MULTIWAY DATA ANALYSIS
R. Coppi and S. Bolasco (Editors)

MULTI-SET NONLINEAR CANONICAL CORRELATION ANALYSIS VIA THE BURT-MATRIX

Robert Tijssen*, Jan de Leeuw**

*Department of Data Theory, Middelstegracht 4, 2312 TW Leiden, University of Leiden, The Netherlands,
**Departments of Psychology and Mathematics, University of California - Los Angeles, 405 Hilgard Avenue, Los Angeles CA 90024, USA.

Canonical Correlation Analysis (CCA) is a data analysis method in which the correlation between two sets (of linear combinations) of categorical variables is maximized. CCA is discussed in geometrical and matrix-algebraic terms as an introduction to a multi-set CCA with optimal scaling properties of category values. Instead of using a rectangular (objects x variables)-data matrix, this technique operates on the square Burt-matrix, which contains the whole of bivariate relations. Basics of the underlying statistical theory are discussed. A real-life application is presented.

1. INTRODUCTION

Canonical Analysis can be regarded as a generalization of regression theory and is found in many disguises amongst multivariate analysis methods (cf. Gittens 1984). *Canonical Correlation Analysis* (CCA) is a CA method offering the product-moment correlation as a measure of resemblance between sets of categorical variables. This measure is derived from both the between and within-set relations. Linear CCA was introduced in classical multivariate statistical analysis by Hotelling (1936). The linearity of classical CCA refers to the fact that its results are invariant under linear transformations of the category values. In the years to follow, Classical CCA has become an established multivariate statistical method for relating two sets of variables (e.g., Thomson 1947). Multi-set extensions of CCA have been developed and succesfully applied during the last decades (cf. Kettenring 1971). More recently, nonlinear generalisations of multi-set CCA have also been introduced (Van der Burg et al. 1988).

2. LINEAR TWO-SETS CCA

The basic strategy of CCA can be described as finding an optimal linear combination, a weighted sum, of the variables in each set, in such a way that the correlation between each set of variables is maximized. A $(n\ x m)$ data matrix H, with measurements of a number of objects or persons i $(i=1,...,n)$ on discretely-valued categorical variables j $(j=1,...,l,...,m)$ with $r_j (r=1,..., k_j)$ categories, is partitioned in k $(k=1,2)$ sets of variables. The first (nxm_1)-sized set of variables will be denoted by a m_1-dimensional vector $\mathbf{h}_1=(\mathbf{h}_{11},...,\mathbf{h}_{1m1})$, the second (nxm_2)-sized set by a m_2-dimensional vector $\mathbf{h}_2=(\mathbf{h}_{21},...,\mathbf{h}_{2m2})$. The space $\mathbb{L}$ spanned by the $m=m_1+m_2$ variables, with dimensionality $s = (1,...,p)$ where $p \leq \min.(m_1,m_2)$, can be partitioned in such a way that the variables in each set span a corresponding linear subspace $\mathbb{L}_1$ and $\mathbb{L}_2$. The aim of CCA can be reformulated as finding directions in $\mathbb{L}_1$ and $\mathbb{L}_2$, the *canonical axes* or *canonical variates*, which are as similar as possible. We therefore need measure to indicate the 'goodness-of-fit'. The cosine of the angle between the *canonical variates*., the *canonical correlation*, is used for this purpose. Linear combinations of the variables in each set are thus formed by the *canonical weights* in the $(m_1 \text{x} s)$ matrix $\mathbf{A}_1$ and the $(m_2 \text{x} s)$ matrix $\mathbf{A}_2$. Conceptually, the general CCA problem of relating two sets of variables can thus be seen as finding weight matrices **A** and weighted columns

$$\mathbf{h}_1\mathbf{a}^s{}_1=\mathbf{h}_{11}\mathbf{a}^s{}_{11} + \mathbf{h}_{12}\mathbf{a}^s{}_{12} + \ldots + \mathbf{h}_{m1}\mathbf{a}^s{}_{1m1} \quad (1a)$$

$$\mathbf{h}_2\mathbf{a}^s{}_2=\mathbf{h}_{21}\mathbf{a}^s{}_{21} + \mathbf{h}_{22}\mathbf{a}^s{}_{22} + \ldots + \mathbf{h}_{m2}\mathbf{a}^s{}_{2m2} \quad (1b)$$

which represent an orthogonal basis for the subspaces $\mathbb{L}_1$ and $\mathbb{L}_2$ of dimensionality $s<p$ which yield an optimal canonical correlation for each dimension.

Shifting the focus from a geometrical point-of-view to an interpretation CCA in matrix-algebraic terms, boils down to finding maximally related axes in subspaces $\mathbb{L}_1$ and $\mathbb{L}_2$ of the common space $\mathbb{L}$, by means of *eigenvalue-eigenvector decomposition* (Wilkinson 1965) of the partitioned datamatrix $\mathbf{H}=(\mathbf{H}_1|\mathbf{H}_2)$. Since the canonical correlation is invariant under linear scaling of $\mathbf{h}_j$ we will require that $\mathbf{h}_j$ is centered with unit-variance (i.e., $\mathbf{u}'\mathbf{h}_j=0$, $\mathbf{h}_j'\mathbf{h}_j=1$; **u** denotes a unity-vector). Solutions for the canonical weights can be found in such a way that the correlations between the weighted combinations $\mathbf{H}_1\mathbf{A}_1$ and $\mathbf{H}_2\mathbf{A}_2$ are optimal. In terms of the sample-based correlation matrix $\mathbf{R}_{1,2} = \mathbf{H}_1'\mathbf{H}_2$, the problem of finding the optimal set of canonical weights is equivalent to solving the following canonical equations (Anderson 1958):

$$[(\mathbf{R}_{1,1})^{-1}\,\mathbf{R}_{1,2}\,(\mathbf{R}_{2,2})^{-1}\,\mathbf{R}_{2,1} - \lambda^2\,\mathbf{I}]\,\mathbf{A}_1=0, \quad (2a)$$

$$[(\mathbf{R}_{2,2})^{-1}\,\mathbf{R}_{2,1}\,(\mathbf{R}_{1,1})^{-1}\,\mathbf{R}_{1,2} - \lambda^2\,\mathbf{I}]\,\mathbf{A}_2=0, \quad (2b)$$

where **I** denotes the identity matrix and λ^2 represents the first and largest eigenvalue of the characteristic equations

$$|\,(\mathbf{R}_{1,1})^{-1}\,\mathbf{R}_{1,2}\,(\mathbf{R}_{2,2})^{-1}\,\mathbf{R}_{2,1} - \lambda^2\,\mathbf{I}\,|=0, \quad (3a)$$

$$|\,(\mathbf{R}_{2,2})^{-1}\,\mathbf{R}_{2,1}\,(\mathbf{R}_{1,1})^{-1}\,\mathbf{R}_{1,2} - \lambda^2\,\mathbf{I}\,|=0. \quad (3b)$$

The largest eigenvalue, either of the product matrix $(\mathbf{R}_{1,1})^{-1}\mathbf{R}_{1,2}(\mathbf{R}_{2,2})^{-1}\mathbf{R}_{2,1}$ or of the matrix $(\mathbf{R}_{2,2})^{-1}\mathbf{R}_{2,1}(\mathbf{R}_{1,1})^{-1}\mathbf{R}_{1,2}$, is now equal to the highest squared canonical correlation in the first dimension. The accompanying first pair of eigenvectors $\mathbf{a}^1{}_1$ and $\mathbf{a}^1{}_2$ of the n-orthonormal basis, will yield the highest canonical correlation between all possible linear combinations of weighted variables within sets. Since the dimension-wise weight vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are interchangeable, via $\mathbf{a}_1=[(\mathbf{R}_{1,1})^{-1}\mathbf{R}_{1,2}\mathbf{x}_2]/\lambda$ and $\mathbf{a}_2=[(\mathbf{R}_{2,2})^{-1}\mathbf{R}_{2,1}\mathbf{a}_1]/\lambda$, only one characteristic equation needs to be solved. The second pair of canonical variates (i.e., $\mathbf{a}^2{}_1$ and $\mathbf{a}^2{}_2$) are uncorrelated with the first pair, etc. The non-paired canonical variates are also mutually uncorrelated between sets.

3. NONLINEAR CCA

A generalization of CCA to enable the incorporation of variables with less restrictive (nonlinear) measurement levels, generally classified either as *nominal* or *ordinal*, dates back to Young et al. (1976). A nonlinear version of two-sets CCA is also proposed by Van der Burg and De Leeuw (1983), with an accompanying alternating least squares program CANALS, producing CCA-results which are invariant under certain nonlinear transformations of the variables. One must however keep in mind that nonlinearily transformed variables do again defines a linear space $\mathbb{L}$. This type of *optimal scaling* (Young 1981) of category values, i.e. rescaled according to the constraints of the respective measurement levels, leads to category values $\mathbf{y}_j = t^j(r_j)$ referred to as *category quantifications* (Gifi 1981). In case of a nonrestricted nominal variable one can visualize the set of possible values $\mathbf{y}_j$ as a k_j-dimensional space $\mathbb{S}_j$. The set of possible transformations of an ordinal variable j defines a regression problem of the datavector $\mathbf{h}_j$ on a polyhedral convex cone $\mathbb{K}^j$ in the k_j-space. Numerical (linear) variables define a regression on a one-dimensional subset of $\mathbb{S}_j$.

CCA now consists of two computational subproblems which are dependent upon each other and can be solved by means of an algorithm which maximizes the canonical correlation, while simultaneously imposing the proper measurement restrictions on variables. Assume a $(n \times k_j)$ binary *indicator matrix* $\mathbf{G}_j$ as a basis for each $\mathbb{S}_j$, with $\mathbf{G}_j\mathbf{u}=\mathbf{u}$ and $g^j{}_{ik}=1$ if $\mathbf{h}_{ij}=\mathrm{k}$; $g^j{}_{ik}=0$ if $\mathbf{h}_{ij}\neq\mathrm{k}$ (cf. De Leeuw 1984). Thus each category defines a binary variable and each individual is represented in only one category per variable. The indicator matrices of the variables j thus contain $\mathbf{k}_j$ independent columns and span an orthogonal basis for each variable. The expression $\mathbf{q}_j=\mathbf{G}_j\,\mathbf{y}_j$ defines a transformed variable (Gifi 1981). Unit-normalized vectors $\mathbf{q}_j$ define a correlation matrix $\mathbf{R}(\mathbf{Q})=\mathbf{R}(\mathbf{q}_1,\ldots,\mathbf{q}_m)$, with elements $\mathbf{r}_{jl}=\mathbf{q}_j'\mathbf{q}_l$. An induced correlation based on the bivariate matrix $\mathbf{C}_{jl}$ can now be defined as $\mathbf{R}(\mathbf{q}_1,\ldots,\mathbf{q}_m)=\mathbf{y}_j'\mathbf{C}_{jl}\mathbf{y}_l$ with $\mathbf{C}_{jl}=\mathbf{G}_j'\mathbf{G}_l$; a diagonal matrix with univariate marginals $\mathbf{D}_{jl}$ and $\mathbf{y}_j'\mathbf{D}_{jl}\,\mathbf{y}_l=1$; $\mathbf{u}'\mathbf{D}_j\,\mathbf{y}_j=0$.

Nonlinear two-sets CCA can now be rephrased as a technique which computes the optimally scaled variables $\mathbf{q}_j$, and canonical variates $\mathbf{Q}_k\mathbf{A}_k$ maximizing an optimality-criterion function $\phi_\mathrm{p}(\mathbf{R}(\mathbf{Q}_k\mathbf{A}_k))$, or at least computes a stationary value of this function. In general, we are searching for all stationary values of $\phi_p(\mathbf{R}(\mathbf{q}_1,\ldots,\mathbf{q}_m))$ on $\mathbb{S}_j$.

4. MULTI-SET NONLINEAR CCA

The generalization of nonlinear two-sets CCA to more than two sets $k(k=1,\ldots,K)$ is, essentially, quite straightforward. Applying this extention to the generalized canonical solution in terms of matrices $\mathbf{Q}=(\mathbf{Q}_1|\ldots|\mathbf{Q}_K)$, results in a geometrical solution with eigenvectors in $\mathbf{A}_k$

'bundling' around a *mean canonical variate*, equal to the sum vector **QA**. The mean canonical variate thus contains the overall information of the canonical variates of each canonical solution and forms a orthogonal basis for the common canonical space $\mathbb{L}$, providing similar interpretations with respect to canonical correlations between the various sets. In *K*-sets CCA the criteria ϕ_p are also based on the maximization or minimization of properties of the eigenvalues of the correlation matrix $\mathbf{R}(\mathbf{Q}_k\mathbf{A}_k)$. Using more than two sets provides us with opportunities to extend the possible ways of computing CCA-solutions. The canonical correlations can now be generalized to measures of relatedness, which are derived under various conditions and optimality criteria, each combination producing somewhat different results (cf. Kettenring 1971; Gifi, 1981; Van de Geer 1984; Meulman 1986). Van der Burg et al. (1988) introduce an alternating least squares multi-set program OVERALS, as a particular generalisation of nonlinear two-sets CCA.

5. MULTI-SET NONLINEAR CCA BASED ON BIVARIATE CROSSTABLES

In some cases one might not have or want to analyze the usual (*nxm*) matrix **H**, but instead an aggregated (*mkxmk*) Burt matrix **C**, with crosstables $\mathbf{C}_{jl}$ between all variables (Burt 1950). In the sequel it will be shown that it is still possible to devise a technique which can perform a *K*-sets nonlinear CCA based on **C** via $\mathbf{R}(\mathbf{Q}_k\mathbf{A}_k)$ (De Leeuw 1983; Tijssen 1985).

It has been shown that a linear CCA-solution is computed on **R**. In case of a nonlinear CCA based on **C** certain requirements must be fulfilled to obtain an optimal CCA-solution, because the induced correlations are only optimal association measures in case of linear and homoscedastical regression between variables. CCA-solutions based on **R**, given the bivariate relations in **C**, which also handle nonlinear variables, thus need an optimal approximation of the linearity property of correlations. Maximizing the largest eigenvalues of an **R**, induced from **C** via category quantifications **y**, is such a method. The resulting **R** will now be as one-dimensional (linear) as possible. Suppose we want to optimize a function of the correlation matrix **R**, written as $\phi(\mathbf{R})$. An example would be the largest eigenvalue of **R**, which leads to Multiple Correspondence Analysis, or the sum of the *p* largest eigenvalues, which leads to Principal Component Analysis. Compare De Leeuw (1986) for a more complete discussion of the statistical concepts behind this data-analytic approach.

In the formalization of this approach, we use the fact that $\mathbf{r}_{jl} = \mathbf{y}'_j \mathbf{C}_{jl} \mathbf{y}_l$, if $\mathbf{y}'_j \mathbf{D}_{jl} \mathbf{y}_l = 1$. The stationary equations for this optimization problem are

$$\Sigma_l \; \delta\phi/\delta\mathbf{r}_{jl} \; \mathbf{C}_{jl} \mathbf{y}_l = \lambda_j \mathbf{D}_j \mathbf{y}_j , \qquad (4)$$

where δ denotes the derivative.

If category values $\mathbf{y}_j$ can be found such that the equality $\mathbf{C}_{jl}\,\mathbf{y}_l = \mathbf{r}_{jl}\,\mathbf{D}_j\,\mathbf{y}_j$ holds for all variables j, then the $\mathbf{y}_j$'s linearize the bivariate regressions. Hence the stationary equations are satisfied no matter how the function $\phi(\mathbf{R})$ is defined.

Studying a K-sets CCA-problem requires a partitioning of **R** according to the allocation of the variables over the sets. Such a division can be accomplished by creating a conveniently arranged (mxm) binary matrix **E**, with elements $\mathbf{e}_{jl} = 1$ if variables belong to the same set and $\mathbf{e}_{jl} = 0$, otherwise. The Hadamard product (Styan, 1973) of **E** and **R** results in a (m*m) matrix - denoted as **E^R**, with $\mathbf{r}_{jl} = \mathbf{e}_{jl}\hat{\ }\mathbf{r}_{jl}$ if the corresponding variables belong to the same set. The nonlinear K-sets CCA-problem is to maximize the sum of the first p eigenvalues of the pair (**R**, **E^R**). This is of the form $\phi(\mathbf{R})$. In order to find the eigenvalues one has to solve the ***generalized eigenvalue problem***:

$$\mathbf{RA} = \lambda^2(\mathbf{E\hat{\ }R})\mathbf{A}, \qquad (5)$$

with the (pxp) diagonal matrix with generalized λ^2-values. The result of this particular eigenvalue problem will provide us with a common basis $\mathbb{L}$ for the sets based on eigenvectors **A** which are now orthonormal with respect to the within-sets correlations, i.e. **A'(E^R)A=I.**

In this paper we shall study a somewhat more general class of criteria ϕ which depend on **R** through the generalized eigenvalues λ^2. One could also use the product of the generalized values, for instance, or the sum of squares of their deviations from unity.

By applying the chain rule, it is well known that

$$\delta\,\lambda^2_s / \delta\,\mathbf{r}_{jl} = [1-(\lambda^2_s\,\mathbf{e}_{jl})]\,\mathbf{a}_{js}\,\mathbf{a}_{ls}. \qquad (6)$$

Combining Eq. (4) and Eq. (6) provides us with two ($\Sigma_j\,\mathbf{k}_j$ x $\Sigma_j\,\mathbf{k}_j$) matrices for each pair of variables:

$$\mathbf{T}_{jl} = \Sigma_s\,\delta\phi/\delta\lambda^2_s\,(\lambda^2_s\,\mathbf{C}_{jl}\,\mathbf{a}_{js}\,\mathbf{a}_{ls}\,\mathbf{e}_{jl}), \qquad (7a)$$

$$\mathbf{U}_{jl} = \Sigma_s\,\delta\phi/\delta\lambda^2_s\,(\mathbf{C}_{jl}\,\mathbf{a}_{js}\,\mathbf{a}_{ls}). \qquad (7b)$$

Hence **T** is based on the within-sets bivariate relations, whereas **U** is based on both the within and between sets bivariate relations.

We can now define stationary equations as

$$\mathbf{Ty} - \mathbf{Uy} = 0, \qquad (8)$$

which suggests an iteration scheme

$$\mathbf{y}^*=\mathbf{y}^{(n+1)}=\mathbf{U}^+\ \mathbf{T}\ \mathbf{y}^{(n)}, \tag{9}$$

with the superscript + denoting the Moore-Penrose inverse, in which alternations between successive steps (n) and ($n+1$) will lead to category quantifications $\mathbf{y}^*$ with (approximate) linearizing properties. The quantifications are normalized in each iteration step, thus inducing an updated optimal correlation matrix $\mathbf{R}^*$ in each step via

$$\mathbf{R}^*=\{\mathbf{r}^*_{jl}\}=\mathbf{y}^*_j{}'\ \mathbf{C}_{jl}\ \mathbf{y}^*_l\,. \tag{10}$$

The resulting quantifications $\mathbf{y}^*$ are thus an optimal function of the bivariate within-sets structure in combination with the total bivariate structure. Categories with a relatively high bivariate frequency will tend to have, on the whole, more similar quantifications.

The iteration process therefore in fact consists of two subprocesses. The inner-iterations [Eq. (9)] produce optimal quantifications $\mathbf{y}^*$ with linearizing properties, which are used in the outer-iteration steps [Eqs. (5), (10)] to compute the corresponding $\mathbf{R}^*$. The generalized eigenvalue-eigenvector decomposition of ($\mathbf{R}^*$,$\mathbf{E}$^$\mathbf{R}^*$) subsequently computes the corresponding eigenvectors $\mathbf{A}$ for the K-sets CCA solution. The elements of the Burt matrix are weighted with the corresponding elements in $\mathbf{A}$, creating updates of $\mathbf{U}$ and $\mathbf{T}$, which provide a new update of $\mathbf{y}^*$. These quantifications are then used to induce a new $\mathbf{R}^*$, etcetera ... until convergence is reached. Assuming this iteration process leads to a stable value of the function $\phi_p(\mathbf{R})$, with a corresponding optimally linearized correlation matrix $\mathbf{R}^*$, one obtains optimal category quantifications in the sense that they induce an optimally linear matrix $\mathbf{R}^*$ from $\mathbf{C}$, while incorporating the K-sets structure in $\mathbf{R}^*$.

The (canonical) correlations between the various obtained quantifications are derived by introducing two m-sized columns of the identity matrix for each variable; the vectors $\mathbf{v}$ and $\mathbf{w}$ with a one in the position of the respective variable(s) in question. The covariances, variances and correlations of the quantifications are now expressed as, respectively:

$$\mathbf{Cov}_{\mathbf{vw}}=\mathbf{A}'_s\,(\mathbf{R}^{*}\wedge(\mathbf{vw}'))\ \mathbf{A}_s\,, \tag{11a}$$

$$\mathbf{Var}_{\mathbf{vv}}=\mathbf{A}'_s\,(\mathbf{R}^{*}\wedge(\mathbf{vv}'))\mathbf{X}_s.=\mathbf{Var}_{\mathbf{ww}}=\mathbf{X}'_s\,(\mathbf{R}^{*}\wedge(\mathbf{ww}'))\ \mathbf{A}_s\,, \tag{11b}$$

$$\mathbf{Cor}_{\mathbf{vw}}=(\text{diag. }\mathbf{Var}_{\mathbf{v}})^{1/2}\ \mathbf{Cov}_{\mathbf{vw}}\ (\text{diag. }\mathbf{Var}_{\mathbf{w}})^{1/2}. \tag{11c}$$

With the appropriate filling of vectors **v** and **w** one can now compute for example: the canonical correlations - with ones in **v** and **w** on the positions indicating the variables of the sets, implying Z_{vw}=A'R*A=I. In a similar fashion one can compute the correlations between a quantified variable and the respective canonical variate (also referred to as *canonical loadings*), reflecting the extent to which a variable contributes in the variate.

6. AN APPLICATION

6.1. INTRODUCTION

An example is given of a data-analytic situation in which the sheer magnitude of the matrix (*n*=2666; *m*=77) matrix may warrant the use of a *K*-sets CCA method based on a smaller Burt-matrix. The data stem from a survey carried out by the Educational Department and the Educational Research Center of the University of Leiden to obtain a comparative inventarisation of different aspects of, in order of increasing educational level, the following three groups of elementary vocational educations: VBO/MLK, IBO and LBO (Van Putten 1987). The study focussed on the assessment of characteristics of the IBO pupils. The Dutch acronym 'IBO' refers to a number of individualized elementary vocational educations, with the aim to provide a suitable education for slow learners from the primary school or special educations, in the pupil age of 12-16 years.

These pupils were subjected to a number of tests, in order to assess (1) cognitive capacities, (2) reading abilities, and (3) arithmetic abilities. They also received (4) a questionnaire with items concerning their attitude on some relevant socio-emotional aspects of the school/classroom-situation, (5) two teachers were asked -independently- to rate the classroom behaviour of each pupil , and (6) background information was obtained on the pupils by means of a parent/caretaker-questionnaire. Each test/questionnaire contained categorical or categorized variables, with a grand total of well over 400 categories.

6.2. DATA PRE-PROCESSING

Data-reduction was achieved by applying the non-metric Principal Component Analysis program PRINCALS (Gifi, 1985) on the above mentioned groups of variables. The results provided us with 'condensed' data in the form of the following independent and relatively clear-cut defined components, on which each pupil obtained a score. These metricized pupil-quantifications were divided into a number of categories (on the average, about 5 categories). In the case of the variables from the groups (1)-(5) low-valued categories indicate the positive/desirable side of the attribute, for instance a high degree of accuracy or no fear of failure, whereas high category scores indicate the opposite qualification. The original survey variables were reduced to the following (composite) variables :

1-Logical reasoning (LR)	2-Workpace (WP)
3-Accuracy (AC)	4-Verbal abilities (VA)
5-Technical reading (TR)	6-Mental arithmetic (MA)
7-Applied arithmetic (AA)	8-Involvement and interest (II)

9-Problematic behaviour (PB)
10-Independence (IN)
11-Motivation (MO)
12-Relations with fellow-pupils (RP)
13-Fear of failure (FF)

The following background variables were selected for the CCA analysis:

14-Previous education (PRE)
15-Promotion previous year (PPY)
16-Sex (SEX)
17-Age (AGE)
18-Occupational family type (OFT)
19-Number of brothers/sisters (NBS)
20-Birth rank among brothers/sisters (BBS)

The above variables were placed in the following sets: (I) a cognitive set (variables 1-7), (II) a socio-emotional set (variables 8-13), and (III) the background information (variables 14-20), respectively.

The pupil scores in the cognitive and socio-emotional set were analyzed with a numerical measurement level. The optimal (linear) properties of these scores provides a plausible basis for such a restriction in the category quantification. The remaining (background) variables are of a more qualitative nature were analyzed with the nominal measurement level, with exception of the pupil's age.

6.3. SOME CCA ANALYSIS RESULTS

The CCA analysis was done for 2 dimensions. The 'fit' of the analysis solution was equal to 1.034, with the eigenvalues .530 and .504 for the first and second dimension, respectively. Both dimensions can thus be regarded as almost equally important and, on the whole, 'explaining' about one half of the existing variance (the maximum fit is equal to 2). The highest canonical correlations were found between the cognitive set and the set of background variables (with values equal to .357 and .405 for the first and second dimension)

A general description of the relations between the separate variables, given the canonical structure, can be given on the basis of the configuration of the correlations between variables and the mean canonical variate (see Fig. 1). Projecting the points on the axes reveals a structure, in which the first dimension is determined by applied arithmetic (AA), previous education (PRE), SEX and, to a lesser extent, by variables such as problematic behaviour (PB). The second dimension is also largely determined by SEX and PRE, but now in combination with technical reading (TR), mental arithmetic (MA), verbal abilities (VA) and motivation (MO).

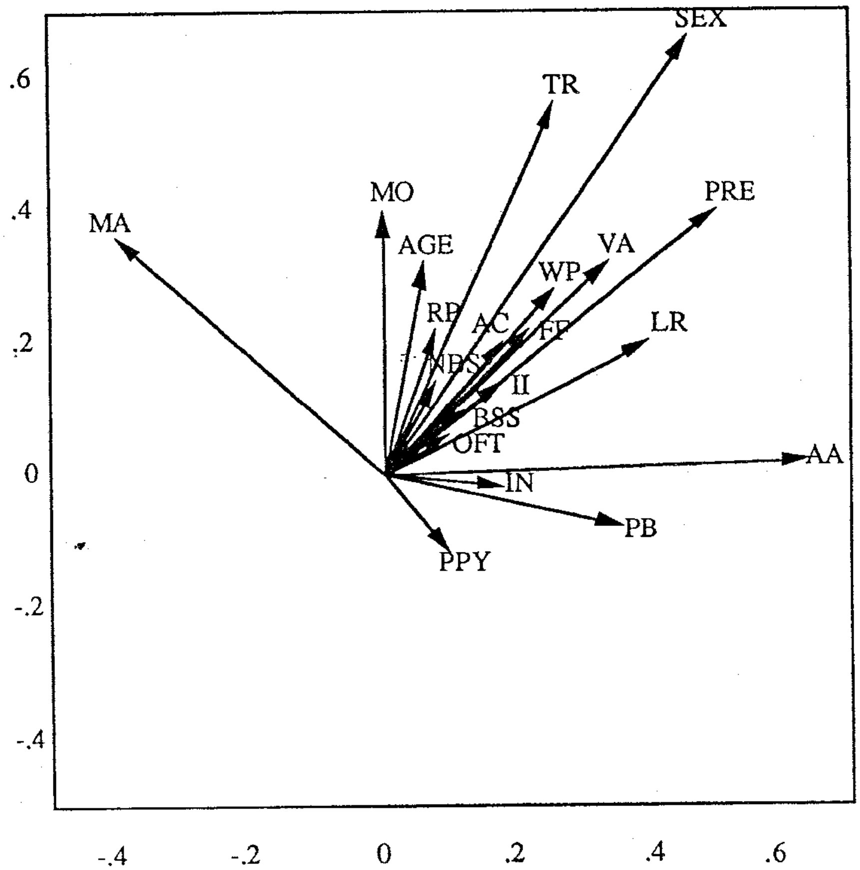

FIGURE 1
Correlations of variables with mean canonical variate.
First dimension - horizontal axis; second dimension-vertical axis

Further insight in the analysis results can be obtained by projecting the linear category quantifications in the mean canonical space (see Fig. 2). These quantifications present an image of the structure of relations on a category-level; for each variable the categories are located on the a straight line through the origin. Both SEX-categories, the 'extreme' categories of PRE and AGE and the lowest- and highest valued category quantifications are shown for the cognitive and socio-emotional variables. With respect to the latter, for example, the variable problematic behaviour (PB) has a label PB+ for non-problematic behaviour and PB- for problematic behaviour. The linear quantification of two (extreme) categories of SEX, PRE and AGE are shown connected with a line.

As for the interpretation of these results, consider the perpendicular lines of AGE and PRE which reveal a boy-girl distinction on the cognitive- and socio-emotional variables which is independent of the previous educational level. However, differences between boys and girls are found in both dimensions, for example: boys are -on the average- older then the girls; tend to have relatively better marks on applied arithmetic-tests (AA+); show less fear of failure (FF+); more independent behavior (IN+); less accuracy (AC-) and are found to have less classroom involvement and interest (II-).

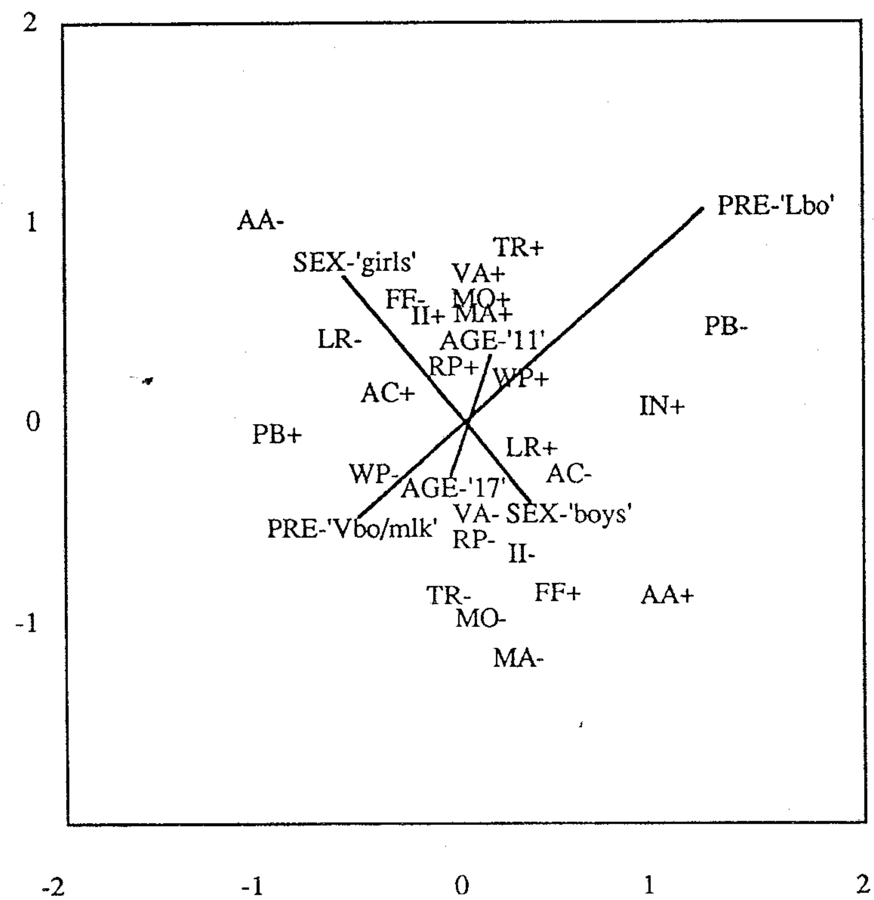

FIGURE 2
Selected rank-one category quantifications.
First dimension-horizontal axis; second dimension - vertical axis

7. CONCLUSIONS

The presented technique based on a joint bivariate analysis seems a potentially fruitful approach to nonlinear K-sets CCA in case of large matrices; when $\sum_j k_j \ll n$ such an algorithm, should be computationally more efficient, in terms of core-memory and CPU processing time. Some pilot applications have indicated that the algorithm is indeed a relatively fast way of computing an - at least locally - stable CCA solution. Although a theoretical proof of convergence is not available (so far), empirical studies of the iteration process have shown that convergence was obtained in all cases. The technique is an interesting alternative to a nonlinear K-sets CCA approach as in the programme OVERALS, which is based on the (objects x variables) data matrix.

REFERENCES

Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* New York: Wiley.
Burt, C. (1950). The factorial analysis of qualitative data. *British Journal of Statistical Psychology*, **3**, 166-185.
Gifi, A. (1981). *Nonlinear multivariate analysis.* Department of Data Theory. University of Leiden. (A revised version will be published by DSWO-Press in 1988).
Gifi, A. (1985). *PRINCALS*, UG-85-03. Department of Data Theory. University of Leiden.
Gittens, R. (1984). *Canonical analysis-A review with applications in ecology.* Berlin: Springer Verlag.
Hotelling, H. (1936). Relations between two sets of variables. *Biometrika*, **28**, 321-377.
Kettenring, J.R. (1971). Canonical analysis of several sets of variables. *Biometrika*, **58**, 433-460.
De Leeuw, J. (1983). *Nonlinear joint bivariate analysis.* Paper presented at the European Meeting of the Psychometric Society. Jouy-en-Josas. France.
De Leeuw, J. (1984). *Canonical Analysis of Categorical Data.* (Doctoral dissertation, University of Leiden, 1973). Leiden: DSWO-Press.
De Leeuw, J. (1986). *Multivariate Analysis with Optimal Scaling.* Paper presented at the International Conference on Advances in Multivariate Statistical Analysis. Calcutta. India.
Meulman, J.J. (1986). *A distance approach to nonlinear multivariate analysis.* Leiden: DSWO-Press.
Styan, G.P. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra and its applications*, **6**, 217-240.
Thomson, G.H. (1947). The maximum correlation of two weighted batteries. *British Journal of Psychology - Statistical Section*, **1**, 27-34.
Tijssen, R.J.W. (1985). *A new approach to nonlinear canonical correlation analysis.* Leiden Psychological Reports - Psychometrics and Research methodology, RR 85-01. University of Leiden.
Van der Burg, E. and De Leeuw, J. (1983). Non-linear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, **36**, 54-80.
Van der Burg, E., De Leeuw, J. and Verdegaal, R. (1988). Nonlinear canonical correlation with m sets of variables. *Psychometrika*, **2**, 171-197.
Van de Geer, J.P. (1984). Linear relations between k sets of variables. *Psychometrika*, **49**, 79-94.
Van Putten, C.M. (1987). Leerlingen in het individueel beroepsonderwijs nader beschouwd. (Doctoral dissertation in Dutch). University of Leiden.
Wilkinson, J.H. (1965). *The algebraic Eigenvalue Problem.* Oxford: Clarendon Press.
Young, F.W., De Leeuw, J. and Takane, Y. (1976). Regression with qualitative and quantitative variables: an alternating least squares method with optimal scaling features. *Psychometrika*, **43**, 279-281.
Young, F.W. (1981). Quantitative Analysis of Qualitative Data. *Psychometrika*, **46**, 357-388.