Computational Statistics & Data Analysis 18 (1994) 141–163
North-Holland

OVERALS

Nonlinear canonical correlation with k sets of variables

Eeke van der Burg
Leiden University, Leiden, Netherlands

Jan de Leeuw
UCLA, Los Angeles, CA, USA

Garmt Dijksterhuis
EIM Small Business Research and Consultancy, Zoetermeer, Netherlands

Abstract: OVERALS is a techinque for canonical correlation analysis with two or more sets of variables. Any three way table can be used as input for the OVERALS program. In OVERALS terminology the ways are called objects, variables and sets. Three measurement levels of the data can be handled: numerical, ordinal and nominal. They can be defined for each variable separately. Also the conditionality of the data is defined variable-wise. The OVERALS technique searches for what is common between sets of variables measured on the same objects. The mathematical model and the algorithm are discussed. In addition an illustration of the technique is given in the form of an application.

1. Introduction

1.a. Characteristics of the input data

1.a.1. Modes and ways

OVERALS is a method for analyzing several sets of data at the same time (Van der Burg, De Leeuw and Verdegaal, 1988). It is not specially made for multiway data. However many multiway tables can be analysed with the OVERALS program. OVERALS is a form of *Canonical Correlation Analysis* (CCA). This

Correspondence to: E. van der Burg, Leiden University, Faculty of Social and Behavioral Sciences, Psychometrics Department, P.O. Box 9555, 2300 RB Leiden, The Netherlands.

0167-9473/94/$07.00
SSDI 0167-9473(93)E0044-5

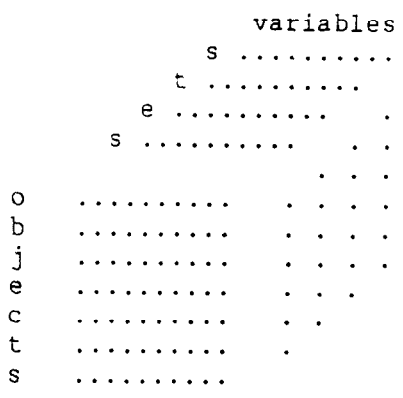

Fig. 1. Three way-table, suited for the OVERALS technique.

method searches for a subspace that several sets of variables, measured on the same objects, have in common. Translated into terminology of multiway data, there are three modes: objects, variables and sets. Thus the data matrix has to be a three-way table as in Figure 1.

For the OVERALS technique it is not necessary, as it is for other multiway methods, that the variables for each set are the same. Or even that there is an equal number of variables for each set. Therefore 'irregularities' in the block are permitted (see 1.b.2).

Naturally many entities can be used as objects, variables and sets, e.g. schoolchildren, schoolresults and schoolclasses or countries, economic characteristics, and years, etc. The OVERALS technique treats the three ways differently, that is it does not make any distinction between modes and ways. Thus every way is taken as a new mode. Consequently, the data are not considered to be symmetric or asymmetric, as (a)symmetry is only under discussion if two ways share one mode.

1.a.2. Type of scales

Several scale types can be handled by the OVERALS technique. A distiction is made between variables measured on *interval* level (*numerical* variables), on *ordinal* level and on *nominal* level. For ordinal data only the order of the categories (per variable) is taken into account, and for nominal data only the classes of objects formed by each variable. This means that, for ordinal variables, each monotone ascending transformation gives the same information about the objects as the original coding. For nominal variables each isomorphic transformation and for numerical variables each linear transformation contains the same information.

For the OVERALS technique the method of *optimal scaling* is chosen to find the best transformations of the variables. In optimal scaling the transformations depend both on the measurement level of the data, and on the model (compare Young, 1981). Or, to be a little bit more precise, the measurement level

	set 1 variables	set 2 variables	set 3 variables	set k variables
o	.....	.....	.....	.....
b	.....	.....	.....	.....
j	.....	.....	.....	.....
e	.....	.....	.....	.....
c	.....	.....	.....	.....
t	.....	.....	.....	.....
s	.....	.....	.....	.....

Fig. 2. Three-way table changed into a two-way table by concatenating the slices for sets.

determines which class of transformations is considered, and the fitted model determines how the transformation is selected from the class of admissible ones.

1.a.3. Conditionality

The data are always interpreted in a variable-oriented way. If one transforms the three-dimensional table into a two-dimensional table by putting the slices for sets next to each other, the data are considered as *column conditional*. This way of concatenating the sets also shows how the input for the OVERALS program must be. A two dimensional table of objects × variables, organized in sets (Figure 2).

Observe again that for OVERALS not all sets need to contain the same variables. Thus flattening out a three-way table is rather a special case of the OVERALS setup.

1.b. Mathematical model

1.b.1. The OVERALS technique for a three-way table measured on interval level

In Figure 2 a data matrix is represented that can serve as input for the OVERALS program. The object of an OVERALS analysis is to find what is in common between the sets of variables. To put it differently, one can ask which *weights* we need to apply to make the weighted sums of all sets of variables as similar as possible to each other. One particular way of measuring similarity of weighted sums uses an unmeasured variable $\boldsymbol{x}$. We want to make the weighted sums as similar to $\boldsymbol{x}$ as possible.

The vector $\boldsymbol{x}$ has n scores, as many as there are objects. For each set there will be m weights, as many as there are variables, and one weighted sum. This formulation of the k-sets canonical correlation problem corresponds to Carroll (1968), who translates it mathematically into maximization of the sum of squared correlations between the unmeasured variable $\boldsymbol{x}$ and weighted sums of variables for each set. Let us collect the weights for each set into a m-vector $\boldsymbol{a}_j$. Denote the data per set by the matrix $\boldsymbol{H}_j$ $(n \times m)$. Then Carroll (1968) searches for

$$\text{maximum} \sum_{j=1}^{k} \text{cor}(\boldsymbol{x}, \boldsymbol{H}_j\boldsymbol{a}_j)^2 \text{ over } \boldsymbol{x} \text{ and } \boldsymbol{a}_j. \tag{1}$$

Van der Burg (1988, p. 101) translates the same k-sets canonical correlation problem into a loss function, minimzing the sum of squared (SSQ) differences between $\boldsymbol{x}$ and the weighted sums while $\boldsymbol{x}$ is unit-normalized, i.e mean zero and variance one. The result is similar to that of Carroll (1968) (see Appendix). It looks as follows:

$$\text{minimize} \sum_{j=1}^{k} \text{SSQ}(\boldsymbol{x} - \boldsymbol{H}_j \boldsymbol{a}_j) \text{ over } \boldsymbol{x} \text{ and } \boldsymbol{a}_j, \tag{2}$$

for unit-normalized $\boldsymbol{x}$ and $\boldsymbol{H}_j$ (per column). Generalization of this expression to allow for more solutions for *object scores* $\boldsymbol{x}$ and weights $\boldsymbol{a}_j$, where a new solution of $\boldsymbol{x}$ is uncorrelated with the preceeding solution(s), gives

$$\text{minimize} \sum_{j=1}^{k} \text{SSQ}(\boldsymbol{X} - \boldsymbol{H}_j \boldsymbol{A}_j) \text{ over } \boldsymbol{X} \text{ and } \boldsymbol{A}_j. \tag{3}$$

for unit-normalized $\boldsymbol{X}$ and $\boldsymbol{H}_j$; $\boldsymbol{X}$ uncorrelated over columns. In (3) $\boldsymbol{X}$ is a $(n \times p)$ matrix, with p the number of solutions (dimensions), and $\boldsymbol{A}_j$ is a $(m \times p)$ matrix. Expression (3) is a formulation of a k-sets CCA-technique. It corresponds to OVERALS for a three-way table (see next paragraph). Other formulations are possible, e.g. Horst (1961) or Kettenring (1971). In (3) only variables measured on interval level are implied. Thus (3) defines a form of linear multivariate analysis, where "linear" refers to a numerical measurement level. In the next paragraph optimal scaling is introduced, resulting in a nonlinear technique.

1.b.2. Optimal scaling

Introducing optimal scaling for the variables (columns of $\boldsymbol{H}_j$) in (3) results in transformed variables $\boldsymbol{Q}_j$ $(n \times m)$ of which the columns satisfy the measurement restrictions. Let us denote the collection of transformations that satisfy the measurement restrictions for a variable $\boldsymbol{h}_t$ (a column of $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_k$) as C_t $(t = 1, \ldots, km)$. C_t contains all linear transformations of $\boldsymbol{h}_t$ for numerical variables, all monotone ascending transformations for ordinal variables, and all isomorphic transformations of $\boldsymbol{h}_t$ for nominal variables). Then expression (3) with optimal scaling is

$$\text{minimize} \sum_{j=1}^{k} \text{SSQ}(\boldsymbol{X} - \boldsymbol{Q}_j \boldsymbol{A}_j) \text{ over } \boldsymbol{X}, \boldsymbol{A}_j \text{ and } \boldsymbol{Q}_j, \tag{4}$$

for unit-normalized $\boldsymbol{X}$ and $\boldsymbol{Q}_j$; $\boldsymbol{X}$ uncorrelated over columns; $\boldsymbol{q}_t \in C_t$ $(t = 1, \ldots, km)$. The vector $\boldsymbol{q}_t$ is a column of $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_k$. Model (4) is described by Van der Burg (1988, p. 101) or Van der Burg and Dijksterhuis (1989). Both these articles refer to Gifi (1981) and Van der Burg et al. (1988). In the latter article a clear description is given of the solutions for the parameters.

Model (4) in a slightly more general form is called OVERALS. OVERALS is more general, because it can handle sets of variables that differ both in number as well as in definition. In this article we refer to (4) as the OVERALS model.

1.b.3. Missing observations
In the OVERALS model (4) no adaptations are made for missing data. In Van der Burg (1988, p. 107) the model with missing options is given. We do not show it here. In practice missing data do not contribute to the loss. Each object that misses an observation on one of the variables of a set, is considered as missing all the scores for that particular set. Thus this object does not contribute to the loss for the corresponding set, but it does contribute for the other sets.

1.b.4. Multiple transformations
The model described by expression (4) deals with one nonlinear transformation for each variable. Sometimes one is interested in more than one transformation. Usually there must be a special reason for this. For instance if one knows that a variable has a nonlinear relationship with some variables and a linear relationship with other variables. In that case one can use copies of a variable in a set (i.e. one variable two or more times repeated, maximally p times). Then each copy gives another transformation.

If p copies are used for a variable and the nominal measurement level is used for all copies, we can compute the so-called multiple nominal transformation of a variable by summing the weighted transformations over copies for each dimension. The multiple nominal transformation is used in (multiple) correspondence analysis. In the usual description of this technique the notion of copies will not be found. It simply says that for a second solution a second transformation is needed. The notion of copies makes it possible to link (multiple) correspondence analysis to OVERALS (Van der Burg et al., 1988).

To distinguish between one nominal transformation (no copies or rather one copy) and a multiple nominal transformation (p copies), we refer to the first type as single nominal and the second type as multiple nominal. Other transformations with copies do not have a special name.

1.b.5. Probalistic assumptions
No probabilistic assumptions are made except for a multinomial distribution of the profile frequencies. If a simple random sample is used, this assumption will be satisfied.

To learn about the significance of OVERALS statistics and to compute confidence intervals and estimate bias, one needs to use randomization techniques, such as the Bootstrap and the Jackknife (Efron, 1982). De Leeuw and Van der Burg (1986) and Van der Burg and De Leeuw (1988) found that randomization techniqes are rather succesful in determining the significance of eigenvalues (OVERALS-fit per dimension). In addition computation of confidence intervals and estimation of bias for eigenvalues using the Bootstrap also was succesful. The Jackknife seemed to perform less satisfactory. Transformations of variables were obviously less stable than eigenvalues.

1.b.6. Possible developments
The algorithm of the OVERALS program is of an Alternating Least Squares (ALS) type (see also section 2). Other types of algorithms are possible. One

alternative is majorization (De Leeuw, 1988). Majorization is used in many multidimensional scaling techniques. In short it replaces the loss function $f(x)$ one is trying to minimize by a more simple loss function $g(x,y)$ that approximates $f(x)$ locally at the current solution y. The more simple function is minimized in each step, after which a new approximation is computed and minimized. A majorization algorithm for OVERALS has been developed, and will be tested out in the future.

In De Leeuw and Van Rijckevorsel (1988) various generalizations of multiple correspondence analysis are suggested. The notion of copies is combined with the notion of bandwidth of a quantification. This produces a system of measurement levels of which both continuous ordinal and continuous numerical variables are special cases, and which incorporates B-splines of various degrees. Incorporating this system in OVERALS produces an even more flexible technique.

Other loss functions than the simple least squares loss function have been tried out. Heiser (1987) has a majorization algorithm for least absolute deviation versions of OVERALS. Work is in progress for a maximum likelihood version, based on a transformed multivariate normal model.

1.c. Links to other techniques

1.c.1. Type of models that can be dealt with by OVERALS

OVERALS is considered as the most general model in the so-called Gifi-system of nonlinear multivariate analysis (Gifi, 1990). It has most Gifi-models as a special case. The name OVERALS is due to this fact. The second part of the name also corresponds to Alternating Least Squares, which refers to the type of algorithm (see section 2).

The models that can be dealt with by OVERALS are PRINCALS (nonlinear principal component analysis), ANACOR (correspondence analysis), HOMALS (multiple correspondence analysis). In addition, the CORALS model (nonlinear canonical correlation analysis) is a special case of OVERALS. If linear techniques are considered, normal PCA is a special case of PRINCALS and consequently of OVERALS. In the next paragraphs we give some explanation to these relationships between OVERALS and the above mentioned models.

When all the sets contain only one variable, we are dealing with a two-way table. If one wonders what is in common between the sets, it corresponds to what is in common between the variables. Then we deal with principal component analysis (PCA). A nonlinear version of PCA (with optimal scaling) is called PRINCALS (Gifi, 1990; SPSS, 1990). Thus OVERALS with one variable per set is PRINCALS. If, in addition, only numerical transformations are used it is normal PCA.

If we restrict not only the number of variables per set to one, but also the transformations to multiple nominal, we get multiple correspondence analysis or HOMALS (Gifi, 1990; SPSS, 1990). This technique is also called dual scaling (Nishisato, 1980). If, in addition, the number of variables is restricted to two,

OVERALS is similar to correspondence analysis or ANACOR (Greenacre, 1984; Gifi, 1990; SPSS, 1990).

Another nonlinear model for PCA is called PRINCIPALS or PRINQUAL (Young, Takane and De Leeuw, 1978; Kuhfeld, Young and Kent, 1987). This model corresponds to OVERALS with one variable per set. However PRINCIPALS/PRINQUAL deals with dicrete and continuous variables, OVERALS and PRINCALS only with discrete variables (see 2.b.2).

Young, De Leeuw and Takane (1976) describe a model for canonical correlation analysis with optimal scaling. They called it CORALS. OVERALS with two sets of variables is similar to CORALS. Van der Burg and De Leeuw (1983) and Gifi (1990) decribe an alternative for the CORALS model, as the parameters of the CORALS model could not be solved easily at that time. This alternative is called CANALS. CANALS is a nonlinear generalization of ordinary canonical correlation analysis for two sets of variables. However it is not similar to CORALS and thus not to OVERALS for two sets (see Appendix).

1.c.2. Other models on k-sets analysis

As mentioned above the CANALS model is very closely related to the OVERALS model for two sets, but is not the same. There are many more CCA-techniques that are related to OVERALS. They are related as they try to find an answer to the same question "what is in common between sets of variables". Names of techniques are for instance SUMCOR, GENVAR, MAXRAT, MAXNEAR. See for a discussion about these models Van de Geer (1984) and Gifi (1990). All these models generalize the two-sets CCA-technique defined by Hotelling (1936). However they emphasize different aspects. Authors who worked on forms of k-sets canonical correlation analysis are e.g. Horst (1961), Carroll (1968), Kettenring (1971), Van de Geer (1984; 1986), Ten Berge (1988) and Gifi (1990).

1.c.3. Models that generalize OVERALS

As mentioned already in 1.c.1. OVERALS is considered as the most general techique in the Gifi-system of nonlinear multivariate analysis. However, time goes on and more general models were invented. Coolen and De Leeuw (1987) discuss a model named PATHALS, a technique for nonlinear path analysis. They show, that OVERALS can be seen as a special case of PATHALS.

Another generalization originates with Van Buuren (1990). He discusses time series techniques with optimal scaling. He shows that the OVERALS loss is a special case of the so-called canonical class loss function of his type of time series analysis (Van Buuren, 1988, p. 133).

A new development in k-sets analysis is filtering the contribution of the sets (in terms of variance accounted for) to get a more stable or better interpretable solution than in OVERALS. Nierop (1989) works with this idea and uses the name set component analysis. He shows that Set Component Analysis has OVERALS as a special case.

1.c.4. Links to other forms of multiway analysis

OVERALS is related to other three-way techniques that generalize principal component analysis. We can compare it, for instance, to the class of three-way models discussed by Kroonenberg (1983). There are some important differences, however. In order to explore them, we now think of the model that says that the OVERALS loss for a particular set of variables is zero. This implies that there exists quantifications $\boldsymbol{Q}_j$, scores $\boldsymbol{X}$, and weights $\boldsymbol{A}_j$, such that $\boldsymbol{Q}_j\boldsymbol{A}_j = \boldsymbol{X}$ for all j (see (4)). Thus we can quantify our variables in such a way that the spaces L_j spanned by the columns of the $\boldsymbol{Q}_j$ have a subspace of dimension p in common, in other words we can quantify our variables such that the dimensionality of the intersection of the L_j is at least p.

If we compare this with the IDIOSCAL or TUCKER2 model, then there we have $\boldsymbol{Q}_j = \boldsymbol{X}\boldsymbol{W}_j\boldsymbol{Y}'$, which has a column rank p and a row rank s (thus $\boldsymbol{X}$ is $n \times p$ and $\boldsymbol{Y}$ is $k \times s$, all sets have k variables). In this formulation the TUCKER2 model is obtained from the OVERALS model by requiring that $\boldsymbol{A}_j = \boldsymbol{A}\boldsymbol{U}_j$ (with $\boldsymbol{A} = \boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{Y})^{-1}$ and $\boldsymbol{U}_j = \boldsymbol{W}_j^{-1}$), i.e. the OVERALS weights themselves satisfy another (smaller) OVERALS model as well. Because in the TUCKER2 situation we often have the same variables in all sets, we can impose further restrictions that also restrict the dual (row) space of our data matrices. Thus the OVERALS model is more general than the TUCKER2 model, both algebraically and from the data analysis point of view (different variables in the sets, and possibility of optimal scaling).

2. An illustration of the characteristics of the computer program

2.a. Description of the algorithm

2.a.1. Goodness-of-fit function

In the foregoing paragraphs the OVERALS-loss has been discussed. The first line of expression (4) shows the loss. This loss is the badness-of-fit for OVERALS. It can be changed into a goodness-of-fit by subtracting it from a constant. From the appendix we can see that the OVERALS-loss can be written as

$$\mathrm{tr}[k\boldsymbol{X}'\boldsymbol{X} - k\boldsymbol{X}'\boldsymbol{P}^*\boldsymbol{X}], \tag{5}$$

where $\boldsymbol{P}^* = \sum_{j=1}^{k} \boldsymbol{P}_j/k$ is a function of the data transformations. Let us denote this by $\boldsymbol{P}^*(\boldsymbol{Q})$. As $\mathrm{tr}[\boldsymbol{X}'\boldsymbol{X}]$ is equal to np, the goodness-of-fit is:

$$knp - \text{loss} = \mathrm{tr}[k\boldsymbol{X}'\boldsymbol{P}^*(\boldsymbol{Q})\boldsymbol{X}] = kn\sum_{i=1}^{p}\mu_i. \tag{6}$$

Thus μ_i corresponds with the fit for the ith solution. This means that the eigenvalues of $\boldsymbol{P}^*(\boldsymbol{Q})$ are the OVERALS-goodness-of-fit measures. The maximum of the eigenvalues is one and the minimum is zero. The sum of the eigenvalues is maximally p and minimally zero. This sum is called FIT in the output of the OVERALS computer program (mark that capitals are used for the

statistics in the computer program output). In addition, $(1-\mu_t)$ is called the MEAN LOSS over sets (per dimension) in the output. The μ's are referred to as EIGENVALUES. The general MEAN LOSS (summed over dimensions) or LOSS corresponds to (p – sum EIGENVALUES) or (p – FIT).

The loss for the solution can also be patitioned over sets. In that case we have for $j = 1, \ldots, k$:

$$\mathrm{tr}\left[\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{P}_j(\boldsymbol{Q})\boldsymbol{X}\right]. \tag{7}$$

This expression summed over sets gives (5). If (7) is divided by n the maximum is p and minimum is zero. Then

$$\mathrm{tr}\left[\boldsymbol{X}'\boldsymbol{X}/n - \boldsymbol{X}'\boldsymbol{P}_j(\boldsymbol{Q})\boldsymbol{X}/n\right] \tag{8}$$

gives the LOSS PER SET, which can even be partitioned over dimensions by computing successivily the diagonal elements of the matrix between square brackets. Averaging (8) over sets gives the MEAN LOSS, i.e. (p – FIT).

The SINGLE FIT for a variable $\boldsymbol{h}_t$ (a column of $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_k$) is defined as

$$\boldsymbol{a}_t'\boldsymbol{a}_t, \tag{9}$$

where the vector $\boldsymbol{a}_t'$ is the row-vector of the weight matrix $\boldsymbol{A}_j$ that corresponds to this particular variable. In paragraph 1.b.2. $\boldsymbol{a}_j$ was defined as a column of $\boldsymbol{A}_j$. We do not use this definition anymore, so that we can redefine $\boldsymbol{a}$. If the squared elements of $\boldsymbol{a}_t$ are taken separately, the contribution of this variable for each dimension is obtained.

If a variable is treated as multiple nominal (thus p single nominal copies), there is much more freedom in scaling the variable than in case of single nominal, ordinal, or numerical. Consequently we can consider the fit if the variable is scaled in this way. This gives the MULTIPLE FIT. When the difference is taken between the MULTIPLE FIT and the SINGLE FIT, we get the SINGLE LOSS. Thus

$$\text{SINGLE LOSS} = \text{MULTIPLE FIT} - \text{SINGLE FIT}. \tag{10}$$

The statistics in (10) can be used to diagnose the qualities of a variable in the analysis. For a mathematical definition of MULTIPLE FIT and SINGLE LOSS, we refer to the appendix (A11 to A13).

2.a.2. Optimization method

The method used to perform the optimization is the so-called Alternating Least Squares method. As we have seen in paragraph 1.b.2 the OVERALS model (4) is formulated as a least squares problem. There are two types of parameters in (4). The first type of parameters is due to the canonical correlation model, i.e. $\boldsymbol{X}$ and $\boldsymbol{A}_j$. The second type of parameters is due to the transformations, namely $\boldsymbol{Q}_j$. These parameters are called scaling parameters. The OVERALS problem is solved for the different sets of parameters alternately, while the other parameters are kept at a constant level. Thus if we know the scaling parameters and the weights, we can solve $\boldsymbol{X}$. Then keep the new $\boldsymbol{X}$ and the $\boldsymbol{Q}_j$ at a constant level,

and compute A_j. The last step is to update the Q_j, while keeping the new X and new A_j at a constant level. This corresponds to one iteration cycle. The iteration process stops if the difference between two consecutive losses is small enough.

2.a.3. Parameter solutions

The solutions for the different parameters are described in detail by Van der Burg (1988, p. 108) and Van der Burg et al. (1988). We discuss the solutions only very briefly here. The parameters X are solved by minimizing (4) with respect to X, this gives:

$$kXM = (I - uu'/n) \sum_{j=1}^{k} Q_j A_j. \tag{11}$$

The vector u consists of n ones. The matrix M is a symmetric matrix of Langrange multiplyers and I is the $(n \times n)$-identity matrix. If the right hand side of (11) is denoted by Z, the matrix $M^2 = Z'Z/n$. Thus to update X first the matrix Z is computed, then M, and next M^{-1} with the help of a routine for eigenvalue decomposition. This solves X as $X = ZM^{-1}$.

The weight parameters could be solved by normal regression, however, we then need the inverse of $(Q_j'Q_j)$, which is a m dimensional eigenvalue problem. To avoid this, the weights are also solved with an alternating least squares algorithm. The product Q_jA_j can always be written as a sum of products of the columns of Q_j and the rows of A_j, i.e. $Q_jA_j = \sum_{t \in I_j} q_t a_t'$. I_j is the set of indices for the variables of set j. If the variables are numbered from 1 to km, $I_j = \{(j-1)m + 1, \ldots, jm\}$. This decomposition can be made for every matrix Q_jA_j. If the loss of (4) is partitioned over sets, the loss part for set j corresponds to

$$\mathrm{SSQ}(X - Q_jA_j) = \mathrm{SSQ}(V_{tj} - q_t a_t'), \tag{12}$$

with $V_{tj} = X - \sum_{s \in I_j - t} q_s a_s'$. $I_j - t$ is the index set I_j without element t. The right hand side of (12) can be minimized over a_t by normal regression. In this case $(q_t'q_t)$ is a scalar which can be easily inverted. Thus

$$a_t = V_{tj}'q_t(q_t'q_t)^{-1}. \tag{13}$$

For every set a_t $(t \in I_j)$ is solved successivily, keeping the other weights at a constant level.

Minimizing (12) over q_t together with the restrictions on q_t mentioned in (4) gives

$$\min \mathrm{SSQ}(V_{tj} - q_t a_t'), \tag{14}$$

with $q_t \in C_t$ and q_t unit-normalized. Van der Burg et al. (1988) show that q_t is solved in three steps. First by normal regression, which gives q_1. Then by linear or monotone regression of h_t on q_1 for numerical and ordinal variables respectivily (q_2), and next by unit-normalizing (q_3). For nominal variables no second step is required. The q_3 is the new update for q_t. Thus

I $\quad q_1 = V_{tj}a_t(a_t'a_t)^{-1}$

II q_2 = linear regression on q_1 for numerical variables
q_2 = monotone regression on q_1 for ordinal variables
$q_2 = q_1$ for nominal variables
III $q_t = q_3$ = unit-normalized q_2 (15)

For every set the q_t are solved successivily, keeping the other q's at a constant level. The monotone regression that is used in the OVERALS program corresponds to Kruskal's secondary approach (Kruskal, 1964). This means that ties, i.e. equal observations within a column of the data matrix, remain tied in the scaling. The primary approach to ties, where untying of ties is possible, is not included in the OVERALS program.

2.a.4. Initial configuration

Two options are possible for the initial configuration of the scaling parameters Q_j. One possibility is a random start configuration. The other possibility is a numerical solution for all the single variables (i.e. variables with single nominal, ordinal or numerical measurement levels) and a multiple nominal solution for multiple variables (variables with a multiple nominal measurement level).

If the numerical/multiple nominal option for the initial configuration is used, the OVERALS computer program solves the problem in the usual alternating least squares way, using two steps. In the first step the single variables start with unit-normalized values and the multiple variables with scores zero. After convergence of the numerical/multiple problem, the OVERALS problem is solved for other measurement levels, using the numerical/multiple nominal solution as start values.

The object scores X always start with a random configuration. The first values of the weights are based on the initial values of X and Q_j, thus no initial values are necessary for the A_j.

2.a.5. Local minima, degeneracies and convergence problems

The OVERALS problem can be translated into an eigenvalue problem, as shown in the Appendix. It concerns the eigenvalue decomposition of matrix P^*. As long as the variables are measured on interval level, the P^* matrix is a constant. However, if single nominal or ordinal restrictions are used the matrix P^* depends on the transformations of the data matrix, thus $P^*(Q)$. For constant P^* the OVERALS iteration process converges to a well defined minimum. However, if the matrix P^* is no longer constant, local minima may occur.

For multiple nominal variables the same holds as for numerical variables. In that case P^* (defined in a slightly different way) is a constant too. We do not explain this here. A discussion on the multiple nominal case is given in Van der Burg et al. (1988).

Very little research is done into the seriousness of the local minima. Using the numerical option for the initial values of variables measured on the ordinal or single nominal measurement level, seems to avoid local minima. This idea is

based more on experience with the OVERALS program, than on systematical investigation.

The OVERALS loss as defined in (4) should go down in every iteration step, independently from the measurement levels used in the program. Sometimes however we find the loss going up. The computer precision must be considered as the reason for this phenomenon.

2.b. Characteristics of the program

2.b.1. Software package and memory requirements

The OVERALS program is included in a software package. It is contained in SPSS for main frames, as well as in SPSS/PC + . It is available in the form of an optional extension named SPSS Categories. In addition, an SPSS version with the extension Categories exists for the Macintosh. All the advantages (and disadvantages) of SPSS also hold for OVERALS For instance, a system file can be used as input, but also raw data. Labels can be given to variables, recoding can be done easily, etc. SPSS has a special user's guide for the extension Categories. It is a small booklet written clearly in typical SPSS style (SPSS, 1990).

The SPSS/PC + version needs 640 K to run OVERALS. To increase workspace 64 K of expanded memory can be used. To install SPSS/PC + (version 3.1 is needed) a hard disk is required. At least 4.5 MB of disk space must be available. (The entire SPSS/PC + system with all of its options and the tutorial requires approximately 14 MB). The program runs. on IBM PC/XT, PC/AT, PS/2, or closely compatible machines. Version 2.0 or later of PC/DOS or MS/DOS is needed and at least one 5.25- or 3.5-inch floppy disk drive.

For the Macintosh version (Mac System 6.0 or higher) 2 MB of memory and a hard drive are necessary. If one wants to run SPSS concurrently under Multi-Finder (in order to do multitasking) 4 MB is needed. For installation of all the SPSS facilities approximately 15 MB of disk space must be available. The Macintoshes suitable for SPSS are Mac Plus, Mac SE, or Mac II series.

2.b.2. Data format and data manipulation

All possibilities for data manipulation that are provided by SPSS may be used for OVERALS data. Therefore it is easy to recode a variable, or to compute a new variable. Specially for continuous variables this recode option is important, as the OVERALS program expects discrete data, i.e. data with not too many different scores per variable. These data manipulations have to be performed before starting the OVERALS analysis.

Apart from being discrete, the data for the OVERALS program have to consist of positive integers. The user has to specify the maximum value for each variable. All negative and zero values or values higher than the maximum are considered as missing values. Fractional values are truncated after the decimal and included in the analysis.

As seen in paragraph 1.b.2. the OVERALS program can handle several types of transformations of the variables using optimal scaling. As the optimal scaling is part of the OVERALS model, we do not consider this as data manipulation.

If a variable is considered as multiple nominal, the program produces the p copies and the p transformations automatically. These transformations are printed in the output as one p-dimensional transformation. Thus the p single transformations are weighted for each dimension and summed over copies. In the Appendix this multiple transformation is referred to as $\boldsymbol{Y}_t$, a $(n \times p)$ matrix. As $\boldsymbol{Y}_t$ has only as many different rows as there are different scores or categories, this matrix is printed in compact form under the name multiple category coordinates.

If an object misses a score on one variable, the scores on the other variables in the same set are also considered as missing for that object (see 1.b.3.). Then all these scores do not contribute to the loss. This seems to be a data manipulation, however it is due to the OVERALS model with missing observations (Van der Burg, 1988, p. 108).

2.c. Input and output of the OVERALS computer program

2.c.1. Input

The SPSS system has to be informed about the format and the names of the variables. For the OVERALS program the user has to specify:

- the names of the variables (with the maximum score),
- the names of the variables in the analysis and their measurement level. Variables that only occur in the first list can be used for plot facilities. Variables in both lists are used for the analysis.
- the number of sets and the number of variables in each set
- the number of observations
- the number of dimensions
- the type of the initial configuration
- the maximum number of iterations
- the convergence criterion value
- print commands (see 2.c.2.)
- plot commands (see 2.c.2.)
- save command (optional) to add object scores to the SPSS system file. A SPSS system file is a data file with information about the data, coded in a special way for SPSS.
- write commands (only for main frames) (see 2.c.2.)
- the data. The data have to be provided in the form of a table of objects $\times$ variables with sets next to each other. Either a raw data file or a SPSS system file can be used.

2.c.2. Available output

The objects scores $\boldsymbol{X}$ form an orthogonal space on to which variables and objects can be projected. The coordinates of the projections of the variables, called component loadings, correspond with the correlations between transformed variables and object scores. These component loadings can be inter-

preted in the same way as component loadings in PCA. Thus for each transformed variable, a vector can be depicted illustrating its importance with respect to the different dimensions of the solution.

The rows of the matrix X provide the scores for each object. If, for every variable, all scores of objects belonging to the same category are averaged, the so-called centroids are obtained. Projecting the centroids of a variable on its vector, gives the projected centroids. For multiple variables p vectors can be used, however in the OVERALS output only the centroids are shown, which are similar to the Y_t-scores (see Appendix) projected on to the space of object scores.

The scores of the transformed variables in unit-normalized form (i.e. q_t) are called the category quantifications in SPSS. The weighted transformed variables (i.e. $q_t a_t'$) are called the single category coordinates. These values are only computed for single variables, and always printed in compact form (see 2.b.2.). For all variables the multiple coordinates Y_t are computed.

Another way of interpreting the projected centroids is by realizing that the projected centroids are similar to the category quantifications projected on to the space of object scores. If we use c_t, a p-vector, for the component loadings of q_t, the projected centroids correspond with $q_t c_t'$.

The output of the OVERALS program always includes
- a table listing measurement levels of each variable by set
- EIGENVALUES and LOSS values by set by dimension.

Optional print output consists of
- marginal frequencies for the variables in the analysis
- history of iterations, i.e. FIT, LOSS and LOSS difference value for every iteration step
- MULTIPLE FIT, SINGLE FIT and SINGLE LOSS per variable
- category quantifications, projected centroids and centroids
- category quantifications, single and multiple category coordinates
- weights and component loadings.

Optional plot output consists of
- component loadings
- object scores [optionally labeled per variable]
- category quantifications (plotted against the original scores) per variable
- all category coordinates and category coordinates per variable
- all centroids and centroids and projected centroids per variable

Optional write output to a SPSS system file (only for main frames)
- variable names, measurement levels, labels, and set numbers
- original scores per category
- category quantifications
- single category coordinates
- multiple category coordinates

- centroids
- projected centroids
- weights
- component loadings.

3. Application

3.a. Vegetable soups

3.a.1. Description of the data

The data to analyse are from sensory research. Forty vegetable soups were rated on five variables by nineteen trained assessors. It concerns odour, taste and mouthfeel attributes (see Table 1). The scores vary from 0 to 98, the assessors used a line scale to indicate their ratings. A score of 0 indicates that an attribute was not present according to the assessor, a score of 98 means that the attribute is very strongly present in a particular soup. Apart from the attributes, the brand of the soup is known, the package and the type (see Table 1). In this application the odour, taste and mouthfeel attributes are analysed. They form a $40 \times 19 \times 5$ multiway table. The other characteristics are used to interpret the solutions and to identify groups.

3.a.2. Objective of the OVERALS analysis

Originally the vegetable soup data were gathered as part of a larger experiment in which chemical and microscopical data also were collected (Cramwinckel and Van Mazijk-Bokslag, 1990). The object of that experiment was to get information about the contents, the sensory quality and the taste of forty different vegetable soups. The main objective of the secondary analysis reported in this paper is not the quality or taste of the soups but the use of the five attributes by the assessors. It is interesting to know on which attribute(s) they agree mostly and which attributes give rise to confusion. In addition, it is interesting to find

Table 1
Sensory variables of the vegetable soup research

Odour and taste attributes
Spiciness (0 = not,...,98 = much)
Vegetables (0 = not,...,98 = much)
Saltiness (0 = not,...,98 = much)
Mouthfeel attributes
Thickness, jelly-like (0 = not,...,98 = much)
Firmness of vegetables (0 = not,...,98 = much)
Package (concentrated in tin, instant in packet, dried in packet, deep-frozen, ready-to-eat in tin)
Type (ordinary, cream, chinese).

out whether the partition of the soups, according to the ratings, makes sense in the light of the type or package of the soup. In order to try to answer these questions the ratings are analysed with OVERALS, using the assessors as sets, the soups as objects and the attributes as variables. Comparison of a solution with ordinal measurement levels and one with numerical measurement levels, also enables us to investigate whether this kind of data should be analysed under metric or nonmetric assumptions.

3.a.3. Pre-treatment of the data

Two assessors (sets) were removed from the data because they had too many missing values, 17 sets remained. As the variables could take 99 different values, they were categorized into a small number of scores (categories). Some preliminary ordinal analyses with different numbers of categories (seven, six and five were tried) showed an extreme outlier in a three-dimensional solution. Removal of the outlier resulted in a new outlier, removal of this outlier again showed a new outlier. The same was encountered with a two-dimensional solution. This phenomenon seemed due to the fact that many of the higher scores and some of the lower ones occurred only once. To cope with the outliers the number of categories was reduced to just three, resulting in less extreme outliers (they did not vanish). This data reduction appears not too severe, as the numerical and the ordinal analyses show comparable and interpretable results. In a previous OVERALS-analysis a partition of the variables into three categories also resulted in an interpretable solution (Van der Burg and Dijksterhuis, 1989). The three categories used for all attributes, are 1 = scores 0 to 25, 2 = scores 26 to 50 and 3 = scores 51 to 98. The third category was taken larger than the first and second category to take account of the skewness. For four variables one category was rescaled to a lower or higher value, as these four categories still contained only one observation.

To perform a numerical analysis (i.e. with only numerical measurement levels) the data also had to be discretisized, as 99 scores are too much (see 2.b.2.). However, a larger number of categories was used as for the ordinal analysis. The scores were recoded into ten categories, corresponding with ten equal sized intervals (category 1 consists of scores 0 to 9, 2 = scores 10 to 19, ..., 10 = scores 90 to 99). The last category was empty for every variable.

3.a.4. Eigenvalues, fit and loss

A three-dimensional solution is obtained in both the ordinal and the numerical analysis. The eigenvalues of the three-dimensional solutions were not really high (see Table 2). The maximum of an eigenvalue is one and the minimum zero. A FIT (= sum eigenvalues) of approximately 1.7 indicates that the solutions found are not very strong. In general high eigenvalues are needed to correspond with much variance in each set. This gives an indication towards differences between the sets probably due to individual differences in the use of the five attributes. As can be seen from Table 2 the two solutions have rather similar eigenvalues, for the moment it seems that the ordinal solution based on just three categories

Table 2
Eigenvalues and fit of a three-dimensional solution with ordinal and numerical measurement levels

Analysis	No. of categories	Eigenvalues			Fit
		dim1	dim2	dim3	
Ordinal	3	0.684	0.533	0.433	1.650
Numerical	10	0.715	0.510	0.429	1.654

is about as good as a numerical solution based on ten categories. A high number of categories does perhaps provide only superfluous information. To see if this conclusion holds, the loss of both solutions is compared. In addition both the component loadings and the object scores are matched.

3.a.5. Component loadings

In order to see how the attributes were used by the assessors, plots of the component loadings are given for each attribute separately. Thus five plots are made instead of one. Dimension 1 and 2, and dimension 1 and 3 are plotted against each other, resulting in 10 separate plots. In Figure 3 these plots are shown for the ordinal solution. Not all vectors are identified by their set number. However, for interpretation of the figure this will do.

A Procrustes rotation (Peay, 1988) to match the three-dimensional configurations of the component loadings for the ordinal and the numerical solution, revealed an almost identity: 97% of the variance of the two configurations was matched. For this reason the results of the numerical solution are not shown. The high match confirmed the idea that the ordinal and numerical solution provide the same information. The matching was performed by means of the Procrustes-PC program of Dijksterhuis and Van Buuren (1990).

Inspection of Figure 3 shows that the assessors agree very much on the attribute thickness. The first dimension depends mostly on this characteristic. The second dimension is dominated by spiciness and firmness. Assessor 3 and, to a lesser extend, assessor 4 must be mentioned here as persons who do not agree with the other assessors about spiciness. The variables corresponding to the attribute 'vegetables' correlate mostly with the second dimension, although the agreement between the assessors is much less on this attribute than on firmness or spiciness. With respect to vegetables assessors 3 and 12 behave rather exceptionally. The third dimension is, for the larger part, determined by saltiness.

Apparently thickness is the attribute on which the assessors agree mostly, followed by spiciness and firmness, and in the last place by saltiness. It seems that 'vegetables' is one of the most difficult attributes to agree about. Especially assessor 3 must be mentioned as a person who interpretes many attributes differently from the other assessors.

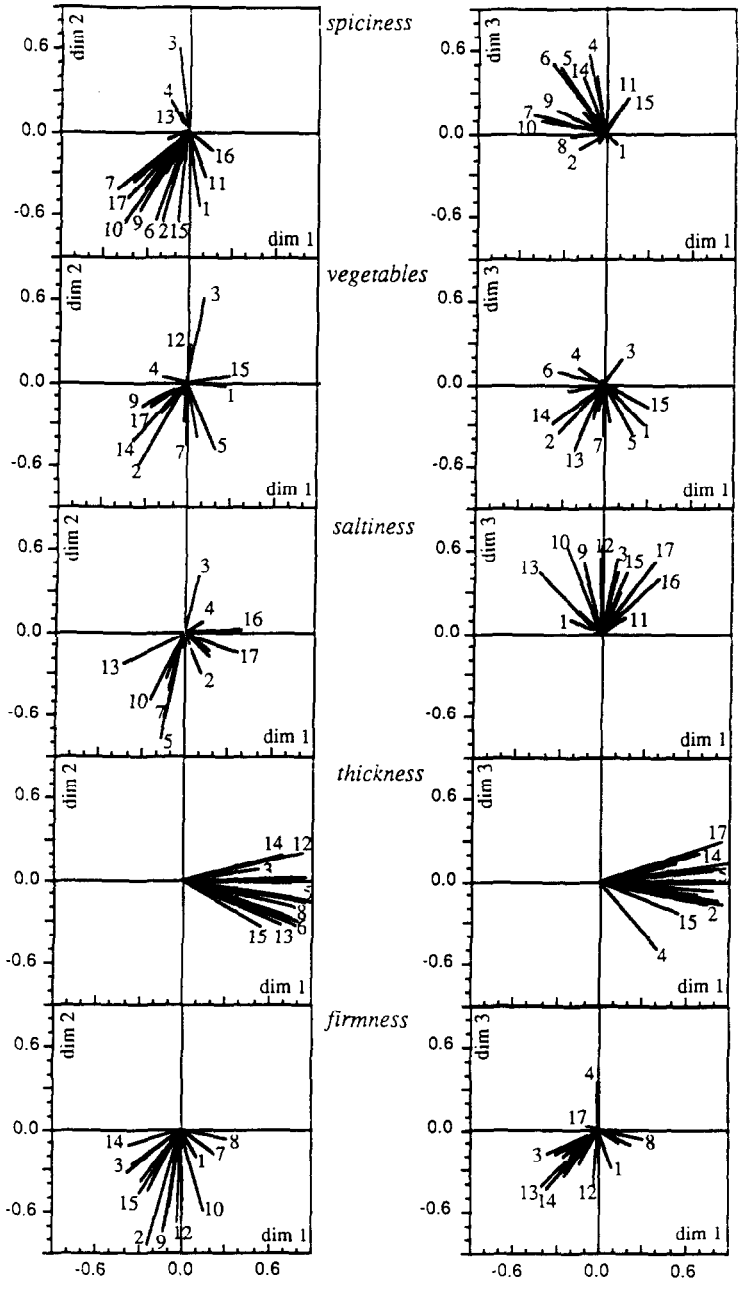

Fig. 3. Component loadings of the ordinal solution for each attribute separately.

3.a.6. Object scores

In Figure 4 the object scores of the ordinal solution are shown. The soups are labeled according to package and type. To check if the ordinal and numerical three-dimensional solutions are alike, again a matching is performed with the help of a Procrustes rotation. This rotation accounts for 98% of the variance in

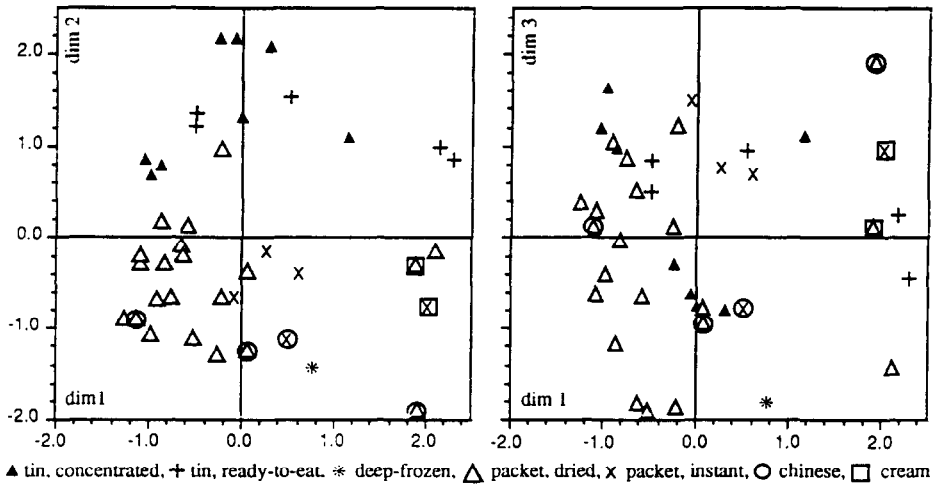

Fig. 4. Object scores of the ordinal solution, labeled by type and package.

the solutions, so only 2% is lost in the matching process. Therefore the solutions are almost identical. Again only the ordinal configuration is shown.

From Figure 4 can be inferred that there does exist an overall difference in taste between soups from tins and from packets (and deep-frozen). This distiction is found along the second dimension, with the tins having higher scores on it than the packets. Since spiciness and firmness were the attributes associated mostly with the second dimension (in negative direction), it seems that soups in tins are less spicy and firm than in packets.

Most of the soups have low scores on the first dimension, six soups have higher scores, they seem more thick than other soups. As may be expected the two cream soups are found between the thicker types. The third dimension is not directly interpretable in terms of package or type of soup. Figure 3 showed that saltiness is the main attribute for the third dimension, so it can be concluded that the judged saltiness of the soups has no clear relation with the package of the soup. The four 'chinese' vegetable soups have low scores on the second dimension in common, so they do posses firm vegetables and are rather spicy. In other respects the chinese soups do not differ much from the other soups. The one deep-frozen soup also seems to have a salty taste and firm vegetables.

Labeling the plot of object scores by brand did not show any regularity. Therefore the names of the soups are not shown in Figure 4.

3.a.7. Conclusion

By analysing k-sets data from the field of sensory research with OVERALS a lot of aspects of the data can be studied. Plotting the component loadings of a variable for all sets, like in Figure 3, provides a useful way to identify the consistency of the use of the variables (in this case attributes) by the sets (assessors in the panel). The relative position of the objects (vegetable soups)

shows the more important (dis)similarities between the objects. The object scores can be compared with the component loadings to see which variables are responsible for congruences or differences between them. In addition, labeling the object scores by external variables (package, type), variables not used in the analysis, also helps to interpret the solution.

It can be concluded from the analysis that it is not useful to analyse the larger number of categories. The results from an ordinal analysis with only three categories and the results from a numerical analysis with ten categories are almost identical. Apparently the higher number of categories does not provide much extra information. An ordinal ten-category solution has not been compared with the numerical solution, as this solution will capitalize on unique patterns in the data.

Appendix

Carroll (1968) solves the following problem

$$\max \sum_{j=1}^{k} \operatorname{cor}(\boldsymbol{x}, \boldsymbol{H}_j\boldsymbol{a}_j)^2. \tag{A1}$$

The solution for $\boldsymbol{a}_j$ is given by normal regression. Thus

$$\boldsymbol{a}_j = (\boldsymbol{H}_j'\boldsymbol{H}_j)^{-1}\boldsymbol{H}_j'\boldsymbol{x}. \tag{A2}$$

Using $\boldsymbol{P}_j = \boldsymbol{H}_j(\boldsymbol{H}_j'\boldsymbol{H}_j)^{-1}\boldsymbol{H}_j'$ gives $\boldsymbol{P}_j\boldsymbol{x} = \boldsymbol{H}_j\boldsymbol{a}_j$. Substituting this in (A1) results in

$$\max \sum_{j=1}^{k} \boldsymbol{x}'\boldsymbol{P}_j\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x}. \tag{A3}$$

Define $\boldsymbol{P}^* = \sum_{j=1}^{k} \boldsymbol{P}_j/k$, then (A3) corresponds to

$$\max k\boldsymbol{x}'\boldsymbol{P}^*\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x}. \tag{A4}$$

Expression (A4) shows that $\boldsymbol{x}$ is proportional to the first eigenvector of $\boldsymbol{P}^*$, and that the maximum is k times the corresponding eigenvalue θ. Thus $\boldsymbol{x}$ satisfies

$$\boldsymbol{P}^*\boldsymbol{x} = \theta\boldsymbol{x}. \tag{A5}$$

The OVERALS-problem for one dimension (2) is

$$\min \sum_{j=1}^{k} \operatorname{SSQ}(\boldsymbol{x} - \boldsymbol{H}_j\boldsymbol{a}_j) \text{ over } \boldsymbol{x} \text{ and } \boldsymbol{a}_j \tag{A6}$$

for $\boldsymbol{u}'\boldsymbol{x} = 0$ and $\boldsymbol{x}'\boldsymbol{x} = n$. The restrictions $\boldsymbol{u}'\boldsymbol{x} = 0$ and $\boldsymbol{x}'\boldsymbol{x} = n$ are similar to the restriction $\boldsymbol{x}$ is unit-normalized. (For $\boldsymbol{u}$ see 2.a.3.) The data matrices are assumed to be unit-normalized (column-wise) too. The solution for $\boldsymbol{a}_j$ is again the regression solution, thus $\boldsymbol{H}_j\boldsymbol{a}_j = \boldsymbol{P}_j\boldsymbol{x}$. Substituting this in (A6) and using the method of undetermined multipliers gives

$$\min \operatorname{tr} k[\boldsymbol{x}'\boldsymbol{x} - 2\boldsymbol{x}'\boldsymbol{P}^*\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{P}^*\boldsymbol{x} - (\boldsymbol{x}'\boldsymbol{x} - n)\phi + \boldsymbol{x}'\boldsymbol{u}\delta], \tag{A7}$$

with ϕ and δ unknown. Solving δ by differentiation of (A7) with respect to $\boldsymbol{x}$, and setting the result to zero, and defining $\mu = 1 - \phi$ gives

$$\boldsymbol{x}\mu = (\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}'/n)\boldsymbol{P}^*\boldsymbol{x}. \tag{A8}$$

As $\boldsymbol{H}_j$ is considered as unit-normalized, the matrix $(\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}'/n)$ (putting the following matrix in deviation from its column means) has no effect on $\boldsymbol{P}^*$. (For $\boldsymbol{I}$ see 2.a.3). Then (A8) turns into:

$$\boldsymbol{P}^*\boldsymbol{x} = \mu\boldsymbol{x}. \tag{A9}$$

Thus $\boldsymbol{x}/\sqrt{n}$ represents the first eigenvector of $\boldsymbol{P}^*$, and $\mu = \theta$. Generalization of (A9) to more solutions gives

$$\boldsymbol{P}^*\boldsymbol{X} = \boldsymbol{X}\boldsymbol{M} \quad \text{with} \quad \boldsymbol{X}'\boldsymbol{X} = n\boldsymbol{I}. \tag{A10}$$

The columns of $\boldsymbol{X}/\sqrt{n}$ represent a rotation of the ordered eigenvectores of $\boldsymbol{P}^*$ and $\boldsymbol{M}$ is the symmetric matrix of Lagrange multipliers with eigenvalues μ_i similar to the eigenvalues of $\boldsymbol{P}^*$. The loss is a function of the sum of the μ's, i.e. loss = $knp\ (1 - \sum_i \mu_i/p)$.

If the variables are measured on interval level $\boldsymbol{P}^*$ is a function of the data, which are constant. However if the data may be transformed $\boldsymbol{P}^*$ is a function of the transformed data, thus $\boldsymbol{P}^*(\boldsymbol{Q})$. In the latter case there is a difference between maximzing the sum of squared eigenvalues (which is the case in CANALS) or the sum of the eigenvalues (OVERALS/CORALS). In the linear case the difference does not exist (Van der Burg and De Leeuw, 1983).

Multiple fit

The quantified variable $\boldsymbol{q}_t$ has $\boldsymbol{a}_t$ as weights for p dimensions. Now if this variable has been considered as multiple nominal, there are p values for $\boldsymbol{q}_t$ (say $\boldsymbol{q}_{t(1)}, \ldots, \boldsymbol{q}_{t(p)}$) and p values for $\boldsymbol{a}_t$ (say $\boldsymbol{a}_{t(1)}, \ldots, \boldsymbol{a}_{t(p)}$). If the weighted quantifications are in the following way:

$$\boldsymbol{Y}_t = \sum_{i=1}^{p} \boldsymbol{q}_{t(i)}\boldsymbol{a}'_{t(i)}, \tag{A11}$$

the multiple nominal quantification for this particular variable is obtained. If the variable were single, we still can compute the $\boldsymbol{Y}_t$. Then we act for a moment as if there are p copies, and compute $\boldsymbol{Y}_t$. The multiple fit for each variable is defined in the following way:

$$\text{MULTIPLE FIT} = \text{tr}\,[\boldsymbol{Y}_t'\boldsymbol{Y}_t]. \tag{A12}$$

As (A12) is formulated for every variable, there are km MULTIPLE FIT's. For each dimension the contribution of a variable to the MULTIPLE FIT can be obtained by computing the elements of the trace separately. The SINGLE LOSS is defined as

$$\text{SINGLE LOSS} = \text{tr}\,[\boldsymbol{Y}_t'\boldsymbol{Y}_t - \boldsymbol{a}_t\boldsymbol{a}_t'] \tag{A13}$$

The expression above holds only for single variables. For a more detailed discussion on multiple nominal transformations we refer to Gifi (1981) and Van der Burg et al. (1988).

Availability of the OVERALS program

The OVERALS program is available as a procedure of the SPSS package. It is part of the optional extension SPSS Categories. This extension consists of programs for conjoint analysis and nonlinear multivariate techniques. The procedures of SPSS Categories are: ORTHOPLAN, PLANCARDS, CONJOINT, ANACOR, HOMALS, PRINCALS and OVERALS.

The address of SPSS in the USA and in Europe is:

SPSS Inc.
444 N. Michigan Avenue
Chicago, Illinois 60611
U.S.A.
Telephone: 312.329.3500

SPSS International B.V.
P.O. Box 115
4200 AC Gorinchem
The Netherlands
Telephone: +31.1830.36711
Telex: 21019 (SPSS NL)
Fax: +31.1830.35839

References

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

Coolen, H. and De Leeuw, J. (1987). Least squares path analysis with optimal scaling. RR-87-03, Leiden: University of Leiden, Department of Data Theory.

Cramwinckel, A.B. and Van Mazijk-Bokslag, D.M. (1989). Chemical, microscopical and sensory research of forty samples of vegetable soup. Internal report (in Dutch). Wageningen: RIKILT Institute.

De Leeuw, J. (1988). Covergence of the majorization method for multidimensional scaling. *Journal of Classification*, **5**, 163–180.

De Leeuw, J. and Van der Burg, E. (1986). The permutational limit distribution of generalized canonical correlations. In: E. Diday, Y. Escoufier, L. Lebart, J.P. Pagès, Y. Schektman and R. Tomassone (Eds.), *Data Analysis and Informatics, IV*, 509–521. Amsterdam: North-Holland.

De Leeuw, J. and Van Rijckevorsel, J.L.A. (1988). Beyond Homogeneity analysis. In: J.L.A. van Rijckevorsel and J. de Leeuw, *Component and correspondence analysis*, 55–80. New York: Wiley.

Dijksterhuis, G.B. and Van Buuren, S. (1989). PROCRUSTES-PC V2.0 Manual. Utrecht: OP&P Sensory Research.

Efron, B. (1982). *The Jackknife, the bootstrap, and other resampling plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, no. 38. Philadelphia: SIAM.

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

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester: Wiley.

Greenacre, M. (1984). *Theory and applications of correspondence analysis.* New York: Academic Press.

Heiser, W.J. (1987). Correspondence analysis with least absulute residuals. *Computational Statistics and Data Analysis*, **5**, 337–356.

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

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

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

Kroonenberg, P.M. (1983). *Three-mode principal component analysis*. Leiden: DSWO Press.

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodnes of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1–27.

Kuhfeld, W.F, Young, F.W. and Kent, D.P. (1987). New developments in psychometric and market research procedures. In: *SUGI Proceedings*, **12**, 111–116. Cary, N.C.: SAS Institute.

Nierop, A. (1989). Generalized set component analysis: a toolbox for multiple sets analysis. RR-89–4. Leiden: University of Leiden, Department of Data Theory.

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

Peay, E.R. (1988). Multidimensional rotation and scaling of configurations to optimal agreement. *Psychometrika*, **53** 2, 199–206.

SPSS (1990). SPSS Categories. Chicago: SPSS Inc.

Ten Berge, J.M.F. (1988). Generalized approaches to the Maxbet problem and the Maxdiff problem, with applications to canonical correlations. Psychometrika, **53**, 487–494.

Van Buuren, S. (1990). Optimal scaling of time series. Leiden: DSWO Press.

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

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

Van der Burg, E. (1988). *Nonlinear canonical correlation and some related techniques*. Leiden: DSWO Press.

Van der Burg, E. and De Leeuw, J. (1983). Non-linear canonical corrlation. *British Journal of Mathematical and Statistical Psychlogy*, **38**, 54–80.

Van der Burg, E. and De Leeuw, J. (1988). Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis. *Applied Stochastic Models and Data Analysis*, vol **4**, 159–172.

Van der Burg, E. and Dijksterhuis, G.B. (1989). Nonlinear canonical correlation of multiway data. In: R. Coppi, and S. Bolasco (Eds.), *Multiway Data Analysis*. Amsterdam: Elsevier Science Publishers.

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

Young, F. (1981). Quantitative analysis of qualitative data. *Psychometrika*, **46**, 347–388.

Young, F, De Leeuw, J. and Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 505–529.

Young, F., Takane, Y. and 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.