RR 89-06

NONLINEAR REDUCED RANK GENERALIZED CANONICAL CORRELATION ANALYSIS INCLUDING COMMON SCALE QUANTIFICATIONS AND DATA WEIGHTS

Ivo A. van der Lans

Department of Data Theory

University of Leiden

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika, 56*, 433-451.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29*, 115-129.

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

Meulman, J. (1988). *Reduced space analysis* (RR 88-01). Leiden: University of Leiden, Department of Data Theory.

Ramsay, J. O. (1977). Monotonic weighted power transformations to additivity. *Psychometrika, 42*, 83-109.

Roskam, E. E. C. I. (1968). *Metric analysis of ordinal data in psychology* [Doctoral dissertation]. Voorschoten: VAM.

Tijssen, R. (1985). *A new approach to nonlinear canonical correlation analysis* (PRM 85-01). Leiden: University of Leiden, Department of Psychology.

Van de Geer, J. P. (1984). Relations among K sets of variables. *Psychometrika , 49*, 79-94.

Van de Geer, J. P. (1986). *Introduction to linear multivariate data analysis* (2 Vols.). Leiden: DSWO Press.

Van den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika, 42*, 207-219.

Van der Burg, E. & De Leeuw, J. (1988). Nonlinear canonical correlation analysis with K sets of variables. In E. van der Burg, *Nonlinear canonical correlation and some related techniques.* Leiden: DSWO Press.

Van der Burg, E., De Leeuw, J. & Verdegaal, R. (1988). Homogeneity analysis with *k* sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika, 53*, 177-197.

Van der Lans, I. A. & Heiser, W. J. (1988). *Nonlinear multiple regression analysis with common scale transformations across predictor variables* (RR 88-10). Leiden: University of Leiden, Department of Data Theory.

Van der Lans, I. A. (1989, July). *Common scale transformations and restrictions upon regression weights in nonlinear multiple regression.* Paper presented at the European Meeting of the Psychometric Society, Leuven, Belgium.

Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika, 46*, 357-387.

Nonlinear reduced rank generalized canonical correlation analysis including common scale quantifications and data weights[1]

Ivo A. van der Lans
Department of Data Theory

Abstract

Generalized canonical correlation analysis (GCCA) is a multivariate data analysis technique for analyzing relations between more than two sets of variables. The nonlinear GCCA technique OVERALS developed by Gifi (1981) and Van der Burg & De Leeuw (1988) optimizes the criterion given for GCCA by Carroll (1968). It does so by estimating optimal model parameters and optimal scores of objects on variables. In this paper we introduce some additional restrictions on model parameters and optimal scores. More precisely, we formulate rank and cone restrictions on canonical weights and common scale restrictions on optimal scores. We also generalize Carroll's weighting factors (Carroll, 1968) by introducing weighting factors which allow us not only to weigh sets differently, but also to assign within each set different weights to different objects.

[1] The author wishes to thank Dré Nierop for pointing out some unmistakable errors in an earlier version of this paper. The author is the only person who is responsible for any remaining errors.

Nonlinear reduced rank generalized canonical correlation analysis including common scale quantifications and data weights

1. Introduction.

In canonical correlation analysis (CCA) (Hotelling, 1936) one looks for p orthogonal linear combinations of variables in each of two sets of variables. These linear combinations are called canonical variables. The correlation between the s^{th} canonical variable from one set and the s^{th} canonical variable from the other set is called a canonical correlation. The objective is to maximize the sum of canonical correlations.

The generalization of CCA to more than two sets is called generalized canonical correlation analysis (GCCA). Several kinds of GCCA can be specified dependent upon the criterion to be optimized (e.g., largest sum of canonical correlations, largest sum of squared canonical correlations) and the orthogonality restrictions imposed. Orthogonality restrictions can be weak or strong (Dauxois & Pousse, 1976). Weak orthogonality restrictions are orthogonality restrictions that operate on the sum across sets of canonical variables. Strong orthogonality restrictions are restrictions that operate on the canonical variables within each set. Different criteria and orthogonality restrictions are discussed by Kettenring (1971), Gifi (1981), and Van de Geer (1984, 1986).

Gifi (1981) and Young (1981), among others, define a branch of multivariate data analysis techniques under the name of nonlinear multivariate analysis. These techniques are nonlinear in the sense that, instead of using *a priori* numerical scores of objects on variables, they estimate an optimal numerical score for each object on each variable, given the criterion of the technique under consideration. Optimal scores have to obey linear, order, and equality restrictions that constitute the data. In addition to linear, order, and equality restrictions we can think of other restrictions like polynomial restrictions, spline restrictions (e.g., Ramsay, 1977), and smoothness restrictions (Heiser, 1985). Finding optimal scores of objects on variables also goes under the name of optimal scaling of variables (cf. Gifi, 1981; Young, 1981). Given *a priori* scores of the objects on variables one can think of optimal scores as if they are obtained by a transformation of these *a priori* scores.

Gifi (1981) and Van der Burg & De Leeuw (1988) give a nonlinear version of GCCA with weak orthogonality restrictions, which they base upon the criterion given for GCCA by Carroll (1968), and which they call OVERALS. OVERALS maximizes the sum of correlations between p canonical variables from each set and p columns of a matrix of comparison scores (Carroll, 1968). This criterion for GCCA has been called the MAXVAR-criterion by Kettenring (1971). The p columns of the matrix of comparison scores are restricted to be orthogonal to each other. This restriction corresponds to weak

orthogonality. Thus OVERALS maximizes the sum of correlations by computing optimal scores of the objects on the variables, optimal comparison scores, and those weights, also called canonical weights, that give the optimal canonical variables.
It should be noted that OVERALS can also be applied if there are only two sets. With nonlinear optimal scaling of variables OVERALS produces a different solution than the one we would obtain by maximizing the sum of canonical correlations. This is because of the different criteria applied.

Nonlinear generalized canonical correlation analysis: OVERALS

As noted OVERALS maximizes the sum of correlations between columns of an *nxp*-comparison matrix **X** and corresponding columns of *nxp*-matrices of canonical variables. Here n is the number of objects and p is the number of dimensions. We have K of these matrices of canonical variables, one matrix of canonical variables from each set. The objective of OVERALS can be written as the minimization of the following least squares loss function:

$$\sigma(\mathbf{X}, \mathbf{Q}, \mathbf{A}) = K^{-1}\Sigma_k \text{ SSQ } (\mathbf{X} - \mathbf{Q}_k\mathbf{A}_k), \qquad (1)$$

where:

K is the number of sets,
$\mathbf{X}$ is an $n \times p$ matrix of comparison scores, with n the number of objects and p the number of dimensions,
$\mathbf{Q}$ is an $n \times m$ partitioned matrix containing (to be estimated) scores on variables, with m the total number of variables,
$\mathbf{Q}_k$ is an $n \times m_k$ matrix containing scaled variables within set k, with m_k the number of variables in set k, thus $\mathbf{Q} = (\mathbf{Q}_1 \mid \mathbf{Q}_2 \mid ... \mid \mathbf{Q}_k \mid ... \mid \mathbf{Q}_K)$,
$\mathbf{q}_{(k)j}$ is an n-vector of optimal scores of n objects on variable j in set k,
$\mathbf{A}$ is an $m \times p$ partitioned matrix of canonical weights,
$\mathbf{A}_k$ is an $m_k \times p$ matrix of canonical weights for the variables in set k, thus $\mathbf{A} = (\mathbf{A}'_1 \mid \mathbf{A}'_2 \mid ... \mid \mathbf{A}'_k \mid ... \mid \mathbf{A}'_K)'$, and
SSQ() denotes the sum of squared elements of the matrix in between the brackets.

Loss function (1) is minimized under the following assumptions and restrictions:

- columns of $\mathbf{X}$ are standardized, zero mean and unit variance, and required to be orthogonal to each other, that is $\mathbf{u}_n'\mathbf{X} = \mathbf{0}$ and $\mathbf{X}'\mathbf{X} = n\mathbf{I}_p$, where $\mathbf{u}_n$ is the

n vector with all elements equal to one and $\mathbf{I}_p$ is the pxp identity matrix. The restriction $\mathbf{X}'\mathbf{X} = n\mathbf{I}_p$ represents weak orthogonality. The restrictions on $\mathbf{X}$ are necessary to avoid degenerate, that is uninteresting or trivial, solutions,

- each $\mathbf{q}_{(k)j}$ is restricted by the data on variable j in set k. As noted these data define linear, order, and/or equality restrictions among elements of $\mathbf{q}_{(k)j}$. For each $\mathbf{q}_{(k)j}$ we require that $\mathbf{q}_{(k)j} \in \Gamma_{(k)j}$, in which $\Gamma_{(k)j}$ is a closed convex cone defined by the linear, order, or equality restrictions,
- columns of $\mathbf{Q}$ are standardized such that $\mathbf{u}'_n\mathbf{q}_{(k)j} = 0$ and $\mathbf{q}'_{(k)j}\mathbf{q}_{(k)j} = n$, for each $\mathbf{q}_{(k)j}$. These restrictions are necessary to select one solution from among an infinite number of solutions that yield the same value of loss function (1) and that differ from each other in trivial aspects only.

It should be noted that in the current formulation loss function (1) accomodates for *single* treatment of variables only. That is, we compute only one set of optimal scores for each variable. Besides single treatment of variables OVERALS allows for multiple treatment of variables. A variable that is treated multiple obtains a separate set of optimal scores for each dimension. *Multiple* treatment of variables in loss function (1) is possible by the introduction of copies of variables in sets (De Leeuw & Van Rijckevorsel, 1988; Van der Burg, De Leeuw & Verdegaal, 1988; Van der Burg & De Leeuw, 1988).

2. Nonlinear reduced rank generalized canonical correlation analysis.

Gifi (1981, p.197) introduces OVERALS primarily as a generalization of homogeneity analysis (see for instance Guttman, 1941; De Leeuw, 1973; and also Benzécri *et al.* 1973 in which homogeneity analysis is treated under the name multiple correspondence analysis). The objective of homogeneity analysis is to minimize the following loss function:

$$\sigma(\mathbf{X}, \mathbf{Y}) = K^{-1}\Sigma_k \, \mathrm{SSQ}(\mathbf{X} - \mathbf{G}_k\mathbf{Y}_k), \qquad (2)$$

where:

$\mathbf{X}$ is an n x p matrix of comparison scores, also called object scores, with n the number of objects and p the number of dimensions,

$\mathbf{Y}$ is an m x p partitioned matrix of category quantifications, with m the total number of categories,

K is the number of categorical variables,

$\mathbf{G}_k$ is the n x n_k indicator matrix for categorical variable k, with n_k the number categories of variable k. Let denote the element in the i^{th} row and r^{th} column of $\mathbf{G}_k$ as $g_{(k)ir}$. Then $g_{(k)ir}$=1 if object i is in category r of variable k and $g_{(k)ir}$ if object i is not in category r of variable k,

$\mathbf{Y}_k$ is an n_kx p matrix of categories quantifications for the categories of variable k, thus $\mathbf{Y} = (\mathbf{Y}'_1 \mid \mathbf{Y}'_2 \mid ... \mid \mathbf{Y}'_k \mid ... \mid \mathbf{Y}'_K)'$.

The loss function of homogeneity analysis is minimized given the restrictions that columns of $\mathbf{X}$ are standardized and orthogonal to each other.

Replacement of the indicator matrices in loss function (2) by matrices of (to be estimated) scores, not necessarily zero or one, gives us loss function (1). In that case the canonical weights, $\mathbf{A}$, in loss function (1) correspond with the category quantifications, $\mathbf{Y}$, in loss function (2).

Whereas we can view upon GCCA as a *generalization* of homogeneity analysis, nonlinear principal components analysis (PCA) (e.g., Gifi, 1981; Kruskal & Shepard, 1974; Roskam, 1968) can be seen as a *special case* of homogeneity analysis (see Figure 1 and, for the moment, forget about the most right entry). That is, nonlinear PCA is equal to homogeneity analysis with rank-one restrictions on each matrix of category quantifications, $\mathbf{Y}_k$ (see for instance De Leeuw & Van Rijckevorsel, 1988). In fact Gifi (1981, 1985) uses an iterative algorithm to minimize the loss function of homogeneity analysis and computes rank restricted category quantifications from the unrestricted category quantifications in an extra step in each iteration to obtain nonlinear PCA solutions. De Leeuw and Van Rijckevorsel (1988) note that it is possible to pose more general rank restrictions on matrices with category quantifications. They note that, besides rank 1 or rank p, we can choose the rank, r_k, of each $\mathbf{Y}_k$ separately (though it will always be in the closed interval from 1 to min(p, n_k-1)). Their loss function for homogeneity analysis with general rank restrictions is:

$$\sigma(\mathbf{X} , \mathbf{B}_1 ,..., \mathbf{B}_K , \mathbf{C}) = K^{-1} \Sigma_k \, SSQ (\mathbf{X} - \mathbf{G}_k \mathbf{B}_k \mathbf{C}'_k), \qquad (3)$$

where

$\mathbf{B}_k$ is an n_kx r_k matrix of quantification parameters,

$\mathbf{C}_k$ is an p x r_k matrix of regression parameters, and

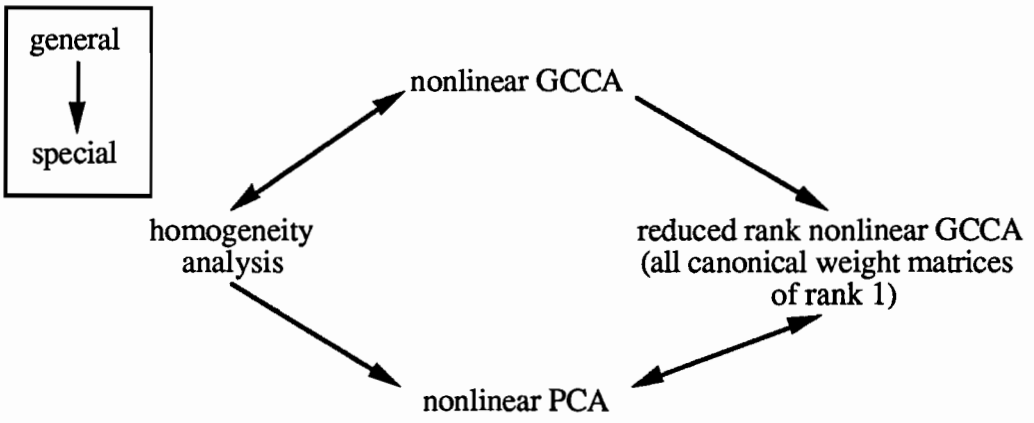
$\mathbf{C}$ denotes ($\mathbf{C}_1 \mid \mathbf{C}_2 \mid ... \mid \mathbf{C}_k \mid ... \mid \mathbf{C}_K$).

The rank restricted category quantifications, $\mathbf{Y}_k$, are decomposed in $\mathbf{B}_k\mathbf{C}'_k$. In the same way in which homogeneity analysis can be generalized to nonlinear GCCA by replacing indicator matrices, $\mathbf{G}_k$, in (2) into matrices of (to be estimated) parameters, $\mathbf{Q}_k$, homogeneity analysis with general rank restrictions can be generalized to, what we call, nonlinear reduced rank GCCA. We write the loss function of nonlinear reduced rank GCCA as:

$$\sigma(\mathbf{X}, \mathbf{Q}, \mathbf{B}_1, ..., \mathbf{B}_K, \mathbf{C}) = K^{-1}\Sigma_k \text{ SSQ } (\mathbf{X} - \mathbf{Q}_k\mathbf{B}_k\mathbf{C}'_k), \quad (4)$$

where $\mathbf{B}_k$ is an $m_k x r_k$ matrix. Note that in this case each $\mathbf{B}_k\mathbf{C}'_k$ gives us a matrix of reduced rank canonical weights, $\mathbf{A}_k$. If all r_k's are equal to 1 then reduced rank nonlinear GCCA can be seen as a generalization of nonlinear PCA in the same way as GCCA can be seen as a generalization of homogeneity analysis. Also, with all r_k's equal to 1 reduced rank nonlinear GCCA can be defined as a special case of nonlinear GCCA in the same way as we can define nonlinear PCA as a special case of homogeneity analysis. Furthermore, it is interesting to notice that nonlinear GCCA can be seen as homogeneity analysis with additivity restrictions upon the category quantifications (Van der Burg, De Leeuw & Verdegaal, 1988). In the same way reduced rank nonlinear GCCA with rank one canonical weights can be seen as a special case of nonlinear PCA. Relations between the four techniques are pictured in Figure 1.

Figure 1. Relations between homogeneity analysis, nonlinear PCA, nonlinear GCCA and nonlinear GCCA with rank one canonical weights.



The definition of $\mathbf{B}_k$ and $\mathbf{C}_k$ in loss function (4) leaves some indeterminacy since we can postmultiply $\mathbf{B}_k$ with any $r_k x r_k$-matrix $\mathbf{T}$ of full rank to obtain a new $\mathbf{B}_k$, without changing the value of the loss function in (4) as long as we postmultiply $\mathbf{C}_k$ with $\mathbf{S}'$ to obtain a new $\mathbf{C}_k$, where $\mathbf{S}$ has the property $\mathbf{TS}=\mathbf{I}_{rk}$. This implies that without loss of generality we can remove this indeterminacy by requiring that $\mathbf{B}'_k\mathbf{Q}'_k\mathbf{Q}_k\mathbf{B}_k=n\mathbf{I}_{rk}$ and that

$\mathbf{C}'_k\mathbf{C}_k$ is diagonal with the first diagonal element as large as possible, the second diagonal element as large as possible given the first diagonal element and so on. That is, given the singular value decomposition $\mathbf{Q}_k\mathbf{B}_k\mathbf{C}'_k=\mathbf{R}\Phi\mathbf{S}'$, we take $\mathbf{B}_k=(\mathbf{Q}'_k\mathbf{Q}_k)^{-1}\mathbf{Q}'_k\mathbf{R}_{rk}$ and $\mathbf{C}_k=\mathbf{S}_{rk}\Phi_{rk}$. Where $\mathbf{R}_{rk}$, $\mathbf{S}_{rk}$, and Φ_{rk} denote the first r_k columns of $\mathbf{R}$ and $\mathbf{S}$, and first r_k rows and columns of Φ respectively. Because we choose the rank of $\mathbf{B}_k\mathbf{C}'_k$ to be less than p, strong orthogonality restrictions can not be imposed.

It is important to notice that in nonlinear GCCA with all variables single the maximum for p is equal to the total number of variables across sets, which is $m=\Sigma_k m_k$. In addition, with all $\mathbf{Q}_k$ fixed the solution is nested, which means that the first dimension of the two-dimensional solution is equal to the one-dimenional solution (occasionally, when two or more dimensions fit equally well, there is some indeterminacy in defining their order). Things are different for reduced rank nonlinear GCCA. Here with all variables single the maximum for p is equal to $\Sigma_k r_k$. However, if we choose $p=\Sigma_k r_k$ then matrices $\mathbf{B}_k$ and $\mathbf{Q}_k$ become arbitrary, except for restrictions imposed by the data and normalization restrictions. That is, any possible choice of $\mathbf{Q}_k$'s and $\mathbf{B}_k$'s will give the same value of the loss function after computation of optimal matrices $\mathbf{X}$ and $\mathbf{C}_k$. Also solutions are not nested anymore, even for fixed $\mathbf{Q}_k$'s.

Another technique, in which we can choose the dimensionality for each set separately is Meulman's (1988) reduced space analysis. In her distance driven approach to GCCA she uses a loss function called STRESS (Kruskal, 1964). However, whereas in the current paper we compute p canonical variables that span an r_k-dimensional space, in her distance driven approach to GCCA one computes r_k canonical variables that satisfy strong orthogonality and from which interobject distances can be derived. The number of orthogonal canonical variables, r_k, can be larger than p in Meulman's approach. There is no direct correspondence between columns of the matrix of comparison scores, $\mathbf{X}$, and columns of the matrices of canonical variables, $\mathbf{Q}_k\mathbf{B}_k$. Permutations of columns of $\mathbf{X}$ and columns of $\mathbf{Q}_k\mathbf{B}_k$ do not affect the value of the stress function.

Having introduced rank restrictions upon canonical weights in this section we postpone an outline of the way in which we compute these weights until Section 6. At this point we want to treat so-called common scale groups (Van der Lans & Heiser, 1988) in nonlinear reduced rank GCCA, followed by cone restrictions upon columns of $\mathbf{B}_k$ and data weights.

$$\min \Sigma_l \; (\mathbb{H}_l \, \mathbf{y}^*_k - \mathbb{H}_l \, \mathbf{y}_k)' \mathbf{V}_k (\mathbb{H}_l \, \mathbf{y}^*_k - \mathbb{H}_l \, \mathbf{y}_k) \qquad (22a)$$

$$\mathbf{y}_k \in \Gamma_{\mathbf{y}k}$$

$$= \min \; (\mathbf{y}^*_k - \mathbf{y}_k)' \Sigma_l \; \mathbb{H}_l \, ' \mathbf{V}_k \mathbb{H}_l \, (\; \mathbf{y}^*_k - \mathbf{y}_k) \qquad (22b)$$

$$\mathbf{y}_k \in \Gamma_{\mathbf{y}k}$$

Equivalent to the procedure used by Van der Lans and Heiser (1988, p. 9-10) which originates from Heiser (1987), we use majorization to find the solution to (22b). That is we replace the projection problem of (22b) in the metric defined by $\Sigma_l \; \mathbb{H}_l \, ' \mathbf{V}_k \mathbb{H}_l$, by iteratively solving the problem of projecting $\mathbf{y}_k^{\text{adj}}$ onto $\Gamma_{\mathbf{y}k}$ in the metric defined by $\mathbf{I}_{nk}$, where:

$$\mathbf{y}_k^{\text{adj}} = \mathbf{y}_k^{\text{previous}} + \{ \; \Sigma_l \; \mathbb{H}'_l \, \mathbf{V}_k \mathbb{x}_l + \Sigma_l \; \mathbb{H}_l \, ' \mathbf{V}_k \mathbb{H}_l \, \mathbf{y}_k^{\text{previous}} \} / \beta_1, \qquad (23)$$

with β_1 the first eigenvalue of $\Sigma_l \; \mathbb{H}_l \, ' \mathbf{V}_k \mathbb{H}_l$.

This projection problem can be easily solved by for instance the "Up and Down Block" algorithm given by Kruskal (1964).

For the ultimate case where all variables of all sets are in one common scale group (see loss function (7)), we can compute updates of $\mathbf{y}$ by projecting the following $\mathbf{y}^{\text{adj}}$ upon the appropriate cone in the metric defined by $\mathbf{I}_{nk}$:

$$\mathbf{y}^{\text{adj}} = \mathbf{y}^{\text{previous}} + \{ \; \Sigma_k \, \Sigma_l \; \mathbb{H}'_l \, \mathbf{V}_k \mathbb{x}_l + \Sigma_k \Sigma_l \; \mathbb{H}_l \, ' \mathbf{V}_k \mathbb{H}_l \, \mathbf{y}^{\text{previous}} \} / \beta_2, \qquad (24)$$

with β_2 the first eigenvalue of $\Sigma_k \Sigma_l \; \mathbb{H}_l \, ' \mathbf{V}_k \mathbb{H}_l$.

Minimization over $\mathbf{B}_k$ for fixed $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{C}$

Again we have K independent minimization problems, each problem corresponding to one set. Before discussing the problem of finding the optimal $\mathbf{B}_k$ for fixed $\mathbb{C}_k$, we discuss the problem of finding optimal unrestricted $\mathbf{B}_k$'s for variable $\mathbf{C}_k$'s. If we denote $\mathbf{J} \; (\; \mathbf{G}_{(k)1} \, \mathbb{y}_k \mid \mathbf{G}_{(k)2} \, \mathbb{y}_k \mid \ldots\ldots \mid \mathbf{G}_{(k)mk} \, \mathbb{y}_k \;)$ by $\mathbb{Q}_k$ then with the inclusion of data weights and for fixed $\mathbb{X}$ and $\mathbb{Y}$ (6) becomes:

$$\sigma_{\mathbf{W}} \; (\; \mathbf{B}_k \, , \mathbf{C} \;) = K^{-1} \Sigma_k \; \text{SSQ}_{\mathbf{V}k} \; (\; \mathbb{X} - \; \mathbb{Q}_k \, \mathbf{B}_k \, \mathbf{C}_k \, '). \qquad (25)$$

The problem of finding the optimal (unrestricted) $\mathbf{B}_k$ and $\mathbf{C}_k$ is is equivalent to redundancy analysis (see for instance Fortier, 1966; Izenman, 1975; Van den Wollenberg, 1977). Minimizing $\mathbf{C}_k$ out of equation (25) for each k the columns of $\mathbf{B}_k$ can be found as the eigenvectors associated with the r_k largest eigenvalues of the matrix $(\mathbb{Q}_k'\mathbf{V}_k\mathbb{Q}_k)^{-1}\mathbb{Q}_k'\mathbf{V}_k\mathbb{X}\mathbb{X}'\mathbf{V}_k\mathbb{Q}_k$. The problem of finding the associated optimal $\mathbf{C}_k$ then is a multivariate multiple regression problem which is treated furtheron. We can choose to normalize $\mathbf{B}_k$ and $\mathbf{C}_k$ in such a way that $\mathbf{B}_k'\mathbf{Q}_k'\mathbf{V}_k\mathbf{Q}_k\mathbf{B}_k=n\mathbf{I}_{r_k}$ or $\mathbf{B}_k'\mathbf{Q}_k'\mathbf{Q}_k\mathbf{B}_k=n\mathbf{I}_{r_k}$, and that $\mathbf{C}'_k\mathbf{C}_k$ is diagonal whichever identification restriction is regarded as most convenient. This can be done by singular value decompositions of either $\mathbf{V}^{1/2}_k\mathbf{Q}_k\mathbf{B}_k\mathbf{C}_k'$ or $\mathbf{Q}_k\mathbf{B}_k\mathbf{C}_k'$.

The problem of finding an optimal $\mathbf{B}_k$ can also be solved for by either alternating between columns, rows or elements of $\mathbf{B}_k$, keeping $\mathbf{C}_k$ fixed. Heiser, Meulman, Van der Lans and Van den Berg (1988) apply an element-wise approach in their program and algorithm MULTIPALS for nonlinear versions of multivariate multiple regression analysis, principal components analysis and redundancy analysis. Since we aim at cone restrictions upon the columns of each $\mathbf{B}_k$, see Section 4, we use a column-wise approach for fitting the $\mathbf{B}_k$'s.

Define $\mathbb{B}_{k\,(-i\,)} = \mathbf{B}_k - \mathbf{b}_{(k)i}\,\mathbf{e}'_i$, (26a)

with $\mathbf{e}_i$ the $i^{\,\text{th}}$ column of the $r_k x r_k$ identity matrix, and $\mathbf{b}_{(k)i}$ the $i^{\,\text{th}}$ column of $\mathbf{B}_k$. Further define

$$\mathbb{x}_l^{\,0} = \mathbb{x}_l - \mathbb{Q}_k\,\mathbb{B}_{k\,(-i\,)}\mathbb{c}_{(k)l}, \tag{26b}$$

with $\mathbb{x}_l$ the $l^{\,\text{th}}$ column of $\mathbb{X}$, and $\mathbb{c}_{(k)l}$ the $l^{\,\text{th}}$ column of $\mathbb{C}_k'$.

Then column i of $\mathbf{B}_k$ can be fitted by taking:

$$\mathbf{b}_{(k)i} = (\Sigma_l\ \mathbb{Q}_k'\mathbf{V}_k\mathbb{Q}_k\ \mathbb{c}^2_{(k)il} + \Sigma_l\ (\mathbb{Q}_k\ \mathbb{c}_{(k)il}\,)'\mathbf{V}_k\,\mathbb{x}_l^{\,0}, \tag{27}$$

where $\mathbb{c}_{(k)il}$ is the $i^{\,\text{th}}$ element of $\mathbb{c}_{(k)l}$.

Instead of computation of (27) we can compute successive updates of $\mathbf{b}_{(k)i}$. This has the advantage that, in order to apply cone restrictions upon, this update can be projected onto the cone in the metric defined by the identity matrix. The updates are computed by:

$$\mathbf{b}_{(k)i}^{\text{adjusted}} = \mathbf{b}_{(k)i}^{\text{previous}} + \{(\Sigma_l\ (\mathbb{Q}_k \mathbf{c}_{(k)il})' \mathbf{V}_k\ \mathbb{x}_l^0) - \Sigma_l\ \mathbb{Q}_k' \mathbf{V}_k \mathbb{Q}_k \mathbf{c}_{(k)il}^2 \mathbf{b}_{(k)i}^{\text{previous}}\} / \beta_3, \tag{28}$$

Where β_3 is the first eigenvalue of the square matrix $(\Sigma_l\ \mathbb{Q}_k' \mathbf{V}_k \mathbb{Q}_k\ \mathbf{c}_{(k)il}^2)$. Again, after finding the optimal $\mathbf{B}_k$ and $\mathbf{C}_k$ we can postmultiply $\mathbf{B}_k$ and $\mathbf{C}_k$ with any $r_k \times r_k$ matrix to satisfy identification constraints, provided that there are no cone restrictions upon columns of $\mathbf{B}_k$. Notice that identification restrictions upon $\mathbf{B}_k$ and $\mathbf{C}_k$ are not necessary during the iterations.

Minimization over $\mathbf{C}_k$ for fixed $\mathbb{X}$, $\mathbb{y}_k$ and $\mathbb{B}_k$

$\mathbf{C}_k$ can be found by taking $\mathbf{C}_k = (\mathbb{B}_k' \mathbb{Q}_k' \mathbf{V}_k \mathbb{Q}_k \mathbb{B}_k)^+ \mathbb{B}_k' \mathbb{Q}_k' \mathbf{V}_k \mathbb{X}$. $\mathbf{C}_k$ can also be computed by alternating between rows or between elements of $\mathbf{C}_k$, updating one at a time keeping the remaining rows or columns fixed. We will not elaborate on that. It might be necessary to ascertain that all elements within each row of $\mathbf{C}_k$ have equal sign. If we compute $\mathbf{C}_k$ element-wise this is rather easy to accomplish. However we neither elaborate on this.

Some remarks about the implementation of the algorithm for common scale sets reduced rank generalized canonical correlation analysis

Based upon the ALS algorithm for nonlinear reduced rank GCCA a (somewhat provisional) program has been written in APL. This program does not incorporate data weights as given in Section 5 yet. Since we use an ALS algorithm we have to assign initial values to all sets of parameters but one. In our APL program initial values are assigned to the parameters in the $\mathbf{y}_k$'s, $\mathbf{B}_k$'s and in $\mathbf{C}$. We start with computing the optimal $\mathbf{X}$ given these initial values. There are two strategies to assign initial values to the parameters. The first strategy is to let the user specify the initial values. The second strategy is to generate random initial values. The current APL program allows the user to choose between the two strategies for each set of parameters, $\mathbf{y}_k$'s, $\mathbf{B}_k$'s and $\mathbf{C}$, separately. By trying out several different starts one should be able to avoid local minima. We expect that local minima will most likely occur when there are order restrictions within $\mathbf{y}_k$'s or within columns of $\mathbf{B}_k$'s.

7. References

Atkinson, A. C. (1985). *Plots, transformations and regression: An introduction to graphical methods of diagnostic regression analysis.* Oxford: Clarendon Press.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972). *Statistical inference under order restrictions.* New York, NY: Wiley.

Benzécri, J. P. et al. (1973). *L'analyse des données* (2 Vols.). Paris: Dunod.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th Annual Convention of the American Psychological Association,* 227-228.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31,* 33-42.

Dauxois, J. & Pousse, A. (1976). *Les analyses factorielles en calcul des probabilités et en statistique: Essai d'étude synthétique* [Thèse de Doct, troisieme cycle]. Toulouse: Université Paul Sabatier.

De Leeuw, J. & Van Rijckevorsel, J. A. L. (1988). Beyond homogeneity analysis. In J. A. L. van Rijckevorsel & J. De Leeuw (Eds.), *Component and correspondence analysis: Dimension reduction by functional approximation.* Chichester: Wiley.

De Leeuw, J. (1973). *Canonical analysis of categorical data* [Doctoral dissertation](2nd ed. 1984). Leiden: DSWO Press.

Feather, N. T. (1982). *Expectations and actions: Expectancy-value models in psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research.* Reading, Mass : Addison-Wesley.

Fortier, J. J. (1966). Simultaneous linear prediction. *Psychometrika, 31,* 369-381.

Gifi, A. (1981). *Nonlinear multivariate analysis.* Leiden: University of Leiden, Department of Data Theory.

Gifi, A. (1985). *PRINCALS* (UG 85-02). Leiden: University of Leiden, Department of Data Theory.

Gittins, R. (1985). *Canonical analysis: A review with applications in ecology.* Berlin: Springer-Verlag.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst et al. (Eds.), *The prediction of personal adjustment* (pp.319-348). New York, NY: Social Science Research Council.

Heiser, W. J. (1985). *Multidimensional scaling by optimizing goodness-of-fit to a smooth hypothesis.* Leiden: University of Leiden, Department of Data Theory.

Heiser, W. J. (1987). Correspondence analysis with least absolute residuals. *Computational Statistics & Data Analysis, 5,* 337-356.

Heiser, W. J., Meulman, J., Van der Lans, I. A. & Van den Berg, G. M. (1988). *Notes on MULTIPALS* (RR 88-09). Leiden: University of Leiden, Department of Data Theory.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika, 28,* 321-327.

Izenman, A. J. (1975). Reduced rank regression for the multivariate linear model. *Journal of Multivariate Analysis, 5,* 248-264.