

# Generalized canonical correlation analysis with missing values

M. van de Velden\*      Y. Takane†

October 20, 2009

Econometric Institute Report EI 2009-28

## Abstract

Two new methods for dealing with missing values in generalized canonical correlation analysis are introduced. The first approach, which does not require iterations, is a generalization of the Test Equating method available for principal component analysis. In the second approach, missing values are imputed in such a way that the generalized canonical correlation analysis objective function does not increase in subsequent steps. Convergence is achieved when the value of the objective function remains constant. By means of a simulation study, we assess the performance of the new methods. We compare the results with those of two available methods; the missing-data passive method, introduced Gifi's homogeneity analysis framework, and the GENCOM algorithm developed by Green and Carroll.

Keywords: Generalized canonical correlation analysis; Missing values

## 1 Introduction

In canonical correlation analysis (Hotelling, 1936) linear combinations of two sets of variables are obtained in such a way that the correlation between the linear combinations is a maximum. Generalizations to a similar approach for more sets of variables have been the topic of several studies (Horst, 1961; Carroll, 1968; Kettenring, 1971). Consequently, several different approaches have been proposed. Kettenring (1971) provides an overview of four different generalizations. In the framework of homogeneity analysis Van der Burg et al. (1988) and Gifi (1990) introduced nonlinear

---

\*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands [vandevelden@ese.eur.nl](mailto:vandevelden@ese.eur.nl)

†McGill University, Department of Psychology, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada [takane@psych.mcgill.ca](mailto:takane@psych.mcgill.ca)

canonical correlation analysis, also referred to by the algorithm name OVERALS, which takes Carroll's generalized canonical correlation analysis as a special case. Here, we will also use the generalization proposed by Carroll (1968). An excellent description of his method, in a similar notation as we employ in this paper, can be found in Steenkamp et al. (1994) who consider the method for marketing applications. An important advantage of Carroll's approach is its computational ease and the fact that the method takes ordinary canonical correlation analysis as a special case.

In generalized canonical correlation analysis several sets of variables are analyzed simultaneously. This makes the method suited for the analysis of various types of data. For example, in marketing research, subjects may be asked to rate a set of objects on a set of attributes. For each individual, a data matrix can then be constructed where the objects are represented row-wise and the attributes column-wise. Using generalized canonical correlation analysis a graphical representation, sometimes referred to as a perceptual map, can be made on the basis of the individuals' observation matrices.

Generalized canonical correlation analysis does not require the objects to be evaluated on the same set of attributes. That is, each individual can use attributes that he/she finds appropriate. Steenkamp et al. (1994) focussed on this flexibility in their analysis of idiosyncratic sets of attributes. Another type of application, considered by Green and Carroll (1988) concerns the derivation of a composite configuration from a set of configurations. For example, multidimensional scaling solutions (perceptual maps) for the same objects from different countries can be used as input data. Generalized canonical correlation analysis can then be applied to the coordinate matrices to obtain a composite configuration. Finally, generalized canonical correlation analysis can be used when, for the same set of subjects, we have data on sets of variables. For example, consumers may be asked to evaluate several aspects of a certain product separately. One set of variables corresponds, for example, to the product's appearance, another set corresponds to the product's functionality and yet another one concerns variables related to the product's price. Hence, data on three sets of variables are obtained that are to be analyzed simultaneously.

Since generalized canonical correlation analysis deals with possibly large sets of data, the possibility of the occurrence of missing values is significant. Some procedures to deal with missings in generalized canonical correlation analysis have been proposed, however, no attempt has been made to compare and evaluate the alternatives. In this paper, we will review two existing procedures and propose two alternative methods. Furthermore, by means of a simulation study, the performance of the methods will be assessed.

Van der Burg et al. (1988) and Gifi (1990) suggested a method for dealing with missing values in nonlinear generalized canonical correlation analysis in which selection matrices are used to discard complete rows containing at least one missing value. Hence, if one element is missing in a row, the complete row is discarded. This method is often applied in the homogeneity analysis framework as set forth in Gifi (1990). It is referred to as the missing-data-passive approach to missing values. Van de Velden and Bijmolt (2006) used an equivalent method for the case

where complete rows were missing. An advantage of the missing-data-passive approach is its computational ease. The solution can be obtained directly by means of an eigenequation. However, discarding complete rows if only one value is missing, clearly implies a considerable loss of information. Moreover, as the data are centered with respect to the fully observed rows, bias may be introduced. For this latter problem, we propose an alternative approach in which a constant term is estimated separately. The proposed method is a generalization of a method first proposed by Shibayama (1995).

The treatment of missing values in generalized canonical correlation analysis was also studied by Green and Carroll (1988). They proposed an iterative procedure in which missing values are imputed by using linear regression. Although the algorithm in practice appears to converge, there is no guarantee that such will always happen. In this paper, we propose an alternative algorithm in which missing values are imputed in such a way that their contribution to the objective becomes as small as possible. The proposed algorithm is designed in such a way that the generalized canonical correlation objective function monotonically decreases. Hence, the new algorithm always converges.

In a simulation study we compare our methods to those proposed by Green and Carroll (1988) and the missing-data-passive approach proposed by Van der Burg et al. (1988) and Gifi (1990).

The paper is organized as follows. In the next section, we briefly introduce generalized canonical correlation analysis. In Section 3, we briefly summarize the missing-data-passive approach and Green and Carroll's method for the treatment of missing values using notation that will later facilitate easy comparison of the methods. Then, in Section 4, we propose two alternatives to the existing methods. A simulation study in which the methods are compared is presented in Section 5. We conclude the paper with a brief summary of our results.

## 2 Generalized Canonical Correlation Analysis

In generalized canonical correlation analysis linear combinations are obtained in such a way that the sum of squared correlations of the linear combinations of the variables with a so-called group configuration is a maximum. Let  $\mathbf{Y}$  denote the unknown group configuration. The order of  $\mathbf{Y}$  is  $m \times k$ , where  $m$  is the number of rows for each observation matrix  $\mathbf{X}_i$  (i.e. the  $i$ th data set) and  $k$  is the dimensionality of the solution. The data matrices  $\mathbf{X}_i$  are first centered. Sometimes, if the variables are for example measured on different scales, they are also standardized. Note that the sizes of the observation matrices  $\mathbf{X}_i$  are  $m \times p_i$  for  $i = 1, \dots, n$ . The dimensionality of the solution,  $k$ , must be chosen by the researcher.

We can formulate as objective

$$\min \phi(\mathbf{Y}, \mathbf{A}_i) = \min \text{trace} \sum_{i=1}^n (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i)' (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i) \quad (1)$$

subject to the restriction

$$\mathbf{Y}'\mathbf{Y} = \mathbf{I}_k. \quad (2)$$

It is known, e.g. Carroll (1968), that for observed  $\mathbf{X}_i$  matrices, the group configuration  $\mathbf{Y}$  can be obtained from the eigenequation

$$\left( \sum_{i=1}^n \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \right) \mathbf{Y} = \mathbf{Y} \mathbf{\Lambda}, \quad (3)$$

where  $\mathbf{\Lambda}$  is a diagonal matrix with as elements the  $k$  largest eigenvalues of  $\sum_{i=1}^n \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i'$  (where we have assumed that the  $\mathbf{X}_i$ 's are of full column rank) and the matrices  $\mathbf{A}_i$  can be calculated as

$$\mathbf{A}_i = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{Y}. \quad (4)$$

An interesting feature of the method is the fact that the sets of variables  $\mathbf{X}_i$  may contain different variables. Hence, the number of variables in each set does not need to be the same. Steenkamp et al. (1994) used this freedom to analyze object evaluations where each individual used their own set of attributes to evaluate objects.

### 3 Existing methods for dealing with missing values in generalized canonical correlation analysis

There exist two methods specifically designed for dealing with missing values in generalized canonical correlation analysis. Before proposing two new methods, we briefly describe the two existing methods. The notation and formulation that we employ will facilitate comparison of all methods.

#### 3.1 The missing-data-passive approach

In the missing-data-passive approach proposed in the context of nonlinear canonical correlation analysis, rows of the data matrices are removed if they contain one or more missing elements. The generalized canonical correlation approach is then applied by only using the observed rows. This method can easily be implemented by introducing a so-called selection matrix. Let  $\mathbf{K}_i$  denote a diagonal matrix with its diagonals either ones or zeros. The ones correspond to rows for which there are no missings in the  $i$ th observation matrix and the zeros correspond to rows of  $\mathbf{X}_i$ , for which at least one value is missing. Obviously, the resulting selection matrices  $\mathbf{K}_i$  are symmetric idempotent, that is,  $\mathbf{K}_i = \mathbf{K}_i' = \mathbf{K}_i \mathbf{K}_i$ . In the missing-data-passive approach, the data are first centered with respect to the fully observed rows. This centering can be achieved by defining:

$$\mathbf{Q}_i = \left( \mathbf{I} - (\mathbf{1}' \mathbf{K}_i \mathbf{1})^{-1} \mathbf{1} \mathbf{1}' \right) \mathbf{K}_i. \quad (5)$$

Inserting the centering and selection matrices into equation (1), we get

$$\min \phi(\mathbf{Y}, \mathbf{A}_i) = \min \text{trace} \sum_{i=1}^n (\mathbf{Y} - \mathbf{Q}_i \mathbf{X}_i \mathbf{A}_i)' \mathbf{K}_i (\mathbf{Y} - \mathbf{Q}_i \mathbf{X}_i \mathbf{A}_i) \quad (6)$$

which we minimize subject to the restriction

$$\mathbf{Y}' \mathbf{K} \mathbf{Y} = \mathbf{I}_k, \quad (7)$$

where

$$\mathbf{K} = \sum \mathbf{K}_i.$$

It is not difficult to see that the resulting group configuration can be obtained from the eigenequation:

$$\mathbf{K}^{-\frac{1}{2}} \left( \sum_{i=1}^n \mathbf{K}_i \mathbf{Q}_i \mathbf{X}_i (\mathbf{X}_i' \mathbf{Q}_i' \mathbf{K}_i \mathbf{Q}_i \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{Q}_i' \mathbf{K}_i \right) \mathbf{K}^{-\frac{1}{2}} \mathbf{Y}_s = \mathbf{Y}_s \mathbf{\Lambda},$$

where  $\mathbf{\Lambda}$  is the diagonal matrix with as elements the  $k$  largest eigenvalues and we have assumed that the  $\mathbf{X}_i$ 's are of full column rank, and  $\mathbf{Y}_s$  is an  $n \times k$  matrix of corresponding orthonormal eigenvectors. Hence, the appropriately standardized group configuration can be obtained as

$$\mathbf{Y} = \mathbf{K}^{-\frac{1}{2}} \mathbf{Y}_s.$$

The matrices  $\mathbf{A}_i$  can be calculated as

$$\mathbf{A}_i = (\mathbf{X}_i' \mathbf{Q}_i' \mathbf{K}_i \mathbf{Q}_i \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{Q}_i' \mathbf{K}_i \mathbf{Y}.$$

## 3.2 Green and Carroll's GENCOM algorithm

Green and Carroll (1988) proposed an iterative procedure for dealing with missing elements in generalized canonical correlation analysis. For the sake of clarity we will briefly reiterate their method here. The basic principle in their approach, which they call GENCOM, is to estimate the missing values using linear regressions of the variables on the group space.

The GENCOM algorithm can be summarized as follows:

1. For each  $\mathbf{X}_i^*$  calculate  $\hat{\mathbf{X}}_i^{(t)}$  by replacing the missing values by the column averages. Thus, for each column, the average is calculated by summing the observed values and dividing this through the total number of observations in a column.
2. Calculate  $\mathbf{Y}^{(t)}$  by applying generalized canonical correlation analysis to the  $\hat{\mathbf{X}}_i^{(t)}$  matrices, and by then adding a column of ones to the configuration matrix. This column of ones serves to estimate the constant in the linear regression model carried out in the next step.

3. For each column of  $\mathbf{X}_i$ , use ordinary least squares to fit:  $\mathbf{x}_{i(j)}^* = \mathbf{Y}^{(t)*} \mathbf{b}_i^{(t)}$ , where  $\mathbf{x}_{i(j)}^*$  is the  $j$ th column of the original data matrix  $\mathbf{X}_i$  after removing the rows corresponding to missing values for that column, and  $\mathbf{Y}^{(t)*}$  is the matrix of corresponding rows of  $\mathbf{Y}^{(t)}$ . Hence, the regression is based on observed values only and  $\mathbf{b}_i^{(t)} = (\mathbf{Y}^{(t)*'} \mathbf{Y}^{(t)*})^{-1} \mathbf{Y}^{(t)*'} \mathbf{x}_{i(j)}^*$ .
4. Construct  $\mathbf{B}_i = [ \mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_{p_i} ]$  and let  $\mathbf{X}_i^{(t)*} = \mathbf{Y}^{(t)} \mathbf{B}_i^{(t)}$ .
5. Calculate  $\hat{\mathbf{X}}_i^{(t+1)}$  by replacing the missing values of the original  $\mathbf{X}_i^*$  matrix with the corresponding elements of  $\mathbf{X}_i^{(t+1)*}$ , whilst keeping the observed values unaltered.
6. Insert  $\hat{\mathbf{X}}_i^{(t+1)}$  in step 2, and repeat until the differences between two subsequent  $\mathbf{Y}^{(t)}$  matrices becomes smaller than a certain convergence criterion.

Note that, like before, index  $i$  indicates different observation matrices, whereas index  $t$  was used to indicate different iterations.

## 4 Alternative methods for the treatment of missing values in generalized canonical correlation analysis

We now propose two alternative methods to deal with missing values in generalized canonical correlation analysis. The first approach is a generalization of the Test Equating method proposed by Shibayama (1988, 1995) and later described by Takane and Oshima-Takane (2003). As shown by Takane and Oshima-Takane (2003), in the context of principal component analysis, the Test Equating method is closely related to the missing-data-passive approach. However, unlike the missing-data-passive approach, in the Test Equating method, means are approximated separately rather than calculated as the simple mean over observed values. In the generalized canonical correlations framework, this implies that mean vectors for each data matrix are approximated (in a least-squares fashion) rather than calculated by simply taking the means of rows after row-wise deletion.

The second approach that we propose, involves imputation of missing values. Like Green and Carroll's GENCOM algorithm, it requires iterations until convergence has occurred. The missing values in this new approach, are imputed in such a way that their influence on the solution becomes as small as possible.

### 4.1 Test Equating method

The Test Equating method, was proposed by Shibayama (1995) in a one-dimensional setting. However, Takane (1995) showed that the method could easily be extended to a  $k$ -dimensional solution similar to principal component analysis. Moreover, Takane and Oshima-Takane (2003) showed that the Test Equating method is closely related

to the missing-data-passive approach in homogeneity analysis (e.g., Meulman, 1982; Gifi, 1990). The difference between the two methods lies in the estimation of a mean term in the Test Equating method. Here, we further generalize the Test Equating method to the generalized canonical correlation analysis case.

To apply the Test Equating method in generalized canonical correlation analysis, we must employ row-wise deletion similar as was the case in the missing-data-passive approach described in Section 3.1. Hence, if a row contains at least one missing value, the complete row will be removed. However, instead of the centering step employed in the missing-data-passive approach, the Test Equating method requires the estimation of a constant term. Thus, in the Test Equating method, the group configuration is approximated by a constant term plus  $k$  linear combinations of the columns of  $\mathbf{X}_i$ . We can formulate this as follows:

$$\min \phi(\mathbf{Y}, \mathbf{A}_i, \mathbf{a}_{i0}) = \min \text{trace} \sum_{i=1}^n (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i - \mathbf{1} \mathbf{a}'_{i0})' \mathbf{K}_i (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i - \mathbf{1} \mathbf{a}'_{i0}). \quad (8)$$

We can solve this minimization problem sequentially. First, differentiation with respect to  $\mathbf{a}_{i0}$  yields as first order condition,

$$\mathbf{a}_{i0} = (\mathbf{1}' \mathbf{K}_i \mathbf{1})^{-1} \mathbf{1}' \mathbf{K}_i (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i).$$

Substituting this into (8), yields, after some manipulations,

$$\min \phi(\mathbf{Y}, \mathbf{A}_i) = \min \text{trace} \sum_{i=1}^n (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i)' \mathbf{Q}_i' \mathbf{K}_i \mathbf{Q}_i (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i), \quad (9)$$

where  $\mathbf{Q}_i$  is as defined in (5). Let

$$\mathbf{P}_i = \mathbf{Q}_i' \mathbf{K}_i \mathbf{Q}_i,$$

it is easily verified that  $\mathbf{P}_i$  is symmetric idempotent, i.e.  $\mathbf{P}_i = \mathbf{P}_i' = \mathbf{P}_i \mathbf{P}_i$ . Solving (9) subject to the constraint

$$\mathbf{Y}' \mathbf{P} \mathbf{Y} = \mathbf{I} \quad (10)$$

where

$$\mathbf{P} = \sum_{i=1}^n \mathbf{P}_i,$$

yields, assuming for the moment that all inverses exist,

$$\mathbf{P}^{-\frac{1}{2}} \left( \sum_{i=1}^n \mathbf{P}_i \mathbf{X}_i (\mathbf{X}_i' \mathbf{P}_i \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{P}_i \right) \mathbf{P}^{-\frac{1}{2}} \mathbf{Y}_s = \mathbf{Y}_s \mathbf{\Lambda},$$

where  $\mathbf{Y}_s$  is an orthonormal matrix of eigenvectors and  $\mathbf{\Lambda}$  is the corresponding diagonal matrix containing the  $k$ -largest eigenvalues in decreasing order. The solution thus becomes:

$$\mathbf{Y} = \mathbf{P}^{\frac{1}{2}} \mathbf{Y}_s$$

$$\mathbf{A}_i = (\mathbf{X}'_i \mathbf{P}_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{P}_i \mathbf{Y},$$

and

$$\mathbf{a}_{i0} = (\mathbf{1}' \mathbf{K}_i \mathbf{1})^{-1} \mathbf{1}' \mathbf{K}_i (\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i).$$

Comparison with the results in Section 3.1 immediately shows the similarity between the two methods. The only difference concerns the standardization with respect to  $\mathbf{P}$  rather than  $\mathbf{K}$ . This computationally minor difference has considerable consequences. In the missing-data-passive approach, the data are taken into deviation from the mean over observed rows. Such a procedure may not be appropriate if the missing values are not at random. If, for example, missing values are related to certain aspects of a variable, certain values will be under- or over represented. Hence, the mean over the observed values is a biased estimator of the constant term. The Test Equating method does account for this problem. Finally, an additional advantage of the Test Equating method is that the constraint (10), ensures that the group configuration is centered. That is,  $\mathbf{Y}' \mathbf{1} = \mathbf{0}$ . In the missing-data-passive approach, this is not the case.

## 4.2 Minimized contribution approach

Green and Carroll (1988) do not give details on numerical properties of their algorithm. There are, however, two important issues concerning the GENCOM algorithm. First of all, although in each step  $\mathbf{Y}$  and  $\mathbf{B}_i$  are optimal with respect to the imputed  $\mathbf{X}_i$  matrices, there is no mechanism ensuring that subsequent  $\mathbf{Y}'$ s become more similar. That is, convergence is not guaranteed. Secondly, the value  $\phi$  will always be at a minimum for a given set of (imputed)  $\mathbf{X}_i$  matrices. However, there is no mechanism that ensures that this value will go down. Consequently, it may occur that the sum of differences between the group configuration  $\mathbf{Y}$  and the linear combinations of the imputed  $\mathbf{X}_i$  matrices is smaller in the first iteration than in the last iteration. (Obviously, if in subsequent steps the change in the  $\mathbf{X}_i$  matrices is small it is plausible that the change in  $\mathbf{Y}$  is also small. Hence, when the imputed values do not change, i.e. when the regression estimates are "stable" the group configuration is likely to be stable as well.)

To resolve these issues we propose a new algorithm that imputes the missing values of the  $\mathbf{X}_i$  matrices in such a way that the value of the objective function does not increase. Based on these imputed  $\mathbf{X}_i$  matrices a new configuration is calculated in the usual way. Thus, the value of the objective function cannot increase in subsequent steps of the iteration process. The algorithm terminates when the value of the objective function remains constant. Our algorithm resembles an algorithm proposed by Ten Berge et al. (1993) for the treatment of missing values in generalized Procrustes analysis.

The new algorithm that we propose is an alternating least-squares algorithm. The imputed values will be chosen in such a way that their contribution to the objective is minimized. To achieve this, we first solve the usual generalized canonical correlation analysis problem with respect to  $\mathbf{Y}$  and  $\mathbf{A}_i$  whilst considering the  $\mathbf{X}_i$  matrices, in which the missing elements are replaced by some initial values, constant.



Next, we will use the same objective function but this time we minimize with respect to the missing values for  $\mathbf{X}_i$  whilst considering  $\mathbf{Y}$  and  $\mathbf{A}_i$  constant. This process is then repeated until the value of the objective function remains constant. As the value of the objective function cannot increase in subsequent steps, convergence is guaranteed.

Recall objective function (1), where the  $\mathbf{X}_i$  matrices may contain missing elements. We will impute values for the missings in such a way that the value of the objective function decreases in each step. Hence, while keeping  $\mathbf{Y}$  and  $\mathbf{A}_i$  fixed we must minimize  $\phi$  with respect to the missing (to be imputed) elements of  $\mathbf{X}_i$ . This problem has not been solved before.

We can formulate the problem in the following way. Let

$$\mathbf{X}_i = \mathbf{X}_i^o + \mathbf{X}_i^m, \quad (11)$$

where  $\mathbf{X}_i^o$  is the  $m \times p_i$  matrix with the observed values and zeros for the non-observed values. The values of  $\mathbf{X}_i^o$  are constant whereas the entries in  $\mathbf{X}_i^m$  that correspond to missing values are the variables with respect to which we carry out the minimization. The other entries, corresponding to observed values, of  $\mathbf{X}_i^m$  will be ignored. Using (11) we get

$$\mathbf{Y} - \mathbf{X}_i \mathbf{A}_i = \mathbf{Y} - \mathbf{X}_i^o \mathbf{A}_i - \mathbf{X}_i^m \mathbf{A}_i = \mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i$$

so that we can express the objective as

$$\min \phi = \min \sum_{i=1}^n \text{trace} (\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i)' (\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i).$$

We want to minimize this function with respect to the variable elements of  $\mathbf{X}_i^m$ . Clearly

$$\text{trace} (\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i)' (\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i) = \text{vec} (\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i)' \text{vec} (\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i),$$

where the  $\text{vec}$  operator transforms a matrix to a vector by stacking the columns. Using a well known relationship between the  $\text{vec}$  operator and the Kronecker product (e.g. Magnus and Neudecker, 1999) we get

$$\text{vec} (\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i) = \text{vec} (\mathbf{Y}_i^*) - \left( \mathbf{A}_i' \otimes \mathbf{I}_m \right) \text{vec} \mathbf{X}_i^m.$$

The matrix  $\mathbf{X}_i^m$ , and hence its vectorization, contains several elements which correspond to observed values. These elements are of no importance and should be kept constant. By employing a selection matrix  $\mathbf{K}_i$  we select only those elements in  $\mathbf{X}_i^m$  which correspond to the missing values. Let the number of missing values in the original  $\mathbf{X}_i$  matrix be  $q_i$ . A  $q_i \times mp$  matrix  $\mathbf{K}_i$ , whose elements are zero or one, is constructed in such a way that

$$\mathbf{K}_i \text{vec} \mathbf{X}_i^m = \mathbf{x}_i.$$

Hence,  $\mathbf{x}_i$  is a  $q_i \times 1$  vector whose elements we want to determine in such a way that  $\phi$  is minimized. It is not difficult to see that  $\mathbf{K}_i' \mathbf{K}_i \text{vec } \mathbf{X}_i^m = \text{vec } \mathbf{X}_i^m$ , so that

$$\text{vec}(\mathbf{Y}_i^* - \mathbf{X}_i^m \mathbf{A}_i) = \mathbf{y}_i - \mathbf{C}_i \mathbf{x}_i,$$

where  $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i^*)$  and  $\mathbf{C}_i = (\mathbf{A}_i' \otimes \mathbf{I}_m) \mathbf{K}_i'$  and we can express the objective function as

$$\min_{\mathbf{x}_i} \phi = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{C}_i \mathbf{x}_i)' (\mathbf{y}_i - \mathbf{C}_i \mathbf{x}_i).$$

This problem can be solved using matrix differentiation. As first-order condition for  $\mathbf{x}_i$  we get

$$\mathbf{C}_i' \mathbf{y}_i = \mathbf{C}_i' \mathbf{C}_i \mathbf{x}_i. \quad (12)$$

Hence, if  $|\mathbf{C}_i' \mathbf{C}_i| \neq 0$ ,

$$\mathbf{x}_i = (\mathbf{C}_i' \mathbf{C}_i)^{-1} \mathbf{C}_i' \mathbf{y}_i. \quad (13)$$

Moreover, if  $|\mathbf{C}_i' \mathbf{C}_i| = 0$ , a vector  $\mathbf{x}_i$  satisfying the first-order condition (12) may be obtained by replacing the inverse of  $\mathbf{C}_i' \mathbf{C}_i$  by its Moore-Penrose inverse.

The updated  $\mathbf{X}_i$  matrices can be obtained by inserting the  $q_i$  elements of  $\mathbf{x}_i$  in the appropriate places.

The algorithm can be summarized as follows:

1. Replace the non-observed values in the original  $\mathbf{X}_i$  matrices by some initial values, for example, the column averages or zeros.
2. Center the imputed  $\mathbf{X}_i$  matrices.
3. Calculate the generalized canonical correlation analysis solution, i.e. the group configuration  $\mathbf{Y}$  and the value of the objective function  $\psi$ , in the usual way using the imputed  $\mathbf{X}_i$  matrices.
4. Use (13) to calculate the vector with missing values  $\mathbf{x}_i$ , and update the  $\mathbf{X}_i$  matrices accordingly.
5. Go back to step 2 and repeat until the difference between two subsequent values for the objective function  $\psi$  is negligible.

The new algorithm will always converge as the value of the objective function (1) decreases monotonically. It may be possible that the attained minimum is an accumulation point. To avoid this, random starts may be considered.

## 5 Simulation study

To investigate the properties of the two existing, as well as the two new approaches, we conduct a simulation study. In the simulation study, synthetic data are generated for several parameter settings so that the methods can be evaluated under various conditions. To assess the performance of the methods, we consider the measures "variance accounted for" (VAF) and the alienation coefficient. In subsection 5.2 we describe these measures and their functions.

### 5.1 Data generation process:

The data generation process can be summarized as follows:

1. For fixed  $m$  and  $k$ , an  $m \times k$  group configuration  $\mathbf{Y}_{true}$  is constructed by drawing from a standard normal distribution and then calculating an orthogonal base.
2. For each observation matrix we draw an  $m \times k$  (standard normal) random matrix multiplied by a factor  $r = 0.125$ , and add this matrix to  $\mathbf{Y}_{true}$ . The resulting matrix is then post-multiplied by a  $k \times p_i$  (uniform) random matrix to obtain the  $i$ th observation matrix  $\mathbf{X}_i$ .
3. For each  $m \times p_i$  observation matrix  $\mathbf{X}_i$ , we draw a matrix indicating which elements are observed and which are missing.
4. We repeat this process  $n$  times leading to  $n$  "observation" matrices  $\mathbf{X}_i$ .

### 5.2 Evaluation criteria and analysis

After generation of the data sets, we apply the four methods described in this paper. To assess the performance of the methods we consider how well the obtained group configuration is able to describe the original data. Steenkamp et al. (1994), proposed the following measure, which they called variance accounted for (VAF). Select the  $j$ th column of  $\mathbf{X}_i$ , say  $\mathbf{x}_{i(j)}$  and calculate the multiple squared correlation coefficient,  $R^2$ , from the linear regression  $\mathbf{x}_{i(j)}^* = \mathbf{Y}^* \mathbf{b}_j + e_{ij}$ , where the superscripted \* indicates that the rows of  $\mathbf{Y}$  and  $\mathbf{x}_{i(j)}$  corresponding to missing rows (i.e. elements) of  $\mathbf{x}_{i(j)}$  have been removed. Repeat this for all columns of  $\mathbf{X}_i$  and for all data matrices. The VAF is defined as the average of all calculated multiple squared correlation coefficients.

In addition, as the true configuration is known, we can also assess how well the solutions "recover" the true configuration. To do this we compare the Euclidean distances between the rows of the true configuration, with the Euclidean distances between the rows of the retrieved configuration. Let  $\mathbf{T}$  denote the matrix with as elements the Euclidean distances between the rows of the true configuration. The  $ij$ th element of  $\mathbf{T}$  is:  $t_{ij} = \sqrt{(\mathbf{y}_i^{true} - \mathbf{y}_j^{true})' (\mathbf{y}_i^{true} - \mathbf{y}_j^{true})}$ , where  $\mathbf{y}_i^{true}$  is the  $i$ th row of  $\mathbf{Y}_{true}$  written as a  $k \times 1$  column vector. Similarly, let  $\mathbf{O}$  denote the matrix with as elements the Euclidean distances between the rows of the derived

configuration:  $o_{ij} = \sqrt{(\mathbf{y}_i - \mathbf{y}_j)'(\mathbf{y}_i - \mathbf{y}_j)}$ , where  $\mathbf{y}_i$  is the  $i$ th row of the obtained group configuration  $\mathbf{Y}$ , written as a  $k \times 1$  column vector. A congruence coefficient, which measures to which degree the Euclidean distances in the two configurations are similar, may be defined as:

$$c = \frac{\text{trace}(\mathbf{T}'\mathbf{O})}{\sqrt{\text{trace}(\mathbf{T}'\mathbf{T}) \text{trace}(\mathbf{O}'\mathbf{O})}}.$$

The congruence coefficient lies between zero and one, where the maximum is attained when the distances in the two configurations are equal. To allow easier discriminations, Borg and Leutner (1985) introduce the following alienation coefficient:

$$a = \sqrt{(1 - c^2)}.$$

Similar to Bijmolt and Wedel (1999), we will use this alienation coefficient, which may be interpreted as a measure of unexplained variance, to assess how well the true configuration is recovered.

### 5.3 Experimental design

In generating the synthetic data sets, we fix the dimensionality of both the true and approximated group configuration at two. We then vary a number of factors that might affect the performance of the methods. Concerning the number of objects (rows) per matrix we consider two cases: Relatively few rows for each set and relatively many rows for each set. We will treat these two cases separately.

#### 5.3.1 Relatively few rows for each set: $m = 14$

This corresponds, for example, to applications in which a set of objects (the rows) are evaluated using a set of attributes (the columns). The number of attributes are fixed to 4 for this case. (A previous simulation study by Van de Velden and Bijmolt, 2006, which considered the situation in which complete rows were missing, showed little effect of the number of columns). For the number of sets we consider two cases:

1. Few (10) sets. This corresponds to the situation in which, for example, different multidimensional scaling configurations are compared
2. Many (100) sets. This corresponds to the situation in which each data matrix represents an observation matrix for an individual.

For the missing values we consider the following scenarios:

- (a) Completely missing at random (CAR)
- (b) Value dependent missings (VDM)
- (c) Row dependent missings (RDM)

In scenario (a), missings occur completely at random. We consider four such cases with probabilities for missing values, in each observation matrix, equal to 5%, 10%, 20% and 40% respectively. Under scenario (b), the situation in which the probability of values to be missing is directly related to the simulated values. This could, for example, occur when certain true values are less desirable and hence reluctantly reported. For this scenario we consider two cases: the elements corresponding to the highest 3 (approximately 5%) or the highest 6 (approximately 10%) values are missing in all observation matrices. Finally, in scenario (c), one set of objects is more likely to generate missings than the other set regardless of the actual values. From a practical point of view, this situation may be particularly interesting in the setting where rows correspond to objects rather than individuals. For example, in the evaluation of objects, it could occur that certain objects are very well known and thus generate fewer missing values. Again we consider two cases: 1) For the elements of the first 4 rows, the probability for a missing value is 20%, for the remaining 10 rows, this probability is 5%. 2) For the elements of the first 4 rows, the probability for a missing value is 40%, for the remaining 10 rows, this probability is 10%.

**Results:** The results of the simulation study with few rows are presented in Tables 1 through 4. We see that, in general, increasing the number of missings, leads to a decrease in fit. This decrease, however, appears to be stronger for the non-iterative approaches. In particular, with many missing values there is a strong decrease. For the situation in which there are many (100) observation matrices (see Tables 3 and 4) both of these methods remove the entire row if it contains a missing value. Several matrices may therefore contain only few rows that are used to calculate the solution. Furthermore, since the number of rows is small (14), for certain sets there will be a missing value in each row, requiring the removal of the entire matrix. If there are 100 observation matrices, observations for all rows exist, making it possible to calculate a solution. However, when there are only 10 sets, this may no longer be the case and the missing-passive and Test Equating method fail to yield coordinates for all rows. A comparison with the other methods is then no longer possible. For the scenarios with 5%, 10% and 20% of the values missing at random, we ignored those cases and calculated the average VAF and alienation coefficient based on the simulations in which a full solution could be calculated. However, for the missing at random scenario with 40% missings, the number of full solutions was too small to obtain meaningful mean measures. We indicated such cases in Tables 1 and 2 by "N.A." (Not Available).

The simulation results for the case with few observation matrices on relatively few observations are presented in Tables 1 and 2.

In Tables 1 and 3 we see that, with respect to the variance accounted for, the Test Equating method outperforms the Missing Passive method in all cases. Furthermore, as conjectured in Section 4.1, the Test Equating method clearly outperforms the Missing Passive approach when the missings are not random. In particular, when missings are related to the values, the fit of the Test Equating configurations is higher. Moreover, for the value dependent missings, the Test Equating method also

Table 1: Average variation accounted for (VAF). Number of rows 14. Number of observation matrices n=10

Missings:	Missing Passive	Test Equating	Gencom	Min. Contribution
CAR 5%	0.74	0.77	0.77	0.84
CAR 10%	0.64	0.77	0.77	0.79
CAR 20%	0.49	0.68	0.77	0.68
CAR 40%	N.A.	N.A.	0.76	0.62
RDM: [0.2,0.05,0.05]	0.63	0.76	0.77	0.75
RDM: [0.4,0.1,0.1]	0.54	0.68	0.77	0.59
VDM, highest 3	0.71	0.74	0.70	0.67
VDM, highest 6	0.71	0.74	0.70	0.67

Table 2: Average alienation coefficients. Number of rows 14. Number of observation matrices n=10

Missings:	Missing Passive	Test Equating	Gencom	Min. Contribution
CAR 5%	0.13	0.13	0.12	0.13
CAR 10%	0.15	0.15	0.14	0.15
CAR 20%	0.28	0.30	0.16	0.25
CAR 40%	N.A.	N.A.	0.29	0.45
RDM: [0.2,0.05,0.05]	0.18	0.18	0.15	0.22
RDM: [0.4,0.1,0.1]	0.30	0.38	0.21	0.53
VDM, highest 3	0.16	0.17	0.23	0.32
VDM, highest 6	0.16	0.17	0.24	0.32

Table 3: Average variation accounted for (VAF). Number of rows 14. Number of observation matrices n=100

Missings:	Missing Passive	Test Equating	Gencom	Min. Contribution
CAR 5%	0.72	0.76	0.76	0.82
CAR 10%	0.62	0.76	0.76	0.79
CAR 20%	0.48	0.76	0.77	0.70
CAR 40%	0.41	0.41	0.76	0.66
RDM: [0.2,0.05,0.05]	0.62	0.76	0.76	0.77
RDM: [0.4,0.1,0.1]	0.51	0.76	0.77	0.67
VDM, highest 3	0.69	0.74	0.70	0.68
VDM, highest 6	0.69	0.74	0.70	0.68

Table 4: Average alienation coefficient. Number of rows 14. Number of observation matrices n=100

Missings:	Missing Passive	Test Equating	Gencom	Min. Contribution
CAR 5%	0.06	0.06	0.06	0.06
CAR 10%	0.07	0.07	0.06	0.06
CAR 20%	0.10	0.09	0.07	0.08
CAR 40%	0.38	0.57	0.19	0.27
RDM: [0.2,0.05,0.05]	0.08	0.09	0.07	0.10
RDM: [0.4,0.1,0.1]	0.14	0.17	0.10	0.25
VDM, highest 3	0.10	0.10	0.17	0.26
VDM, highest 6	0.10	0.10	0.17	0.26

performs better than both iterative procedures.

Comparison of Tables 1 and 3 reveal that, for all methods, the fit decreases slightly when more sets are involved in the analysis. With fewer data, it becomes easier to fit noise. On the other hand, Tables 2 and 4 show that recovery of the true configuration improves (i.e. the alienation coefficients decrease) when more sets are available.

Only for the completely at random case and few (5%-10%) missings, the Minimized contribution approach seems to work best, both in terms of fit as well as recovery. However, the effect of having more missings, appears to be quite strong for this method. Variance accounted for decreases considerably and alienation coefficients tend to be much higher than those for the other methods.

When there are relatively many row dependent missings, the non-iterative procedures as well as the minimized contribution approach, are outperformed by the GENCOM algorithm. This is especially the case when there are relatively few sets. This is probably caused by the significant loss of information due to the row-wise deletion.

### 5.3.2 Relatively many rows for each set: $m = 100$

This setting corresponds to applications where the rows correspond to cases. The columns represent variables. Each matrix has observations on sets of variables. In a sense, this could be considered the conventional generalized canonical correlation analysis case. For the number of sets we again consider two cases: 4 sets and 8 sets. Furthermore, the number of columns (i.e. variables per set) is also varied in this setting:

1. The number of columns for each set is obtained by drawing from a normal distribution with mean 4 and standard deviation 2, and rounding the number to the nearest integer, with a minimum value of 2. Hence, the expected value for the number of columns is slightly higher than 4:  $E[p] > 4$
2. The number of columns for each set is obtained by drawing from a normal distribution with mean 8 and standard deviation 2, and rounding the number to the nearest integer, with a minimum value of 2. Hence, the expected value for the number of columns is slightly higher than 8:  $E[p] > 8$

For the missing values in this setting we consider the following scenarios:

- (a) Completely missing at random (CAR).
- (b) Value dependent missings (VDM).

For the completely missing at random scenario we consider the same 4 cases as before, i.e. , missings occur with probabilities 5%, 10%, 20% and 40% In the value dependent missings we now consider four scenarios, the highest 1.25%, 2.5%, 5% and 10% of the values are missing. The case where the missings are related to the rows seems less appropriate in this setting and is therefore not studied.

**Results:** The results of the simulation study with sets of  $m = 100$  rows, are presented in Tables 5 through 8. As in the case with few rows, the Missing Passive and Test Equating method, may yield sets that do not contain data on certain rows. Hence, we cannot calculate a solution for the complete configuration. This occurs quite frequently as we have relatively few sets. Again, we indicated this in the Tables by writing "N.A."

In Tables 5 through 8, we see that the variance accounted for in this scenario is generally lower than the situation with few sets whereas the alienation coefficients are quite a bit larger. The most striking difference with the results for the case with few rows, however, is the performance of the minimized contribution approach. Except for the situation in which there are few value dependent missings, this approach yields poor results. Upon closer inspection of the Missing Passive approach, this appears to be caused by a type of degeneracy in which the imputed values (in absolute value) become extremely large. These values dominate the solution as they lead to low values for the objective function. The resulting configurations, however,



Table 5: Average variation accounted for (VAF). Number of rows 100. Number of observation matrices  $n=4$

Missings:	$E[p] >$	Missing Passive		Test Equating		Gencom		Min. Contribution	
		4	8	4	8	4	8	4	8
CAR 5%		0.49	0.43	0.51	0.47	0.52	0.44	0.25	0.12
CAR 10%		0.45	0.39	0.48	0.40	0.52	0.46	0.21	0.07
CAR 20%		N.A.	N.A.	N.A.	N.A.	0.53	0.48	0.24	0.08
CAR 40%		N.A.	N.A.	N.A.	N.A.	0.55	0.49	0.14	0.35
VDM, highest 1.25%		0.51	0.50	0.52	0.51	0.49	0.47	0.44	0.35
VDM, highest 2.5%		0.51	0.49	0.52	0.50	0.49	0.47	0.43	0.35
VDM, highest 5%		0.48	0.46	0.48	0.47	0.47	0.30	0.27	0.10
VDM, highest 10%		0.48	0.46	0.48	0.46	0.47	0.29	0.27	0.10

Table 6: Average alienation coefficients. Number of rows 100. Number of observation matrices  $n=4$

Missings:	$E[p] >$	Missing Passive		Test Equating		Gencom		Min. Contribution	
		4	8	4	8	4	8	4	8
CAR 5%		0.39	0.43	0.39	0.43	0.41	0.49	0.78	0.87
CAR 10%		0.43	0.46	0.43	0.46	0.42	0.45	0.88	0.96
CAR 20%		N.A.	N.A.	N.A.	N.A.	0.44	0.45	0.95	0.84
CAR 40%		N.A.	N.A.	N.A.	N.A.	0.49	0.48	0.78	0.54
VDM, highest 1.25%		0.37	0.38	0.37	0.38	0.45	0.47	0.55	0.70
VDM, highest 2.5%		0.37	0.38	0.37	0.38	0.44	0.47	0.56	0.70
VDM, highest 5%		0.40	0.41	0.40	0.41	0.47	0.67	0.86	0.91
VDM, highest 10%		0.40	0.41	0.40	0.41	0.47	0.68	0.86	0.91

Table 7: Average variation accounted for (VAF). Number of rows 100. Number of observation matrices  $n=8$

Missings	$E[p] >$	Missing Passive		Test Equating		Gencom		Min. Contribution	
		4	8	4	8	4	8	4	8
CAR 5%		0.42	0.36	0.46	0.43	0.47	0.44	0.32	0.15
CAR 10%		0.37	0.28	0.44	0.37	0.47	0.44	0.21	0.10
CAR 20%		0.29	N.A.	0.37	N.A.	0.47	0.45	0.27	0.11
CAR 40%		N.A.	N.A.	N.A.	N.A.	0.48	0.45	0.16	0.27
VDM, highest 1.25%		0.43	0.42	0.45	0.44	0.44	0.43	0.37	0.29
VDM, highest 2.5%		0.43	0.42	0.45	0.44	0.44	0.43	0.37	0.29
VDM, highest 5%		0.39	0.38	0.39	0.41	0.42	0.30	0.24	0.10
VDM, highest 10%		0.39	0.38	0.39	0.41	0.42	0.31	0.24	0.10

Table 8: Average alienation coefficients. Number of rows 100. Number of observation matrices  $n=8$

Missings	$E[p] >$	Missing Passive		Test Equating		Gencom		Min. Contribution	
		4	8	4	8	4	8	4	8
CAR 5%		0.30	0.34	0.30	0.34	0.30	0.34	0.57	0.82
CAR 10%		0.34	0.42	0.34	0.42	0.32	0.34	0.82	0.94
CAR 20%		0.46	N.A.	0.45	N.A.	0.33	0.34	0.94	0.93
CAR 40%		N.A.	N.A.	N.A.	N.A.	0.40	0.38	0.90	0.53
VDM, highest 1.25%		0.29	0.30	0.29	0.30	0.34	0.35	0.50	0.69
VDM, highest 2.5%		0.29	0.30	0.29	0.30	0.34	0.36	0.51	0.69
VDM, highest 5%		0.32	0.33	0.32	0.33	0.35	0.57	0.86	0.91
VDM, highest 10%		0.32	0.33	0.32	0.33	0.36	0.56	0.86	0.91

are quite poor in terms of variance accounted for and recovery (as measured by the alienation coefficient).

In this setting, the Test Equation method does not appear to perform better than the Missing-data-passive approach and both methods perform similar as the GENCOM algorithm. The reason for this is that the mean of the observed rows (as done in Missing Passive) may be a reasonable estimate as it is based on a fairly large number of rows. Even when missings are value dependent, the overall effect becomes small due to the relatively large remaining number of observed rows. On the other hand, it should be noted that in this setting, where we typically only consider relatively few sets, the Missing Passive and Test Equating method often fail to yield a solution when the amount of missings increases.

## 6 Summary and conclusions

Generalized canonical correlation analysis is a mathematically simple, yet versatile technique with potential applications in many fields of research. In generalized canonical correlation analysis, linear combinations of sets of variables are obtained in such a way that the sum of squared distances between the linear combination and an overall group configuration becomes minimal. When the data sets contain

missing values, two procedures exist: Missing-data-passive, in which rows for which a missing value exists, are removed from the data, and GENCOM, an iterative approach proposed by Green and Carroll (1988), where missing values are imputed based on linear regression estimates. In this paper, we introduced two new methods for dealing with missing values in generalized canonical correlation analysis. The first approach, the Test Equating method, does not require iterations. Like the missing-data-passive method, it removes rows that contain missing values. In the Test Equating method, a constant term is estimated whereas the missing-data-passive method considers the data by removing the mean of the observed rows. When missings do not occur completely at random, the latter procedure yields biased results.

In the second new approach, the minimized contribution approach, missing values are imputed in a such a way that the generalized canonical correlation analysis objective function is minimized. Unlike the missing-data-passive and Test Equating method, no data is discarded in this method. Instead, an iterative procedure is employed to obtain the optimal values.

For an appraisal of the existing and new methods, we conducted a simulation study in which various parameters were varied. The simulation study was designed to mimic two quite different types of data that can be analyzed by generalized canonical correlation analysis: 1) Few cases per set. 2) Many cases per set (and relatively few sets). It was found that, when there are few cases per set, performance of the four methods is similar. However, when few values are missing completely at random, the minimized contribution approach performs best. On the other hand, when missings are related to the actual values, the Test Equating method performs best. For applications with many rows and few sets, it was found that the imputation methods are outperformed by the noniterative methods. Moreover, the new minimized contribution method performed poorly. The poor performance appears to be caused by a type of degeneracy in which extremely large (in absolute value) numbers are imputed for the missings. These large imputed value dominate the solution and lead to a "low" value of the objective. However, the quality of such solutions, as measured by variance accounted for (calculated using only the observed values) is low and the fit with respect to the true underlying configuration, as measured by the alienation coefficient, is poor. In addition, in this setting with many rows and few sets, differences between the Missing Passive and Test Equating methods are rather small, with a slight advantage for the Test Equating method. The situation does not change much when more sets are introduced.

In light of the outcomes of the simulation study as well as the computational aspects of the methods, we conclude that the Test Equating method is perhaps the best choice for dealing with missing values in generalized canonical correlation analysis. Only when we have few cases, many sets and few (<5%) missings completely at random, the minimized contribution approach should perhaps be considered as alternative. Finally, it should be noted that in some cases, in particular those with relatively many missings and few sets, the row-wise deletion used in the Missing-data-passive and Test Equating method may prevent these methods from finding a solution for all rows. Thus, a choice should be made between the minimized contri-

bution approach and the Gencom algorithm. Given the generally poor performance of the minimized contribution approach, we suggest using the GENCOM algorithm in such cases.

## References

- Bijmolt, T. H., Wedel, M., 1999. A comparison of multidimensional scaling methods for perceptual mapping. *Journal of Marketing Research* 36, 277–285.
- Borg, I., Leutner, D., 1985. Measuring the similarity between MDS configurations. *Multivariate Behavioral Research* 20, 325–334.
- Carroll, J. D., 1968. Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the American Psychological Association*, 227–228.
- Gifi, A., 1990. *Nonlinear Multivariate Analysis*. Wiley, Chichester.
- Green, P. E., Carroll, J. D., 1988. A simple procedure for finding a composite of several multidimensional scaling solutions. *Journal of the Academy of Marketing Science* 16, 25–35.
- Horst, P., 1961. Relations among  $m$  sets of measures. *Psychometrika* 26, 129–149.
- Hotelling, H., 1936. Relations between two sets of variates. *Biometrika* 28, 321–377.
- Kettenring, J. R., 1971. Canonical analysis of several sets of variables. *Biometrika* 58, 433–451.
- Magnus, J., Neudecker, H., 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, Chichester.
- Meulman, J., 1982. *Homogeneity Analysis of Incomplete Data*. DSWO Press, Leiden.
- Shibayama, T., 1988. Kessokuchi o fukumu testo sukoa no tahenryoukaiseki (multivariate analysis of test scores with missing data)., unpublished Doctoral Dissertation, Faculty of Education, University (in Japanese).
- Shibayama, T., 1995. A linear composite method for test scores with missing values. *Niigata daigaku kyouikugakubu kiyou (Memoirs of the Faculty of Education, Niigata University)* 36, 445–455.
- Steenkamp, J.-B. E. M., Van Trijp, H. C. M., Ten Berge, J. M. F., 1994. Perceptual mapping based on idiosyncratic sets of attributes. *Journal of Marketing Research* 31, 15–27.
- Takane, Y., 1995. *Seiyakutsuki Shuseibunbunsekihou (Constrained Principal Component Analysis)*. Asakurashoten, Tokyo.

- Takane, Y., Oshima-Takane, Y., 2003. Relationships between two methods for dealing with missing data in principal component analysis. *Behaviormetrika* 30, 145–154.
- Ten Berge, J. M. F., Kiers, H. A. L., Commandeur, J. J. F., 1993. Orthogonal procrustes rotation for matrices with missing values. *British Journal of Mathematical and Statistical Psychology* 46, 119–134.
- Van de Velden, M., Bijmolt, T. H., 2006. Generalized canonical correlation analysis of matrices with missing rows: A simulation study. *Psychometrika* 71, 323–331.
- Van der Burg, E., De Leeuw, J., Verdegaal, R., 1988. Homogeneity analysis with  $k$  sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika* 53, 177–197.