Statistique et analyse des données

Renate Meyer

Canonical correlation analysis as a starting point for extensions of correspondence analysis

Statistique et analyse des données, tome 16, nº 1 (1991), p. 55-77

<http://www.numdam.org/item?id=SAD_1991__16_1_55_0>

Statistique et Analyse des Données
1991 - Vol. 16 n° 1 p. 55 - 77

CANONICAL CORRELATION ANALYSIS AS A STARTING POINT FOR EXTENSIONS OF CORRESPONDENCE ANALYSIS

Renate MEYER
Department of Medical Statistics and Documentation
of the RWTH Aachen

Pauwelsstr. 30, 5100 Aachen, Federal Republic of Germany

Résumé.

L'analyse des correspondances d'un tableau de contingence à deux dimensions peut être considérée comme un cas particulier de l'analyse canonique de deux ensembles de variables indicatrices. Ainsi, la manière la plus naturelle de généraliser l'analyse des correspondances à une table à trois ou plusieurs dimensions consiste à appliquer l'analyse canonique convenablement généralisée au cas de plusieurs variables indicatrices.

Dans cet article nous présentons quatre généralisations différentes de l'analyse canonique à $Q \geq 3$ *ensembles de variables aléatoires et nous appliquons ces méthodes aux variables indicatrices. Ainsi nous obtenons quatre extensions différentes de l'analyse des correspondances. La détermination des variables canoniques nous mène à des problèmes aux valeurs propres généralisés. Nous présentons un nouvel algorithme itératif pour résoudre ces problèmes et nous prouvons sa convergence globale. En fait, alors que la première extension, équivalente à "l'analyse des correspondances multiple", est souvent appliquée grâce à ses calculs directs, nous montrons que les trois autres extensions fournissent des interprétations beaucoup plus significatives des représentations graphiques. Les problèmes calculatoires sont surmontés par notre algorithme.*

Mot clés. tableau de contingence, analyse des correspondances, analyse canonique, problème aux valeurs propres.

Classification AMS : 62H17, 62H20, 65U05

Abstract.

Correspondence analysis (CA) *of two-way contingency tables can be regarded as a special case of canonical correlation analysis with two sets of indicator variables. Thus, it is obvious to extend* CA *to three-way and higher-dimensional contingency tables by appropriately generalized canonical correlation analysis techniques applied to the set of indicator variables at hand.*

In this paper four different generalizations of canonical correlation analysis to $Q \geq 3$ *sets of random variables are proposed, their applications to indicator variables are studied, and the resulting extensions of* CA *to* Q*-dimensional contingency tables are presented. The determination of canonical variates leads to generalized eigenvalue problems. We will present a new iterative procedure for their solution and prove the global convergence of this algorithm. Actually, whereas the first extension, which is equivalent to "multiple correspondence analysis", is widely used in practice because of its straightforward computations, we think that the other three extensions yield more meaningful interpretations of the resulting graphical representations. The arising computational difficulties have been surmounted by our algorithm.*

Keywords : contingency tables, correspondence analysis, canonical correlation analysis, eigenvalue problem.

1. INTRODUCTION.

Many attempts have been made to trace back accurately the historical development of the multivariate statistical analysis method called *correspondence analysis*. For a sketch of the historical development as well as a correspondence analysis of co-citations see van Rijckevorsel (1987). As described in detail by Greenacre (1984), there are a number of possible equivalent approaches to correspondence analysis all of them leading to the same mathematical method, which is essentially a singular value decomposition of an appropriately normalized data matrix :

- **reciprocal averaging**, see Hirschfeld (1935), Richardson and Kuder (1933), Horst (1935), Benzécri (1964) and Hill (1974).

- **dual or optimal scaling**, see Guttman (1941), Hayashi (1950), Bock (1960) and Nishisato (1980).

- **simultaneous linear regressions,** see Lingoes (1964).

- **discriminant analysis,** see Fisher (1940), Lebart et al. (1984).

- **weighted least squares approximation** of a matrix by a matrix of reduced rank, see Eckart & Young (1936), Greenacre (1984).

- **generalized principal components analysis,** see Benzécri (1977), Lebart et al. (1984).

- **canonical correlation analysis,** see Hotelling (1933), McKeon (1966).

The geometric form of the above methods originated in France in a linguistic context - already indicated by the French term "*analyse des correspondances* " - and was developped by J.P. Benzécri in the early 1960s. The interpretation of the graphical representation of a contingency table is primarily based on the *transition formulae* and "*le principe barycentrique* ".

Given a higher-way contingency table, we face an abundance of possible analysis techniques. A first attempt may consist of reducing this problem to the two-dimensional case. For example, given a three-way contingency table of order $I \times J \times K$, one can analyze K separatate $I \times J$ tables for each category of the third variable. Gifi (1990) concatenated these "sliced" contingency tables to three two-way tables having order $I \times (J \times K)$, $J \times (I \times K)$ and $K \times (I \times J)$ respectively, called "*tableaux multiples* " in the French literature (Benzécri (1980)). Clearly, these techniques don't take into account the possibly complex three-dimensional interaction structure.

Another simple generalization of correspondence analysis is the so called *multiple correspondence analysis,* see Benzécri (1977), Greenacre (1984), or *homogeneity analysis,* see Gifi (1990), which has been proposed to deal with the situation of $Q \geq 2$ categorical variables cross-classified in a multidimensional contingency table by directly applying the correspondence analysis algorithm to the corresponding $N \times (J_1+J_2 ...+J_Q)$ indicator matrix, where N is the number of observations and J_q the number of categories of variable q $(q = 1,2,...,Q)$. Although there have been other proposals, for example by Masson (1974), Leclerc (1980), Van de Geer (1986), this method of multiple correspondence analysis has been used in all practical applications without any exception worth mentioning, certainly because of the straightforward und uncomplicated calculations. Nevertheless, as mentioned for example by Greenacre (1988), this technique is not a natural generalization of the geometric approach and, besides several other shortcomings, does not yield meaningful interpretations of the resulting graphical representation of generalized canonical correlation vectors. Thinking of multiple correspondence analysis as the weighted least squares approximation of the Burt matrix, Greenacre (1988) particularly criticizes the fitting of subtables on the diagonal of the Burt matrix, the resulting inflation of total variation and underassessment of the variation explained by a principal axis thereby. Therefore he suggests an alternative generalization of correspondence analysis which fits only the off-diagonal subtables of the Burt matrix by an alterning least squares algorithm. However, he admits the lack of a specific geometric interpretation of the graphical display.

As mentioned previously, correspondence analysis can be viewed as a special case of canonical correlation analysis, custom-made to two sets of indicator variables that define a two-dimensional contingency table. Thus in order to generalize correspondence analysis to Q-dimensional contingency tables, $Q > 2$, the obvious thing to do is to first extend canonical correlation analysis to Q sets of variables and then to apply this extension to Q sets of indicator variables. By this way, we will arrive at several and, in particular, some new generalizations of correspondence analysis, since there are several possible and quite different ways of extending canonical correlation analysis.

The ensuing questions will be : *Do these different extensions of canonical correlation analysis yield different extensions of correspondence analysis or do they coincide when applied to indicator variables ?* In case they do not coincide :

- *Which of these extensions is equivalent to the usual multiple correspondence analysis ?*

- *How can the joint display of the "generalized" canonical correlation vectors be interpreted ?*
  or equivalently, since some type of *transition formulae* will define the basis for any joint representation

- *What different kinds of transition formulae do we obtain ?*

In mathematical literature we find several approaches of extending canonical correlation analysis to more than two sets of variables, one as early as 1961 by Horst and five possible extensions proposed by Kettenring (1971), his SUMCOR method being equivalent to Horst's technique, his MAXVAR method being equivalent to a generalization of Carroll (1968). The solution of the MAXVAR problem can be calculated directly, whereas the SUMCOR method as well as all other methods require iterative procedures whose convergence properties have not yet rigorously been proved.

In this paper we will follow the approaches of Kettenring (or Carroll and Horst, respectively) in generalizing canonical correlation analysis to $Q > 2$ sets of variables and show the convergence of a new iterative procedure for solving the generalized eigenvalue problems that characterize the generalized canonical correlation vectors. (This will include a convergence proof of Horst's iterative procedure.) Last not least we are going to study the application of these generalizations of canonical correlation analysis to indicator variables and try to answer the questions raised above with a strong emphasis upon the geometrical aspect of these techniques.

2. CANONICAL CORRELATION ANALYSIS OF TWO SETS OF VARIABLES.

Given two sets of random variables $\{X_{11},X_{12},\dots,X_{1J_1}\}$ and $\{X_{21},X_{22},\dots,X_{2J_2}\}$, the objective of canonical correlation analysis is to find a pair of linear combinations $U = a_1'X_1$ and $V = a_2'X_2$ with

$$X_1 = (X_{11},X_{12},\dots,X_{1J_1})', \qquad a_1 = (a_{11},a_{21},\dots,a_{1J_1})',$$
$$X_2 = (X_{21},X_{22},\dots,X_{2J_2})', \qquad a_2 = (a_{21},a_{22},\dots,a_{2J_2})',$$

such that U and V have largest possible correlation

$$\rho = \frac{\mathrm{Cov}(a_1'X_1,a_2'X_2)}{\{\mathrm{Cov}(a_1'X_1)\mathrm{Cov}(a_2'X_2)\}^{1/2}}. \tag{2.1}$$

Since the correlation coefficient is scale - and translation - invariant, we consider only solutions with zero expectation and unit variance. Having determined the *first pair of canonical correlation variales* U *and* V further linear combinations of the two given variable sets may be sought which again have maximal correlation but are constrained to be uncorrelated to the preceding solutions. For ease of notation, we will consider only the first canonical correlation variables here.

In practical applications where two centered data matrices $Z_1 = (x_{ni}^{(1)} - \bar{x}_{.i}^{(1)}) \in \mathbb{R}^{N \times J_1}$ and $Z_2 = (x_{ni}^{(2)} - \bar{x}_{.i}^{(2)}) \in \mathbb{R}^{N \times J_2}$ corresponding to N realizations of X_1 and X_2 are given, the optimization problem is to maximize the empirical correlation coefficient, i.e. to determine

$$(2.2) \qquad \frac{1}{N} \hat{a}_1' Z_1' Z_2 \hat{a}_2 = \max_{a_q' Z_q' Z_q a_q = 1, q=1,2} \frac{1}{N} a_1' Z_1' Z_2 a_2 .$$

The solutions, called *canonical correlation vectors*, are given by

$$(2.3) \qquad \hat{a}_1 = d_1 (Z_1' Z_1)^+ Z_1' b + [I_{J_1} - (Z_1' Z_1)^+ Z_1' Z_1] w_1 ,$$

$$(2.4) \qquad \hat{a}_2 = d_2 (Z_2' Z_2)^+ Z_2' b + [I_{J_2} - (Z_2' Z_2)^+ Z_2' Z_2] w_2$$

with b the eigenvector corresponding to the largest eigenvalue λ_1^2 of the matrix $\frac{1}{2}(Z_1((Z_1'Z_1)^+Z_1' + Z_2(Z_2'Z_2)^+Z_2')$, standardized according to $\frac{1}{N}$ b'b = 1, with arbitrary vectors $w_1 \in \mathbb{R}^{J_1}$, $w_2 \in \mathbb{R}^{J_2}$, and with constants $d_q := \{\frac{1}{N} b' Z_q (Z_q' Z_q)^+ Z_q' b\}^{-1/2}$ for q = 1,2. (I_k denotes the k-dimensional identity matrix, Z^+ the Moore-Penrose inverse of the matrix Z.) This results from our theorem 3.1. for Q = 2.

Let us briefly review how canonical correlation analysis is used for exploring the interrelationship between two categorical variables A and B with J_1 resp. J_2 categories. In this case, our data are given by the partitioned $N \times (J_1+J_2)$ indicator matrix $W = [X_1, X_2]$ with components $x_{ni}^{(1)} = 1$ (respectively $x_{nj}^{(2)} = 1$) if category i of variable A (respectively category j of variable B) is observed in the nth observation, and 0 otherwise. Applying canonical correlation analysis to this binary matrix yields, on the one hand, the values of the canonical correlation coefficients ρ (characterizing the strength of interrelationship between A and B), and the components $\hat{a}_{1i}$, $\hat{a}_{2j}$ of the vectors $\hat{a}_1$ resp. $\hat{a}_2$ which can be interpreted as an optimal scaling of the (qualitative) J_1 resp. J_2 alternatives of A resp. B. On the other hand, as shown for example by Greenacre (1984), this procedure is equivalent to a correspondence analysis of the $J_1 \times J_2$-contingency table $P = (p_{ij})$ that contains the observed relative frequencies p_{ij} of the J_1 resp. J_2 categories of A and B leading to a meaningful geometrical representation. In fact, the basic objective of correspondence analysis, conceived as a generalized principal components analysis, is to find a lowdimensional subspace that minimizes the weighted sum of squared distances from the *row profiles* $p_i := (\frac{p_{i1}}{p_{i.}},\ldots,\frac{p_{iJ_2}}{p_{i.}})'$, $i = 1,2,\ldots,J_1$, to their projection points $\tilde{p}_i$ on this subspace, i.e.

$$\sum_{i=1}^{J_1} p_{i.} \|p_i - \tilde{p}_i\|^2_{D_c^{-1}} \rightarrow \min$$

where $D_c := \mathrm{diag}(p_{.1},p_{.2},\ldots,p_{.J_2})$ and $\|x\|^2_{D_c^{-1}} := x' D_c^{-1} x$ is a generalized euclidean distance. Considering, e.g., a one-dimensional approximation for ease of illustration, we find that the optimum representation of the row profiles (neglecting the trivial eigenvalue $\lambda^2 = 1$) is given by the J_1 components $\hat{a}_{11},\hat{a}_{12},\ldots,\hat{a}_{2J_1} \in \mathbb{R}^1$ of $\hat{a}_1$, and similarly by the J_2 components $\hat{a}_{21},\hat{a}_{22},\ldots,\hat{a}_{2J_2} \in \mathbb{R}^1$ of $\hat{a}_2$ for the column profiles. So, the J_1 components of the first canonical correlation vector $\hat{a}_1$ are the coordinates of the J_1 projection points of the row profiles with respect to an orthogonal basis of the optimal one-dimensional subspace. By analogy the canonical correlation vectors $\hat{a}_2$ are solutions of the dual problem concerning the column profiles. The representation of both row and column profiles in one joint display may be justified and distances between row column profiles may be interpreted by the *transition formulae* :

(2.5) $$\hat{a}_{1i} = \frac{1}{\lambda} \sum_{j=1}^{J_2} \frac{p_{ij}}{p_{i.}} \hat{a}_{2j} ,$$

(2.6) $$\hat{a}_{2j} = \frac{1}{\lambda} \sum_{i=1}^{J_1} \frac{p_{ij}}{p_{.j}} \hat{a}_{1i} ,$$

i.e. the coordinates of the jth column profile are centroids (weighted by the jth column profile) of the coordinates of the row profiles and stretched along the first principal axis by a factor $\frac{1}{\lambda}$ and vice versa.

3. CANONICAL CORRELATION ANALYSIS OF Q SETS OF VARIABLES.

Following the canonical correlation analysis approach, CA can be extended to the case of Q-dimensional contingency tables (with $Q \geq 2$) by applying a suitably generalized canonical correlation analysis to the Q sets of indicator variables at hand. We will investigate four different extensions of canonical analysis to the case of $Q \geq 2$ sets of variables which all reduce to the classical canonical correlation analysis when $Q = 2$. Consider $Q \geq 2$ random vectors $X_q \in \mathbb{R}^{J_q}$, $q = 1,2,...,Q$, whose interrelationship and dependence structure has to be analysed. Given N realizations of these Q random vectors as entries in Q (centered) data matrices Z_q, $q = 1,2,..,Q \in \mathbb{R}^{N \times J_q}$, the empirical covariance matrix of Q linear combinations $a_q' X_q$, $q = 1,2,...,Q$, can be partitioned in the following way :

$$S(a) = \hat{\mathrm{Cov}}(a_1' X_1,...,a_Q' X_Q) = \frac{1}{N} \begin{bmatrix} a_1' Z_1' Z_1 a_1 & \cdots & a_1' Z_1' Z_Q a_Q \\ \vdots & \ddots & \vdots \\ a_Q' Z_Q' Z_1 a_1 & \cdots & a_Q' Z_Q' Z_Q a_Q \end{bmatrix}.$$

The first extension (Carroll (1968)) to be examined here, consists in finding Q linear combinations $a_q' X_q$, standardized each according to $\mathrm{Var}(a_q' X_q) = 1$, and an additional unit variance variable Y that maximize the sum of squared correlation coefficients between the Q linear combinations and the auxiliary variable. The empirical version of this optimization problem can be formulated in the following way ($b \in \mathbb{R}^N$, containing the N values of the variable Y, $a \in \mathbb{R}^J$, $J := \Sigma_q J_q$) :

$$(3.1) \qquad F_1(a,b) := \frac{1}{QN^2} \sum_{q=1}^{Q} \{b' Z_q a_q\}^2 \rightarrow \max_{a,b}$$

subject to $\frac{1}{N} b'b = 1$ and $\frac{1}{N} a_q' Z_q' Z_q a_q = 1$, $q = 1,...,Q$.

Theorem 3.1. Define $A_1 := \frac{1}{Q} \Sigma_{q=1}^{Q} Z_q (Z_q' Z_q)^+ Z_q'$ and let $\hat{b}$ denote the eigenvector corresponding to the largest eigenvalue λ_1^2 of A_1, standardized according to $\frac{1}{N} \hat{b}' \hat{b} = 1$. Then the solution of (3.1) are given by $\hat{b}$ and

$$(3.2) \qquad \hat{a}_q = d_q (Z_q' Z_q)^+ Z_q' \hat{b} + [I_{J_q} - (Z_q' Z_q)^+ Z_q' Z_q] w_q, \quad q = 1,2,...,Q$$

with arbitrary J_q-dimensional vectors w_q and a constant $d_q := \{\frac{1}{N} b' Z_q (Z_q' Z_q)^+ Z_q' b\}^{-1/2}$.

Proof. In order to maximize $F_1(a,b)$ w.r.t. $a_1,a_2,\dots,a_Q$ for a fixed b we use the Cauchy-Schwarz inequality :

$$F_1(a,b) := \frac{1}{QN^2} \sum_{q=1}^{Q} \{b'Z_q a_q\}^2 \overset{\text{Cauchy-Schwarz}}{\leq} \frac{1}{QN^2} \sum_{q=1}^{Q} (a_q'Z_q'Z_q a_q)(b'b).$$

Equality holds if and only if for all $q = 1,2,\dots,Q$ a constant c_q exists such that :

$$Z_q a_q = c_q\, b.$$

The (correctly standardized) least-squares solutions of this system of linear equations are given by (cf. Rao, Mitra (1971), theorem 2.3.1) :

$$\tilde{a}_q = \frac{c_q((Z_q'Z_q)^+Z_q'b + [I_{J_q} - (Z_q'Z_q)^+Z_q'Z_q]w_q)}{\{\frac{c_q^2}{N}\, b'Z_q(Z_q'Z_q)^+Z_q'b\}^{1/2}}$$

with arbitrary vectors $w_q \in \mathbb{R}^{J_q}$. Substituting $\tilde{a} = [\tilde{a}_1',\tilde{a}_2',\dots,\tilde{a}_Q']'$ in the objective function F_1 yields :

$$\max_{a:1/N\, a_q'Z_q'Z_q a_q=1\ \forall q} F_1(a,b) = F_1(\tilde{a},b)$$

$$= \frac{1}{QN^2} \sum_{q=1}^{Q} d_q^2\,\{b'Z_q(Z_q'Z_q)^+Z_q'b\}^2$$

$$= \frac{1}{N}\, b'A_1 b.$$

By a well known result, the functions $b'A_1b$, with b satisfying $\frac{1}{N}\, b'b = 1$, is maximized by the appropriately standardized eigenvector $\hat{b}$ corresponding to the largest eigenvalue λ_1^2 of the matrix A_1 which completes the proof. □

Note that the application of this technique to Q indicator matrices corresponding to a Q-dimensional contingency table is the same as the usual *multiple correspondence analysis* with respective standardization of the score vector (Greenacre (1984)), whereas the following three generalization induce different extensions of correspondence analysis. Using the approach of Kettenring (1971), these three Q-dimensional generalization of canonical correlation analysis can be formulated in the following way : We search for linear combinations $a_q'X_q$, q = 1,2,...,Q, each having unit variance, which either
– maximize the sum of correlations between each pair, i.e.

SUMCOR

(3.3) $$f_2(a) := \frac{1}{N} \sum_{q=1}^{Q} \sum_{q^*=1}^{Q} a_q'Z_q'Z_{q^*}a_{q^*} = a'A_2(a)a \rightarrow \max_{1/N\, a_q'Z_q'Z_qa_q=1\ \forall q}$$

– or maximize the sum of squared correlations, i.e.

SSOCOR

(3.4) $$F_3(a) := \frac{1}{N} \sum_{q=1}^{Q} \sum_{q^*=1}^{Q} \{ a_q'Z_q'Z_{q^*}a_{q^*} \}^2 = a'A_3(a)a \rightarrow \max_{1/N\, a_q'Z_q'Z_qa_q=1\ \forall q}$$

– or minimize the generalized variance of S(a), i.e.

GENVAR

(3.5) $$F_4(a) := \det\{S(a)\} = a'A_4(a)a \rightarrow \max_{1/N\, a_q'Z_q'Z_qa_q=1\ \forall q}$$

where $a = [a_1',a_2',\dots,a_Q']' \in \mathbb{R}^J$, $J = \sum_{q=1}^{Q} J_q$, and the matrices $A_i(a)$ are defined by :

$$A_2(a) := A_2 := \frac{1}{N} \begin{bmatrix} Z_1'Z_1 & \cdots & Z_1'Z_Q \\ \vdots & \ddots & \vdots \\ Z_Q'Z_1 & \cdots & Z_Q'Z_Q \end{bmatrix}$$

$$A_3(a) := \frac{1}{N^2} \text{blockdiag} \left[\sum_{q=1}^{Q} Z_1'Z_q a_q a_q' Z_q' Z_1, \ldots, \sum_{q=1}^{Q} Z_Q' Z_q a_q a_q' Z_q' Z_Q \right] ,$$

$$A_4(a) := \frac{1}{N} \text{blockdiag} \{ Z_1' T_{(1)} S_{(1)}^{-1} T_{(1)}' Z_1, \ldots, Z_Q' T_{(Q)} S_{(Q)}^{-1} T_{(Q)}' Z_Q \},$$

where $S_{(q)}$ is the $(Q-1) \times (Q-1)$–matrix obtained from $S(a)$ by omitting the qth row and qth column and $T_{(q)}$ defined by

$$T_{(q)} := \frac{1}{N} [Z_1 a_1, \ldots, Z_{q-1} a_{q-1}, Z_{q+1} a_{q+1}, \ldots, Z_Q a_Q] .$$

Using this notation, the common structure of all three optimization problems becomes evident, but we should realize that the matrices $A_3(a)$ and $A_4(a)$ are really dependent on the vector a, whilst this is not the case for $A_2(a) = A_2$. All three optimization problems lead to some kind of a (generalized) eigenvalue problem :

Theorem 3.2. Any solution $\hat{a}$ of the ith optimization problem, $i \in \{2,3,4\}$ is an eigenvector of the *"generalized eigenvalue problem"* :

(3.5) $$A_i(\hat{a})\hat{a} = D_\lambda B\hat{a}$$

(3.6) $$\frac{1}{N} \hat{a}_q' Z_q' Z_q \hat{a}_q = 1 \qquad q = 1,2,\ldots,Q$$

with the blockdiagonal matrices

$$D_\lambda = \text{blockdiag}[\lambda_1 I_{J_1}, \ldots, \lambda_Q I_{J_Q}] ,$$
$$B = \frac{1}{N} \text{blockdiag} [Z_1'Z_1, \ldots, Z_Q'Z_Q] .$$

Proof. The proofs are straightforward and can be formulated by applying the Cauchy-Schwarz inequality in the same way as in the proof of theorem 3.1.

Remarks.

1. Kettenring (1971) proposed a method of successive cyclic iterations to find the maximum of the functions F_2, F_3 and F_4. Each cycle of his algorithm consists of Q separate steps. In cycle i and step q^*, $1 \le q^* \le Q$, the current canonical correlation

vectors $a_q^{(i)}$, $q = 1,...q^*-1$, and $a_q^{(i-1)}$, $q = q^* +1,...,Q$, that have been determined so far, are considered as fixed and $a_{q^*}^{(i)}$ is selected to maximize the sum of correlations (resp. the sum of squared correlations or the squared multiple correlation) of $a_{q^*}^{(i)'} X_{q^*}$ with the other current variables. For the SUMCOR method e.g., $a_{q^*}^{(i)}$ in cycle i and step q^* ist defined as

$$a_{q^*}^{(i)} := \left(\frac{Z_{q^*}'Z_{q^*}}{N}\right)^{-1} \left[\sum_{q=1}^{q^*-1} \frac{Z_{q^*}'Z_{q^*}}{N} a_q^{(i)} + \sum_{q=q^*+1}^{Q} \frac{Z_{q^*}'Z_{q^*}}{N} a_q^{(i-1)}\right]$$

and standardized according to $a_{q^*}^{(i)'} \frac{Z_{q^*}'Z_{q^*}}{N} a_{q^*}^{(i)} = 1$.

2. Yanai (1986) applied the SUMCOR and SSQCOR extensions to $Q = 15$ categorical variables in order to compare their results with the usual *multiple correspondence analysis*. The well-known Newton-Raphson algorithm was used to maximize the functions F_2 and F_3.

3. Lafosse (1989) aims at maximizing the sum of squared correlations between each of the Q variables and the sum of the remaining variable, i.e.

$$\sum_{q^*}^{Q} \rho^2 (a_{q^*}' X_{q^*}, \sum_{q \neq q^*} a_q' X_q) \rightarrow \max.$$

Without giving a convergence proof he proposes an algorithm which might be considered as an intermediate form of Carroll's and Horst's algorithms. Cycle i and step q^* of this iteration consists of the following calculations :

$$s_{q^*}^{(i)} := \frac{1}{N} \sum_{q=1}^{q^*-1} Z_q a_q^{(i)} + \frac{1}{N} \sum_{q=q^*+1}^{Q} Z_q a_q^{(i-1)} ,$$

$$s_{q^*}^{(i)} := \frac{s_{q^*}^{(i)}}{\|s_{q^*}^{(i)}\|} ,$$

$$a_{q^*}^{(i)} := \left(\frac{Z_{q^*}'Z_{q^*}}{N}\right)^{-1} Z_{q^*}' s_{q^*}^{(i)} .$$

4. AN ALGORITHM FOR GENERALIZED EIGENVALUE PROBLEMS.

In this chapter we will present a new iterative procedure for the solution of the generalized eigenvalue problem (3.5), (3.6). We consider the following l_p-norm optimization problem ($1 < p \leq 2$), which has been investigated by Häussler (1984) in the context of robust l_p-discrimination for the special case of a linear function ϕ and $Q = 1$: Let $\phi : \mathbb{R}^J \to \mathbb{R}^L$, $\phi(a) = R(a)'a$, be a function of the partitioned vector

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_Q \end{bmatrix} , \quad a_q \in \mathbb{R}^{J_q}, \ q = 1,2,\ldots,Q,$$

with a $(J \times L)$-matrix $R(a) := [r_1(a),\ldots,r_L(a)]$. We want to maximize the l_p-norm of $\phi(a)$:

$$(4.1) \qquad F(a) := \|\phi(a)\|_p^p = \sum_{l=1}^{L} |r_l(a)'a|^p \to \max_{\|v_q(a_q)\|_p^p = \Sigma_{k=1}^{K_q} |s_{qk}'a|^p = 1 \ \forall q}$$

with respect to $a \in \mathbb{R}^J$ subject to the restriction that the l_p-norm of $v_q(a_q) = S_q'a_q$ is equal to 1 for $q = 1,2,\ldots,Q$, where $S_q := [s_{q1},\ldots,s_qK_q]$ is a given $(J_q \times K_q)$-matrix with full rank K_q. By defining the $J \times J$-matrices

$$A(a) := \sum_{l=1}^{L} |r_l(a)'a|^{p-2} r_l(a) r_l(a)' ,$$

$$B_q(a) := \text{blockdiag}[0,\ldots,\sum_{k=1}^{K_q} |s_{qk}' a_q|^{p-2} s_{qk}s_{qk}',\ldots,0], \quad q = 1,2,\ldots,Q$$

(4.1) can be reformulated as

$$(4.2) \qquad F(a) := a'A(a)a \to \max_{a'B_q(a)a=1 \ q=1,2,\ldots,Q} .$$

A straightforward derivation shows :

Theorem 4.1. If the function F is strictly convex and continuously differentiable with $\text{grad}_a F(a) = cA(a)a$, where $c > 0$, a necessary condition for $\hat{a}$ to solve (4.1), resp. (4.2), is the existence of constants $\lambda_1, \lambda_2,\ldots,\lambda_Q \in \mathbb{R}$ such that

(4.3) $$A(\hat{a})\hat{a} = \sum_{q=1}^{Q} \lambda_q B_q(\hat{a})\hat{a} ,$$

(4.4) $$\hat{a}'B_q(\hat{a})\hat{a} = 1 \quad q = 1,2,\ldots,Q .$$

Obviously, (4.3) and (4.4) reduce to our original problems SUMCOR, SSQCOR and GENVAR by setting $p = 2$ and substituting $A(a)$ by $A_2(a)$, $A_3(a)$ or $A_4(a)$, respectively, as well as $B_q(a)$ by $B_q := \text{blockdiag}[0,\ldots\frac{1}{N} Z_q'Z_q,\ldots,0]$ for $q = 1,2,\ldots,Q$. We propose (and have used) the following iterative procedure for solving generalized eigenvalue problems of the type (4.3) :

Theorem 4.2. The following algorithm :

Choose $a^{(0)} \in \mathbb{R}^J$ subject to $a^{(0)'}B_q(a^{(0)})a^{(0)} = 1$, $\forall\, q = 1,2,\ldots,Q$. For $i \geq 1$ solve the following system of linear equations for $x^{(i+1)}$:

(4.5) $$\sum_{q=1}^{Q} B_q(a^{(i)})x^{(i+1)} = A(a^{(i)})a^{(i)}$$

and define

(4.6) $$D^{(i+1)} := \text{blockdiag}[\ldots,\{x_q^{(i+1)'}B_q(a^{(i)})x_q^{(i+1)}\}^{-1/2} I_{J_q},\ldots]$$

(4.7) $$a^{(i+1)} := D^{(i+1)}\, x^{(i+1)}$$

is globally convergent, i.e. it converges to a solution of (4.3) and (4.4) for any starting vector $a^{(0)} \in \mathbb{R}^J$.

The following proof of theorem 4.2. follows the same lines as the proof given in Watson (1985) who considered the simpler case of a *linear* function ϕ . We will need some preliminary lemmas.

Lemma 4.3. For any real numbers α, β, $\beta \neq 0$, and $1 < p < 2$, the following inequality holds :

$$|\alpha|^p - |\beta|^p - \frac{1}{2}p|\beta|^{p-2}(\alpha^2 - \beta^2) \leq 0.$$

Proof. Setting $x := \frac{\alpha}{\beta}$ the inequality in lemma 4.3 is equivalent to

$$f(x) := |x|^p - \frac{p}{2}x^2 + \frac{p}{2} \le 1 .$$

It is easily shown by differentiation that this function $f(x)$ obtains its global maximum at $x = 1$ with $f(1) = 1$.

Lemma 4.4. Let the sequence $\{a^{(i)}\}$ be defined as in theorem 4.2. Then the following two statements are equivalent for $i \in \mathbb{N}_0$:

i) $a^{(i)} = a^{(i+1)}$

ii) $a^{(i)}$ is a solution of the generalized eigenvalue problem (4.3), (4.4), i.e.

$$A(a^{(i)})a^{(i)} = \sum_{q=1}^{Q} \lambda_q B_q(a^{(i)})a^{(i)} , \tag{4.8}$$

$$a^{(i)'}B_q(a^{(i)})a^{(i)} = 1, \quad q = 1,2,...,Q \tag{4.9}$$

for some real numbers $\lambda_1,...,\lambda_Q$.

Proof. We will first show the implication i) $\rightarrow$ ii). If $a^{(i+1)} = a^{(i)}$ for some $i \in \mathbb{N}_0$, we get

$$x^{(i+1)} = D^{(i+1)-1} a^{(i+1)} = D^{(i+1)-1}a^{(i)}$$

and therefore :

$$\begin{aligned} A(a^{(i)})a^{(i)} &\overset{(4.5)}{=} \sum_{q=1}^{Q} B_q(a^{(i)})D^{(i+1)-1}a^{(i)} \\ &\overset{(4.6)}{=} \sum_{q=1}^{Q} \{x^{(i+1)'}B_q(a^{(i)})x^{(i+1)}\}^{1/2} B_q(a^{(i)})a^{(i)} \\ &= \sum_{q=1}^{Q} \lambda_q B_q(a^{(i)})a^{(i)} \end{aligned}$$

with $\lambda_q := \lambda_q^{(i)}$.

On the other hand, if $a^{(i)}$ is a solution of (4.3) and (4.4), then

$$\sum_{q=1}^{Q} B_q(a^{(i)})x^{(i+1)} \overset{(4.5)}{=} A(a^{(i)})a^{(i)} \overset{(4.3)}{=} \sum_{q=1}^{Q} \lambda_q B_q(a^{(i)})a^{(i)} ,$$

and since $B_q(a^{(i)}) > 0$ for all q, we conclude from the definition of B that $x_q^{(i+1)} = \lambda_q a_q^{(i)}$ for all q. This yields ?

$$a_q^{(i+1)} \overset{(4.6)(4.7)}{=} \{x^{(i+1)'} B_q(a^{(i)}) x^{(i+1)}\}^{-1/2} x_q^{(i+1)}$$

$$= \{\lambda_q^2 a^{(i)'} B_q(a^{(i)}) a^{(i)}\}^{-1/2} \lambda_q a_q^{(i)}$$

$$\overset{(4.4)}{=} a_q^{(i)}$$

for all q which proves the implication ii) $\to$ i). □

Proof of theorem 4.2.

In order to prove the convergence of the iterative procedure defined in theorem 4.2. we will first show that for $i \in \mathbb{N}_0$ $F(a^{(i+1)}) \geq F(a^{(i)})$ with equality if and only if $a^{(i+1)} = a^{(i)}$. To accomplish this, we will need the following two inequalities :

(4.10) $$a^{(i+1)'} B_q(a^{(i)}) a^{(i+1)} \geq 1 \quad \forall\, q = 1,2,\ldots,Q,\ i \in \mathbb{N}_0$$

(4.11) $$a^{(i+1)'} B_q(a^{(i)}) a^{(i)} \leq 1 \quad \forall\, q = 1,2,\ldots,Q,\ i \in \mathbb{N}_0$$

In the case $p = 2$, the matrix $B_q(a^{(i)})$ is independent of $a^{(i)}$ (by definition), so with the restriction (4.7) imposed on each element of the sequence $\{a^{(i)}\}_{i \in \mathbb{N}_0}$ we have $a^{(i+1)'} B_q(a^{(i)}) a^{(i+1)} = a^{(i+1)'} B_q(a^{(i+1)}) a^{(i+1)} = 1$.

If $1 < p < 2$, setting $\alpha := s'_{qj} a_q^{(i+1)}$ and $\beta := s'_{qj} a_q^{(i)}$ in lemma 4.3. and summing over j gives the required result (4.10).

Since all $a' B_q(a) a$ are convex functions, we have

$$a^{(i+1)'} B_q(a^{(i+1)}) a^{(i+1)} \geq a^{(i)'} B_q(a^{(i)}) a^{(i)} + p[a^{(i+1)'} B_q(a^{(i)}) a^{(i)} - a^{(i)'} B_q(a^{(i)}) a^{(i)}]$$

which implies $a^{(i+1)'} B_q(a^{(i)}) a^{(i)} \leq 1$.

Using the definition of the sequence $\{a^{(i)}\}_{i \in \mathbb{N}_0}$ as well as both inequalities (4.10) and (4.11), we get

(4.12) $$a^{(i+1)'} A(a^{(i)}) a^{(i)} \geq a^{(i)'} A(a^{(i)}) a^{(i)}.$$

Since F is a convex function with gradient $cA(a)a$, we have

$$F(a^{(i+1)}) \geq F(a^{(i)}) + c[a^{(i+1)'} A(a^{(i)}) a^{(i)} - a^{(i)'} A(a^{(i)}) a^{(i)}].$$

By (4.12) we conclude that $F(a^{(i+1)}) \geq F(a^{(i)})$ with equality iff $a^{(i)} = a^{(i+1)}$ due to strict convexity. Combining this result with lemma 4.4. and using the continuity of $A(a)$ and $B_q(a)$, it appears that any limit of the sequence $\{a^{(i)}\}_{i\in \mathbb{N}_0}$ is necessarily a solution of the generalized eigenvalue problem (4.3), (4.4).

Since the compact set $L := \{x = [x_1', x_2', \dots, x_J']' \in \mathbb{R}^J \,|\, x'B_q(x)x = 1 \;\forall\, q = 1,2,\dots,Q\}$ contains the sequence $\{a^{(i)}\}_{i\in \mathbb{N}_0}$, we know by the theorem of Bolzano-Weierstraß that there exists a subsequence $\{a^{(i_n)}\}_{n\in \mathbb{N}_0} \subset \{a^{(i)}\}_{i\in \mathbb{N}_0}$ converging in L. Define

$$a^* = \lim_{n\to\infty} a^{(i_n)}.$$

As we have shown, the continuous, bounded function F is strictly increasing, which implies the convergence of $\{F(a^{(i_n)})\}_{n\in \mathbb{N}_0}$ and $\{F(a^{(i_n+1)})\}_{n\in \mathbb{N}_0}$. By continuity of F, setting $a^{**} := \lim_{n\to\infty} a^{(i_n+1)}$, we derive $F(a^*) = F(a^{**})$, which implies $a^* = a^{**}$ due to strict convexity. □

Remark. Having determined the first generalized canonical correlation vectors $a^{(1)}$, successive vectors $a^{(m)}$, $m \geq 2$, orthogonal to all previous ones, may be calculated by redefining the matrices $A_i^{(m)}(a)$, $i = 2,3,4$ in the following way :

$$A_i^{(m)}(a) := \left(I - \sum_{k=1}^{m-1} \text{blockdiag}\left[\frac{Z_1'Z_1}{N} a_1^{(k)} a_1^{(k)'}, \dots, \frac{Z_Q'Z_Q}{N} a_Q^{(k)} a_Q^{(k)'} \right] \right) A_i(a).$$

5. APPLICATION AND CONCLUDING REMARKS.

It is now evident that correspondence analysis for $Q \geq 2$ qualitative variables can be defined by applying one of the four previously mentioned generalizations of canonical correlation analysis to the corresponding binary $N \times (J_1 + \dots + J_Q)$ indicator matrix $W = [X_1, \dots, X_Q]$. This will be illustrated by an example of a three-dimensional contingency table, which has been previously analyzed by van der Heijden (1985) using loglinear models as well as ordinary correspondence analysis. The data have been collected by the German Office for Statistics in Western Germany during 1974-1977 (see also Heuer (1979)). The table cross-classifies 43210 suicides by gender (men, women), age (10-15, 15-20,..., 85-90, 90+) and method of suicide (MATT - poisoning by solid or liquid matter, GASH - by toxification with gas at home, GASO - by toxification with other gas, HANG - by hanging, strangling, suffocating, DROWN - by drowning, GUNS - by guns and explosives, KNIFE - by knifes etc., JUMP - by jumping, OTHER - by other methods). Table 1 shows the first and second canonical correlation vectors derived by each of the four proposed extensions of correspondance analysis.

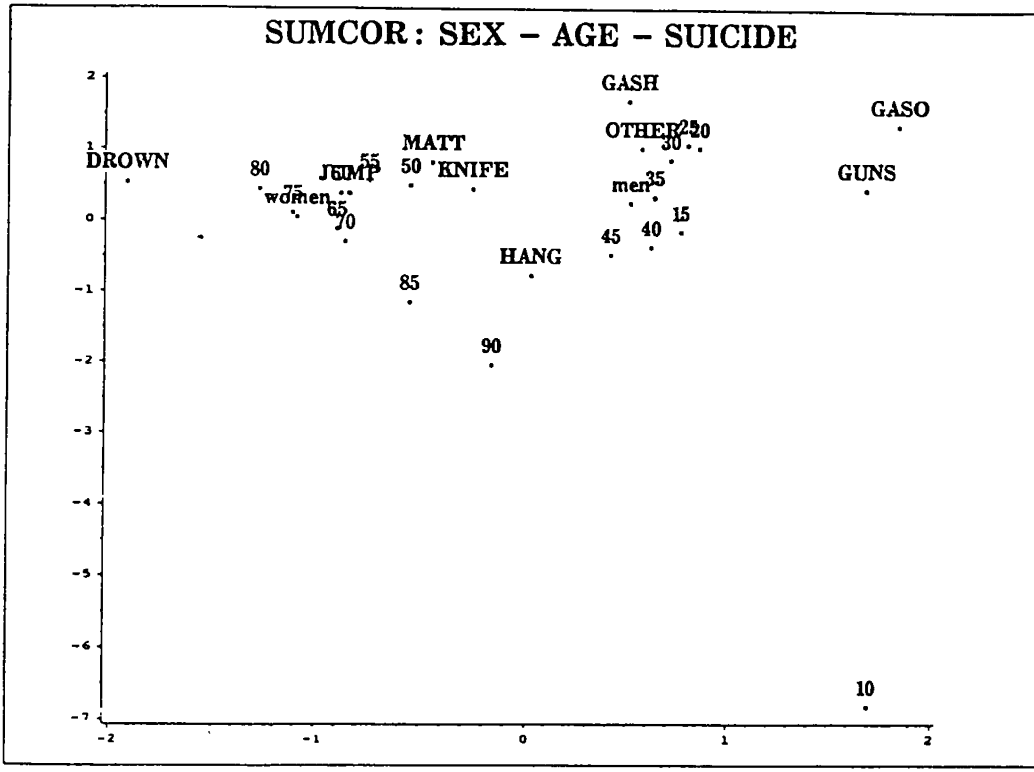

Figure 1 : 2-dimensional display of SUMCOR-scores.

As the four methods yielded quite similar results for this contingency table, only the graphical display of the scores obtained by the SUMCOR method is shown in Figure 1. It is easily seen that whereas relatively more men commit suicide when they are young, relatively more women commit suicide when they are older. In the younger age groups, suicide is committed extraordinarily often by gas at home and other gas, by guns and other methods, while drowning, jumping, poisoning, hanging, and stabbing is prevalent in the older age groups.

As the transition formulae define the rationale for the interpretation of the joint display, we comment on the transition formulae of the four different generalized correspondence analysis techniques.

Table of first and second generalized canonical correlation vectors for SEX-AGE-SUICIDE-Example

Variable	Category	MAXVAR $a^{(1)}$	MAXVAR $a^{(2)}$	SUMCOR $a^{(1)}$	SUMCOR $a^{(2)}$	SSQCOR $a^{(1)}$	SSQCOR $a^{(2)}$	GENVAR $a^{(1)}$	GENVAR $a^{(2)}$
Gender -	men	0.757	-0.757	0.513	0.251	0.697	-0.301	0.448	0.527
	women	-1.3211	1.3211	-1.096	0.050	-0.742	-0.601	-1.109	0.347
Age	10-15	1.299	-5.901	1.698	-6.826	2.375	-5.402	2.063	-5.201
	15-20	1.017	0.444	0.765	-0.158	0.574	0.345	0.600	0.511
	20-25	1.294	1.562	0.852	1.027	0.529	1.300	0.683	1.336
	25-30	1.230	1.530	0.796	1.071	0.479	1.311	0.625	1.339
	30-35	1.093	1.188	0.711	0.856	0.463	1.061	0.664	0.981
	35-40	0.923	0.635	0.634	0.335	0.460	0.608	0.623	0.543
	40-45	0.808	-0.141	0.618	-0.372	0.529	0.004	0.629	-0.079
	45-50	0.533	-0.303	0.416	-0.480	0.336	-0.047	0.324	0.106
	50-55	-0.572	0.137	-0.553	0.489	-0.717	0.751	-0.821	1.074
	55-60	-0.813	0.042	-0.748	0.549	-0.883	0.742	-0.946	0.986
	60-65	-1.005	-0.316	-0.883	0.383	-0.952	0.520	-0.992	0.685
	65-70	-1.096	-0.924	-0.901	-0.119	-0.869	0.020	-0.886	0.092
	70-75	-1.067	-1.239	-0.862	-0.300	-0.723	-0.258	-0.634	-0.423
	75-80	-1.338	-0.935	-1.117	0.107	-1.048	0.092	-0.956	-0.023
	80-85	-1.501	-0.668	-1.278	0.439	-1.222	0.322	-1.046	0.068
	85-90	-0.793	-1.808	-0.553	-1.150	-0.339	-0.985	-0.193	-1.256
	90+	-0.412	-2.675	0.164	-2.042	0.198	-1.819	0.404	-2.263
Suicide	MATT	-0.438	0.984	-0.448	0.816	-0.491	0.232	-0.681	0.109
	GASH	0.918	1.086	0.503	1.691	0.585	0.694	0.371	0.698
	GASO	2.556	1.082	1.841	1.354	1.655	0.435	1.425	0.920
	HANG	0.043	-1.185	0.024	-0.778	-0.158	-1.275	0.189	-1.318
	DROWN	-2.287	0.043	-1.921	0.527	-1.693	-0.087	-1.633	-0.735
	GUNS	2.289	-0.123	1.684	0.439	1.484	-0.320	1.575	0.194
	KNIFE	-0.150	-0.443	-0.257	0.439	-0.158	-0.308	0.028	-0.429
	JUMP	-0.980	0.464	-0.842	0.382	-0.899	-0.232	-0.982	-0.661
	OTHER	0.882	1.089	0.569	1.020	0.451	0.348	0.370	0.563

Table 1 : Results of MAXVAR-, SUMCOR-, SSQCOR- AND GENVAR-Extensions.

Given a Q-dimensional contingency table $(p_{j_1 j_2 \ldots j_Q})$ let D_q be the $J_q \times J_q$ diagonal matrix containing the observed relative frequencies of the J_q categories of the qth variable and P_{qq*} the 2-dimensional $J_q \times J_{q*}$ marginal table cross-classifying the variables q and q*. The *transition formulae* derived by the usual multiple *correspondence analysis* takes the form (Greenacre (1984)) :

$$\hat{a}_q = \frac{1}{\mu} \sum_{q^* \neq q} D_q^{-1} P_{qq^*} \hat{a}_{q^*} \ , \quad q = 1,2,\ldots,Q, \tag{5.1}$$

i.e. the scaling vector $\hat{a}_q$ of the J_q column profiles corresponding to the qth variable is the sum of centroids $D_q^{-1} P_{qq*} \hat{a}_{q*}$ of the scaling vectors of the remaining column profiles (each weighted by the corresponding row profile of the 2-dimensional marginal table-cross-classifying variable q and q*) and stretched out by a constant factor $\frac{1}{\mu}$ (μ a multiple of λ_1^2 in theorem 3.1). This formula makes sense in the special case $Q = 2$, whereas for $Q \geq 3$ it has at least two shortcomings :

(i) Each variable q is expanded by a constant factor $\frac{1}{\mu}$, independently of the strength of relationship between this variable and the remaining $Q - 1$ variables. The *transition formulae* obtained by the SUMCOR method take this into account, stretching the qth variable by a factor $\frac{1}{\mu_q}$, a multiple of the qth *generalized eigenvalue* $\lambda_q = \sum_{q*} \hat{a}_q' P_{qq*} \hat{a}_{q*}$, as can easily be seen by applying formula (4.3) to the situation of Q sets of indicator variables.

(ii) For fixed q, the formula (5.1) uses equal weights for different q* when summing up the respective barycenters of the scaling vectors of the remaining profiles $D_q^{-1} P_{qq*} \hat{a}_{q*}$, whilst one might expect each barycenter to be differentially weighted by a mass w_{qq*} according to the degree of dependence between the respective variables q and q*. The SSQCOR method, as well as the GENVAR method lead to *transition formulae* consistent with this idea :

$$\hat{a}_q = \frac{1}{\mu_q} \sum_{q^* \neq q} w_{qq^*} D_q^{-1} P_{qq^*} \hat{a}_{q^*} \ ,$$

with weights $w_{qq*} := \hat{a}_q' P_{qq*} \hat{a}_{q*}$, and a factor μ_q proportional to the qth generalized eigenvalue $\lambda_q = \sum_{q*} (\hat{a}_q' P_{qq*} \hat{a}_{q*})^2$ for SSQCOR and $w_{qq*} := \sum_{r \neq q} s_{rq*} (\hat{a}_r' P_{rq} \hat{a}_q)$, and a factor μ_q proportional to the qth *generalized eigenvalue* $\lambda_q = \sum_r \sum_{q*} (\hat{a}_q' P_{qr} \hat{a}_r)(\hat{a}_q' P_{qq*} \hat{a}_{q*}) s_{rq*}$ (where s_{rq*} is an element of the matrix $S_{(q)}^{-1}$ for GENVAR.

Obviously, the latter three extensions of correspondence analysis yield more meaningful theoretical interpretations of the resulting graphical representations than the usual *multiple correspondence analysis*. The availability of fast parallel computers, with matrix oriented languages at one's disposal, facilitate the implementation of the algorithms presented above, and therefore the necessity of extra computation for the SUMCOR, SSQCOR or GENVAR methods can not be considered as a disadvantage in comparison with the straightforward *multiple correspondence analysis* and therefore should not deter the statistical practitioner or data analyst from applying these methods. In our opinion this computational *"handicap"* is by far compensated by the gain in interpretability as well as by the possibility of detecting structures of interdependence between the cross-classified categorical variables which would be hidden otherwise.

Acknowledgement.

I would like to thank Prof. H.H. Bock for directing my attention to these extensions and for helpful discussions.

References

Bellman, R. (1960) Introduction to Matrix Analysis. *McGraw-Hill Book Company, New York.*

Benzécri, J.P. (1964) Cours de linguistique mathématique. *Publications de l'Institut de Statistique de l'Université de Paris, 13.*

Benzécri, J.P. (1977) Sur l'analyse des tableaux binaires associés à une correspondance multiple. *Cahiers de l'Analyse des Données 2, 55-71.*

Benzécri, J.P. (1980) Pratique de l'analyse des données. *Dunod, Paris.*

Bock, R.D. (1960) Methods and application of optimal scaling. *The University of North Carolina, Psychometric Laboratory Research Memorandum 25.*

Carroll, J.D. (1968) Generalization of Canonical Correlation Analysis to three or more sets of variables. *Proc. 76th annual convention of the APA, 227-228.*

Eckart, C., Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika 1, 211-218.*

Fisher, R.A. (1940) The precision of discriminant functions. *Ann. Eugen. 10, 422-429.*

Gifi, A. (1990) Nonlinear Multivariate Analysis. *Department of Data Theory, University of Leiden, The Netherlands.*

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

Greenacre, M.J. (1988) Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika 75, 457-467.*

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

Häussler, W.M. (1984) Computational Experience with an EV Algorithm for robust L_p-Discrimination. *Comp. Stat. Quarterly 1, 233-244.*

Hayashi, C. (1950) On the quantification of qualitative data from the mathematico-statistical point of view. *Ann. Inst. Statist. Math. 2, 35-47.*

Heuer, J. (1979) Selbstmord bei Kindern und Jugendlichen. *Stuttgart, Ernst Klett Verlag.*

Hill, M.O. (1974) Correspondence analysis : a neglected multivariate method. *Appl. Satist. 23, 340-354.*

Hirschfeld, H.O. (1935) A connection between correlation and contingency. *Cambridge Philosophical Soc. Proc. 31, 520-524.*

Horst, P. (1935) Measuring complex attitudes. *J. Social Psychol. 6, 369-374.*

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

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol. 24, 417-441, 498-520.*

Kettenring, J.R. (1971) Canonical Analysis of several sets of variables. *Biometrika 58, 433-451.*

Krafft, 0. (1978) Lineare statistische Modelle und optimale Versuchpläne. *Vandenhoeck & Ruprecht, Göttingen.*

Lafosse, R. (1989) Proposal for a Generalized Canonical Analysis, In : R. Coppi, S. Bolasco (editors), Multiway Data Analysis. *North Holland.*

Lebart, L., Morineau, A., Warwich, K. (1979) Traitement des Données Statistiques. *Dunod, Paris.*

Lebart, L. Morineau, A., Fenelon, J.P. (1984) Multivariate Descriptive Statistical Analysis. *New York : Wiley.*

Leclerc, A. (1980) Quelques Propriétés optimales en analyse de données en terme de corrélation entre variables. *Mathématiques et Sciences Humaines 18, 51-67.*

Lingoes, J.C. (1964) Simultaneous linear regression : an IBM 7090 program for analyzing metric/nonmetric or linear/nonlinear data. *Behav. Res. 9, 61-94.*

Masson, M. (1974) Processus linéaires et analyse des données non linéaires. *Thèse de 3ème cycle, Université Pierre et Marie Curie, Paris.*

Masson, M. (1980) Méthodologies générales de traitement statistique de l'information de masse. Paris : *Cedic-Fernand Nathan.*

McKeon, J.J. (1966) Canonical analysis : some relations between canonical correlation, factor analysis, discriminant function analysis, and scaling theory. *Psychometric monograph n° 13, Psychometric Society.*

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

Rao, C.R., Mitra, S.K. (1971) Generalized inverse of matrices and its applications. *Wiley, New York.*

Richardson, M. Kuder, G.F. (1933) Making a rating scale that measures. *Personnel J. 12, 36-40.*

Van de Geer, J.P. (1986) Relations among k sets of variables with geometrical representation and application to categorical variables. In : Multidimensional Data Analysis (Leeuw, Heiser, Meulman, editors). DSWO *Press, Leiden.*

Van der Heijden, P.G.M. (1985) Correspondence analysis used complementary to loglinear analysis. *Psychometrika 50, 429-447.*

Van Rijckevorsel, J.L.A. (1987) The Application of Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis. *SSWO Press, Leiden.*

Watson, G.A. (1985) On the Convergence of EV Algorithms for Robust l_p-Discrimination. *Comp. Stat. Quarterly 4, 307-314.*

Yanai, H. (1986) Some Generalizations of Correspondence Analysis in Terms of Projection Operators. In : E. Diday et al. (editors), Data Analysis and Informatics IV. *North Holland.*