Methods of Operations Research 43, 1980

NON LINEAR PRINCIPAL COMPONENTS ANALYSIS WITH B-SPLINES

Jan DE LEEUW
Afdeling DATATHEORIE, Rijksuniversiteit Leiden, the Netherlands,
Jan VAN RIJCKEVORSEL and Hans VAN DER WOUDEN
Afdeling Onderzoek Assistentie, Erasmus Universiteit Rotterdam, the Netherlands.

ABSTRACT

A non linear principal components technique is presented. B-Splines are used as a basis for the non linear transformations. The method is a generalization of correspondence analysis, which uses discrete step functions as a basis. B-Splines give a smoother approximation than step functions and they are computationally more efficient than orthogonal polynomials. An application on the Thurstone cylinder data is presented.

Keywords: B-Splines, Principal Components Analysis, Data Analysis, Analyse des Correspondances.

1. INTRODUCTION

The technique described here is a generalization of correspondence analysis (Benzécri, 1973; Hill, 1974; Gifi, 1981; Bouroche & Saporta, 1980) which uses discrete step functions as a basis to approximate continuous non linear transformations. The approximations are optimal in the sense that they maximize the sum of first few dominant eigenvalues of the correlation matrix. The advantage is that the data do not have to fit the omnipresent linear mould which in many cases of practical data analysis is far too restricted in its assumptions.
Optimal continuous non linear transformations are very difficult to compute except for special cases like the bivariate

normal distribution (Van Rijckevorsel, Bettonvil & De Leeuw, 1980).
We propose here an approximation based on B-Splines (De Boor, 1978) which is a generalization of the step function approach. The advantages are that the approximation is still non linear but smoother than step functions. The jumps at the breakpoints in the data are not continuous for step functions, but they are continuous if we use spline functions with order > 1.
It seems to be the case that the computational loss of efficiency by using second order B-Splines is negligable, especially for a limited number of breakpoints.
The example shown is a version of the Thurstone cylinder problem commonly used in literature to gauge non linear principal components techniques (Kruskal & Shepard, 1974; Young, Takane & De Leeuw, 1978; Gifi, 1981).

2. PCA

2.1. PCA of numerical variables

There are a number of different ways in which the equations of PCA can be derived (Rao, 1980; Okamoto, 1968; Le Roux & Rouanet, 1979). We prefer an approach using least squares loss functions.
Suppose $h_1, \ldots, h_m$ are real valued random variables defined on a common probability space. We assume that $AVE(h_j) = 0$, $VAR(h_j) = 1$, where AVE( ) and VAR( ) are expectation and variance.
We want to find p new random variables $x_1, \ldots, x_p$, on the same space and an m x p matrix $A = \{a_{ju}\}$, $u = 1, \ldots, p$, such that

$$(1) \qquad \sigma(X;A) = \frac{1}{m} \sum_j VAR(h_j - \sum_u a_{ju}x_u)$$

is as small as possible. To identify the solution we require that

(2) $$\mathrm{COV}(x_u, x_v) = \delta^{uv}$$

where COV(,) is covariance, and we require that A'RA is diagonal, where $R = \{r_{j\ell}\}$ is the matrix of correlations of the h_j, or $r_{j\ell} = \mathrm{COR}(h_j, h_\ell)$.
The solution to this minimization problem is well known since the time of Pearson (cf MacDonell, 1901). If $R = K\Lambda^2K'$ is the eigendecomposition of R, with eigenvalues decreasing along the diagonal, then the optimal solution for A is $A = K_p\Lambda_p$, the truncated eigensolution involving only the p largest eigenvalues. The optimal solution for X is

(3) $$x_u = \lambda_u^{-1} \sum_{j=1}^{m} k_{ju} h_j$$

The minimum value of the loss function is

(4) $$\sigma_{min} = \frac{1}{m} \sum_{q=p+1}^{m} \lambda_q^2$$

which is the sum of the m - p smallest eigenvalues of the correlation matrix.

2.2. Non linear principal components analysis

We now generalize the results of section 1 by defining the loss function

(5) $$\sigma(\Phi; X; A) = \frac{1}{m} \sum_{j=1}^{m} \mathrm{VAR}(\phi_j(h_j) - \sum_{u=1}^{p} a_{ju} x_u),$$

where the ϕ_j are real valued functions satisfying
$\mathrm{AVE}(\phi_j(h_j)) = 0$, $\mathrm{VAR}(\phi_j(h_j)) = 1$.
The idea is that in non linear principal components analysis the loss function is minimized over X and A, as before, but also over the vector of transformations Φ. This form of non linear principal components analysis is useful if we are not exactly sure about the scale level of our variables h_j, that is if our variables are not strictly numerical, and also if we expect that a non linear transformation of the variables

is called for to produce a better and more stable fit to the bilinear or inproduct approximations of PCA. It is possible to generalize non linear PCA even further, by requiring that the transformations ϕ_j are restricted to be in specified cones of transformations K_j, but in this paper we only treat the unrestricted case.
It is clear that minimizing $\sigma(\Phi;X;A)$ over all the three sets of variables amounts to the same thing as maximizing the sum of the first p eigenvalues of $R(\Phi)$, where

$$(6) \qquad \{R(\Phi)\}_{j\ell} = \mathrm{COR}(\phi_j(h_j),\ \phi_\ell(h_\ell)).$$

This follows trivially from the theory of the previous section.

3. NON LINEAR TRANSFORMATIONS FOR p = 1

We want to find non linear transformations $\phi_j(h_j)$ with $E(\phi_j(h_j)) = 0$ of real valued random variables h_j, $j = 1, \ldots, m$ and a new random variable x with $E(x) = 0$ in such a way that x resembles in a least squares sense the transformations $\phi_j(h_j)$, $j = 1, \ldots, m$ as much as possible. Accordingly we can formulate the following loss function for non linear transformations that is to be minimized over a real valued random variable x and non linear transformations ϕ_j with

$$(7) \qquad \phi_j \in \mathcal{L}_j = \{\ \phi : \mathrm{AVE}(\phi_j(h_j)) = 0\ ;\ \mathrm{AVE}(\phi(h_j)^2) < \infty\ \}$$

Minimize σ over x and ϕ

$$(8) \qquad \sigma(x,\phi) = \frac{1}{m} \sum \mathrm{VAR}(x - \phi_j(h_j))$$

with normalizations $\mathrm{AVE}(x) = 0$ and $\mathrm{VAR}(x) = 1$.
Define $\sigma(x;*) = \min\ \{\ \sigma(x;\phi)\ :\ \phi\ \}$ then according to the definition of conditional expectation this leads to

$$(9) \qquad \sigma(x;*) = 1 - \frac{1}{m} \sum \mathrm{VAR}(E(x|h_j)).$$

We are thus looking for a random variable x such that the average correlation ratio of x with $\phi_j(h_j)$, $j = 1, \ldots, m$ is as large as possible. This ratio is the squared correlation coefficient if our transformations ϕ were restricted to be linear.

Another interpretation is possible if we define $\sigma(*;\phi) = \min \{ \sigma(x;\phi): x \}$. The minimum is attained for x proportional to $\sum \phi_j(h_j)$.

It follows that minimizing $\sigma(*;\Phi)$ over Φ is equivalent to maximizing the sum of all covariances of the $\phi_j(h_j)$, while keeping the sum of the variances to a constant. This is also equivalent to maximizing the dominant eigenvalue of the correlation matrix $R(\Phi)$.

Define for every space $\mathcal{L}_j$, cf (7), a complete orthonormal basis g_{js}, $s = 1, 2, \ldots$, such that for every s and t

$$(10) \qquad \mathrm{COV}(g_{js}(h_j), g_{jt}(h_j)) = \delta^{st},$$

where δ^{st} is the Kronecker delta. Then $(C_{st})_{j\ell}$ is defined as the covariance between the transformations $g_{js}(h_j)$ and $g_{\ell t}(h_\ell)$.

Define C as the supermatrix

$$C = \begin{bmatrix} C_{11} & \cdots & \cdots & \cdots & \cdots \\ \vdots & \ddots & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \vdots & C_{st} & & C_{ss} & \\ \vdots & & & & \ddots \\ \vdots & & & & & \ddots \end{bmatrix}$$

with an infinite number of submatrices C_{st}, $s = 1, 2, \ldots$ and $t = 1, 2, \ldots$, all of order m.

We can express the transformations of every variable as

(11) $$\phi_j(h_j) = \sum a_{js} g_{js}(h_j)$$

with linear weights a_{js} and hence

(12) $$\mathrm{COV}(\phi_j(h_j), \phi_\ell(h_\ell)) = \sum_s \sum_t a_{js} a_{\ell t} \mathrm{COV}(g_{js}(h_j), g_{\ell t}(h_\ell))$$

The sum of covariances between transformed variables is

(13) $$\sum_s \sum_t a_s' C_{st} a_t$$

and the sum of variances is

(14) $$\sum_s a_s' a_s$$

This leads to the following characteristic equation

(15) $$\sum_t C_{st} a_t = \lambda a_s$$

We thus have to compute the dominant eigenvalue of an infinitely large supermatrix C while holding the sum of the variances constant.

4. THE VERY SPECIAL CASE WITH p = 1 AND SIMULTANEOUS DIAGONABILITY

In addition to p = 1 we now suppose that

(16) $$\rho_{j\ell st} = \delta^{st} \rho_{j\ell s}$$

with

(17) $$\rho_{j\ell st} = \mathrm{COV}(g_{js}(h_j), g_{\ell t}(h_\ell))$$

Thus elements of the different bases are uncorrelated if they do not have the same rank number in the basis. Another way to put this is that the orthogonal bases are chosen in such a way that they diagonalize all bivariate marginal distribu-

tions of the h_j simultaneously. Such a choice of bases is not possible in general, but a number of sufficient conditions for the existence of these bases can be easily derived from the work of Lancaster (1969) and his pupils.
A very interesting sufficient condition is that the distribution of the h_j is multivariate normal, in this case we can use for the g_{js} the Hermite-Chebyshev polynomials, and we find the relation

$$(18) \qquad \rho_{j\ell st} = \delta^{st}(\rho_{j\ell})^{(s)} ,$$

where the $\rho_{j\ell}$ are the ordinary population correlation coefficients, and where (s) is a power. Thus Hermite-Chebyshev polynomials g_{js} and $g_{\ell s}$ have correlation equal to the ordinary (linear) correlation to the power s. The doubly infinite matrix of which the eigenproblem must be solved consists of an infinite sequence of diagonal blocks of order m, the first one with elements $\rho_{j\ell}$, the second one with elements $\rho^2_{j\ell}$, and so on. The eigenvalues of the supermatrix are the eigenvalues of the diagonal blocks.

It follows from the work of Styan (1973) that the largest eigenvalue of all eigenvalues of the infinite matrix is the largest eigenvalue of the ordinary population correlation matrix. Thus for the multinormal distribution the optimal transformations are all linear, and non linear and linear PCA amount to the same thing if p = 1. It is of some interest in this case to study the remaining eigenvalues. For the second largest one there are two possibilities. It is either the second largest eigenvalue of the matrix with elements $\rho_{j\ell}$, in which case the second best solution is also linear, or it is the largest eigenvalue of the matrix with elements $\rho^2_{j\ell}$, in which case the second best transformations are all quadratic (Gifi, 1981, ch 11).

This becomes interesting if we realize that the same results are automatically true for variables which are not multinormal, but which can be made multinormal by separate trans-

formations of each of the variables. In this case the optimal transformation first transforms to multinormality, and then applies the Hermite-Chebyshev polynomials. Situations corresponding to this model seem to occur quite often in real world situations, and because a dominant first eigenvalue of the correlation matrix implies that usually the dominant eigenvalue of the matrix of squared correlations is larger than the second eigenvalue of the correlation matrix, this means that the two dimensional solutions often show quadratic plots, sometimes called 'horse shoes' (Hill, 1974).

5. THE VERY SPECIAL CASE WITH $p = 1$ AND FINITE BASES

If the bases can be chosen to be finite, for example because the h_j are discrete variables assuming only a finite number of values, then things also simplify dramatically. Suppose basis g_{ju} has k_j elements. The matrix $\rho_{j\ell st}$ is now of order $\sum k_j$, and can be thought of as consisting of m^2 matrices, one for each pair (j,ℓ), the one corresponding with pair (j,ℓ) has k_j rows and k_ℓ columns.
If we choose the g_{ju} equal to indicator functions for discrete variables assuming values $1, \ldots, k_j$, then this reduces to correspondence analysis (Hill, 1974; Bouroche & Saporta, 1980; Benzécri, 1973; Gifi, 1981). In this case $g_{ju}(h_j) = 1$ if $h_j = u$ and $g_{ju}(h_j) = 0$ otherwise, with $u = 1, \ldots, k_j$.
In our general framework correspondence analysis corresponds to choosing $p = 1$ and using indicator matrices, either because we have discrete variables or because we have discreticized continuous variables.
It is general practice in correspondence analysis to compute at least two solutions to the eigenequations, in order to be able to make plots. In many cases, but certainly not in all, these two dimensional solutions will be horse shoes, especially if the variables are designed to measure a one-dimensional attitude or aptitude or preference or utility

scale. From our point of view computing two solutions with correspondence analysis must be interpreted to computing two solutions to the PCA problem with $p = 1$, not one solution to the problem with $p = 2$. But we emphasize that our interpretation of correspondence analysis is certainly not the only possible one (cf Gifi, 1981, chapters 3 and 4).

6. WHAT IF $p > 1$?

If $p > 1$ then our optimization problem is not an eigenproblem anymore, and our form of PCA becomes very similar to the one outlined by Kruskal and Shepard (1974), Young, Takane and De Leeuw (1978), Tenenhaus (1977). In this case it is most useful to go back to the original loss function (5), and to apply an alternating least squares algorithm. In its simplest form such an ALS-algorithm minimizes the loss function over X and A for Φ fixed at its current value, then minimizes the loss function over Φ with X and A fixed at the current value, and it then proceeds by alternating these two steps. Minimizing over X and A is a linear PCA problem, and we know how to solve this, minimizing over Φ amounts to computing the conditional expectation of the weighted sum

$$\sum_{u=1}^{p} a_{ju} x_u$$

given the h_j. Again in the general case computing such a conditional expectation may be far from simple, and we need special assumptions in order to proceed. Simultaneous diagonability does not help very much in this case, and consequently we can only derive some additional useful results in this case if we can choose finite bases. This is exactly the case studied by the authors mentioned earlier in this section, who sometimes introduce the additional generalization that the optimal transformations must be in a given convex cone of transformations.

One computational remark is important here. In stead of an ALS-algorithm with two steps it seems wiser to use an ALS-

algorithm with three steps. The step which computes X and A for given Φ is split into a step which computes X for given A and Φ, and a step which computes A for given X and Φ. This amounts to the same thing as using one or more Bauer-Rutishauser inner iterations (Rutishauser, 1969, 1970) before computing new currently optimal transformations.

7. FINITE BASES

We have seen that a convenient finite basis are the indicator functions, which can be used without loss of generality or precision if the variables are discrete and assume only a finite number of values. If the h_j are continuous we can also use indicator functions if we first discreticize the h_j in a finite number of intervals. This means, in other terminology, that we approximate the optimal transformations by using a finite basis for the subspace of step functions. Step functions usually give a poor approximation of continuous functions, not necessarily in terms of precision of approximation, but in terms of data analytical interpretation. In examples in which we know the optimal transformations (for example multinormal populations with different continuous distortions) it is often difficult to 'recover' the precise nature of the optimal transformations from the plot of the step functions (Gifi, 1981, ch 11).
As a consequence we should like to have a more flexible finite basis to approximate continuous transformations, because for practical purposes it is always necessary to approximate infinite or very-large dimensional spaces by using relatively low-dimensional subspaces. Fortunately a natural and powerful generalization of the indicator functions is available, and has been studied a great deal in recent years. We mean the B-Splines (De Boor, 1978).

8. THE APPROXIMATION WITH B-SPLINES

To generalize the indicator functions we use an orthonormal

basis of spline functions. Such a spline function $\$_{j,k,\tau}$ for variable j is any linear combination of B-Splines of order k for the knot sequence τ. Thus

$$(19) \qquad \$_{j,k}(h_j) = \sum_{s=1}^{r_j} a_j B_{js}^{(k)}(h_j)$$

We have dropped the index τ because the knot sequence is fixed and r_j is the number of intervals + the order (=k) of B-Splines.

Define the inner product matrices

$$(20) \qquad (U_{j\ell}^{(k)})_{st} = \mathrm{COV}(B_{js}^{(k)}(h_j),\ B_{\ell t}^{(k)}(h_\ell)),$$

and $F_j^{(k)} = \mathrm{diag}\ (U_{jj}^{(k)})$. Define the supermatrices

$$U^{(k)} = \begin{Bmatrix} U_{11}^{(k)} & \cdots & \cdots & \cdots & \cdots \\ \vdots & \ddots & & & \\ \vdots & & U_{j\ell}^{(k)} & & \\ \vdots & & & \ddots & \\ \vdots & & & & U_{mm}^{(k)} \end{Bmatrix}$$

and $F^{(k)} = \{F_1^{(k)}, \ldots\ldots, F_m^{(k)}\}$.

If we use B-Splines of order 1 these matrices reduce to some familiar forms: First, if the partitioning induced by the number of knots is made increasingly finer $F_j^{(1)}$ becomes the univariate density of variable j and $U_{j\ell}^{(1)}$ becomes the bivariate density of variables j and ℓ.

Second, if we use a limited number of knots the submatrix $F_j^{(1)}$ will contain the univariate marginals of the intervals between adjoining knots and the submatrix $U_{j\ell}^{(1)}$ will contain the bivariate marginals of such intervals between variables j and ℓ. Further we can interprete the first order B-Splines as a sequence of indicator functions, each of them indicating whether a datapoint belongs to a certain interval between two knots or not. If we write these indicator functions as columns

of an indicator matrix $G_j^{(1)}(n \times r_j)$ where r_j equals the number of knots, then we can say

$$U_{j\ell}^{(1)} = G_j^{(1)} G_\ell^{(1)} \tag{21}$$

and

$$F_j^{(1)} = G_j^{(1)\prime} G_j^{(1)} \tag{22}$$

Using the linear weights a_j, $j = 1, \ldots, m$, the respective sums of variances and covariances now become

$$\sum \sum a_j' \, U_{j\ell}^{(1)} a_\ell \tag{23}$$

and

$$\sum a_j' \, F_j^{(1)} a_j \tag{24}$$

and the characteristic equation is hence

$$\sum U_{j\ell}^{(1)} a_\ell = \lambda \, F_j^{(1)} a_j \tag{25}$$

The equivalence with analyse des correspondances and principal components analysis is clear (cf Hill, 1974, p. 342; Gifi, 1981) provided we use these first order B-Splines. We will refine the approximation by using second order B-Splines but we still can use a kind of pseudo indicator functions with 2 non zero elements per row. $F_j^{(2)}$ becomes a banded symmetric matrix and is not diagonal anymore. The sparseness of $G_j^{(1)}$ is partly gone but especially for $k \leq 2$ we will still have the computational efficiency of the B-Splines combined with the straightforward simple generalization of the step function approach. The use of B-Splines takes away the disadvantage of the crudeness or discontinuity of the step functions at the breakpoints (cf Hill, 1974, p. 349) while maintaining the advantage of not having to use strict linear combinations of untransformed variables.

Solving (25) for $k \geqslant 1$ simply consists of computing the eigenvalues of a symmetric positive semidefinite matrix, which can be done with standard methods. The Bauer-Rutishauser method is to be preferred here too if one is only interested in a few dominant eigenvalues. The algorithm based on this principle is identical to Richardson's method of reciprocal averages (Horst, 1936; Hill, 1973; De Leeuw, 1976; Nishisato, 1980). A production computer program along these lines, written in ANSI fortran IV, that fits discrete step functions is HOMALS (Gifi, 1981; De Leeuw & Van Rijckevorsel, 1980). In the terminology of Gifi (1981, ch 1) we can say that correspondence analysis fits multiple transformations. "Multiple" means that a variable is transformed differently for every dimension. Thus we choose $p > 1$, and compute a number of solutions, each defining a different transformation.
In applying PCA with $p > 1$ we use a single approach. We compute only one transformation for all dimensions per variable. In fitting B-Splines in PCA we want to approximate Φ with a least squares (=LS) approximation by splines of a low order. Computationally this leads to a banded and linear system of equations to be solved. We used an algorithm to solve this system in the substep of the ALS structure to approximate Φ. This least squares approximation algorithm with splines is extensively discussed by De Boor (1978, p. 249).
The authors are currently working on an algoritm and computerprogram that combines the approximation with B-Splines ($k \geqslant 1$) and the simultaneous iteration method.

9. AN EXAMPLE

The data are a version of Thurstone's cylinder data (Thurstone, 1947, p. 117). They consist of 20 objects and 10 variables. They describe the properties of cylinders. These cylinders vary in two dimensions, which are the first 2 variables. The other 8 variables are monotone functions of these 2 dimensions. These functions are shown in table 1.

var 1	$= a_i$
var 2	$= b_i$
var 3	$= 2(\pi b_i)^{\frac{1}{2}}$
var 4	$= 2a_i(\pi b_i)^{\frac{1}{2}}$
var 5	$= a_i b_i$
var 6	$= (2\pi)^{-1} a_i b_i^2$
var 7	$= (2\pi)^{-\frac{1}{2}} a_i b_i^{-\frac{1}{2}}$
var 8	$= a_i b_i^{-1}$
var 9	$= a_i^{-1} b_i$
var 10	$= 2a_i b_i^{-2}$

Table 1. Thurstone's cylinder data

The first two variables are a sample of size 20 from two independent uniform random variables on the unit interval. The meaning of these variables (in terms of cylinders) is not relevant in our case. We simply use the fact that a centered log transformation of all variables gives a matrix of exact rank two. The aim of the techniques discussed here is to recover these transformations. The idea is that non linear principal components analysis will fit this matrix perfectly in two dimensions, and that correspondence analysis will not accomplish this result because the second eigenvalue is not the second root of the correlation matrix $R(\Phi)$.
The fit in table 2 is defined as the sum of the first two eigenvalues of $\frac{1}{m}R(\Phi)$. The maximum fit is thus equal to 1.

Technique used	Type of transformations approximating Φ	Total fit for p = 2	eigenvalues	
metric PCA	single linear	.87	.60	.27
ordinal PCA	single monotone	.94	.66	.28
splines-PCA k=1	single LS splines k=1	.89	.61	.28
splines-PCA k=2	single LS splines k=2	.98	.68	.30
splines-PCA k=3	single LS splines k=3	.99	.69	.29
Correspondence-analysis+splines	multiple splines k=2	.77	.77	.66

Table 2. Fit, eigenvalues and transformation-types for several techniques.

The linear fit is surprisingly high compared with the ordinal fit, which on the other hand is clearly inferior to the single

spline approaches for $k > 1$. The gain in using higher degree splines ($k > 2$) is negligible. The special role of the second eigenvalue of the multiple spline transformations becomes clear if one realizes that, in order to obtain perfect fit, this value should maximally be .23 (i.e. 1 - .77). It is somewhat unwise to apply multiple transformations in such a "single" case, where an underlying "correct" transformation and a two-dimensional structure are known to exist. A much wiser application of a more dimensional correspondence analysis in a similar context on multinormally distributed variables is discussed by Lafaye de Micheaux (1978). Together with the superiority of the splines approach is this the main conclusion of this example.

REFERENCES

J.P. Benzécri e.a., Analyse des données (2 vols.) (Dunod, Paris, 1973)

C. de Boor, A practical guide to splines (Springer, Berlin, 1978)

J.M. Bouroche et G. Saporta, L'analyse des données (Presses Universitaires de France, Paris, 1980)

A. Gifi, Non-linear multivariate analysis (Department of Datatheory, University of Leiden, 1981)

M.O. Hill, Reciprocal averaging: an eigenvector method of ordination, J. Ecol. 61 (1973) 237-251.

M.O. Hill, Correspondence analysis: a neglected multivariate method, J. of the Royal Statistical Society (Series C) 23 (1974) 340-354.

P. Horst, Obtaining a composite measure from a number of different measures of the same attribute, Psychometrika 1 (1936) 53-60.

J.B. Kruskal and R.N. Shepard, A nonmetric variety of linear factor analysis, Psychometrika 39 (1974) 123-157.

D. Lafaye de Micheaux, Approximation d'analyse economiques non-linéaires de variables aléatoires (Dissertation, Université de Nice, 1978)

H.O. Lancaster, The chi-squared distribution (Wiley, New York, 1969)

J. de Leeuw, Homals (Paper read at the Psychometric Society Meeting, Murray Hill, N.Y., 1976)

J. de Leeuw and J.L.A. van Rijckevorsel, Homals & Princals, some generalizations of principal components analysis, in: E. Diday et al. (eds.), Data analysis and informatics (North Holland, Amsterdam, 1980)

W.R. MacDonnell, On criminal anthropology and the identification of criminals, Biometrika 1 (1901) 177-227.

S. Nishisato, Analysis of categorical data: dual scaling and its applications (University of Toronto Press, Toronto, 1980)

M. Okamoto, Optimality of principal components, in: P.R. Krishnaiah (ed.), Multivariate Analysis II (Academic Press, New York, 1968)

C.R. Rao, Matrix approximation and reduction of dimensionality in multivariate statistical analysis, in: P.R. Krishnaiah (ed.), Multivariate Analysis V (North Holland Publ. Comp., Amsterdam, 1980)

B. le Roux and H. Rouanet, L'analyse statistique des protocolles multidimensionelles: analyse en components principales, Publ. Inst. de Statistique de l'Université de Paris 24 (1979) 47-74.

H. Rutishauser, Computational aspects of F.L. Bauer's simultaneous iteration method, Numer. Math. 13 (1969), 4-13.

H. Rutishauser, Simultaneous iteration method for symmetric matrices, Numer. Math. 16 (1970) 205-223.

J.L.A. van Rijckevorsel, B. Bettonvil and J. de Leeuw, Recovery and stability in non linear principal components analysis (Paper presented at the European meeting of the Psychometric Society, Groningen, 1980)

G.P. Styan, Hadamard products and multivariate statistical analysis, Linear Algebra and its Applications 6 (1973) 217-240.

M. Tenenhaus, Analyse en composantes principales d'un ensemble de variables nominales et numériques, Revue de Statist. Appl. 25 (1977) 39-56.

L.L. Thurstone, Multiple Factor Analysis (University of Chicago Press, Chicago, 1947)

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 43 (1978) 279-281.