CHAPTER 2

THE METHODS UNDER INVESTIGATION

2.1 INTRODUCTION

Two types of methods, used for the simultaneous analysis of two or more groups (three SCA–methods and one ML–method), will be presented separately, starting with the SCA–methods. To ease the presentation, the following definitions are useful:

p = number of groups,
n_k = number of individuals in group k, with $N = \sum_{k=1}^{p} n_k$,
m = number of variables,
q = number of components ($q<m$),
$\mathbf{X}_k$ = $n_k \times m$ matrix with deviation scores in group k,
$\mathbf{Z}_k$ = $n_k \times m$ matrix with standard scores in group k,
$\mathbf{C}_k$ = covariance matrix of $\mathbf{X}_k$,
$\mathbf{R}_k$ = correlation matrix of $\mathbf{X}_k$, $\mathbf{R}_k = n_k^{-1}\mathbf{Z}_k'\mathbf{Z}_k$,
$\mathbf{W}_k$ = weights matrix ($m \times q$) for group k,
$\mathbf{P}_k$ = pattern matrix ($m \times q$) for group k,
$\mathbf{S}_k$ = structure matrix ($m \times q$) for group k.

For the factor analysis methods, additional definitions will be presented separately to avoid confusion.

The SCA–methods will be introduced using standard scores, while the ML–method will be introduced using deviation scores. This is done because this is the way the methods are usually presented.

2.2 SIMULTANEOUS COMPONENTS ANALYSIS METHODS

Because the SCA–methods are all generalizations of PCA, first a joint formulation of PCA, when performed on p groups separately, will be

given. When one has two or more populations, separate PCA's can be performed on each group, by minimizing the function

$$f_{sep}(\mathbf{W}_1,\ldots,\mathbf{W}_p,\mathbf{P}_1,\ldots,\mathbf{P}_p) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}_k\mathbf{P}_k' \|^2, \tag{2.1}$$

for a given q ($<m$). This function will serve as a basis for the discussion of the SCA-methods in the sequel.

2.2.1 SCA-W

In their search for a generalization of PCA, Millsap and Meredith (1988) started from the position that PCA is essentially a method that generates linear combinations (components) of the variables that explain as much variance as possible in these variables. Furthermore, they paired this position with the concept of stationary (invariant) component weights. If the same component weights are used in each group, and the variables are the same for each group, then the components have the same meaning in each group, provided that the weights matrix is used for interpretation of the components. The resulting method, called SCA-W, (Simultaneous Components Analysis with an invariant Weights matrix), generalizes PCA by computing one matrix of weights, defining components with the same meaning across groups, such that the total amount of variance explained by these components is a maximum. Millsap and Meredith (1988) suggested a gradient method to find the optimal weights in SCA-W, but this method appeared to have certain disadvantages, which can be avoided by adopting an alternating least squares algorithm (ALS) for SCA-W, as proposed by Kiers and Ten Berge (1989) and implemented by Kiers (1990).

Starting from function (2.1), the step towards SCA-W is very small. For each population the same weights matrix $\mathbf{W}$ is used, so SCA-W minimizes

$$f_w(\mathbf{W},\mathbf{P}_1,\ldots,\mathbf{P}_p) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}\mathbf{P}_k' \|^2. \tag{2.2}$$

It should be noted that the single weights matrix will, in general, not lead to standardized component scores in all groups. In case standardization is desired, one should employ the relaxed constraint that

the weights matrices must be columnwise proportional. This, however, leads to the same optimization problem, and is therefore not discussed any further.

2.2.2 SCA-P

In search of a generalization of PCA, Levin (1966) capitalized on the property that PCA produces a loadings matrix **A** such that **AA'** resembles the correlation matrix of the variables. Accordingly, he suggested a generalized method based on carrying out a PCA on the average of the correlation matrices.

Levin's idea for a generalization of PCA to two or more groups will now be related to the idea of keeping the pattern matrix **P** fixed across groups. In the definition of SCA-P, the solutions for the different groups are constrained in such a way that the pattern matrices are equal, or equivalently, perfectly congruent columnwise across all groups. Thus, SCA-P minimizes

$$f_p(\mathbf{W}_1,...,\mathbf{W}_p,\mathbf{P}) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}_k\mathbf{P}' \|^2, \tag{2.3}$$

over $\mathbf{W}_1,...,\mathbf{W}_p$ and **P**. Function (2.3) reaches a minimum when the matrix **P** ($m \times q$) contains the first q eigenvectors of the correlation matrix of $(\mathbf{Z}_1'|\cdots\cdots|\mathbf{Z}_p')'$, see Kiers and Ten Berge (1994a). They also showed that, in case all groups have the same size n, this solution is the same as the one found with the method proposed by Levin (1966). In fact, Levin (1966) minimizes the function

$$g(\mathbf{P}) = \| \bar{\mathbf{R}} - \mathbf{P}\mathbf{P}' \|^2, \tag{2.4}$$

in which the unweighted mean of the correlation matrices is used, while SCA-P uses the correlation matrix of $(\mathbf{Z}_1'|\cdots\cdots|\mathbf{Z}_p')'$, which is the weighted mean of the correlation matrices. Finally, it is worthwhile noting that at the minimum of function (2.3), $\mathbf{W}_1=\mathbf{W}_2=...=\mathbf{W}_p=\mathbf{W}=\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}$, so instead of (2.3) we minimize

$$f_p(\mathbf{W},\mathbf{P}) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}\mathbf{P}' \|^2, \tag{2.5}$$

so SCA-P could also be named SCA-WP, because both the weights and the pattern matrix are kept equal over groups (Kiers & Ten Berge, 1994a). However, we will refer to this method as SCA-P.

2.2.3 SCA-S

In SCA-S, the explained variance is maximized over solutions in which the structure matrices are perfectly congruent columnwise across all groups. The method SCA-S can thus be viewed as a method that requires components to have, across groups, the same direction in the space of the variables.

The structure matrix for group k is defined as $\mathbf{S}_k = \mathbf{R}_k\mathbf{W}_k$. If the columns of the matrix $\mathbf{S}_k$ are required to be congruent columnwise across all groups, the matrix $\mathbf{S}_k$ should equal $\mathbf{SD}_k$ for a certain matrix $\mathbf{S}$ and for a certain diagonal matrix $\mathbf{D}_k$, $k=1,...,p$. Therefore, SCA-S can be described mathematically as the method that minimizes function (2.1), subject to the constraint that

$$\mathbf{W}_k = \mathbf{R}_k^{-1}\mathbf{SD}_k, \tag{2.6}$$

for certain matrices $\mathbf{S}$ and (diagonal) $\mathbf{D}_k$, assuming that $\mathbf{R}_k$ is nonsingular, $k=1,...,p$. Rewriting function (2.1), using constraint (2.6), gives

$$f_s(\mathbf{S},\mathbf{D}_1,...,\mathbf{D}_p,\mathbf{P}_1,...,\mathbf{P}_p) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{R}_k^{-1}\mathbf{SD}_k\mathbf{P}_k' \|^2. \tag{2.7}$$

Kiers and Ten Berge (1994a) have presented an ALS algorithm that minimizes function (2.7).

2.3 A SIMULTANEOUS FACTOR ANALYSIS METHOD

2.3.1 Simultaneous Factor Analysis in Several Populations, using a Maximum Likelihood estimation procedure (SIFASP-ML)

Jöreskog (1971) proposed a factor analysis model for the

simultaneous analysis of two or more groups. Let $\mathbf{x}_k = (x_{k1},...,x_{km})'$ be a vector of order $m\times1$, representing the variable scores in group k, $k=1,...,p$. Jöreskog regards $\mathbf{x}_k$ as a random vector with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ $(m\times m)$. It is assumed that a factor analysis model holds in each population, so that $\mathbf{x}_k$ can be accounted for by q common factors $\boldsymbol{\xi}_k = (\xi_{k1},...,\xi_{kq})'$ and m unique factors $\boldsymbol{\delta}_k = (\delta_{k1},...,\delta_{km})'$, as

$$\mathbf{x}_k = \boldsymbol{\mu}_k + \boldsymbol{\Lambda}_k\boldsymbol{\xi}_k + \boldsymbol{\delta}_k, \tag{2.8}$$

with $\mathcal{E}(\boldsymbol{\xi}_k)=0$, $\mathcal{E}(\boldsymbol{\delta}_k)=0$ and where $\boldsymbol{\Lambda}_k$ is a matrix of factor loadings $(m\times q)$. The usual factor analysis assumptions then imply that

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k\boldsymbol{\Lambda}_k' + \boldsymbol{\Theta}_k, \tag{2.9}$$

where $\boldsymbol{\Phi}_k$ is the $q\times q$ factor covariance matrix (covariance matrix of $\boldsymbol{\xi}_k$) and $\boldsymbol{\Theta}_k$ is the $m\times m$ diagonal covariance matrix of unique factors $\boldsymbol{\delta}_k$ in group k, $k=1,...,p$. In the sequel, only deviation scores will be considered. Therefore, the vector $\boldsymbol{\mu}_k$ is specified to be the null vector.

In simultaneous factor analysis covariance matrices of two or more groups are analyzed. This means that, for each set of two or more groups, three matrices have to be estimated: $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$, $k=1,...,p$. The simultaneous factor analysis model may specify that certain parameters in the matrices $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$ have assigned values (fixed parameters) and that some set of unknown elements in the matrices $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$ are the same for all k (equality constraints). The most common situation, which is the one investigated in the present study, is when the whole factor pattern $\boldsymbol{\Lambda}_k$ is assumed to be invariant across groups. So, for the present study, the model described in (2.9) is modified to

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}\boldsymbol{\Phi}_k\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_k. \tag{2.10}$$

Let $\mathbf{C}_k$ be the sample covariance matrix of group k with n_k-1 degrees of freedom. If we assume that $\mathbf{x}_k$ has a multinormal distribution it follows that the matrix $\mathbf{C}_k$ has a Wishart distribution with covariance matrix $\boldsymbol{\Sigma}_k$ and n_k degrees of freedom. The maximum likelihood estimates for the matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$, $k=1,...,p$ are obtained by minimizing the function

$$F = \sum_{k=1}^{p} \left(\frac{n_k}{N} \right) F_{ML}(\mathbf{C}_k, \mathbf{\Sigma}_k), \tag{2.11}$$

where F_{ML} is the log likelihood function

$$F_{ML}(\mathbf{C}_k, \mathbf{\Sigma}_k(\theta_j)) = \log|\mathbf{\Sigma}_k(\theta_j)| + \mathrm{tr}(\mathbf{C}_k \mathbf{\Sigma}_k(\theta_j)^{-1}) - \log|\mathbf{C}_k| - m, \tag{2.12}$$

where the matrix $\mathbf{\Sigma}_k(\theta_j)$ is the estimate of the sample covariance matrix $\mathbf{C}_k$ using model parameters θ_j. In order to fit the model to the data in the ML–sense it is usually necessary that the model is identified. How to make sure that the model is identified will be discussed in the next subsection.

2.3.2 Identification of parameters

Suppose that the matrix $\mathbf{\Lambda}$ in the model, presented in equation (2.10), is replaced by a matrix $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{T}^{-1}$ and $\mathbf{\Phi}_k$ is replaced by $\mathbf{\Phi}_k^* = \mathbf{T}\mathbf{\Phi}_k\mathbf{T}'$, $k=1,\ldots,p$, where $\mathbf{T}$ is an arbitrary nonsingular matrix of order $q \times q$. Then the estimated matrix $\hat{\mathbf{\Sigma}}_k$ remains the same so that function (2.11) is unaltered. Because the matrix $\mathbf{T}$ has q^2 independent elements, at least q^2 independent conditions must be imposed on the parameters in the matrices $\mathbf{\Lambda}$ and $\mathbf{\Phi}_k$, $k=1,\ldots,p$, to make these uniquely identified. To ensure identification (in the present situation), in each column of the loadings matrix $\mathbf{\Lambda}$ (of order $m \times q$), $q-1$ elements are set to zero and one element is given the value of one. The ones in the columns must be put on different rows. Specifically, this means that a $q \times q$ block of the matrix $\mathbf{\Lambda}$ is defined to be the identity matrix (ID–block). The rows of the ID–block may be placed arbitrarily among the rows of the matrix $\mathbf{\Lambda}$. It must be noted, however, that some choices of the position of the rows of the ID–block in the matrix $\mathbf{\Lambda}$ may lead to suboptimal solutions of SIFASP–ML. In the present experiments, the rows of the ID–block are spread across the matrix $\mathbf{\Lambda}$. Further attention will be given to this topic in the description of the first experiment in which SIFASP–ML is used (Chapter 6), when discussing specific identification constraints. The remaining elements of the matrix $\mathbf{\Lambda}$ are completely free. The matrix $\mathbf{\Phi}_k$ is restricted to be symmetrical and the matrix $\mathbf{\Theta}_k$ is set to be diagonal in each group.

In SIFASP–ML, the identification constraints are merely employed to ensure that the ML fitting procedure works. Once a solution is obtained, we leave it free to rotate to any more favorable position.

In the present study, Jöreskog and Sörbom's structural equation modeling program LISREL 8 (1993) will be used to estimate the parameters in the SIFASP model, using the maximum likelihood estimation method (SIFASP–ML).

2.4 ROTATIONS USED TO FACILITATE INTERPRETATION OF THE SOLUTIONS, FOUND WITH THE METHODS OF ANALYSIS

The methods SCA–P, SCA–W, SCA–S and SIFASP–ML are, just as PCA, insensitive to both orthogonal and oblique rotations (nonsingular transformations) of the pattern, weights (not defined for SIFASP–ML) and structure matrices. For the four different methods, three different rotational techniques are used.

The solutions, found by the methods SCA–P and SIFASP–ML, are rotated using the Varimax rotation (Section 2.4.1). Note that the solutions, found by the methods SCA–P and SIFASP–ML, both contain one pattern matrix and p structure matrices. The Varimax rotated pattern matrix contains as many near zero and large elements as possible, and as few as possible elements in between. Applying this rotation to the pattern matrix, found by the methods SCA–P and SIFASP–ML, therefore generally increases the interpretability of the resulting factors, compared to the unrotated factors. The columns of the pattern matrix are rescaled to unit length before and after rotation, which makes the complete procedure the Harris and Kaiser's (1964) Independent Cluster rotation (HKIC, see Kiers & Ten Berge, 1994b).

To consider the rotation, used for SCA–W, it must firstly be noted that comparison of the solutions, found by the different methods, with the true model will primarily be based on the pattern and the structure matrices found, and *not* on the weights matrix. The reasons for not using the weights matrix in the comparison will be explained in Chapter 3. For

the solution, found by SCA–W, one may think a straightforward procedure to apply would also be the HKIC rotation, because this procedure can be seen as a method to rotate the pattern and the weights matrix simultaneously to a simple pattern and simple weights, respectively (Kiers & Ten Berge, 1994b). However, this only holds when the HKIC is applied to a single group. When applied to the solution found by SCA–W, that is, one weights matrix and p pattern matrices, the p pattern matrices are not rotated to a simple pattern. Because the pattern and the structure matrices of the solutions, found with the different methods of analysis, will be used for comparing the methods of analysis, and *not* the weights matrix, this rotation is not satisfactory in the present context.

In order to give SCA–W an opportunity that is comparable to the opportunity given to the other methods for simplifying the solution found, it is necessary to use a rotational procedure that rotates all the pattern matrices, found by SCA–W, to simple form, instead of the weights matrix. A rotational procedure that accomplishes this was proposed by Hakstian (1976, see Section 2.4.2).

As far as SCA–S is concerned, one may apply the Varimax rotation to the structure matrix, because this rotation, among other things, aims at finding some large structure elements, which facilitates interpretation. However, interpretation is facilitated even more by having components that are strongly related to certain non–overlapping subsets of variables. Therefore, in the present study, a rotational method that accomplishes this, devised by Kiers and Ten Berge (1994a), was chosen for the rotation of the structure matrix, found by SCA–S (Section 2.4.3).

2.4.1 HKIC Rotation of the SCA–P and the SIFASP–ML solution

For SCA–P, a Varimax rotation is performed on the normalized pattern matrix, found by minimizing (2.3), using a matrix $\mathbf{T}_V$, resulting in a rotated pattern matrix $\mathbf{P}_V = \mathbf{P}(\text{diag}(\mathbf{P}'\mathbf{P}))^{-\frac{1}{2}}\mathbf{T}_V$. To arrive at a correctly scaled pattern matrix for each group, $\mathbf{P}_V$ is multiplied by a diagonal matrix $\mathbf{D}_k = (\text{diag}(\mathbf{W}_V'\mathbf{R}_k\mathbf{W}_V))^{\frac{1}{2}}$, where $\mathbf{W}_V$ is the rotated weights matrix $\mathbf{W}_V = \mathbf{W}\mathbf{T}_V$, which in turn is rescaled for each group as $\mathbf{W}_k = \mathbf{W}_V\mathbf{D}_k^{-1}$. The rotated

structure and component correlation matrices for each group are then calculated as $\mathbf{S}_k = \mathbf{R}_k\mathbf{W}_k$ and $\mathbf{\Phi}_k = \mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k$, respectively, the usual expressions for obliquely rotated components (see Section 1.2).

For all solutions found with SIFASP-ML we have

$$\hat{\mathbf{\Sigma}}_k = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Phi}}_k\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Theta}}_k. \tag{2.13}$$

A Varimax rotation is performed on the orthonormalized pattern matrix $\hat{\mathbf{\Lambda}}(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{-\frac{1}{2}}$, using a matrix $\mathbf{T}_V$. Next, the associated factor covariance matrix is scaled to unit diagonal by pre- and postmultiplication with $\mathbf{D}_k = (\text{diag}(\mathbf{T}_V'(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\hat{\mathbf{\Phi}}_k(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\mathbf{T}_V))^{-\frac{1}{2}}$. The resulting matrix with factor correlations is given as $\hat{\mathbf{\Phi}}_k^* = \mathbf{D}_k\mathbf{T}_V'(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\hat{\mathbf{\Phi}}_k(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\mathbf{T}_V\mathbf{D}_k$ and the associated rotated pattern matrix is $\hat{\mathbf{\Lambda}}_k^* = \hat{\mathbf{\Lambda}}(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{-\frac{1}{2}}\mathbf{T}_V\mathbf{D}_k^{-1}$. The rotated structure matrix, containing correlations between the variables and the factors, for the SIFASP-ML solution is then calculated as $\hat{\mathbf{S}}_k^* = \hat{\mathbf{\Lambda}}_k^*\hat{\mathbf{\Phi}}_k^*$.

2.4.2 Hakstian's Rotation of the SCA-W solution

Hakstian's rotation procedure maximizes the sum of Varimax functions for the p pattern matrices, using the same rotation matrix for all p pattern matrices. Before the maximization, the supermatrix $\mathbf{P}_o$, containing the p pattern matrices one below the other, is made columnwise orthonormal as a form of standardization of the method. Hakstian's rotation method applied to such a supermatrix is a direct generalization of the HKIC procedure. Let $\mathbf{T}_H$ be the rotation matrix. Then $\mathbf{P}_H = \mathbf{P}_o\mathbf{T}_H$ is the supermatrix containing all rotated pattern matrices one below the other. The rotated weights matrix, structure matrix and component correlation matrix are then calculated as $\mathbf{W}_H = \mathbf{W}\mathbf{T}_H\mathbf{D}_H$, $\mathbf{S}_k = \mathbf{R}_k\mathbf{W}_H$ and $\mathbf{\Phi}_k = \mathbf{W}_H'\mathbf{R}_k\mathbf{W}_H$, respectively, with $\mathbf{D}_H = (\text{diag}(\mathbf{W}_H'\mathbf{R}_k\mathbf{W}_H))^{-\frac{1}{2}}$.

2.4.3 Simple Structure Rotation of the SCA-S solution

For SCA-S, the rotational freedom is exploited in a simple structure rotation technique, designed by Kiers and Ten Berge (1994a). The idea is to find a transformation matrix $\mathbf{T}_S$ that yields an optimal number of high correlations of the components with certain subsets of variables, across

all sets. To achieve this, the sums of certain subsets of squared elements of the rotated structure matrix are maximized, searching for non–overlapping subsets of variables, each associated with a different component. Kiers and Ten Berge (1994) have solved the maximization problem by using a monotonically convergent ALS algorithm. The rotated structure matrix is given as $\mathbf{S}_S = \mathbf{S}\mathbf{T}_S$. Using a diagonal matrix $\mathbf{D}_k = (\text{diag}(\mathbf{S}_S'\mathbf{R}_k^{-1}\mathbf{S}_S))^{-\frac{1}{2}}$, the rotated weights matrix, pattern matrix and component correlation matrix are then calculated as $\mathbf{W}_k = \mathbf{R}_k^{-1}\mathbf{S}_S\mathbf{D}_k$, $\mathbf{P}_k = \mathbf{S}_S(\mathbf{S}_S'\mathbf{R}_k^{-1}\mathbf{S}_S)^{-1}\mathbf{D}_k^{-1}$ and $\mathbf{\Phi}_k = \mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k$, respectively.

2.5. MEASURES OF FIT FOR THE SCA–METHODS AND SIFASP–ML

The fit measures used in the SCA–methods and the ML–methods, used in the present study, are not comparable, because the variants of the two types of methods currently used pertain to completely different aspects of fit. The SCA–methods presented maximize the amount of explained variance of the variables. Consequently, the amount of explained variance of the variables is used as a measure of fit. The ML–methods presented, however, try to maximize the likelihood function of the parameters in a postulated model, or equivalently, minimize the chi–square value, which is directly calculated from the likelihood function (2.11).

For each SCA–method, the calculation of the variance explained by each component separately and the variance explained in each group will be presented. The amount of explained variance is defined as "The total amount of variance to be explained minus the residual variance". For SIFASP–ML, the chi–square measure is presented. Other fit measures for SIFASP–ML will be presented in Section 4.3, together with the descriptions of how these measures will be used as indicators of the optimal dimension.

2.5.1 Explained variances of the SCA–W solutions

Because the matrix $\mathbf{S}_k$ contains correlations between variables and

components, on the diagonal of $\mathbf{S}_k'\mathbf{S}_k$ are the variances explained by each component separately in group k. The explained variances per component, averaged over groups, are on the diagonal of $\sum_{k=1}^{p} n_k(\mathbf{S}_k'\mathbf{S}_k)/N$. Using (1.2), the total explained variance in group k is given by

$$\mathrm{EV}_W = \mathrm{tr}(\mathbf{W}'\mathbf{R}_k^2\mathbf{W}(\mathbf{W}'\mathbf{R}_k\mathbf{W})^{-1}) = \mathrm{tr}(\mathbf{P}_k\mathbf{\Phi}_k\mathbf{P}_k') = \mathrm{tr}(\mathbf{S}_k\mathbf{\Phi}_k^{-1}\mathbf{S}_k'). \tag{2.14}$$

The total amount of explained variance, averaged over groups is given by $\sum_{k=1}^{p} n_k \mathrm{tr}(\mathbf{S}_k\mathbf{\Phi}_k^{-1}\mathbf{S}_k')/N$.

2.5.2 Explained variances of the SCA-P solutions

Let $\mathbf{K}\Lambda\mathbf{K}' = \sum_{k=1}^{p} n_k\mathbf{R}_k/N$ be an eigendecomposition. Then, in SCA-P we have $\mathbf{P} = \mathbf{K}_q$ and hence $\mathbf{W} = \mathbf{K}_q$. Furthermore, let $\mathbf{k}_j$ be the j^{th} column of the matrix $\mathbf{K}$, and λ_j the j^{th} eigenvalue. The variance, explained by the j^{th} component in group k is calculated as $\mathbf{k}_j'\mathbf{R}_k\mathbf{k}_j$. This is because the fitted data can be written as $\mathbf{Z}_k\mathbf{K}_q\mathbf{K}_q' = \mathbf{Z}_k\mathbf{k}_1\mathbf{k}_1' + \ldots + \mathbf{Z}_k\mathbf{k}_q\mathbf{k}_q'$, and the total explained variance for group k can be written as $n_k^{-1}\|\mathbf{Z}_k\mathbf{K}_q\mathbf{K}_q'\|^2 = n_k^{-1}\|\mathbf{Z}_k\mathbf{k}_1\mathbf{k}_1'\|^2 + \ldots + n_k^{-1}\|\mathbf{Z}_k\mathbf{k}_q\mathbf{k}_q'\|^2 = \mathbf{k}_1'\mathbf{R}_k\mathbf{k}_1 + \ldots + \mathbf{k}_q'\mathbf{R}_k\mathbf{k}_q$. The variance, explained by component j, averaged over groups is given as $(1/N)\sum_{k=1}^{p} n_k\mathbf{k}_j'\mathbf{R}_k\mathbf{k}_j = \lambda_j$. The total amount of explained variance, averaged over groups, is given by $\sum_{j=1}^{q}\lambda_j$.

2.5.3 Explained variances of the SCA-S solutions

When $\mathbf{W}_k$ is chosen such that $\mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k$ has unit diagonal, and hence the matrix $\mathbf{S}_k$ contains correlations between variables and components, on the diagonal of $\mathbf{S}_k'\mathbf{S}_k = \mathbf{W}_k'\mathbf{R}_k^2\mathbf{W}_k$ are the explained variances per component for group k. The explained variance per component, averaged over groups, are on the diagonal of $\sum_{k=1}^{p} n_k\mathrm{diag}(\mathbf{W}_k'\mathbf{R}_k^2\mathbf{W}_k)/N$. The total explained variance for group k is given by (1.2), with $\mathbf{\Phi} = \mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k = \mathbf{S}'\mathbf{R}_k^{-1}\mathbf{S}$, as

$$\mathrm{EV}_S = \mathrm{tr}(\mathbf{S}(\mathbf{S}'\mathbf{R}_k^{-1}\mathbf{S})^{-1}\mathbf{S}'). \tag{2.15}$$

The explained variance, averaged over groups is given by $\sum_{k=1}^{p} n_k \mathrm{tr}(\mathbf{S}(\mathbf{S}'\mathbf{R}_k^{-1}\mathbf{S})^{-1}\mathbf{S}')/N$.

2.5.4 The Chi-Square measure for SIFASP-ML

The χ^2 goodness-of-fit measure is defined as $N-p$ times the minimum of the fit function (2.11). The degrees of freedom are given as

$$df = \tfrac{1}{2}p(m)(m+1) - t, \tag{2.16}$$

where p is the number of groups, m the number of variables per group and t is the total number of independent parameters estimated in the model.

2.6 FIRST COMPARISON OF THE METHODS

2.6.1 Comparisons of PCA-sep, SCA-W, SCA-P and SCA-S

Ten Berge, Kiers and van der Stel (1992) compared SCA-W with the method, consisting of separate PCA's for each group (PCA-sep), and found that the loss in explained variance for SCA-W was relatively small. The maximal difference in percentage of explained variance between the components found with PCA-sep and SCA-W, for the one component solution, was found to be 12.5%. This difference was found in the so-called 'reversed eigenvalue case': Two covariance matrices were constructed, using 2 components, so that they had the same eigenvalues, in opposite order. For covariance matrices, constructed using three and four components (thus constructing covariance matrices of rank 3 and 4, respectively), the maximal difference in the reversed eigenvalue case, for the one dimensional solution, was found to be 10.1% and 8.58%, respectively. In a large number of randomly constructed data, the maximal discrepancies in percentages of explained variance between the methods PCA-sep and SCA-W were recorded for the one dimensional solution. For covariance matrices simulated, using 2, 3 and 4 components, these were 12.0, 8.7 and 6.8%, respectively.

Other interesting results, found by Ten Berge, Kiers and van der Stel (1992), were that SCA-W shows an average positive sampling bias in explained variance (over 14 replications), overestimating the amount of explained variance (in the population) with maximal 2.07%, and that even

at a sample size of n=50, drawn from a population of size n=2283, components were surprisingly well recovered, with average congruence values (over 14 replications) for the two components in the specific population at hand of .97 and .94, respectively.

A very important mathematical relation between the methods PCA–sep, SCA–W, SCA–P and SCA–S was discovered by Kiers and Ten Berge (1994a). They showed that the four methods PCA–sep, SCA–W, SCA–P and SCA–S are related hierarchically in that they account for (weakly) decreasing amounts of variance. Hence, one can view the methods as increasingly restrictive variants of PCA–sep. Kiers and Ten Berge formulated this as the following theorem: Let f_{sep}, f_w, f_p and f_s, for the time being, denote the total amounts of variance explained by optimal q–dimensional solutions found with the methods PCA–sep, SCA–W, SCA–P and SCA–S, respectively. Then, for any set of data $\mathbf{Z}_1,...,\mathbf{Z}_k$,

$$f_{sep} \geq f_w \geq f_p \geq f_s. \tag{2.17}$$

For the proof of this theorem, see Kiers and Ten Berge (1994a).

Ten Berge (1994) has shown that when f_w equals f_p, f_p also equals f_s. Ten Berge also showed that when we have correlation matrices $\mathbf{R}_k$ which have the same q dominant eigenvectors, the three SCA–methods and PCA–sep will explain the same amount of variance when q components are drawn. The remaining number of eigenvectors (ranging from zero to m–q) may be chosen completely different for each correlation matrix.

2.6.2 Comparison of the SCA–methods with SIFASP–ML

The method SIFASP–ML is inferior to the other methods with respect to the specifications necessary to start an analysis: Whereas the SCA–methods do not need any description of components in advance, SIFASP–ML does, in order to ensure identification (see Section 2.3.2). This is a serious drawback of the method, in the sense that one has to know part of the solution to be able to arrive at the solution. In the present study, this means that we have to give (part of) the solution away, in order to be able to test SIFASP–ML. One can, of course, reason that in analyzing scores on a set of variables (taken from two or more

groups) a researcher practically always has an expectation or theory on the basis of which to start the analysis. For this reason, SIFASP–ML is given the headstart it requires. The question whether this headstart is too much of an advantage for SIFASP–ML will be answered negatively.