Properties of derived response variables with special relations of conditional linear independence to a set of explanatory variables* (Cox & Wermuth, 1992; Wermuth & Cox 1993)

Let again, as above, Y be an $p \times 1$ vector of variables and X be an $q \times 1$ vector of variables both measured in deviations from their means, but assume—in contrast to the situation appropriate for computing canonical variables—that only transformations for the response variables Y are considered while the variables X are thought of as being explanatory variables with a strong individual identity so that they should remain untransformed. An $q \times 1$ response Y^* is desired such that each new response Y_i^* has linear conditional independence of all explanatory variables except one, i.e. $Y_i^* \perp (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_q) \mid X_i$. If $p = q$ the new vector Y^* is obtained by requiring that the matrix of regression coefficients of Y^* on X is the identity matrix so that in particular the regression of Y_i^* on X involves only X_i. If however $p > q$ a unique solution, in a reasonable sense optimal, is only achieved after first reducing Y to the $q \times 1$ vector $\tilde{Y}$ of variables in the canonical regression of Y on X. The matrix of non-zero regression coefficients of $\tilde{Y}$ on X can be written as $Z_+^T \Sigma_{yx} \Sigma_{xx}^{-1}$ so that the derived responses become

$$Y^* = \Sigma_{xx}(Z_+^T \Sigma_{yx})^{-1} \tilde{Y} = \Sigma_{xx}(Z_+^T \Sigma_{yx})^{-1} Z_+^T Y. \tag{14}$$

It follows from the diagonal form given in (5) for $C^T = \Sigma_{yx}$ that these derived responses can only be obtained if all canonical correlations are non-zero, that is if $r = q$ so that W_+ is a $q \times q$ matrix of full rank and $(Z_+^T \Sigma_{yx})^{-1} = W_+ \Lambda^{-1}$. The covariance matrix of Y^* and X is then:

$$\begin{pmatrix} \Sigma_{y^*y^*} & \Sigma_{y^*x} \\ \cdot & \Sigma_{xx} \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} W_+ K^{-1} W_+^T \Sigma_{xx} & \Sigma_{xx} \\ \cdot & \Sigma_{xx} \end{pmatrix}.$$

This implies in particular that the joint correlation matrix of derived responses and explanatory variables coincides with the correlation matrix of canonical variables if (1) the explanatory variables are uncorrelated so that Σ_{xx} is diagonal and (2) all canonical correlations are distinct so that $W_+ = I_q$.

In the case $r = q$ it follows from (12) and (10) that further equivalent expressions for the derived responses are

$$Y^* = \Sigma_{xx}(W_+ K^{-1} W_+^T)\Sigma_{xy} \Sigma_{yy}^{-1} Y = \Sigma_{xx}(\Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx})^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} Y,$$

the second expression may be simpler to use for some purposes.

References

Chambers, J. (1977). *Computational methods for data analysis*. Wiley, New York.

Cox, D. R. & Wermuth, N. (1992). On the calculation of derived variables in the analyses of multivariate responses. *J. Multivar. Anal.* **42**, 162–170.

Dempster, A. P. (1969). *Elements of continuous multivariate analysis*. Addison-Wesley, Reading.

de Leeuw, J. (1988). Multivariate analysis with linearizable regressions. *Psychometrika* **53**, 437–454.

Eckart, C. & Young, G. (1931). The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–217.

Gower, J. C. (1966). Some distance properties of latent roots and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Proc. Camb. Phil. Soc.* **31**, 520–524.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educat. Psychol.* **24**, 417–441, 498–520.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321–377.

Kato, T. (1966). *Perturbation theory for linear operators*. Springer, New York.

Lanczos, C. (1958). Linear systems in self-adjoint form. *Amer. Math. Month.* **65**, 665–679.
Paige, C. C. (1985). The general linear model and the generalized singular value decomposition. *Lin. Algeb. Applicat.* **70**, 269–284.
Rao, C. R. (1973). *Linear statistical inference and its applications.* Wiley, New York.
Schmidt, E. (1907). Entwicklung willkürlicher Funktionen aus Systemen vorgeschriebener. *Math. Ann.* **63**, 433–476.
Schwerdtfeger, H. (1960). Direct proof of Lanczos' decomposition theorem. *Amer. Math. Month.* **67**, 855–860.
Searle, S. R. (1982). *Matrix algebra useful for statistics.* Wiley, New York.
Wermuth, N. & Cox, D. R. (1993). Derived variables calculated from similar joint responses: some characteristics and examples. *Comp. Statist. Data Anal.* (to appear).
Zaanen, A. C. (1953). *Linear analysis.* North-Holland, Amsterdam.

Received April 1992, in final form April 1993

Nanny Wermuth, Psychologisches Institut der Universität Mainz, 55099 Mainz, Germany.

Appendix

In order to prove the singular value decomposition of a rectangular matrix by following arguments of Schmidt (1907), we first give some notation and facts.

An $s \times s$ symmetric matrix M, i.e. a matrix which is equal to its transpose ($M = M^T$), is said to be *positive definite* if for all non-null $s \times 1$ vectors x the quadratic form x^TMx is positive and it is termed *non-negative definite* if $x^TMx \geqslant 0$.

Rank of MM^T. The rank $r \leqslant \min\{s, t\}$ of the matrix product MM^T of a rectangular $s \times t$ matrix M is equal to the number of linearly independent rows or columns of M since $\{x \mid MM^Tx = 0\} = \{x \mid (M^Tx)^T(M^Tx) = 0\} = \{x \mid Mx = 0\}$.

Orthogonal decomposition and diagonal form of a symmetric matrix. For every $s \times s$ symmetric matrix M it is possible to find an orthogonal $s \times s$ matrix U, i.e. a matrix which pre- or post-multiplied by its transpose gives the identity matrix ($I_s = UU^T = U^TU$) and a diagonal matrix $K = \operatorname{diag}(\kappa_1, \ldots, \kappa_s)$ such that $M = UKU^T$, $U^TMU = K$, where $\kappa_1, \ldots, \kappa_s$ are the not necessarily distinct eigenvalues of M and the columns of U are the corresponding eigenvectors, i.e. U satisfies $MU = UK$. This implies in particular that there is an orthogonal transformation $x = Uy$ such that $x^TMx = y^TU^TMUy = \Sigma\, \kappa_i y_i^2$, so that the eigenvalues of a non-negative definite matrix M are either zero or positive and the eigenvalues of a positive definite matrix M are all positive; if M is positive definite its inverse M^{-1} exists.

Square root of a non-negative definite matrix. For every $s \times s$ non-negative definite matrix M a *square root* $M^{1/2}$ can be found which returns M if squared, i.e. $M^{1/2}M^{1/2} = M$. In terms of the orthogonal decomposition of M it is defined as $M^{1/2} = U \operatorname{diag}(\sqrt{\kappa_1}, \ldots, \sqrt{\kappa_s})U^T$.

Lemma (Singular value decomposition: decomposition of a rectangular matrix with column-orthogonal matrices and its diagonal form.) For every $s \times t$ matrix M of rank r is is possible to find an $s \times r$ column-orthogonal matrix U_+, i.e. $U_+^T U_+ = I_r$, an $t \times r$ column-orthogonal V_+, i.e. $V_+^T V_+ = I_r$, and an $r \times r$ positive definite diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ such that

$$M = U_+ \Lambda V_+^T, \quad U_+^T M V_+ = \Lambda, \tag{A1}$$

where $\lambda_1, \ldots, \lambda_r$ are the singular values of M, i.e. the positive square roots of the non-zero eigenvalues of MM^T, and V_+ and U_+ satisfy

$$V_+\Lambda = M^TU_+, \quad U_+\Lambda = MV_+. \tag{A2}$$

Proof. The matrix MM^T is symmetric and non-negative definite, hence there exists an orthogonal decomposition such that $MM^TU = UK_s$, where K_s can be written as

$$K_s = \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix}, \quad K = \operatorname{diag}(\kappa_1, \ldots, \kappa_r) > 0,$$

and r is the rank of M. After partitioning $U = (U_+, U_0)$ so that U_+ corresponds to the positive eigenvalues and U_0 corresponds to the zero eigenvalues we get

$$MM^T U_+ = U_+ K, \quad U_+^T U_+ = I_r \tag{A3}$$

and after defining $V_+ = M^T U_+ K^{-1/2}$ and $K^{1/2} = \Lambda$ we have

$$(U_+^T M) V_+ = U_+^T (MM^T U_+) K^{-1/2} = U_+^T U_+ K^{1/2} = \Lambda,$$

and V_+ is column-orthogonal since

$$V_+^T V_+ = K^{-1/2} U_+^T (MM^T U_+) K^{-1/2} = K^{-1/2} U_+^T U_+ K^{1/2} = I_r.$$

The matrix V_+ can be completed to an $t \times t$ orthogonal matrix $V = (V_+, V_0)$ by solving the equations $MV_0 = 0$ and $V_0^T V_0 = I_{t-r}$ for the $t \times (t-r)$ column-orthogonal V_0. Such a solution exists because the rank of M is r, i.e. the equation $Mv = 0$ has $t - r$ linearly independent solutions which form the columns of V_0 after orthonormalizing, i.e. after rescaling so that they are orthogonal and have length one. Then

$$M = U(U^T M V) V^T = (U_+, U_0) \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_+^T \\ V_0^T \end{pmatrix} = U_+ \Lambda V_+^T,$$

which completes the proof of (A1). The first statement of (A2) holds by definition of V_+ and the second follows from this definition replacing $M^T U_+$ by $V_+ K^{1/2}$ in the first statement of (A3). □