

NONLINEAR PRINCIPAL COMPONENT ANALYSIS

JAN DE LEEUW

ABSTRACT. Two quite different forms of nonlinear principal component analysis have been proposed in the literature. The first one is associated with the names of Guttman, Burt, Hayashi, Benzécri, McDonald, De Leeuw, Hill, Nishisato. We call it *multiple correspondence analysis*. The second form has been discussed by Kruskal, Shepard, Roskam, Takane, Young, De Leeuw, Winsberg, Ramsay. We call it *nonmetric principal component analysis*. The two forms have been related and combined, both geometrically and computationally, by Albert Gifi. In this paper we discuss the relationships in more detail, and propose an alternative algorithm for nonlinear principal component analysis which combines features of both previous approaches.

This paper was originally presented at *COMPSTAT 1982* in Toulouse, France. It was published previously in H. Caussinus, P. Ettinger, and R. Tomassone (eds) *COMPSTAT 1982*, Wien, Physika Verlag, 1982, 77-86.

1. INTRODUCTION

We suppose that the reader is familiar with multiple correspondence analysis and with nonmetric principal component analysis. For multiple correspondence analysis we refer to Cazes et al. [1977] and to Lebart et al.

2000 *Mathematics Subject Classification*. 62H25.

Key words and phrases. correspondence analysis, nonmetric scaling, multidimensional scaling, eigenvalues, principal component analysis.

[1977] or to Hill [1974] and Nishisato [1980]. We only discuss a comparatively simple case, the many interesting generalizations developed especially in Toulouse could in principle also be fitted into our framework. For nonmetric principal component analysis we refer to Kruskal and Shepard [1974] or Young et al. [1978]. Again we study a simple special case of this technique. For previous attempts to integrate the two approaches we refer to De Leeuw and Van Rijckevorsel [1980], De Leeuw et al. [1981] and Gifi [1981].

2. NONMETRIC PRINCIPAL COMPONENT ANALYSIS

Suppose $\mathcal{L}_1 \cdots, \mathcal{L}_m$ are closed subspaces of a separable Hilbert space \mathcal{H} , with inner product $\langle \cdot, \cdot \rangle$, norm $\|\cdot\|$, and unit sphere \mathcal{S} . For each choice of elements $y_j \in \mathcal{L}_j \mathcal{S}$, the intersection of \mathcal{L}_j and \mathcal{S} , we can compute the matrix $R(y_1, \cdots, y_m)$ with elements $r_{j\ell}(y_1, \cdots, y_m) = \langle y_j, y_\ell \rangle$. This matrix is a *correlation matrix*, in the sense that it is positive semi-definite and has diagonal elements equal to unity. The problem of *nonmetric principal component analysis* (NCA) is to find $y_j \in \mathcal{L}_j \mathcal{S}$ in such a way that the sum of the p largest eigenvalues of the matrix $R(y_1, \cdots, y_m)$ is maximized (or, equivalently, the sum of the $m-p$ smallest eigenvalues is minimized). Observe that for different choices of p this defines a different problem. In some cases we shall not only be interested in solutions that maximize our criterion, but we shall be interested in all solutions of the stationary equations corresponding with the maximization problem.

It is clear that NCA generalizes ordinary principal component analysis, in which the subspaces \mathcal{L}_j are one-dimensional. Our formulas deal with a somewhat restricted form of NCA, because most of the literature we have mentioned treats the more general case in which \mathcal{L}_j is a closed convex cone.

Tenenhaus [1982] even discusses the case of a possibly infinite number of convex cones. There is another serious restriction of generality in most of our paper: we suppose that all \mathcal{L}_j are finite-dimensional. For notational convenience we shall even suppose that $\mathbf{dim}(\mathcal{L}_j)$ is the same for all j , but this last assumption is in no way essential. We use orthonormal bases for each of the \mathcal{L}_j and collect them in the $n \times q$ matrices G_j . Here q is the common dimensionality of the \mathcal{L}_j and n is the dimensionality of \mathcal{H} . We do not necessarily assume that n is finite. If n is not finite, then G_j is simply an ordered q -tuple of elements of \mathcal{H} , and all “matrix operations” we use have their obvious interpretations.

Thus $y_j = G_j \alpha_j$ is in $\mathcal{L}_j \mathcal{S}$ if and only if the q -vector α_j satisfies $\alpha_j' \alpha_j = 1$. Then $r_{j\ell}(y_1, \dots, y_m) = \alpha_j' C_{j\ell} \alpha_\ell$, where $C_{j\ell}$ is the $q \times q$ matrix defined by $C_{j\ell} = G_j' G_\ell$. Observe that for all j it is true that $C_{jj} = I$, the identity of order q . It is also convenient to define the $n \times mq$ supermatrix $G = (G_1 \mid \dots \mid G_m)$ and the $mq \times mq$ supermatrix $C = G'G$. The supermatrix C is sometimes called the *Burt table* of the NCA problem, observe that it depends on the choice of the bases.

A useful operation for our purposes is the *direct sum* of a number of matrices [McDuffee, 1946, p. 81]. If X and Y are matrices of dimension $a \times b$ and $c \times d$, then the direct sum is the $(a + c) \times (b + d)$ matrix

$$X \dot{+} Y = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}.$$

The extension to direct sums of more than two matrices is obvious. We now define A as the $mq \times rn$ matrix $\alpha_1 \dot{+} \dots \dot{+} \alpha_m$. In this notation $R(y_1, \dots, y_m) = A'CA$, provided that $A'A = I$.

It follows that $\sigma_p(y_1, \dots, y_m)$, the sum of the p largest eigenvalues of the correlation matrix $R(y_1, \dots, y_m)$ has the representation

$$(1) \quad \sigma_p(y_1, \dots, y_m) = \max\{\mathbf{tr} T'A'CAT\}$$

with T varying over the $rn \times p$ matrices satisfying $T'T = I$. Thus maximization of $\sigma_p(y_1, \dots, y_m)$ over $y_j \in \mathcal{L}_j\mathcal{S}$ amounts to the same thing as maximization of $\mathbf{tr} T'A'CAT$ over all $m \times p$ matrices T satisfying $T'T = I$ and all $mq \times m$ matrices A of the form $A = \alpha_1 + \dots + \alpha_m$ also satisfying $A'A = I$. If A and T satisfy these restrictions, and U is the $mq \times p$ matrix AT , then $U'U = I$, and U consists of m submatrices U_j , of dimension $q \times p$, and of the form $U_j = \alpha_j t'_j$, with t_j row j of T . Thus U is *blockwise of rank one*, each subspace \mathcal{L}_j defines a block.

We can now define

$$(2) \quad \sigma_p(\mathcal{L}_1, \dots, \mathcal{L}_m) = \max \mathbf{tr} U'CU$$

with U varying over all orthonormal $mq \times p$ matrices which are blockwise of rank one. The NCA problem is to compute $\sigma_p(\mathcal{L}_1, \dots, \mathcal{L}_m)$ and to find the maximizer U of the required form.

We now derive the stationary equations of the NCA problem. From (1) we must have

$$(3) \quad RT = T\Omega,$$

with Ω a symmetric matrix of order p . For identification purposes we can actually assume that Ω is diagonal. Also remember that $R = A'CA$. Differentiation of (1) with respect to the α_j gives the equations

$$(4) \quad \sum_{\ell=1}^m \gamma_{j\ell} C_{j\ell} \alpha_\ell = \theta_j \alpha_j$$

where $\Gamma = \{\gamma_{j\ell}\}$ is defined by $\Gamma = TT'$, where the θ_j are undetermined multipliers, and where the α_j satisfy $\alpha'_j\alpha_j = 1$. It follows from (3) and (4) that at a solution $\Theta = \mathbf{diag}(T\Omega T')$, and thus $(tr)(\Theta) = \mathbf{tr}(\Omega)$. Most algorithms for NCA are based on alternating solution of (3) for T , given A , and solution of (4) for A , given T .

Although (3) and (4) can be used to construct convergent algorithms for finding stationary points [Gifi, 1981] they give little insight into the mathematical structure of the NCA problem. It is not at all clear, for example, if (3) and (4) have more than one solution. and if so, how these different solutions are related. There is one fortunate exception. If $p = 1$, i.e. if we choose $y_j \in \mathcal{L}_j\mathcal{S}$ in such a way that the largest eigenvalue of $R(y_1, \dots, y_m)$ is maximized. then the requirement that U is blockwise of rank one is no restriction at all, and thus the problem is maximizing $u'Cu$ over $u'u = I$, where we have written u instead of U , because U is of dimensions $mq \times 1$. The solutions to the NCA problem with $p = 1$ are consequently the eigenvectors of C . A little reflection shows that $p = m - 1$, i.e. the solution for $y_j \in \mathcal{L}_j\mathcal{S}$ minimizing the smallest eigenvalue of $R(y_1, \dots, y_m)$ gives the identical result. If v is an eigenvector of C , $v'v = 1$, eigenvalue μ , then we can partition it into blocks v_j , each with q elements. If a block is nonzero we set $\alpha_j = \frac{v_j}{v'_jv_j}$, if a block is zero α_j is an arbitrary unit length vector. Moreover $t_j = \sqrt{v'_jv_j}$, i.e. $u_j = v_j$. Remember that in this case T is $m \times 1$, its elements are written simply as t_j . Also $\theta_j = \mu v'_jv_j$ and $\omega = \mu$. It is of some interest to observe that μ is not always the largest eigenvalue of the corresponding matrix R with elements $r_{j\ell} = \alpha'_jC_{j\ell}\alpha_\ell$. More precisely: if μ is the largest eigenvalue of C then $\omega = \mu$ is also the largest eigenvalue of the corresponding R , and the same thing is true for the smallest eigenvalue of C . But no such thing is true for the intermediate eigenvalues.

3. MULTIPLE CORRESPONDENCE ANALYSIS

The last paragraph brings us naturally to *multiple correspondence analysis* (MCA). This is defined simply as the computation of some or all of the eigenvalues and eigenvectors of C . Accordingly each solution of

$$(5) \quad Cv = \mu v$$

will be called an MCA solution. In a more restricted sense we can also call the problem of computing

$$(6) \quad \rho_p(\mathcal{L}_1, \dots, \mathcal{L}_m) = \max \mathbf{tr} V'CV$$

with V varying over all orthonormal $mq \times p$ matrices the MCA problem. This makes it more similar to the NCA problem. The main difference between the two, from this point of view, is that the MCA problem is *nested*, i.e. the solution for $p - 1$ are the first $p - 1$ dimensions of the solution for p . Although each choice of p defines a different MCA problem the various solutions are closely related. In fact they can all be constructed from the mq solutions of (5). Another obvious difference between NCA and MCA is that computing (6) gives p different elements of each $\mathcal{L}_j\mathcal{S}$, and thus p different correlation matrices. Each of these correlation matrices has at least one eigenvalue equal to the corresponding μ of (5), and usually only one. Computing (2) on the other hand gives only one solution for the y_j , thus only one correlation matrix, of which the p largest eigenvalues are equal to $\Omega = U'CU = T'RT$. In Gifi's terminology there is *multiple selection* from $\mathcal{L}_j\mathcal{S}$ in MCA, there is *single selection* from $\mathcal{L}_j\mathcal{S}$ in NCA (for *selection* Gifi also uses *transformation* or *quantification*. depending on the context, cf infra).

Before we study the relations between NCA and MCA in more detail, we must make one important comment. Interpreting NCA and MCA as generalizations of ordinary principal component analysis is only one possibility. We then use the fact that these techniques can be interpreted as optimizing some property of the correlation matrix before performing the actual principal component analysis. This particular interpretation is natural in some contexts, for example if \mathcal{L}_j is defined as a subspace of possible transformations or quantifications of an element of \mathcal{H} . Sometimes, however, the notion of quantification or transformation does not make much sense, because the variable is purely nominal and a one-dimensional arrangement of its possible values is somewhat far-fetched. In these cases it is usually more sensible to interpret NCA and MCA as multidimensional scaling techniques which map objects or individuals into low-dimensional space in such a way that some criterion defined in terms of distance and/or separation is optimized. Such interpretation are possible and fruitful, they are discussed in detail by Gifi [1981] and Heiser [1981].

In the previous section we have already one relationship between NCA and MCA. If $p = 1$ then NCA and MCA are the same, basically because $m \times q$ matrices are always blockwise of rank one. If $p > 1$ then comparing (2) and (6) shows directly that $\rho_p(\mathcal{L}_1, \dots, \mathcal{L}_m) \geq \sigma_p(\mathcal{L}_1, \dots, \mathcal{L}_m)$ with equality if and only if the eigenvectors corresponding with the p largest eigenvalues of C are blockwise of rank one. In the next section we shall try to find out if this condition for equality is ever likely to be met in practice. We shall also discuss another problem connected with MCA. The $m \times q$ solutions to (5) give $m \times q$ correlation matrices, and thus $m \times q$ principal component analyses. Gifi calls this *data production* in order to distinguish it from *data reduction*. A pertinent question is if the $m \times q$ correlation matrices are related

in some simple way, so that it is easy to see where the redundancy in this representation is.

4. RELATIONSHIPS IN AN IMPORTANT SPECIAL CASE

Suppose that there exist α_j and $r_{j\ell}$ such that $\alpha'_j \alpha_j = 1$ for all j and

$$(7) \quad C_{j\ell} \alpha_\ell = r_{j\ell} \alpha_j$$

for all j, ℓ . Suppose moreover that T is of order m , satisfies $TT' = T'T = I$, and $RT = T\Omega$, with Ω of order m and diagonal. Thus (7) is the critical assumption, it defines α_j and $r_{j\ell}$, T is then defined simply as the matrix of eigenvectors of R . Now define $U = AT$, with $A = \alpha_1 + \dots + \alpha_m$. By using (7) it is now simple to verify that $CU = U\Omega$. Thus U constructed in this way defines m solutions to (5), i.e. to MCA or to NCA with $p = 1$. The m solutions to (5) moreover induce the same correlation matrix R . If we select p columns from T , then these p columns obviously satisfy (3). We can also verify that they satisfy (4), if taken together with the α_j , and that consequently we can construct $\binom{m}{p}$ solutions to the NCA equations (3) and (4) if (7) can be satisfied. All these NCA solutions have the same R .

Now suppose v is another solution of (5), not one of the m solutions constructed by using the α_j of (7). By orthogonality $\sum v'_j U_j = \sum v'_j \alpha_j t'_j = 0$, which is possible only if $v'_j \alpha_j = 0$ for all j . Thus, in the terminology suggested by Dauxois and Pousse [1976], v is not only *weakly* but actually *strongly orthogonal* to the m columns of U . In particular this implies that if there is a second solution of (7), i.e. if $C_{j\ell} \beta_\ell = s_{j\ell} \beta_j$ for all j, ℓ , then $\alpha'_j \beta_j = 0$ for all j . Such a second solution of (7) again defines m solutions to (5) and $\binom{m}{p}$ solutions to (3) and (4). In total, of course, (7) can only have

q solutions, all strongly orthogonal, each of them corresponding with m solutions of (5), and each of them having a single induced correlation matrix R . If we find q solutions to (7) then we have found mq , and consequently all, solutions to (5). We have also found $q\binom{m}{p}$ solutions to (3) and (4), but this need not be all possible solutions.

We now investigate if (7) is a realistic condition, with interesting interpretations. In the first place we repeat (7) in words: it says that if two of the matrices $C_{j\ell}$ have a subscript in common, then they have a singular vector in common. Another interpretation of (7) is also quite useful. Suppose α_j and $r_{j\ell}$ satisfy (7) and define $y_j = G_j\alpha_j$. Then (7) says that the projection of y_ℓ on \mathcal{L}_j is equal to the projection of y_ℓ on the one-dimensional subspace through y_j . In the next section this interpretation will be related to linearity of the regression between two random variables.

A condition stronger than (7) is that there exist m matrices K_j orthonormal, of order q , and diagonal matrices $D_{j\ell}$ such that

$$(8) \quad C_{j\ell}K_\ell = K_jD_{j\ell}$$

This is equivalent to the condition that (7) has q different solutions, it says that if two matrices $C_{j\ell}$ have a subscript in common then they must have all their singular vectors in common. By using familiar results this can be reduced to the condition that several matrices must commute [Bellman, 1960, p. 56], but in itself this commutation result does not give much insight. We know already that if (8) is true then we can construct all mq solution to (5). We now show how this can be done systematically.

Suppose (8) is satisfied. Let $K = K_1 + \dots + K_m$. Then $K'CK$ has submatrices $D_{j\ell} = K_j' C_{j\ell} K_\ell$, and as we know these submatrices are all diagonal. Moreover they are all of order q , there are m^2 of these $D_{j\ell}$, they all have

q nonzero elements at most. It is now possible to construct a permutation matrix P such that $P'DP = P'K'CKP = E$ is of the form $E = E_+ \cdots + E_q$, where each of the E_t is a symmetric matrix of order m . Thus D consists of m^2 matrices, each with at most q nonzero elements, and E consists of q matrices, each with at most m^2 nonzero elements. Element j, ℓ of matrix E_t is equal to element t, t of matrix $D_{j\ell}$. As the next step in our solving of (5) we construct $L = L_1 + \cdots + L_q$ with each L_t orthonormal of order m , such that L_t diagonalizes E_t . Thus $F = L'EL = L'P'DPL = L'P'K'CKPL$ is of the form $F = F_1 + \cdots + F_q$, with all of the F_t diagonal. It follows that KPL diagonalizes C , thus KPL contains the eigenvectors of C , and the matrices F_t contain eigenvalues in some order. If we look somewhat closer to KPL , we find that it consists of mq submatrices, all of order $q \times m$, and all of rank one. Conversely this particular block structure of the eigenvectors of C , together with the rank one property of the blocks, implies that (8) is satisfied. It also implies that there is no data production, there are only q different induced correlation matrices, not mq . We also know that this block structure makes it possible to find $q \binom{m}{p}$ solutions to (3) and (4). It is a useful exercise to find out what happens to the results of this section if the \mathcal{L}_j have different dimensionalities, making some $C_{j\ell}$ rectangular instead of square.

5. EVEN MORE SPECIAL CASES

We now apply the three-step or KPL -diagonalization in some cases in which we can easily show that it works. If $m = 2$ we use the singular value decomposition $C_{12} = K_1 D_{12} K_2'$. Here D_{12} is diagonal of order q , with singular values on the diagonal. Thus the E_t are correlation matrices of order two, each E_t has one singular value as off-diagonal element. The F_t are diagonal of order two, the two elements are one plus the singular value and one minus

the singular value. This is the familiar result linking multiple correspondence analysis with two variables with ordinary correspondence analysis of the cross table of the two variables. We emphasize that in ordinary correspondence analysis we often use a very special basis of indicator functions or step functions. This basis has some attractive properties, we mention for example the centroid principle of Benzecri, and we mention the fact that (7) can be interpreted as the assumption that all bivariate regressions can be linearized by suitable choice of the $y_j \in \mathcal{L}_j\mathcal{S}$. The stronger assumption (8) means that all bivariate distributions can be diagonalized simultaneously.

Now consider the case $q = 2$ with a basis of normalized indicators for each of the \mathcal{L}_j . Thus $C_{j\ell}$ is a 2×2 table with bivariate frequencies, divided by square root of the product of the univariate marginals. These $C_{j\ell}$ satisfy (8), one column of K_j is proportional to the square roots of the marginals, the other columns must be orthogonal to it and is consequently also determined completely by the marginals. It follows that $E = E_1 + E_2$ where E_1 is of order m and has all elements equal to one and where E_2 is likewise of order m and has as its elements the matrix of phi-coefficients or point correlations of the binary variables. Thus F_1 has one element on the diagonal equal to m and $m - 1$ diagonal elements equal to zero. The elements of F_2 are the eigenvalues of the matrix of phi-coefficients. This is the familiar result that multiple correspondence analysis or nonmetric principal component analysis of binary variables is the same thing as ordinary component analysis of phi-coefficients. A matrix E_1 of order m , with all elements +1, occurs whenever we use normalized indicators as a basis. Thus there is always a trivial eigenvalue equal to m and $m - 1$ trivial eigenvalues equal to zero in this case.

As a final special case suppose we have m standard normal variables, and suppose \mathcal{L}_j is the linear space of all polynomial transformations of degree not exceeding $q - 1$. As a basis of each of the \mathcal{L}_j we take the Hermite-Chebyshev polynomials of degree $0, \dots, q - 1$. It is well known that transform s of variable j and transform t of variable ℓ are uncorrelated (orthogonal) if $s \neq t$. If $s = t$ then their correlation is $\rho_{j\ell}^s$ where $\rho_{j\ell}$ is the correlation in the original multinormal distribution. Thus (8) is satisfied, E_1 has all elements equal to one, E_2 has elements $\rho_{j\ell}$, E_3 has elements $\rho_{j\ell}^2$ and so on. The eigenvalues of the MCA problem are those of E_1 and those of E_2 and so on. The largest nontrivial one is the largest eigenvalue of E_2 , the smallest nontrivial one is the smallest of E_2 [Styan, 1973]. but the order of the others is undecided. In fact the second largest MCA, eigenvalue can be the largest of E_3 , in which case all transformations on the second dimension are quadratic functions of the first (horse-shoe or Guttman effect). But the second largest MCA eigenvalue can also be the second largest eigenvalue of E_2 , the second set of transformations is then linear with the first. For the $q \binom{m}{p}$ NCA solutions we can compute from this representation the transformations in two dimensions are both linear, or both quadratic, and so on. NCA rank one restrictions make sure that both or all p transformations come from the same E_t [Gifi, 1981, Ch. 11]. The multinormal example has been generalized in many directions by Lancaster and his pupils [Lancaster, 1969]. It is remarkable that in all three special cases condition (8) is satisfied, and *KPL*-diagonalization consequently works. Our results indicate that in general NCA solutions can easily be interpreted in a principal component analysis framework, the same thing is consequently true for MCA solutions with $p = 1$. But multidimensional MCA solutions are difficult to interpret in this framework if (8) is not approximately true. The rest of our paper studies if (8) is approximately true in a wide variety of examples.

6. AN ALGORITHM

In order to investigate in how far (8) is satisfied in practical examples, i.e. in how far *KPL*-diagonalization is possible we have written an APL-algorithm PREHOM. A FORTRAN-version is currently being developed. The technique is fairly simple: if $F = L'P'K'CKPL$, then we want to maximize the sum of squares of the diagonal elements of F on the condition that $L = L_1 \dot{+} \dots \dot{+} L_q$ and $K = K_1 \dot{+} \dots \dot{+} K_m$, where all matrices in the direct sums are square orthonormal. Observe that P is a fixed permutation matrix. For fixed K we maximize the criterion by choosing L_r as the eigenvectors of E_r , the maximum value is the sum of squares of the elements of E_r , which means that we maximize our criterion by maximizing the sum of squares of all diagonal elements of all $D_{j\ell} = K'_j C_{j\ell} K_\ell$. This is done by a straightforward generalization of the Jacobi-procedure which cycles through all the relevant plane rotations. As an initial estimate of K_j we use the eigenvectors of $\sum C_{j\ell} C_{\ell j}$, summed over $\ell \neq j$. This initial estimate is usually very good, but for precise convergence the program usually needs another five to ten cycles of plane rotations.

We have analyzed a large number of examples with PREHOM, although APL-restrictions imply that our examples cannot be too large (C must have order less than 35). The conclusion of the examples is partly as we expect it to be: if the number of subspaces (variables) is small (three or four) then *KPL*-diagonalization works very well, and if the average number of categories (dimensionalities) is close to two then *KPL*-diagonalization also works well. Generally *KPL*-diagonalization seems to work best for ordinal or rating scale type variables, and much less well for purely nominal variables. It came as a surprise to us that even in the least favourable circumstances *KPL*-diagonalization still works quite well, and it can predict

the first three MCA-solutions (largest eigenvalues) and the last three MCA-solutions (smallest eigenvalues) very well.

PREHOM can be used to predict MCA-results very well. Our experience [Gifi, 1981, Ch. 13] suggest that NCA is often very similar to MCA with $p = 1$, and thus PREHOM can be supposed also to approximate NCA very well. It follows from these two conclusions that in ordinal and rating scale examples NCA and MCA actually compute the same solutions, but in a different order. The NCA eigenvalues can be found in the list of MCA eigenvalues, but they are only very rarely the p largest ones. If we want to relate MCA and NCA in the same dataset we can most efficiently do this by using PREHOM or a similar program.

There are four actual examples in the tables on the next page. Example 1 are three variables with 3 + 4 + 3 categories, 100 observations, from Burt's classical MCA-paper. The variables are somewhat between nominal and ordinal. Example 2 are 6 variables with 2 + 5 + 3 + 2 + 5 + 2 categories, 25 observations, describing screws, nails, and tacks, taken from John Hargigan's book on cluster analysis. Most of the variables are clearly nominal. Example 3 has 5 ordinal variables with 2 + 4 + 5 + 4 + 6 categories, about 30000 observations, data from a school career survey by the Dutch Central Bureau of Statistics. Example 4 are 8 rating scales with three categories each, 110 observations, taken from the Dutch Parliament Survey. All four tables have four columns: in the first column the eigenvalues of the E_t are given (without the trivial ones), the second column has the same numbers but ordered, the third column has the actual MCA eigenvalues, and the fourth column has diagonal elements of $V'KPL$, cosines between actual and predicted MCA eigenvectors.

REFERENCES

- R. Bellman. *Introduction to Matrix Analysis*. McGrawHill, New York, 1960.
- P. Cazes, A. Baumerdier, S. Bonnefous, and J.P. Pagès. Codage et Analyse des Tableaux Logiques. Cahiers du B.U.R.O. 27, 1977.
- J. Dauxois and A. Pousse. *Les Analyses Factorielles en Calcul des Probabilités et en Statistique: Essai d'Étude Synthétique*. PhD thesis, Université Paul-Sabatier, Toulouse, France, 1976.
- J. De Leeuw and J. Van Rijckevorsel. HOMALS and PRINCALS. In E. Diday et al., editor, *Data Analysis and Informatics*. North Holland Publishing Company, Amsterdam, 1980.
- J. De Leeuw, J. Van Rijckevorsel, and Van Der Wouden H. Nonlinear Principal Component Analysis Using B-Splines. *Methods of Operations Research*, 23:211–234, 1981.
- A. Gifi. Nonlinear Multivariate Analysis. Technical report, Department of Data Theory, University of Leiden, 1981.
- W.J. Heiser. *Unfolding Analysis of Proximity Data*. PhD thesis, University of Leiden, 1981.
- M.O. Hill. Correspondence Analysis: a Neglected Multivariate Method. *Applied Statistics*, 23:340–354, 1974.
- J.B. Kruskal and R.N. Shepard. A Nonmetric Variety of Linear Factor Analysis. *Psychometrika*, 39:123–157, 1974.
- H.O. Lancaster. *The Chi-squared Distribution*. Wiley, New York, New York, 1969.
- L. Lebart, A. Morineau, and N. Tabard. *Techniques de la Description Statistique*. Dunod, 1977.
- C.C. McDuffee. *The Theory of Matrices*. Chelsea, New York, 1946.

S. Nishisato. *Analysis of Categorical Data: Dual Scaling and its Applications*. University of Toronto Press, Toronto, Canada, 1980.

G.P. Styan. Hadamard Products and Multivariate Statistical Analysis. *Linear Algebra and its Applications*, 6:217–240, 1973.

M. Tenenhaus. Multiple Correspondence Analysis and Duality Schema: a Synthesis of Different Approaches. *Metron*, 40:289–302, 1982.

F.W. Young, Y. Takane, and J. De Leeuw. The Principal Components of Mixed Measurement Level Multivariate Data: an Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 45:279–281, 1978.

KPL	KPL↓	MCA	cos
2.61	2.61	2.61	1.00
0.39	1.90	1.90	1.00
0.00	1.42	1.43	1.00
1.90	0.78	0.78	0.99
0.78	0.58	0.58	0.99
0.33	0.39	0.49	0.78
1.42	0.33	0.22	0.77
0.58	0.00	0.00	0.98

TABLE 1. Burt Data

KPL	KPL↓	MCA	cos
3.68	3.68	3.73	0.99
1.40	2.04	2.21	0.84
0.68	1.83	1.97	0.59
0.23	1.49	1.67	0.69
0.01	1.40	1.18	0.59
0.01	0.68	0.77	0.57
2.04	0.67	0.51	0.75
0.67	0.51	0.50	0.63
0.29	0.29	0.33	0.75
1.49	0.23	0.11	0.89
0.51	0.17	0.00	1.00
1.83	0.01	0.00	1.00
0.17	0.01	0.00	1.00

TABLE 2. Hartigan Data

KPL	KPL↓	MCA	cos
2.61	2.61	2.62	1.00
1.01	1.99	1.98	1.00
0.84	1.84	1.84	1.00
0.41	1.01	1.03	0.71
0.13	1.01	1.02	0.52
1.99	1.00	1.00	0.61
1.00	1.00	1.00	0.91
0.75	1.00	1.00	0.54
0.26	0.99	0.98	0.66
1.84	0.98	0.97	0.68
1.00	0.84	0.83	0.96
0.98	0.75	0.75	0.99
0.18	0.41	0.41	0.97
1.01	0.26	0.27	0.90
0.99	0.18	0.19	0.92
1.00	0.13	0.10	0.87

TABLE 3. CBS data

KPL	KPL↓	MCA	cos
3.87	3.87	3.88	1.0
1.37	2.09	2.42	0.9
0.82	1.37	1.40	0.7
0.55	1.31	1.28	0.7
0.43	1.14	1.16	0.8
0.39	1.03	1.05	0.6
0.37	0.84	0.90	0.8
0.19	0.82	0.72	0.7
2.09	0.67	0.66	0.8
1.31	0.57	0.61	0.8
1.14	0.55	0.51	0.8
1.03	0.43	0.39	0.5
0.84	0.39	0.33	0.7
0.67	0.37	0.27	0.6
0.57	0.34	0.22	0.7
0.34	0.19	0.19	0.9

TABLE 4. Rating data