288

Nonlinear Canonical Correlation Analysis with k Sets of Variables

Research Report 87-8

Eeke van der Burg
Jan de Leeuw

department of EDUCATION

Division of Educational Measurement and Data Analysis

University of Twente

NONLINEAR CANONICAL CORRELATION ANALYSIS WITH K SETS OF VARIABLES

Eeke van der Burg

Jan de Leeuw[1]

[1] The present address of the author is: University of California Los Angeles, Departments of Psychology and Mathematics, 405 Hilgard Ave., Los Angeles, CA 90024, USA.

Abstract

The multivariate technique OVERALS is introduced as a nonlinear generalization of canonical correlation analysis (CCA). This is done by first introducing two sets CCA, which is familiar from multivariate analysis text books. Then two sets CCA is expanded to a *k* sets technique. Next optimal scaling with single transformations is introduced. Finally multiple transformations are added.

Keywords: canonical correlation, optimal scaling, nonlinear transformation, *k* sets, homogeneity analysis.

Introduction

CCA for two sets of variables is a standard technique, first described by Hotelling (1936). A nonlinear version was introduced by Young, De Leeuw & Takane (1976), Gifi (1981, chap. 6) and Van der Burg and De Leeuw (1983). CCA with k sets of variables is a generalization of two sets CCA. Many generalizations are possible. Descriptions can be found in Steel (1951), Horst (1961), Carroll (1968), Kettenring (1971), Gifi (1981, chap. 6), De Leeuw (1984), Van de Geer (1984, 1986, part IV), and Meulman (1986, chap 4 & 6). Nonlinear CCA with k sets of variables is a generalization of linear CCA with k sets, but also of nonlinear CCA with two sets. Which generalizations are used depends on subjective choices. In this paper the approach of Gifi (1981, chap. 6) is followed. This is also discussed by Verdegaal (1985, 1986), and Van der Burg, De Leeuw & Verdegaal (1986). Gifi, as well as Verdegaal, and Van der Burg *et al.* introduce OVERALS as homogeneity analysis (multiple correspondence analysis) with k sets of variables. Conceptually this is a very reasonable approach. However historically it is not. CCA is typically represented as a technique with a vector interpretation of the variables. Homogeneity analysis uses a completely different interpretation. Therefore OVERALS is introduced in this paper as a nonlinear generalization of k sets CCA (cf. Van der Burg *et al.*, 1984).

Two sets of variables

Two sets CCA is a technique that computes linear combinations of sets of variables which correlate in an optimal way. Generalized (or *k* sets) CCA does the same for *k* sets. Nonlinear CCA relates sets of nonlinearly transformed variables in an optimal way.

Let us consider the case in which ther are only two sets in detail. Suppose that the sets of variables are denoted by $\mathbf{H}_1$ ($n \times m_1$) and $\mathbf{H}_2$ ($n \times m_2$). Between brackets the dimensions of the matrix are given. Each column $\mathbf{h}$ of $(\mathbf{H}_1, \mathbf{H}_2)$ corresponds to a variable. It consists of n scores for objects (observations). The number of variables in the first and second set is denoted by m_1 and m_2. respectively. Suppose the p weights for each variable are collected as row vectors of the matrices $\mathbf{A}_1$ ($m_1 \times p$) and $\mathbf{A}_2$ ($m_2 \times p$). Assuming that each vector $\mathbf{h}$ is standardized (i.e. mean zero and variance one) CCA is:

(1) minimize $\mathrm{tr}(\mathbf{H}_1\mathbf{A}_1 - \mathbf{H}_2\mathbf{A}_2)'(\mathbf{H}_1\mathbf{A}_1 - \mathbf{H}_2\mathbf{A}_2)$
over $\mathbf{A}_1$ and $\mathbf{A}_2$ subject to the conditions that
$\mathbf{A}_1'\mathbf{H}_1'\mathbf{H}_1\mathbf{A}_1 = n\mathbf{I}$ and $\mathbf{A}_2'\mathbf{H}_2'\mathbf{H}_2\mathbf{A}_2 = n\mathbf{I}$

The weighted linear combinations $\mathbf{H}_1\mathbf{A}_1$ and $\mathbf{H}_2\mathbf{A}_2$ are called *canonical variates* (in the literature also the term canonical variables is used). For each set they are uncorrelated and have variance one (I refers to the identity matrix). Thus the canonical variates are orthogonal. Formulation (1) results in a maximum sum of correlations between the canonical variates (of the different sets). Usually these correlations are called the *canonical correlations*. Note that (1) is formulated as a least squares minimization problem. This is not the usual approach, but it does result in a solution identical to the classical two sets CCA solution.

K sets of variables

If there are k sets of variables instead of two sets, we have $(\mathbf{H}_1, \ldots, \mathbf{H}_k)$ and $(\mathbf{A}_1, \ldots, \mathbf{A}_k)$, with $\mathbf{H}_t$ (n x m_t) and $\mathbf{A}_t$ (m_t x p) for t=1,...,k. Again canonical variates are wanted which are maximally related. In this case many canonical correlations have to be considered. One way to deal with this is to require that the canonical variates are as similar as possible to $\mathbf{X}$, a p-dimensional orthogonal representation of the objects, also called *object scores*. Then the orthogonality condition on the canonical variates for each of the sets is usually dropped (otherwise the problem is more difficult to solve). This gives:

(2) minimize $\Sigma_{t=1}^{k} \mathrm{tr}(\mathbf{X} - \mathbf{H}_t\mathbf{A}_t)'(\mathbf{X} - \mathbf{H}_t\mathbf{A}_t)$

over $\mathbf{X}$ and $\mathbf{A}_1, \ldots, \mathbf{A}_k$ subject to the conditions that

$\mathbf{X}'\mathbf{X} = n\mathbf{I}$ and $\mathbf{u}'\mathbf{X} = 0$, with $\mathbf{u}$ a vector (dimension n) consisting of only ones. Note that restricting (2) to two sets results in (1). In that case $\mathbf{X}$ can be expressed in terms of $\mathbf{H}_1\mathbf{A}_1$ and $\mathbf{H}_2\mathbf{A}_2$, in addition the condition $\mathbf{X}'\mathbf{X}=n\mathbf{I}$ can be translated into the usual constraints for canonical variates. Thus expression (2) is a proper generalization of (1). Also note that we are dealing again with a least squares problem (both for $\mathbf{X}$ and $\mathbf{A}_t$), which can be solved alternatingly. Carroll (1968), De Leeuw (1973), Saporta (1975), and Gifi (1981, chap. 6) use a similar formulation for generalized CCA, consequently they use the same optimization criterion and the same constraints.

In other approaches different criteria are used. Define $\mathbf{R}$ (kp x kp) as the correlation matrix of $(\mathbf{H}_1\mathbf{A}_1,\ldots,\mathbf{H}_k\mathbf{A}_k)$. Steel (1951), Horst (1961), Kettenring (1971) and Van de Geer (1984, 1986) optimize properties of $\mathbf{R}$ over choice of weights $\mathbf{A}$. Their optimality criteria vary from the determinant of $\mathbf{R}$ to the largest or smallest eigenvalue of $\mathbf{R}$, and to the sum of the (squared) elements of $\mathbf{R}$.

In addition to different criteria, we also find that, in the literature *successive* and *simultaneous* solutions are distinguished (Ten Berge, 1977; Van de Geer, 1984, 1986, part IV). In the simultaneous case all p solutions (dimensions) are computed at once. This also holds for formulation (2). However Van der Burg *et al.* (1986) prove that the matrix $\mathbf{X}$ corresponds to an eigenvector solution, so that (2) also gives successive solutions. The successive approach solves for each dimension in succession, imposing additional restrictions with regard to previous solutions. Restrictions always refer to *orthogonality constraints*, which can be either *weak* or *strong* (Dauxious & Pousse, 1976). Weak constraints require orthogonal sums of canonical variates (or object scores). Strong constraints imply canonical variates to be orthogonal within sets. For a more detailed discussion of optimizing criteria, simultaneous or successive solutions, and orthogonality constraints we refer to Gifi (1981, chap. 6).

A completely different approach to k sets canonical correlation analysis is found in Meulman (1986, chap. 2 and 4). Starting from (2), k sets CCA is reformulated in distance terms, so that the relation to multidimensional scaling becomes clear.

Optimal scaling

In this section k sets CCA with optimal scaling is discussed. Two different formulations are used.

Several authors apply *nonlinear transformations* in multivariate techniques. This can be done in the form of *optimal scaling* (Young, 1981). Optimal scaling means that for each variable a nonlinear transformation is permitted, such that it maximizes the analysis criterion. Naturally the transformations are restricted by measurement constraints.

Thus combining the CCA problem with measurement restictions gives CCA with optimal scaling.

The optimal scaling is included in k sets CCA in the following manner (cf. Young *et al.*, 1976; Van der Burg & De Leeuw, 1983). Instead of using the original variables $\mathbf{h}$ (columns of $\mathbf{H}_1,\ldots,\mathbf{H}_k$), transformed variables $\mathbf{q}$ (columns of $\mathbf{Q}_1,\ldots,\mathbf{Q}_k$) are used, which are optimally scaled (and thus satisfy the measurement restrictions). The matrices $\mathbf{Q}_t$ and $\mathbf{H}_t$ have the same size. Geometrically this means that instead of considering a variable as a *vector*, a variable is considered as a *cone of vectors* of which one is chosen (such that the analysis criterion is maximized). This cone is completely defined by measurement restrictions (De Leeuw, 1977). Satisfying measurement restrictions for *numerical* variables means that $\mathbf{q}$ is a linear transformation of $\mathbf{h}$. For *ordinal* variables it means that $\mathbf{q}$ is a monotone transformation of $\mathbf{h}$ (Kruskal & Shephard, 1974, secondary approach) and for *nominal* variables it means that $\mathbf{q}$ is equivalent with $\mathbf{h}$ (i.e. similar observations or *ties* of $\mathbf{h}$ correspond to ties in $\mathbf{q}$). Thus k sets CCA with optimal scaling is

(3) minimize $\Sigma_{t=1}^{k}\mathrm{tr}(\mathbf{X} - \mathbf{Q}_t\mathbf{A}_t)'(\mathbf{X} - \mathbf{Q}_t\mathbf{A}_t)$
over $\mathbf{X}$ and $\mathbf{A}_1,\ldots,\mathbf{A}_k$ subject to the conditions that
$\mathbf{X}'\mathbf{X} = n\mathbf{I}$, $\mathbf{u}'\mathbf{X} = 0$ and $\mathbf{q} = \mathbf{f}(\mathbf{h})$ with $\mathbf{f} \in C(\mathbf{h})$,

where $C(\mathbf{h})$ refers to the set of permissible transformations of $\mathbf{h}$, and $\mathbf{f}$ refers to a transformation. As for each variable only one transformation is employed. (3) defines k sets CCA with *single transformations*. Furtheron it will be shown that multiple transformations are also possible, which leads to k sets CCA with single and/or multiple transformations. This technique is called OVERALS, therefore one can refer to k sets CCA with optimal scaling (3) as OVERALS with single transformations.

Considering a variable as a collection of category scores, which means automatically that variables are supposed to be *discrete*, makes measurement restrictions perhaps more clear. When transformations are defined with respect to categories (obtaining *category quantifications*), ties are automatically maintained. Then nominal transformations do not employ additional restrictions, ordinal transformations require the category quantifications to be a monotone transformation of the original category scores, and numerical transformations yield a linear transformation of the original category scores. In this paper the terms transformations (of variables) and category quantifications are used both. They mean exactly the same thing, although the corresponding interpretations are perhaps different. In fact interpreting a variable as a collection of observations (instead of category scores) opens the possibility of continuous transformations. These are discussed by Kruskal & Shephard (1974) in the context of multidimensional scaling (type I transformations), by De Leeuw, Young & Takane (1976) with regard to additive structure analysis, by Young, Takane & De Leeuw, (1978), De

Leeuw & Van Rijckevorsel (1980) and Kuhfeld, Young & Kent (1987) in the framework of nonlinear principal component analysis. In this paper continuous transformations are not considered, therefore both interpretations are similar and it does not matter which terminology is used.

If variables are considered to be discrete they can be characterized completely in terms of their category scores. Let us consider variable $\mathbf{y}_j$, the j'th column of matrix $(\mathbf{Y}_1,\ldots,\mathbf{Y}_k)$, and let $\mathbf{G}_j$ (n x k_j) be an indicator matrix for the k_j categories of variable $\mathbf{y}_j$ (i.e. element (i,r) of $\mathbf{G}_j$ is one if object i belongs to category r, and zero otherwise), in addition let $\mathbf{c}_j$ be the corresponding k_j–vector of category scores. (Note that k (number of sets) and k_j (number of categories for variable $\mathbf{h}_j$) have a different meaning). Then variable $\mathbf{h}_j$ can be written as

(4) $$\mathbf{h}_j = \mathbf{G}_j\mathbf{c}_j.$$

An example may clarify the notation. Suppose a data matrix of 8 objects and 3 variables is given with numbers of categories 2,4,3 respectively (Fig.1). Then the corresponding indicator matrices are $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$, and the vectors of category scores are $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_3$ (Fig. 1). Multiplication of $\mathbf{G}_1$ with $\mathbf{c}_1$ gives $\mathbf{h}_1$, of $\mathbf{G}_2$ with $\mathbf{c}_2$ gives $\mathbf{h}_2$, etc..

Insert Table 1 about here

A transformation $\mathbf{q}_j$ of $\mathbf{h}_j$ can be expressed analogously in terms of the category quantifications $\mathbf{z}_j$ and the indicator matrix $\mathbf{G}_j$. Then

(5) $$\mathbf{q}_j = \mathbf{G}_j\mathbf{z}_j.$$

Now suppose matrix $\mathbf{Q}_t$ corresponds to $(\mathbf{q}_r, \ldots, \mathbf{q}_s)$, with $J(t)=(r, \ldots, s)$ the indices corresponding to set t. Then matrix $\mathbf{Q}_t\mathbf{A}_t$ can be rewritten as

(6) $$\mathbf{Q}_t\mathbf{A}_t = \Sigma_{j\in J(t)}\mathbf{G}_j\mathbf{z}_j\mathbf{a}_j',$$

with $\mathbf{a}_j'$ rows of $\mathbf{A}_t$. In this expression $\mathbf{z}_j$ and $\mathbf{a}_j$ are vectors of parameters and $\mathbf{G}_j$ is fixed. The matrix $\mathbf{Y}_j=\mathbf{z}_j\mathbf{a}_j'$ defines a line in the p-dimensional space with the category quantifications $\mathbf{z}_j$ on this line, and with direction cosines proportional to the weights $\mathbf{a}_j$. With the help of (6) OVERALS with single transformations (3) can be rewritten as

(7) minimize $\Sigma_{t=1}^{k}\mathrm{tr}(\mathbf{X} - \Sigma_{j\in J(t)}\mathbf{G}_j\mathbf{Y}_j)'(\mathbf{X} - \Sigma_{j\in J(t)}\mathbf{G}_j\mathbf{Y}_j)$
over $\mathbf{X}$ and $\mathbf{Y}_j$ subject to the conditions that
$\mathbf{u}'\mathbf{X} = 0$, $\mathbf{X}'\mathbf{X} = n\mathbf{I}$ and
$\mathbf{Y}_j = \mathbf{z}_j\mathbf{a}_j'$ with $\mathbf{z}_j \in C_j$ and $\mathbf{a}_j$ unrestricted.

The notation C_j is used for the set of permissible quantifications of the category scores $\mathbf{c}_j$ (similar, of course, to the set of permissible transformations of variable $\mathbf{h}_j$). Note that (7) does not yield results different from (3). Only the formulations differ.

Multiple transformations

What happens if the condition that $\mathbf{Y}_j$ takes the form $\mathbf{z}_j\mathbf{a}_j'$ is dropped for some variables? Then the transition from single transformations to *multiple transformations* is made. Geometrically it means that category quantifications are no longer required to be on a line. Thus the vector interpretation (or vector-in-cone interpretation) is abandoned. This results in OVERALS, which is k sets CCA with single and/or multiple transformations (Gifi, 1981, chap. 6). Thus OVERALS is

(8) minimize $\Sigma_{t=1}^{k}\mathrm{tr}(\mathbf{X} - \Sigma_{j\in J(t)}\mathbf{G}_j\mathbf{Y}_j)'(\mathbf{X} - \Sigma_{j\in J(t)}\mathbf{G}_j\mathbf{Y}_j)$
over $\mathbf{X}$ and $\mathbf{Y}_j$ subject to the conditions that
$\mathbf{u}'\mathbf{X} = 0$ and $\mathbf{X}'\mathbf{X} = n\mathbf{I}$, and for some variables
$\mathbf{Y}_j = \mathbf{z}_j\mathbf{a}_j'$ with $\mathbf{z}_j \in C_j$.

Expression (8) is the definition of OVERALS, also given by Van der Burg *et al.* (1986). They introduce OVERALS from a different point of view. They start with homogeneity analysis or multiple correspondence analysis (Guttman, 1941; Benzécri *et al.*, 1973; De Leeuw, 1973; Nishisato, 1980; Gifi, 1981; Greenacre, 1984) and then introduce sets of variables by additive restrictions on interactive codings (per set of variables). Next they combine this additive homogeneity analysis with optimal scaling. In this paper another route is followed. Starting from classical CCA, the additivity of variables within sets comes in naturally. Linear combinations of variables are generalized to linear combinations of optimally scaled variables. Then this optimal scaling is further generalized to multiple transformations. From a geometrical point of view the latter step does not seem natural, as the vector interpretation of a variable is dropped. However it is necessary to link OVERALS with single transformations to homogeneity analysis so that the interrelations between the various techniques become more clear. Because of this link of OVERALS with homogeneity analysis OVERALS can also be viewed as multidimensional scaling with sets of variables, and also as *k* sets principle component analysis. If the data contain no set structure additivity restrictions on interactive variables per set are automatically satisfied. Therefore OVERALS with one variable per set is the same as nonlinear principal component analysis (De Leeuw & Van Rijckevorsel, 1980; Gifi, 1984). If, in addition, only multiple transformations are used we are back to homogeneity analysis.

Copies of variables

One way of avoiding the complex step from single to multiple transformations is by introducing *copies* (De Leeuw, 1984). This means that the same variable may occur more than once in a set. For each copy a single transformation is formed so that a variable is represented by several vectors instead of one. Thus a vector interpretation of the variables can be maintained, although it is more easy to imagine one vector than two or more. An example may help. Suppose the variable *age* correlates with *health* in a linear way but correlates with *smoking behaviour* nonlinearly (e.g. young and old do not smoke, and people between thirty and sixty smoke relativily much). Then, if *age* must be correlated with *health* and smoking *behaviour* at the same time (in one set), it may be usefull if *age* is represented by two copies. For a formal proof that a multiple transformation corresponds to p transformed copies, we refer to Van der Burg *et al.* (1986).

If we compare an OVERALS algorithm with copies to an algorithm with multiple transformations there are several advantages. In the first place there is the possibility of determining how many copies are wanted. In the case of multiple transformations always p quantifications are used. A second advantage is, that the measurement level of each copy can be fixed separately. If each dimension of a multiple transformation is considered as a separate solution, it appears from the definition of OVERALS (8), that only nominal transformations are possible. The tie restrictions are automatically satisfied, when variables are considered as a collection of categories. For multiple transformations, however, no further restrictions were formulated in (8). For that reason multiple transformations are often referred to as *multiple nominal quantifications*.

A disadvantage may be that the link to homogeneity analysis becomes rather confusing. Fundamental in homogeneity analysis is the p-dimensional representation of objects and categories. If the number of copies is smaller than p and larger than one, we are dealing with a constrained representation, which is unusual in homogeneity analysis.

A second disadvantage, and maybe a more serious one, is the fact that a user must determine beforehand how many copies and which measurement levels must be used in an OVERALS analysis. This may be very difficult.

For OVERALS with single and multiple transformations an algorithm is implemented in a computer program (also called OVERALS) (Verdegaal, 1986; Van der Burg *et al.*, 1986). For OVERALS with copies no special computer program exists. Of course the OVERALS computer program can be used for it. Then only single measurement levels must be employed, in addition copies of the variables must be added to the input data matrix.

Missing data

The OVERALS algorithm does not contain any provisions for dealing with missing values. Gifi (1981, chap. 6) and Verdegaal (1985) treat the case of incomplete data. This is also discussed here to make the paper more complete.

Suppose a variable $\mathbf{h}_j$, belonging to set $\mathbf{H}_t$, misses one or more observations. Define matrix $\mathbf{M}_t$ (n x n), a diagonal matrix showing which observations are missing for the variables of set t. $\mathbf{M}_t$ is defined as follows: element (i,i) is 1 if no variable of set t misses observation i and zero if one of the variables of set t misses observation i. Thus it does not matter which of the variables $\mathbf{h}_j$, $j \in J(t)$, misses an observation. Let $\mathbf{M}_*$ denote the average over the missing matrices, i.e. $\mathbf{M}_* = \Sigma_t \mathrm{M}_t / k$. OVERALS for incomplete data is defined as follows

(9) minimize $\Sigma_{t=1}^{k} \mathrm{tr}(\mathbf{X} - \Sigma_{j \in J(t)} \mathbf{G}_j \mathbf{Y}_j)' \mathbf{M}_t (\mathbf{X} - \Sigma_{j \in J(t)} \mathbf{G}_j \mathbf{Y}_j)$
over $\mathbf{X}$ and $\mathbf{Y}_j$ subject to the conditions that
$\mathbf{u}'\mathrm{M}_*\mathbf{X} = 0$ and $\mathbf{X}'\mathbf{M}_*\mathbf{X} = n\mathrm{I}$, and for some variables
$\mathbf{Y}_j = \mathbf{z}_j \mathbf{a}_j$ with $\mathbf{z}_j \in C_j$.

The solutions for $\mathbf{X}$ and $\mathbf{Y}_j$ are given without a proof. The way they are obtained is very similar to the way for 'nonmissing' OVERALS. Van der Burg *et al.*(1986) discuss these solutions extensively. Gifi (1981, chap. 6) treats solutions for 'missing' OVERALS briefly, and Verdegaal (1985) is rather detailed. Using a hat (^) for a new update the solution for object scores is

$$(10) \qquad \hat{\mathbf{X}} = (1/k)\mathbf{M}_*^{-1}\mathbf{J}(\Sigma_{t=1}^{k}\mathbf{M}_t\Sigma_{j\in J(t)}\mathbf{G}_j\mathbf{Y}_j)\Phi^{-1},$$

with

$$(11) \qquad \mathbf{J} = \mathbf{I} - (\mathbf{M}_*\mathbf{u}\mathbf{u}')/(\mathbf{u}'\mathbf{M}_*\mathbf{u})$$

the centering operator, and Φ a symmetric matrix of Lagrange multipliers. The solution for multiple category quantifications is

$$(12) \qquad \hat{\mathbf{Y}}_j = (\mathbf{G}_j'\mathbf{M}_t\mathbf{G}_j)^+\mathbf{G}_j'\mathbf{M}_t(\mathbf{X} - \Sigma_{r\in J(t)}\mathbf{G}_r\mathbf{Y}_r + \mathbf{G}_j\mathbf{Y}_j).$$

where $(\ )^+$ refers to the Moore Penrose inverse. If the matrices $\mathbf{D}_j$ are defined as

$$(13) \quad \mathbf{D}_j = \mathbf{G}_j'\mathbf{M}_t\mathbf{G}_j,$$

the single category quantifications and weights are obtained from the multiple quantifications in the same way as for 'nonmissing' OVERALS. Note that substitution of $\mathbf{I}$ for $\mathbf{M}_t$ in (9) gives (8), which is the definition of OVERALS for complete data. The definition of OVERALS for incomplete data corresponds to the definitions of nonlinear principal component analysis (PRINCALS) and homogeneity analysis

(HOMALS) for incomplete data (Gifi, 1981, chap. 3 & 5; Meulman, 1982). It means that in case of no set structure (i.e. $J(t)=t$) those techniques are exactly the same as OVERALS for incomplete data.

Application

An example is taken from Van Rijckevorsel (1987). It concerns the car data originating from the American consumers report, April 1980 (Table 2). Winsberg & Ramsey (1983) also used this file. The data contain some basic characteristics of 33 popular cars available in the U.S, like price, engine size, weight, and fuel consumption. Discretizations are used according to the knots given by Van Rijckevorsel (1987, page 37).

Insert Table 2 about here

As the variables of the car data represent three kinds of characteristics, the data were considered as consisting of three sets of variables. The first set contains price (P), the second one consists of the 'psysical' characteristics weight (W) and size (S), and the third set of the performance characteristics, miles/gallon in the city (MGC or C) and on the highway (MGH or H)).

Insert Table 3 about here

Several analyses were done. Firstly all the variables were taken as ordinal. Although the variables seem to be measured on interval level, we are not sure about the linearity of the relations between the variables, therefore we used an ordinal measurement level. A two dimensional solution was considered. The *generalized canonical correlations* are shown in Table 3. The *transformations* of the variables are given in Fig. 1 (first row); horizontally the original *category scores* of each variable are plotted and vertically *the category quantifications*. The *component loadings* are given in Fig. 2 and *object scores* in Fig. 3. Cars are labeled by 1 to 9, and A to X (cf. Table 2) . For an interpretation of OVERALS results we refer to Van der Burg at al. (1986). Component loadings and object scores belong to the same space, however they are plotted next to each other because of a difference in scale.

Figures 2 and 3 show that the first dimension (horizontally) is dominated by price, weight and MGC. The second dimension gives differences in MGH and size. As the transformation of size is rather flat for the higher values (see Fig. 1), no difference is expected between the smaller and the larger cars except for the very small ones. Small cars lie in the direction opposite of the S—vector in Fig. 3. There we find 1,2,3,5,8,9,F,G. They all have a size under 100 and a fuel consumption on the highway which is rather low. At the right of Fig. 3, we find I,K,L,M,O,P,Q,S,T,U,V,W,X, which are the more expensive cars (above 6700 dollars), the havier ones (above 3000 kilos), and the less economically driving cars (MGC less than 14). High in the plot we find the 'medium' cars 6,B,C,D,E,H,J,N. Somewhat lower cars 4,7,A are found, which have a rather small size, and use comparativily little fuel. The latter characteristic also holds for car R.

A second analysis was performed with copies of weight and size (taken as nominal). We wondered if higher prices always correspond with bigger, heavier, and less economically driving cars. The generalized canonical correlations were .892 and .603, which is a little better than in analysis I (see Table 3). The transformations of the variables are shown in Fig. 1 (second row). Weight and size contain an ordinal and a nominal transformation, corresponding to W1/S1, and W2/S2 respectivily. The component loadings can be found in Fig. 4, and the object scores in Fig. 5. The S2—vector is very short, therefore the nominal version of size (S2) is of no importance. Thus the higher canonical correlations are mainly caused by the nominal weight variable (W2). The first dimension is still dominated by price and MGC, now combined with size (S1), and weight (W1). The second dimension shows MGH (H), together with the nominal weight (W2). Transformation of W2 is rather capricious. Cars of middle weight (2500–3000) get a negative category quantification, while the others get a positive quantification. This means that middle weight cars can be found in the direction opposite to the W2-vector, and that the other cars lie in the direction of W2. Indeed cars 6,B,C,E,H,J,N,R (Fig. 5) have a medium weight. Again R is placed relativily low as it uses comparatively little fuel on the highway. The first dimension is very similar in both analyses. It discriminates cheap cars from medium priced and expensive cars. This correlates with size and MGC. Thus the little, cheap and economically driving cars are found under left in Fig. 5, similar to Fig.3. Only cars 4,7,A have grown very close in Fig. 5 to the cheap cars as they are rather light. In analysis I the engine size dominated the second dimension. In analysis 2 it is the weight.

Insert Figures 4 and 5 about here

A third analysis was performed, but now MGC and MGH were copied and considered to be nominal. Results are shown in Fig. 3 (third row), Fig. 6 (component loadings) and Fig. 7 (object scores). The canonical correlations do not differ much different from those of analysis II. The first dimension does a little better (Table 3). In Fig. 7 the first dimension is dominated by price, weight and MGH. The second dimension is determined by MGC and size. This solution is in fact more similar to analysis I than to analysis II. The same clusters are found. Fig. 7 shows that the group 4,7,A has moved away from the cheap cars as in Fig. 3, because these cars use more fuel in the city than the small cars do (but still a moderate amount). Although some small cars are also expensive in use, e.g. cars 1,2,G (with MGC=18.3, 18.6, 19.7) these cars show more resemblance (in size and weight) to the small ones than 4,7,A do. Therefore the nominal transformation of MGC jumps a lot (Fig. 1). Categories 4 and 6 are quantified lowest, categories 1 and 5 in the middle, and categories 2 and 3 highest. Cars K and T still lie in the position between big cars and medium cars. The upper cluster is now determined by MGC, especially objects with a score between 9 and 15 can be found in the upper part. These objects are: 6,B,C,D,E,H,J,K,N,R and T.

Insert Figures 6 and 7 about here

All analyses discriminate between small, medium,and big cars. Differences between the solutions are found mainly in the second dimension. The accent of the second dimension depends on which variables are copied. However all solutions show the medium cars at the top. Analysis I highlights size and MGH, analysis II weight and MGH and analysis III MGC and size. All solutions show that car 1 (audi 4000) is expensive for a small car. Cars 4,7,A (Datsun 510, Honda Accord, Plymouth Horizon) have characteristics of small and medium cars. Cars K,T (Chev Malibu, Chev Impala) use comparativily little fuel in the city, but for the rest they are like big cars. Cars M,I (Dodge Diplomat, AMC Eagle) are good representatives for big cars. Typical small cars are 2,3,5 (Chev Chevette, Datsun 210, Dodge Colt), and typical medium cars are 6,B,H (Ford Mustang, Plymouth Sapparo, Toyota Corona).

As the relations between the variables in this data set are strongly ordinal, the solutions with nominal copies do not reveal completely new aspects. Only a difference in accentuation is obtained, showing unexpected irregularities in car charisteristics.

Discussion

The OVERALS technique described in this article fits very well in the tradition of *k* sets CCA, as far as such a tradition exists. The earliest predecessor is Carroll (1968), who compares his work with that of Horst (1961). Some years later Kettenring (1971) also discussed several *k* sets methods. Of those early articles on *k* sets CCA only Carroll's method is based on minimization of a loss between object scores and canonical variates of all sets together. In this article we used a similar formulation for OVERALS, and we extended it.

Earlier the OVERALS technique has been treated as a type of homogeneity analysis (Gifi, 1981, chap. 6; De Leeuw, 1984; Verdegaal, 1985, 1986; Van der Burg *et al.*, 1986). Treated in this way it fits perfectly in the Gifi–system of nonlinear multivariate analysis (Gifi, 1981; De Leeuw, 1984). This system combines the ideas of correspondence analysis and homogeneity analysis with the priciples of optimal scaling. Optimal scaling, however, can also easily be built into *linear* multivariate analysis, offering a system that is not as general as the Gifi–system. However, the extension using copies fills the gap between the two systems. It also offers new possibilities with regard to data analysis.

Multivariate analysis with optimal scaling and copies has the advantage that the old idea of a variable representing a vector is maintained. Traditionally CCA is based on the vector interpretation. Therefore this article fits better in the CCA–tradition than the interpretation of Van der Burg *et al.* (1986) does.

In addition, the method of copies gives the possibility of handling nonlinearities more carefully than *k* sets homogeneity analysis does. From the homogeneity point of veiw, one can only decide that a variable is represented either by a vector, or by unrestricted (in the sense of a measurement level) categories. With the help of copies one can be more subtle and use as many copies as are needed from the data analysis standpoint, imposing measurement restrictions on each copy separately. Especially if one prefers to keep the vector interpretation, and if interrelations are expected to vary from variable to variable, the method of copies provides a nice tool for data analysis.

References

Benzécri, J.P. et al. (1973). l'Analyse des Données (Data Analysis) (2 vols). Paris: Dunod

Carroll, J.D. (1968). Generalization of canonical correlation analysis to three or more sets of varaibles. Proceedings 76'th APA Convention, 227–228.

Dauxious, J. & Pousse, A. (1976). Les Analyses Factorielles en Calcul des Probabilités et en Statistique: Essay d'Etude Synthetique. Doctoral Dissertation (Thèse de Doct., 3e cycle). Toulouse: Université Paul Sabatier.

De Leeuw, J. (1973). Canonical Analysis of Categorical Data. Dissertation, University of Leiden. Reissued, Leiden: DSWO Press, (1983).

De Leeuw, J. (1984). The Gifi-system of nonlinear multivariate analysis. In E. Diday et al. (eds), Data Analysis and Informatics III. Amsterdam: North Holland Publishing Company.

De Leeuw, J. & Van Rijckevorsel, J.J.L. (1980). HOMALS and PRINCALS, some generalizations of principal component analysis. In E. Diday et al. (eds). Data Analysis and Informatics. Amsterdam: North Holland Publising Company.

De Leeuw, J., Young, F. & Takane, Y. (1976). Additive structure analysis in qualitative data: an alternating least squares method with optimal scaling features. Psychometrika, 29, 55–65.

Gifi, A. (1981). Nonlinear Multivariate Anlysis. University of Leiden, Department of Data Theory. Reissued, Leiden: DSWP Press (1987).

Greenacre, M. (1984). Theory and Applications of Correspondence Analysis. New York: Academic Press.

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. New York: Social Science Research Council.

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

Horst, P. (1961). Relations among *m* sets of measures. Psychometrika, 26, 129–149.

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

Kuhfeld, W.F , Young, F.W. and Kent, D.P. (1987). New developments in psychometric amd market research procedures. To appear in: SAS Users Group International Proceedings.

Meulman, J. (1982). Homogeneity Analysis of Incomplete Data. Leiden: DSWO Press.

Meulman, J. (1986). A Distance Approach to Nonlinear Multivariate Analysis. Leiden: DSWO Press.

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

Saporta, G. (1975). Liaisons Entre Plusieurs Ensembles de Variables et Codage de Données Qualitatives. Doctoral Dissertation (Thèse de Doct., 3e cycle). Paris: University of Paris VI

Steel, R.G.D. (1951). Minimum generalized variance for a set of linear functions. Annals of Mathematical Statistics, 22, 456–460.

Ten Berge, J.M.F. (1973). Optimizing Factorial Invariance. Doctoral Dissertation. Groningen: University of Groningen

Van de Geer, J.P. (1984). Linear relations between k sets of variables. Psychometrika, 49, 79–94.

Van de Geer, J.P. (1986).

Van der Burg , E. & De Leeuw, J. (1983). Non–linear canonical corrlation. British Journal of Mathematical and Statistical Psychlogy, 36, 54–80.

Van der Burg , E., De Leeuw, J. & Verdegaal, R. (1984). Non–linear canonical corrlation with m sets of variables. University of Leiden, Department of Data Theory. Research Report RR–84–12.

Van der Burg , E., De Leeuw, J. & Verdegaal, R. (1986). Homogeneity analysis with *k* sets of variables: an alternating least squares method with optimal scaling features. University of Twente: Dept: of Education. RR–86–5. To be published in Psychometrika.

Van Rijckevorsel, J.L.A. (1987). The Application of Fuzzy Coding and Horshoes in Multiple Correspondence Analysis. Leiden: DSWO Press.

Verdegaal (1985). Meer Sets Analyse voor Kwalitatieve Gegevens (Multi-set Analysis of Qualitative Data). University of Leiden, Department of Data Theory. Research Report RR-85-14.

Verdegaal, R. (1986). OVERALS. University of Leiden, Department of Data Theory. User's Guide UG-86-01.

Winsberg, S. and Ramsey, J.O. (1983). Monotone spline transformations for dimension reduction. Psychometrika, 48, 575-595.

Young, F. (1981). Qualitative Analysis of Qualitative Data. Psychometrika, 46, 247-388.

Young, F, De Leeuw, J. & Takane, Y. (1976). Regression with Qualitative and Quantitative variables: and alternating least sqiares method with optimal scaling features. Psychometrika, 41, 505-529.

Young, F., Takane, Y. & De Leeuw, J. (1978). The Principal Components of Mixed Measurement Multivariate Data: an Alternating Least Squares Method with Optimal Scaling Features. Psychometrika, 43, 279-281.

Table 1

Data matrix **H**, corresponding indicator matrices G_1, G_2, G_3, and vectors of category scores $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$.

	H			G_1		G_2				G_3		
	1	2	3	a	b	p	q	r	s	u	w	v
1	a	p	v	1	0	1	0	0	0	0	0	1
2	a	r	u	1	0	0	0	1	0	1	0	0
3	b	s	w	0	1	0	0	0	1	0	1	0
4	b	q	u	0	1	0	1	0	0	1	0	0
5	a	p	w	1	0	1	0	0	0	0	1	0
6	b	r	u	0	1	0	0	1	0	1	0	0
7	b	q	v	0	1	0	1	0	0	0	0	1
8	b	r	w	0	1	0	0	1	0	0	1	0

a, b	<— $\mathbf{c}_1$
p, q, r, s	<— $\mathbf{c}_2$
u, v, w	<— $\mathbf{c}_3$

Table 2

The car data from the American consumers report, April 1980, with knot squence used for discretization.

CAR TYPE	CODE	PRICE	SIZE	MGC	MGH	WEIGHT
Audi 4000	1	7700	98	18.3	35.8	2190
Chev Chevette	2	5100	98	18.6	35.9	2120
Datsun 210	3	4750	86	25.2	40.7	2020
Datsun 510	4	5950	119	21.9	42.7	2430
Dodge Colt	5	4900	98	24.2	41.2	1880
Ford Mustang	6	5800	141	15.6	29.8	2610
Honda Accord	7	6500	110	19.9	5.7	2290
Honda Civic CVCC	8	4100	91	22.7	39.7	1760
Mazda GLC	9	4150	85	26.9	47.2	1980
Plymouth horizon	A	5400	105	21.6	38.3	2150
Plymouth Sapparo	B	6500	159	13.8	26.6	2790
Pontiac Sunbird	C	4950	151	14.6	28.8	2700
Toyota Cellica	D	6650	134	14.5	26.6	2410
Toyota Corona	E	5700	134	16.6	34.9	2570
Toyota Tercel	F	4350	92	23.3	39.5	1970
VW Rabbit	G	6100	97	19.7	45.7	1870
AMC Concord	H	6850	152	15.6	29.2	2910
AMC Eagle	I	8200	258	10.2	22.8	3360
Chev Citation	J	7200	173	14.4	31.6	2660
Chev Malibu	K	7200	232	13.9	22.8	3180
Dodge Aspen	L	6700	225	10.4	23.8	3330
Dodge Diplomat	M	7550	318	10.4	22.2	3540
Ford Fairmont	N	6350	140	13.1	29.6	2800
Mercury Monarch	O	7150	302	11.5	20.9	3510
Olds Cutlass	P	7550	231	11.9	25.5	3240
Pontiac LeMans	Q	7400	231	11.9	25.5	3220
Pontiac Phoenix	R	7350	151	17.8	37.9	2550
Buick Regal	S	7400	231	9.8	22.4	3280
Chev Impala	T	8150	231	12.1	26.1	3590
Chev Monte Carlo	U	7500	305	10.3	22.2	3380
Dodge St. Regis	V	8100	318	11.2	21.5	3730
Mercury Marquis	W	8550	302	11.2	20.6	3720
Pontiac Catalina	X	8100	231	10.8	22.8	3610
knots		4000	50	9	20	1500
		5000	100	12	25	2000
		6000	150	15	30	2500
		7000	200	18	35	3000
		8000	250	21	40	3500
		9000	300	24	45	4000
			350	27	50	

Table 3

generalized canonical correlations		
I	0.884	0.553
II	0.892	0.603
III	0.901	0.599

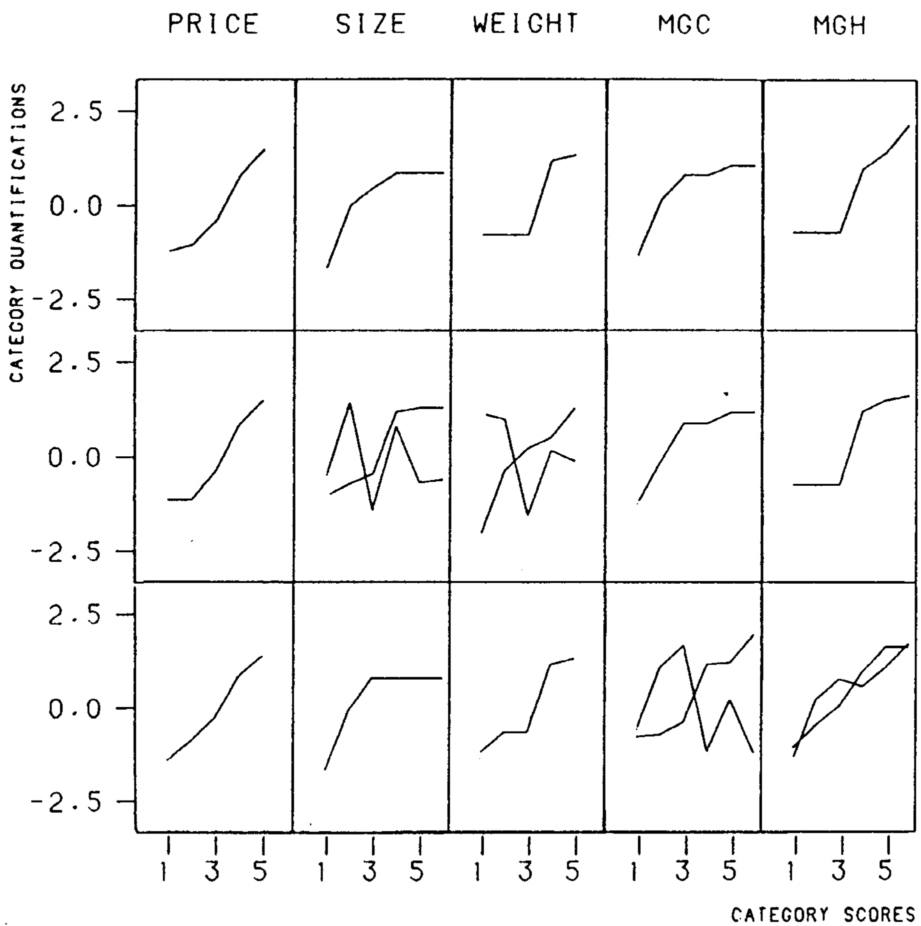
PRICE
SIZE
WEIGHT
MGC
MGH
CATEGORY QUANTIFICATIONS
2.5
0.0
-2.5
2.5
0.0
-2.5
2.5
0.0
-2.5
1 3 5
1 3 5
1 3 5
1 3 5
1 3 5
CATEGORY SCORES

Figure 1

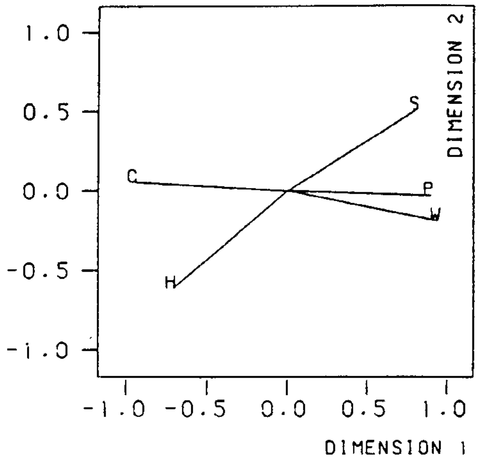

Figure 2

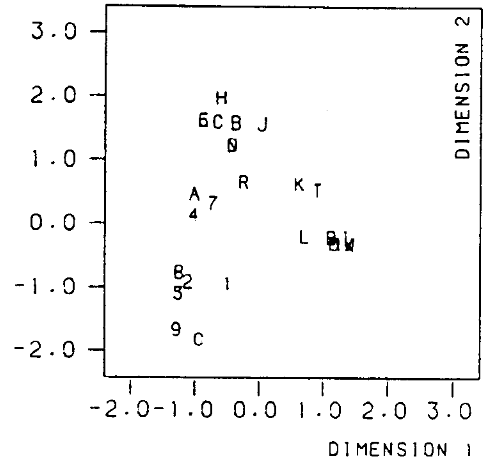

Figure 3

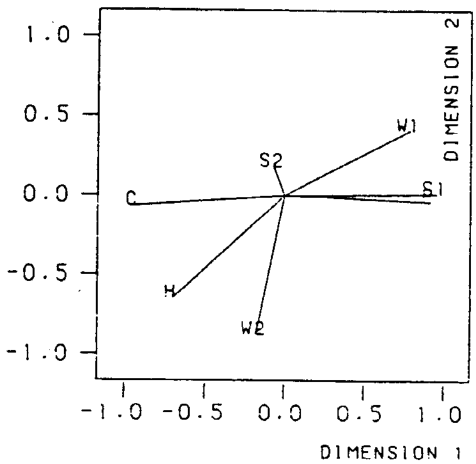

Figure 4

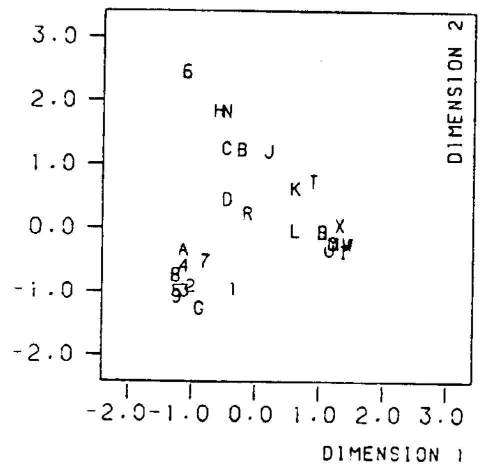

Figure 5

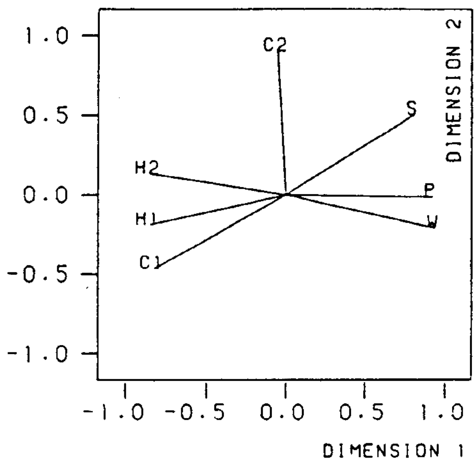

Figure 6

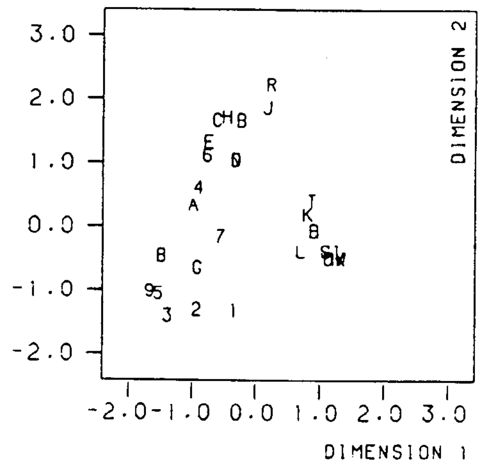

Figure 7

Figure Captions

Figure 1. Transformations of the variables from three analyses. Horizontally the original scores. Vertically the category quantifications

Figure 2. Component loadings, Analysis I

Figure 3. Object scores, Analysis I.

Figure 4. Component loadings, Analysis II

Figure 5. Object scores, Analysis II.

Figure 6. Component loadings, Analysis III

Figure 7. Object scores, Analysis III.