A General Framework for Multivariate Analysis with Optimal Scaling: The **R** Package aspect

Patrick Mair
WU Wirtschaftsuniversität Wien

Jan de Leeuw
University of California, Los Angeles

Abstract

In a series of papers De Leeuw developed a general framework for multivariate analysis with optimal scaling. The basic idea of optimal scaling is to transform the observed variables (categories) in terms of quantifications. In the approach presented here the multivariate data are collected into a multivariable. An *aspect* of a multivariable is a function that is used to measure how well the multivariable satisfies some criterion. Basically we can think of two different families of aspects which unify many well-known multivariate methods: Correlational aspects based on sums of correlations, eigenvalues and determinants which unify multiple regression, path analysis, correspondence analysis, nonlinear PCA, etc. Non-correlational aspects which linearize bivariate regressions and can be used for SEM preprocessing with categorical data. Additionally, other aspects can be established that do not correspond to classical techniques at all. By means of the R package **aspect** we provide a unified majorization-based implementation of this methodology. Using various data examples we will show the flexibility of this approach and how the optimally scaled results can be represented using graphical tools provided by the package.

Keywords: aspect, lineals, optimal scaling, bilinearizability, R.

1. Introduction

Gifi (1990) offers a comprehensive collection of nonlinear multivariate methods based on optimal scaling. The starting point of the underlying analysis is a 0/1 dummy matrix based on the data which are considered as categorical. Subsequently, a loss function involving the (unknown) object and category scores is established. During the iterations we stretch/squeeze the variables and compute category scores such that they are optimal in the sense of a minimal loss function. This procedure is referred to as *optimal scaling*. The simplest method is *homogeneity analysis* which is also known as *multiple correspondence analysis*. By imposing restrictions on the quantification ranks we get *nonlinear principal component analysis* and

by defining sets *nonlinear multi-set canonical correlation analysis*. More detailed descriptions with a strong computational background and, correspondingly, the presentation of the R (R Development Core Team 2009) package **homals** are given in De Leeuw and Mair (2009a).

In this paper we somewhat extend the Gifi approach in terms of a general system for multivariate analysis which we denote as the *aspect* framework. The theoretical foundation of this approach was already given by De Leeuw (1988b), some of it also in De Leeuw (1982), and further discussed in De Leeuw (1988a), De Leeuw (1993), and De Leeuw, Michailidis, and Wang (1999). Essentially, we can subdivide this framework into two parts: optimization of functions of the correlation matrix on the one hand, and optimization of non-correlational loss functions on the other hand. In the first case we speak of *correlational aspects*. In the second case, where we mainly focus on linearizing regressions (LINEALS), we speak of *non-correlational aspects*. Eventually we get optimally scaled unidimensional category quantifications (scores). These scores and the corresponding correlation matrix, respectively, can be used for further multivariate modeling.

As De Leeuw *et al.* (1999) point out, in many situations (e.g., social and behavioral sciences) we do not know exactly how to express the variables. For instance, we can think of regression examples where we can apply logarithmic, exponential, or square root transformations to the data. In the Gifi (and in our aspect) framework, the transformations are unknown: We scale (i.e., transform or quantify) the categories with respect to a certain criterion. In the particular aspect approach we select an target function, investigate how this function varies over all feasible transformations of the variables, and, finally, quantify the categories accordingly.

In the next sections we start with some basic definitions, present correlational aspects and their optimization by means of majorization. Then, as a particular non-correlational aspect, we focus on LINEALS, discuss implications of this approach and focus on LINEALS as a *structural equation models* (SEM) preprocessing tool for categorical data. The applied part of the paper focuses on the R package **aspect**—available from the Comprehensive R Archive Network at http://CRAN.R-project.org/package=aspect—by means of various examples.

2. Basic definitions and notations

2.1. Cones in n-space: The theoretical framework

Let us assume that our data are collected in a $n \times m$ matrix $H = (h_1, h_2, \ldots, h_m)$ with $i = 1, \ldots n$ and $j = 1, \ldots, m$. From a general perspective, multivariate analysis involves computations on m random variables $H_1, H_2, \ldots, H_m$, which we can collect into a so called *multivariable*. An *aspect* ϕ of a multivariable is a well-defined function that is used to measure how well the multivariable satisfies some desirable criterion (De Leeuw *et al.* 1999, p. 529). We can think of many different criteria; thus, many different aspects can be defined as we will show below.

So far we did not mention that our approach will rescale the categories of each h_j by means of optimal scaling. This implies that we are not only interested in the variables as we observe them but rather in transformations of the observed data (monotonic, polynomial, "splinical"). If one such transformation or re-expression of the variable suits our criterion better than the original expression, we use the transformed version (De Leeuw 1988a).

A convenient mathematical formulation is offered by the concept of convex cones. In general, a convex cone is a subspace of a vector space. Let us assume that the observed h_j and its transformations are in a known closed convex cone $\mathcal{K}_j$, which is a subset of $\mathbb{R}^n$. A set $\mathcal{K}_j$ is called a convex cone iff any non-negative combination of elements from $\mathcal{K}_j$ remains in $\mathcal{K}_j$. In our multivariate setting, we have the convex cones $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$. Various cones are important: $\mathcal{K}$ can be a polyhedral convex cone of monotone transformations, a linear subspace of low-order polynomial transformations, a linear subspace for imputation of missing data, or a whole space for latent variables. A thorough historical discussion of the cone concept in connection with well-known multivariate techniques can be found in De Leeuw (1988b).

2.2. Indicator matrix and category scores: Practical issues

As mentioned above, we emphasize optimal scaling of categorical variables where each variable H_j has number of categories k_j. For each variable we define an indicator or dummy matrix G_j of dimension $n \times k_j$. Within each row i we have one non-zero element indicating in which category of variable j observation i falls. These indicator matrices can be collected in an indicator supermatrix $G = (G_1 | \ldots | G_m)$ of dimension $n \times K$ where $K = \sum_{j=1}^{m} k_j$.

The contingency table of variables j and l is $C_{jl} = G_j^\top G_l$. Similar to G we can define the $K \times K$ supermatrix C which is known as the *Burt matrix*. These notations bring us closer to correspondence analysis (CA) as described in De Leeuw and Mair (2009b). We think of CA and related methods mainly in an analytical way, as opposed to the (more common) geometrical approach.

The crucial part of the whole computation is the determination of the optimally scaled category scores y_j for each variable. The y_j are vectors of length k_j which are normalized, either to $y_j^\top D_j y_j = 1$ or to $y_j^\top D_j y_j = N$ where D_j is the diagonal matrix with the univariate margins on the diagonal (i.e., $D_j = C_{jj}$). Moreover we require $y_j \in \mathcal{K}_j$, where $\mathcal{K}_j$ is a suitable cone in $\mathbb{R}^{k_j}$. Consequently,

$$s_{jl} = y_j^\top C_{jl} y_l \tag{1}$$

are the elements of the covariance matrix $S(\mathbf{Y})$ of the transformed variables. In order to exclude trivial solutions we have to impose normalization. This implies that, rather than using $S(\mathbf{Y})$, for correlational aspects we will use the correlation matrix $R(\mathbf{Y})$ with its elements r_{jl}.

An additional task in optimal scaling are restrictions on the scale levels of the variables. Basically, we regard each variable as categorical. A nominal scale level involves no restrictions on the resulting scores. An ordinal scale level requires monotonicity and numerical variables additionally require equidistance between the scores. For Gifi methods the formulas are given in De Leeuw and Mair (2009a) and, consequently, implemented in the **homals** package. In our aspect framework the task it quite simple. Regardless whether we have correlational or non-correlational aspects, all that is needed to impose ordinal scale levels, is to perform isotone regression within each iteration of the optimization. We use the Pool-Adjacent-Violators Algorithm (PAVA; see e.g. Barlow, Bartholomew, Bremner, and Brunk 1972) which is implemented in the **isotone** package (De Leeuw, Hornik, and Mair 2009). This guarantees monotonically increasing category scores.

3. Correlational aspects and their optimization

3.1. Types of correlational aspects

First we examine aspects directly defined on the correlation matrix. Formally, we study problems of the form $\phi(R(\mathbf{Y}))$. If we have only $m = 2$ variables we only have one correlation coefficient and, therefore, only one aspect we can study: The variation of the correlation coefficient under the choice of category quantification. References and discussions for the $m = 2$ case can be found in De Leeuw (1983b).

Here we focus on the general case of a multivariate dataset with m variables. Basically, the aspect ϕ defines the criterion to be optimized and $\mathbf{Y} = (y_1, y_2, \ldots, y_m)$ with y_j as the vector of category scores to be computed. From a general point of view the optimization problem can be formulated as

$$\phi(R(\mathbf{Y})) \to \max! \tag{2}$$

In other words: We scale the observed variables (categories) in a manner such that an aspect of choice on the correlation matrix, which is based on these scores, is maximized.

Now we specify various aspects based on $R(\mathbf{Y})$ which are referred to as *correlational aspects*. Note that the following listing is just a collection of possible correlational aspects. In fact, the **aspect** package allows the user to specify his own aspects. For the following convex correlational aspects which are pre-specified in the package, we also give the first order derivatives needed for optimization. The aspect specification in the package (by means of the aspect argument; see Section 5) is given in parenthesis.

- *Sum of correlations* (`"aspectSum"`): We can define a very simple aspect by taking the sum of the correlations. Optionally the correlations can be transformed by an even power $q > 1$. Formally, the aspect and its derivative can be expressed as

$$\phi(R(\mathbf{Y})) = \sum_{l<j} (r_{jl})^q \tag{3a}$$

$$\frac{\partial \phi}{\partial r_{jl}} = q(r_{jl})^{q-1} \tag{3b}$$

  where r_{jl} are the elements of the correlation matrix R. Trivially, for $q = 1$ the elements of the matrix of partial derivatives $\partial \phi / \partial R$ are all 1. This special case corresponds to the SUMCOR model proposed by Horst (1961b, 1965). The $q = 2$ case reflects the SSQCOR method by Kettenring (1971).

- *Sum of absolute correlations* (`"aspectAbs"`): This aspect is based on the sum of absolute correlations (again, with the option of including power q which now does not have to be even) and can be expressed as

$$\phi(R(\mathbf{Y})) = \sum_{l<j} |r_{jl}|^q \tag{4a}$$

$$\frac{\partial \phi}{\partial r_{jl}} = q \operatorname{sgn}(r_{jl}) |r_{jl}|^{q-1}. \tag{4b}$$

  Of course, we need to take suitable measures if one of the correlations is zero, because in that case the derivative will not exist.

- *Eigenvalue aspects* (`"aspectEigen"`): The basic definition of an eigenvalue aspect is to maximize the largest eigenvalue λ of R. The finding of such quantifications is the same as finding the dominant eigenvalue in MCA. In literature it was pointed out repeatedly (see e.g. De Leeuw 2006) that the first MCA dimension should be distinguished clearly from the others because the remaining solutions provide additional (suboptimal) solutions of the stationary equations. In other words, maximizing the first eigenvalue implies that we want to scale the variables in a way that they are as uni-dimensional as possible. In the psychometric literature this approach is sometimes referred to as MAXVAR (Horst 1961a, 1965). Further, as a multidimensional generalization, we can look at the sum of the first p eigenvalues

$$\phi(R(\mathbf{Y})) = \sum_{j=1}^{p} \lambda_j(R) = \max_{Z^\top Z = I} tr(Z^\top R Z) \tag{5a}$$

$$\frac{\partial \phi}{\partial R} = Z Z^\top \tag{5b}$$

where Z is the $m \times p$ matrix of the first p eigenvectors. This generalization is used in nonlinear PCA (cf. De Leeuw 2006; De Leeuw and Mair 2009a). The aspect is convex because it is the pointwise maximum of a family of linear functions.

- *Determinant aspects* (`"aspectDeterminant"`): This aspect which is related to the multi-normal negative log-likelihood, corresponds to the GENVAR criterion by Steel (1951). It can be expressed as

$$\phi(R(\mathbf{Y})) = -\log(\det(R)) \tag{6a}$$

$$\frac{\partial \phi}{\partial R} = R^{-1}. \tag{6b}$$

The aspect $\log(\det(R))$ is concave, because it is the minimum over Γ of the linear functions $\log(\det(\Gamma)) + tr(\Gamma^{-1} R)$, and thus its negative is again convex. This aspect can also be defined using covariances instead of correlations (De Leeuw 1988b). The covariance-based version is related to the Box-Cox approach in regression (Box and Cox 1964) and can be used for structural relation models. The latter relationship is studied extensively in Meijerink (1996).

- *Squared multiple correlations* (`"aspectSMC"`): One variable j is defined as target (outcome, response) variable. The squared multiple correlation (SMC) of the $m-1$ remaining variables and its derivative can be written as

$$\phi(R(\mathbf{Y})) = \max_{b_j}(1 - b_{(j)}^\top R b_{(j)}) \tag{7a}$$

$$\frac{\partial \phi}{\partial R} = -b_{(j)} b_{(j)}^\top \tag{7b}$$

where $b_{(j)} = b_1, b_2, \ldots, b_m$ is a vector of weights and restricted to have $b_j = 1$. This apect is the multiple regression aspect (cf. De Leeuw 1986). Again it is convex, because it is a pointwise maximum of linear functions.

- *Sum of squared multiple correlations* (`"aspectSumSMC"`): The previous aspect can be extended in a way such that we compute SMC's between various combinations of indicators and target variables and take the sum of the resulting SMC's. Let us define $\mathcal{J}$

as an arbitrary index set for $j = 1, \ldots, m$. Thus, (7) generalizes to

$$\phi(R(\mathbf{Y})) = \max_{b_j} \sum_{j \in \mathcal{J}} (1 - b_{(j)}^\top R b_{(j)}) \tag{8a}$$

$$\frac{\partial \phi}{\partial R} = \sum_{j \in \mathcal{J}} -b_{(j)} b_{(j)}^\top. \tag{8b}$$

This aspect can be used for path analysis where $\mathcal{J}$ is determined by targets in the path model. Note that it is not necessary for $\mathbf{b}_{(j)}$ to contain all remaining predictors but rather those determining a path to target j. If $\mathcal{J}$ is the full index set and each predictor is regarded in each SMC (in other words we take the sum over all m SMCs), this results in *image analysis* discussed in Guttman (1953). The full index set version is implemented in our package.

These are the correlational aspects pre-specified in the **aspect** package. More aspects can be found in De Leeuw (1988b) and De Leeuw *et al.* (1999). As mentioned above, the package allows for the specification of user-defined aspects by means of functions which must return the function value of $\phi(R(\mathbf{Y}))$ and the derivatives $\partial \phi / \partial R$.

3.2. Optimization using majorization

The aspect optimization problem is tackled by a majorization algorithm (De Leeuw 1994). Here we will just quote briefly the most important results from De Leeuw (1988b) and De Leeuw *et al.* (1999) which are relevant for our aspect framework.

Majorization is a very general optimization approach. It is based on the principle that we have to find a minimum of a complicated function $f(x)$. The idea is to find a simple surrogate function $g(x, y)$, which for fixed y, is larger than $f(x)$: We say that g majorizes f. This surrogate function is then subject of optimization. In our correlational aspect setting $\phi(R(\mathbf{Y}))$ has to be maximized. We have to make sure that ϕ is convex. As we indicated above, most of the interesting aspects fulfill these assumptions (De Leeuw 1988b).

Within each majorization step k we have to update the category scores y_j for a particular variable j at a time, recompute the aspect including it derivatives $\partial \phi / \partial r_{jl}$ and then update y_j for the subsequent variable j. At the optimum, the following stationary equation holds (De Leeuw 1988a):

$$\sum_{l=1}^{m} \frac{\partial \phi}{\partial r_{jl}} C_{jl} y_l = \lambda_j D_j y_j. \tag{9}$$

If the convex aspect is not differentiable we can substitute any subgradient for the partial derivatives. The corresponding Langrange multipliers λ_j are

$$\lambda_j = \sum_{l=1}^{m} \frac{\partial \phi}{\partial r_{jl}} r_{jl}. \tag{10}$$

Based on theorems in De Leeuw (1988b) and De Leeuw *et al.* (1999) the update for the

category quantifications for variable j is given by

$$\widetilde{y}_j^{(k+1)} = \sum_{\substack{l=1 \\ l \neq j}}^{m} \frac{\partial \phi}{\partial r_{jl}} C_{jl} y_l^{(k)}, \tag{11a}$$

$$y_j^{(k+1)} = \text{NORM}_j \left(\text{PROJ}_j \left(\widetilde{y}_j^{(k+1)} \right) \right), \tag{11b}$$

where PROJ_j is the projection on the cone $\mathcal{K}_j$, i.e.,

$$\text{PROJ}_j \left(\tilde{y}_j^{(k+1)} \right) = \underset{y \in \mathcal{K}_j}{\operatorname{\mathbf{argmin}}} \left(y - \tilde{y}_j^{(k+1)} \right)^\top D_j \left(y - \tilde{y}_j^{(k+1)} \right), \tag{12}$$

and NORM_j is the normalization such that $\left(y_j^{(k+1)} \right)^\top D_j y_j^{(k+1)} = n$, i.e.,

$$\text{NORM}_j \left(y_j^{(k+1)} \right) = \frac{y_j^{(k+1)}}{\sqrt{\left(y_j^{(k+1)} \right)^\top D_j y_j^{(k+1)}}} \sqrt{n}. \tag{13}$$

Within each iteration k we update the category scores using as given above. We perform these steps for each variable j and convergence is reached when $|\phi(R(\mathbf{Y}^{(k+1)})) - \phi(R(\mathbf{Y}^{(k)}))| < \varepsilon$. Convergence is proved (for convex aspects) in De Leeuw (1988b). Eventually, we get optimal scored categories (optimal with respect to the corresponding aspect) which we can organize as $n \times m$ score matrix. This matrix, or the corresponding correlation matrix based on Pearson correlations, can be used for subsequent analyzes.

4. Non-correlational aspects: Linearizing regressions

4.1. Bilinearizability

Before we describe the LINEALS approach we elaborate general statements about the property of *bilinearizability*. Bilinearizability means that we can find transformations of the variables such that all bivariate regressions are exactly linear. In the following paragraphs we present some conditions, implications, and consequences of bilinearizability. In these discussions we must distinguish between bilinearizability as a population characteristic and as a sample characteristic: The population can be exactly bilinearizable and the sample can be exactly bilinearizable. If a distribution is bilinearizable, the discretized version will not be bilinearizable in general. However, in both cases of exact bilinearizability it does not matter which aspect we use. We should also note that in general the scaling depends on the other variables in the multivariable (just like regression coefficients depend on the other predictors in the system) unless there is exact bilinearizability of all variables in the system.

Having $m = 2$ variables only, linearization is achieved if (see Hirschfeld 1935)

$$Cy_2 \quad = rD_1 y_1 \tag{14a}$$
$$C^\top y_1 \quad = rD_2 y_2. \tag{14b}$$

This equation system can be solved by *singular value decomposition* (SVD) of $D_1^{-1/2}CD_2^{-1/2}$. In fact, this is what simple CA does (see e.g. De Leeuw and Mair 2009b). In other words, if we perform optimal scaling for the $m = 2$ case (e.g., using any aspect), we always achieve bilinearizability. This is true for any discrete bivariate distribution, either population or sample.

In the $m > 2$ case things become more complicated. For this case, requiring bivariate linearity amounts to (cf. De Leeuw 1982)

$$C_{jl}y_l = r_{jl}D_jy_j, \tag{15}$$

where j and l represent variable indices. The crucial question at this point is: Under which conditions can we achieve bivariate linearization for the $m > 2$ case? To examine this question we can summarize the following statements about bilinearizability, based on elaborations in de Leeuw's preceeding papers :

- The $m = 2$ case: As mentioned above, simple CA linearizes the regressions by means of SVD.

- Binary variables: This case is trivial since having two categories only we always achieve linearization (Bekker and De Leeuw 1988, p. 27).

- Bilinearizable distributions: Assuming bilinearizability of the population distribution is weaker than assuming multivariate normality or multivariate ellipticity (De Leeuw 1988a, p. 446). Bilinearizability also applies to mixtures of standardized multinormals or to Yule's *strained multinormal*[1] family. For Yule's family, aspect-based techniques "unstrain" the distribution by finding the inverse transformations which makes the distribution of the variables multinormal (De Leeuw 1993; De Leeuw *et al.* 1999). However, a further topic of research is to define a general class of bilinearizable distributions.

- Complete bilinearizability: In this even more restrictive case, bilinearizability holds for each dimension of an MCA solution. The conditions are that there exist matrices Y_j with $Y_j^\top D_j Y_j = I$ such that

$$C_{jl}Y_l = D_jY_j\Gamma_{jl}, \tag{16}$$

  with Γ_{jl} diagonal. The continuous multinormal satisfies these equations with Y_j the Hermite polynomials.

- In a data matrix with bilinearizability it does not matter which aspect we choose because they all give the same transformations (De Leeuw 2006, p. 120). Or, more precisely, if a system of transformations satisfies the stationary equations for one correlational aspect, it satisfies the stationary equations for other aspects as well.

- If bilinearizability holds in the population, the sample correlation coefficients computed by maximizing an aspect have the same standard errors as the sample correlation coefficients computed from known scores (De Leeuw 2006, p. 120). This property will be examined more detailed in Section 4.3.

---

[1]The strained multinormal was introduced by Yule (1912, p. 613). This distribution implies that the observed variables are smooth monotonic (i.e., strained) transformations of multinormals.

The most important implication is based on expression (9): If we substitute the bilinearizability condition (15) into (9) we find that the linearizing y's satisfy the stationary equation with the corresponding expression (10) for the Lagrange multiplier. This proves the following: If linearizing transformations exist, they will be found by optimal scaling techniques (De Leeuw 1988a, p. 447). Of course if they exist in the population distribution, then they will exist approximately in the sample. But if bilinearizability can be achieved, the corresponding scores can be found by any aspect optimization. The degree of linearization can be examined by means of regression plots (see Section 5), or by bootstrap approaches similar to Van der Burg and De Leeuw (1988). Regression plots can also be used as diagnostics to study deviations, for example, from multivariate normality.

A concluding remark concerns the importance of bivariate linear regressions within the aspect framework. As mentioned above, it guarantees that different aspects lead to the same quantifications. As a theoretical consequence, this implies that classical asymptotic tests and estimates can still be used on the transformed data. We will further investigate this issue in Section 4.3 when using LINEALS for SEM preprocessing.

4.2. LINEALS formulation and its optimization

In the listing above we state that if bilinearizability is achieved it does not matter which aspect we choose. However, we do not know a priori whether we find linearizing scores or not. Therefore, a somewhat natural aspect is desirable which optimizes a bilinearizability criterion. Such a criterion can be established by means of the following two components (De Leeuw 1988a): On the one hand we have the elements r_{jl} of the Pearson correlation matrix R which measure the linear association between the variables. On the other hand we can use Pearson's correlation ratio

$$\eta_{jl}^2 = y_j^\top C_{jl} D_l^{-1} C_{lj} y_j \tag{17}$$

for judging non-linear relationships. Note that the correlation ratio is not symmetric. In the case of exact bilinearizability, $\eta_{jl}^2 - r_{jl}^2 = 0$. Now we can define the (non-correlational) aspect

$$\phi(\mathbf{Y}) = \sum_{j=1}^{m} \sum_{l=1}^{m} (\eta_{jl}^2 - r_{jl}^2) \to \min! \tag{18}$$

which transforms the variables by minimizing the sum of differences between the correlation ratios and the squared correlation coefficients. Equation 18 is also known as LINEALS (De Leeuw 1988a; Van Rijckevorsel 1987). Stating the loss function in this way tackles the linearizing regressions problem in a direct manner unlike correlational aspects or various homals versions.

Eventually, the value of the LINEALS loss function can be used as total discrepancy index. The smaller the value the closer we achieve bilinearizability. Again, in the case of perfect bilinearizability we get the same optimal scaling with any method of choice and the discrepancy index becomes 0. Perfect bilinearizability is not realistic for real-life applications. In fact, De Leeuw *et al.* (1999) mention that we may be able to find approximate bilinearizability in ordinal data but it is very unlikely in purely nominal variables. However, minimization of (18) leads to a lower discrepancy compared to other optimal scaling techniques. Therefore, if the aim is to achieve bilinearization, LINEALS should be considered. The difference between the squared correlation coefficients and squared correlation ratio's can also be used to diagnose deviations from linearity.

Minimizing (18) is relatively simple and no majorization is required. Basically we optimize (18) over one y_j at the time, keeping all others fixed at current values. Each of these sub-problems is a generalized eigenvalue problem (De Leeuw 1988a). Note that in the current implementation there are no cone restrictions on the transformations.

We start with normalized scores $\mathbf{Y}^{(0)}$, compute $r_{jl}^{(0)}$ and $\eta_{jl}^{(0)}$, and, finally, $\phi(\mathbf{Y}^{(0)})$ according to (18). The last computation reflects the loss update within each iteration k.

To update the category scores y_j for a fixed variable j we compute the $k_j \times k_j$ matrix

$$\widetilde{U}_j^{(k)} = \sum_{\substack{l=1 \\ l \neq j}}^{m} C_{jl} \left(D_l^{-1} - 2y_l^{(k)} \left(y_l^{(k)} \right)^{\top} \right) C_{lj} \tag{19}$$

and its normalized version

$$U_j^{(k)} = \text{NORM}(\widetilde{U}_j^{(k)}). \tag{20}$$

Let $v_j^{(k)}$ denote the last eigenvector of $U_j^{(k)}$. The score update is given by

$$y_j^{(k+1)} = \left(v_j^{(k)} \right)^{\top} D_j^{1/2}. \tag{21}$$

Analogous to the majorization for correlational aspects in Section 3.2, these computations are performed for each variable j separately until $|\phi(\mathbf{Y}^{(k+1)}) - \phi(\mathbf{Y}^{(k)})| < \varepsilon$.

4.3. LINEALS as SEM preprocessing

In the case of categorical (ordinal) data, SEM software such as **EQS** (Bentler 1995) and **LISREL** (Jöreskog and Sörbom 2005) computes the polychoric correlation matrix based on the indicators. **EQS** uses the Lee, Poon, and Bentler (1992) approach whereas **LISREL** provides the **PRELIS** module (Jöreskog 2005). Note that polychoric correlations require an underlying normality assumption.

Meijerink (1996) combines optimal scaling with SEM. He proposes an ML estimation for monotonously transformed numerical indicators. The assumption is that by means of the nonlinear transformation multinormality is achieved. This 1-step estimation approach unifies Gifi-based optimal scaling and path modeling.

The starting point of our 2-step LINEALS-SEM approach is the computation of the induced correlations based on the resulting scores y_j:

$$\rho_{jl} = \frac{y_j^{\top} C_{jl} y_l}{\sqrt{y_j^{\top} D_j y_j} \sqrt{y_l^{\top} D_l y_l}}. \tag{22}$$

Let F denote a corresponding probability distribution. Applying the delta method for determining the standard errors, De Leeuw (1988b) shows that if the population distribution satisfies bilinearizability, the derivatives of y_j and y_l with respect to F vanish from the expression. This generalizes results from Steiger and Browne (1984) who show that if the scores y_j and y_l are chosen in a way such that they maximize the correlation coefficient, the distribution of the optimal correlation coefficient (i.e., our induced correlation) is the same as the distribution of the ordinary correlation coefficient between $G_j y_j$ and $G_l y_l$ with the scores

considered as fixed numbers (De Leeuw 1988b). Hence, the asymptotic distribution of the correlation coefficients can be computed as if the optimal transformations are known a-priori (i.e., non-random).

In order to avoid any distribution assumption we can use the formulas given in Isserlis (1916) to compute the fourth moment weight matrix which includes standard errors and covariances (see also De Leeuw 1988a, p. 450):

$$
\begin{aligned}
w_{jl,uv} &= r_{jl,uv} - \tfrac{1}{2}r_{jl}(r_{jj,uv} + r_{ll,uv}) - \tfrac{1}{2}r_{uv}(r_{uu,jl} + r_{vv,jl}) \\
&\quad + \tfrac{1}{4}r_{jl}r_{uv}(r_{jj,uu} + r_{jj,vv} + r_{lluu} + r_{llvv})
\end{aligned}
\tag{23}
$$

with

$$
r_{jl,uv} = \frac{s_{jl,uv}}{\sqrt{s_{jj}s_{ll}s_{uu}s_{vv}}},
\tag{24a}
$$

$$
s_{jl,uv} = n^{-1}\sum_{i=1}^{n}(y_{ij} - \overline{y}_j)(y_{il} - \overline{y}_l)(y_{iu} - \overline{y}_u)(y_{iv} - \overline{y}_v).
\tag{24b}
$$

In (24a), the s-terms represent the sample variances whereas in (24b), the $y_{i\cdot}$ represent the i-th element of the resulting score matrix and $\overline{y}_\cdot$ the corresponding mean.

Now assume that we estimate the induced correlation matrix with LINEALS which, in turn, acts as SEM input. If the SEM software offers an asymptotical distribution free (ADF) approach to estimate the parameters (Browne 1984), we get consistent and efficient estimates of the structural parameters if the population distribution of the observed variables is bilinearizable. As mentioned above, this theory covers elliptic distributions, mixtures of standard multinormals, Yule's strained multinormals, etc. Relaxations of the multinormal assumption affecting correlation-based methods is studied in De Leeuw (1983a). Elaborations concerning the connection with the likelihood theory for multinormal distributed data can be found in De Leeuw (1988b).

5. The R package aspect

5.1. Description of main functionalities

According to the explanations in Sections 3 and 4 the **aspect** package consists of two main functions: `corAspect()` and `lineals()`. The `corAspect()` function for computing correlational aspects provides the argument `aspect` which allows the user to specify the type of correlational aspect according to the listing in Section 3.1 where we also gave the string specifications for the argument. If additional parameters are needed (e.g., exponent, number of eigenvalues, target variable, etc.), they can be passed through "..." (see corresponding help file). Alternatively, the user can specify his own correlational aspect and pass the whole function to the `aspect` argument. Again, an example can be found in the `corAspect()` help file. In addition, by means of the `level` argument the scale level for each variable can be defined separately (i.e., mixed scale levels are allowed).

Both functions return the final value of the target function, the number of iterations, the resulting (optimally scored) category quantifications, the correlation matrix, and the Burt matrix. Note that `corAspect()` returns additionally the eigenvalues of the correlation matrix

and `lineals()` returns the matrix of correlation ratios. Both functions return an object of class "aspect" where `print`, `summary`, and `plot` methods (i.e., transformation plot and regression plot) are provided.

5.2. Correlational aspects: Internet terminal data

In this section we give examples for correlational aspects using part of the data collected in Wurzer (2006). The dataset is about the use of public Internet terminals in Vienna. In February/March 2005, 218 passers-by in 15 different places of Vienna where a multimedia station was nearby were asked to complete a questionnaire. The eight items we use are the following:

- (1) Do you know at least one place where you can find such a terminal? (yes/no)

- (2) Have you already used such a terminal? (yes/no)

- (3)–(7) How often do you use the Internet on each of the following locations: home, work, cafe, terminal, cellphone? (5 items, each on a 5-point scale; see below)

- (8) Which of the following descriptions fits you best? (I'm here on vacation/I am from here/I'm here on business travel)

The 5-point items have the following categories: daily (1), almost daily (2), several times a week (3), several times a month (3), once a month (4), less frequently (5). As we see, the first two items are nominal (dichotomous), the next five are ordinal, and the last one is nominal (polytomous) again.

The first example uses the largest eigenvalue aspect which implies that we are basically doing a multiple CA scaling. The level restrictions on the variables are imposed according to their scale levels.

```
R> library("aspect")
R> data("wurzer")
R> res.cor <- corAspect(wurzer, aspect = "aspectEigen",
+    level = c(rep("nominal", 2), rep("ordinal", 5), "nominal"))
R> plot(res.cor, plot.type = "transplot", ylim = c(-3.5, 6))
```

To visualize the category transformations we produce transformation plots for each variable. They are given in Figure 1. On the x-axis we have the original (observed) scores, on the y-axis the optimal category scores. We see clearly what the different level restrictions imply: monotonicity for the ordinal variables and no restrictions for the nominal variables.

Next, we compute the SMC aspect by setting item 2, i.e., "terminal used", as target variable.

```
R> res.smc <- corAspect(wurzer, aspect = "aspectSMC",
+    level = c(rep("nominal", 2), rep("ordinal", 5), "nominal"), targvar = 2)
R> plot(res.smc, plot.type = "transplot", ylim = c(-4, 2))
```

Again, we produce transformation plots which are given in Figure 2. We have the same level restrictions but it is obvious that we get different category quantification since we optimize with respect to a different criteria.
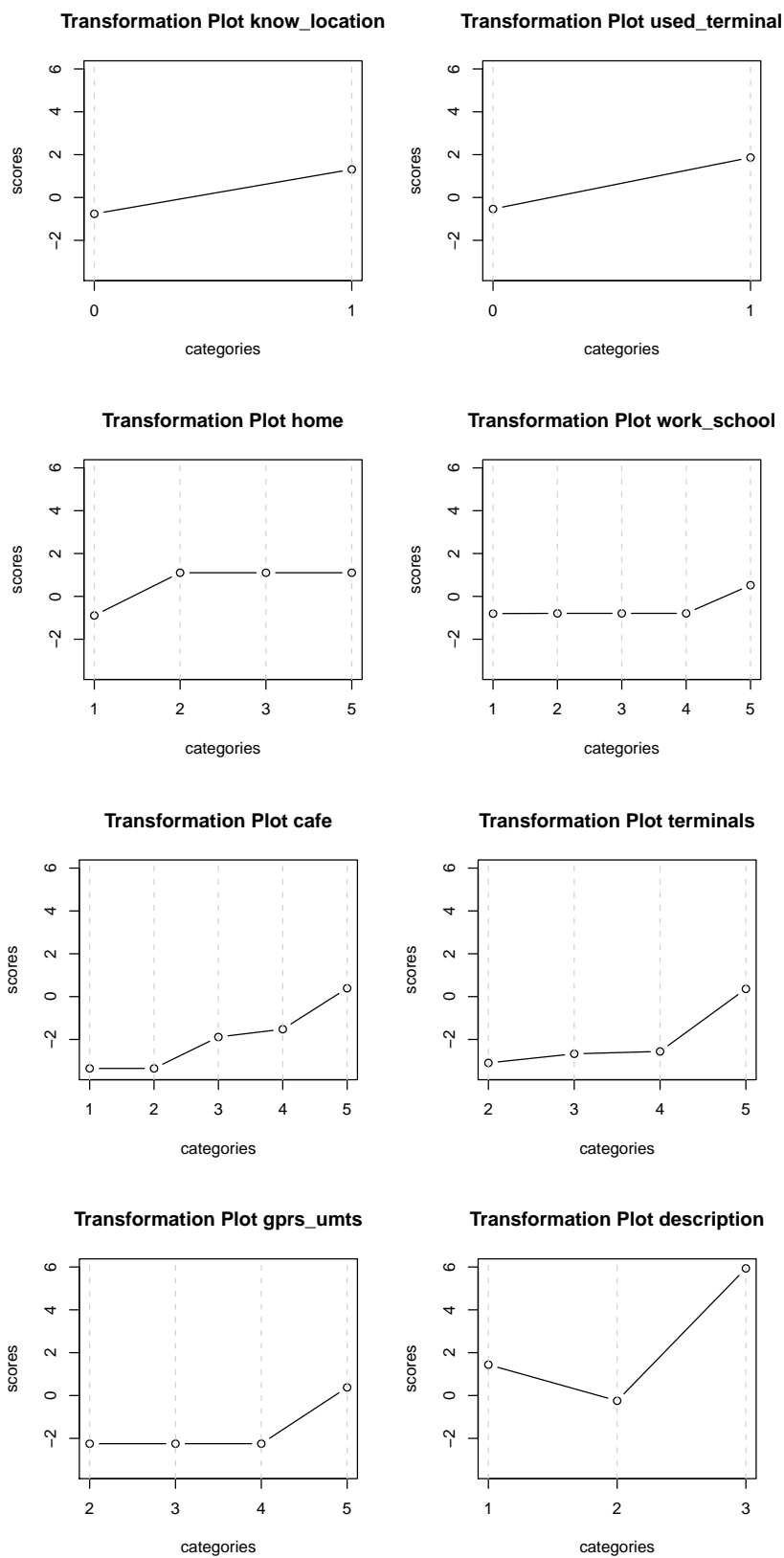
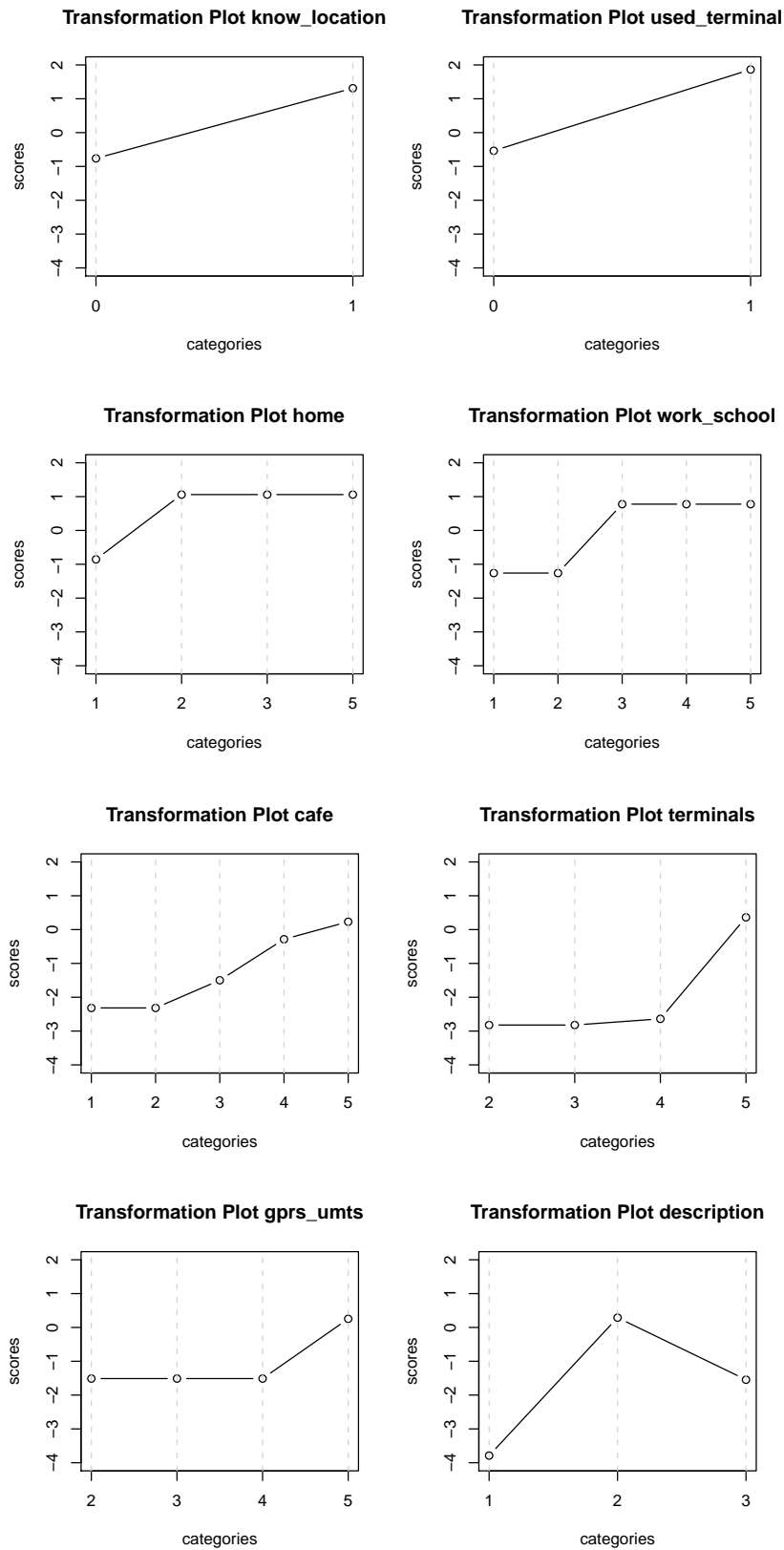Figure 1: Transformation plots for eigenvalue aspect.

Figure 2: Transformation plots for SMC aspect.

Finally, the package allows also for the specification of a user-defined aspect. As an example, we define the following non-sense aspect with arguments `r` as the correlation matrix, `a` and `b` as additional parameters:

```
R> myAspect <- function(r, a = 1, b = 1) list(a * b * r,
+    matrix(a * b, nrow = nrow(r), ncol = ncol(r)))
R> res.my <- corAspect(wurzer, aspect = myAspect, a = 2, b = 4)
```

Note that in the aspect definition the first list element corresponds to function value and the second one to the first derivative with respect to `r`.

5.3. LINEALS-SEM example

The data we use to show how LINEALS can be used within an SEM context (as described in Section 4.3) were collected by Duncan, Duncan, and Hops (1998). The duration of this longitudinal study was 5 years and the number of subjects (adolescents) $n = 1204$. At 4 points in time the participants were asked to rate cigarette and marijuana consumption on a 5-point scale:

- (1) Never consumed.

- (2) Previous but no use over the last 6 months.

- (3) Current use of less than four times a month.

- (4) Current use of between 4 and 29 times a month.

- (5) Current use of 30 or more times a month.

Note that this dataset, which is included in the package, provides also the amount of alcohol comsumption. For illustrational issues we only use marijuana (`POT_T1`, ..., `POT_T4`) and cigarettes (`CIG_T1`, ..., `CIG_T4`).

```
R> library("sem")
R> library("polycor")
R> data("duncan")
R> duncan.pc <- duncan[, 1:8]
R> head(duncan.pc)
```

```
  POT_T1 POT_T2 POT_T3 POT_T4 CIG_T1 CIG_T2 CIG_T3 CIG_T4
1      4      5      5      5      3      4      2      2
2      1      1      1      1      1      1      1      1
3      1      1      1      3      2      3      3      4
4      1      1      1      3      2      3      3      4
5      3      3      5      3      2      3      2      3
6      3      3      5      3      2      3      2      3
```

On these data we will fit a 2-factor model using the classic polychoric solution on the one hand, and our LINEALS approach on the other hand. We start with the computation of the correlation matrices of the indicators. The polychoric correlation we estimate by means of the **polycor** package (Fox 2007) which uses ML. For both cases we first print out the resulting correlation matrices.

```
R> dpc.polychor <- as.data.frame(apply(duncan.pc, 2,
+    function(ccol) cut(ccol, c(0, 1, 2, 3, 4, 5))))
R> Rp <- hetcor(dpc.polychor, ML = TRUE)$correlations
R> round(Rp, 3)
```

	POT_T1	POT_T2	POT_T3	POT_T4	CIG_T1	CIG_T2	CIG_T3	CIG_T4
POT_T1	1.000	0.880	0.839	0.751	0.648	0.594	0.554	0.461
POT_T2	0.880	1.000	0.914	0.832	0.610	0.635	0.627	0.536
POT_T3	0.839	0.914	1.000	0.905	0.582	0.613	0.657	0.589
POT_T4	0.751	0.832	0.905	1.000	0.513	0.576	0.607	0.642
CIG_T1	0.648	0.610	0.582	0.513	1.000	0.930	0.904	0.813
CIG_T2	0.594	0.635	0.613	0.576	0.930	1.000	0.942	0.870
CIG_T3	0.554	0.627	0.657	0.607	0.904	0.942	1.000	0.909
CIG_T4	0.461	0.536	0.589	0.642	0.813	0.870	0.909	1.000

```
R> res.lin <- lineals(duncan.pc, level = "ordinal")
R> Rl <- res.lin$cormat
R> round(Rl, 3)
```

	POT_T1	POT_T2	POT_T3	POT_T4	CIG_T1	CIG_T2	CIG_T3	CIG_T4
POT_T1	1.000	0.820	0.757	0.700	0.524	0.472	0.440	0.362
POT_T2	0.820	1.000	0.891	0.799	0.521	0.534	0.520	0.447
POT_T3	0.757	0.891	1.000	0.881	0.511	0.534	0.551	0.495
POT_T4	0.700	0.799	0.881	1.000	0.477	0.523	0.534	0.533
CIG_T1	0.524	0.521	0.511	0.477	1.000	0.898	0.852	0.755
CIG_T2	0.472	0.534	0.534	0.523	0.898	1.000	0.925	0.842
CIG_T3	0.440	0.520	0.551	0.534	0.852	0.925	1.000	0.899
CIG_T4	0.362	0.447	0.495	0.533	0.755	0.842	0.899	1.000

Comparing these two correlation matrices we get the interesting result that the LINEALS correlations R_l are uniformly smaller than their polychoric counterparts R_p. This has to be examined in more detail in simulation studies (see Discussion) but we can make the following statements. Under population bilinearizability in a discretized multinormal the LINEALS correlations are consistent estimates of the maximal correlation coefficients. The polychorics are consistent estimates of the correlation coefficients in the continuous multinormal, which are actually also the maximal correlation coefficients for the continuous multinormal. The maximal correlations coefficients for the continuous normal are larger than those for the discretized normal. Thus, in the discretized multinormal case, polychorics are larger than LINEALS correlations.

Before fitting the 2-factor model, we examine bilinearizability of the LINEALS solution which has a loss value of .248. Having 8 variables, we have 28 bivariate combinations. For illustration purposes we pick out 3 of them including the regression plot for the observed categories

```
R> plot(res.lin, plot.type = "regplot", plot.var = c(1, 2))
R> plot(res.lin, plot.type = "regplot", plot.var = c(1, 5))
R> plot(res.lin, plot.type = "regplot", plot.var = c(5, 6))
```

Figure 3 represents the bivariate regressions for marijuana consumption T1 vs. T2, for cigarette consumption T1 vs. T2, and for marijuana consumption T1 vs. cigarette consumption T1. On the left hand side we have the unscaled solution based on the original categories. Let us first look at the red lines. For the scores on the x-axis we compute the conditional expectations in y-direction (i.e., the expected scores conditional on the x-value). If we condition on the y-axis and compute the expected scores in x-direction, we get the blue lines. The plot to the left represent the unscaled solutions with the underlying frequency table. The plots on the right hand side show clearly how optimal scaling stretches and squeezes the initial category scores. In the case of perfect bilinearizability we would get straight lines.

Now we start to specify our 2-factor model using the **sem** package (Fox 2006). The marijuana responses form the first factor; the cigarette responses the second factor. The loadings are denoted by λ, the measurement-error variances by θ, the variances of the latent variables by ϕ, and the correlation between the latent variables by ρ.

```
R> dpc.ram <- specify.model(textConnection("
    POT -> POT_T1, NA, 1
    POT -> POT_T2, lam12, NA
    POT -> POT_T3, lam13, NA
    POT -> POT_T4, lam14, NA
    CIG -> CIG_T1, NA, 1
    CIG -> CIG_T2, lam22, NA
    CIG -> CIG_T3, lam23, NA
    CIG -> CIG_T4, lam24, NA
    POT_T1 <-> POT_T1, th11, NA
    POT_T2 <-> POT_T2, th12, NA
    POT_T3 <-> POT_T3, th13, NA
    POT_T4 <-> POT_T4, th14, NA
    CIG_T1 <-> CIG_T1, th21, NA
    CIG_T2 <-> CIG_T2, th22, NA
    CIG_T3 <-> CIG_T3, th23, NA
    CIG_T4 <-> CIG_T4, th24, NA
    POT <-> POT, phi1, NA
    CIG <-> CIG, phi2, NA
    POT <-> CIG, rho1, NA
    "))
```

Now for both correlation matrices R_p and R_l we fit the same 2-factor model.

```
R> sem.pol <- sem(dpc.ram, Rp, N = 1204)
R> sem.lin <- sem(dpc.ram, Rl, N = 1204)
```

Note that the **sem** package does not provide ADF estimation but rather uses ML assuming multinormality. This is somewhat unfortunate for the LINEALS approach since we did not
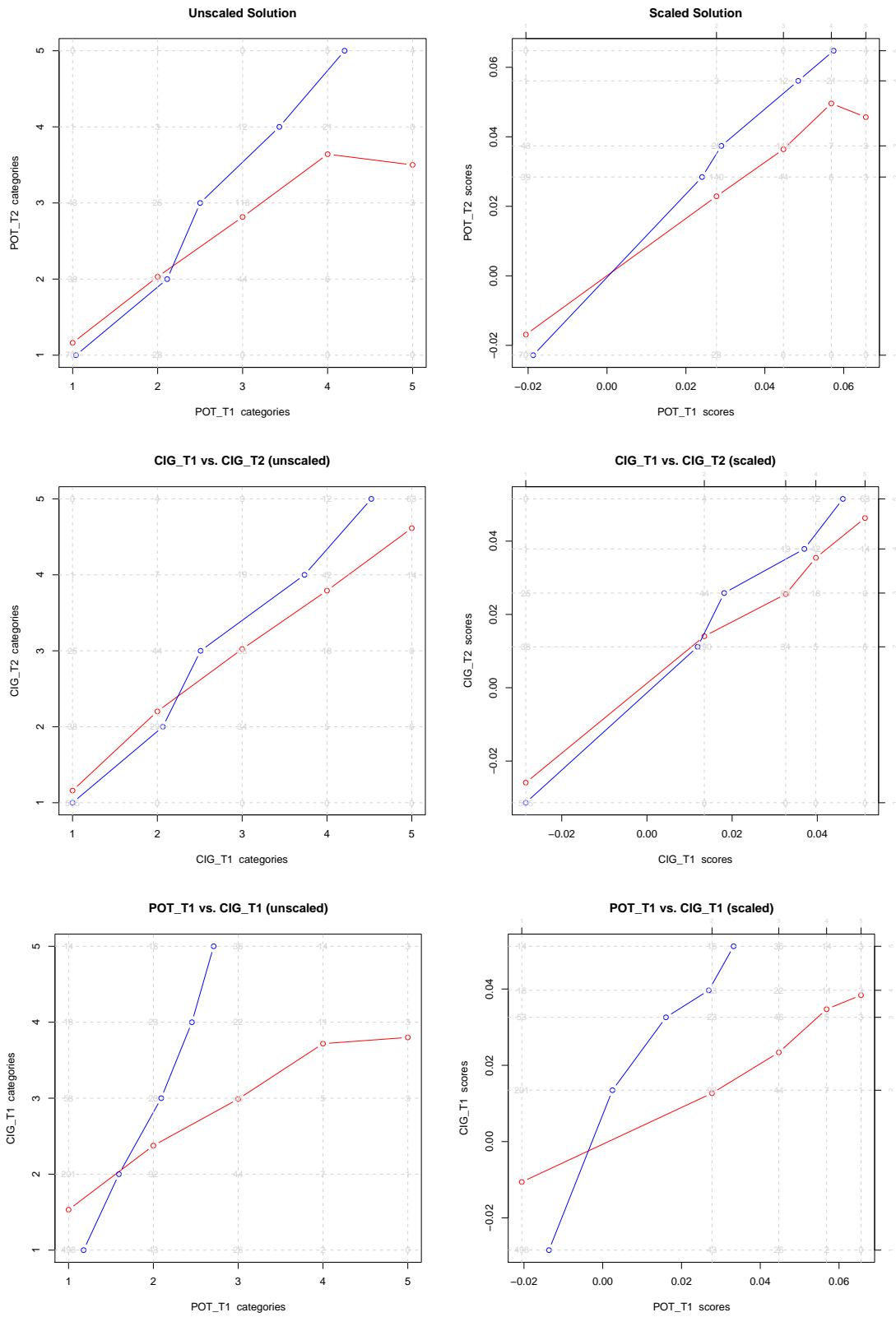
Figure 3: Regression plots for marijuana and cigarette bilinearizability.

| | Polychoric | | LINEALS | |
Parameter	Estimate	Standard error	Estimate	Standard error
$\lambda_{11}{}^*$	1.000	–	1.000	–
λ_{12}	1.078	0.0208	1.142	0.0279
λ_{13}	1.117	0.0206	1.190	0.0284
λ_{14}	1.040	0.0226	1.105	0.0292
$\lambda_{21}{}^*$	1.000	–	1.000	–
λ_{22}	1.038	0.0136	1.071	0.0180
λ_{23}	1.042	0.0138	1.082	0.0187
λ_{24}	0.970	0.0172	1.002	0.0213
θ_{11}	0.237	0.0115	0.343	0.0160
θ_{12}	0.113	0.0071	0.143	0.0086
θ_{13}	0.049	0.0054	0.070	0.0067
θ_{14}	0.174	0.0085	0.198	0.0098
θ_{21}	0.124	0.0065	0.198	0.0102
θ_{22}	0.057	0.0045	0.079	0.0064
θ_{23}	0.050	0.0043	0.061	0.0059
θ_{24}	0.176	0.0084	0.194	0.0098
ϕ_1	0.763	0.0402	0.657	0.0391
ϕ_2	0.549	0.0407	0.802	0.0405
ρ	0.549	0.0307	0.428	0.0272

Table 1: SEM parameter for polychoric and LINEALS.

want to impose any distribution assumption on our transformed data. In addition, using ML, the standard errors and p-values arising from the analysis of the polychoric correlation are known to be incorrect, given that ML or the assumption of a Wishart distribution for the correlations is assumed. Therefore, a better way would be to use SEM software such as **EQS** which does ADF or some robust ML. Assuming that **EQS** is installed, the **REQS** package (Mair, Wu, and Bentler 2010) would help to carry out such an analysis.

However, since we are somewhat reluctant to leave the R environment, we have to deal with this shortcoming. However, for the purpose of illustration also ML does the job and the resulting parameter estimates and their standard errors are given in Table 1.

By looking at common fit indices such as the *comparative fit index* (CFI; Bentler 1990) and the *root mean squared error of approximation* (RMSEA) we get a $\text{CFI}_p = 0.875$ and $\text{RMSEA}_p = 0.290$ for the polychoric R_p. With LINEALS preprocessing and the resulting R_l input matrix we get $\text{CFI}_l = 0.932$ and $\text{RMSEA}_l = 0.189$. Taking into account common rules of thumb (e.g., CFI > 0.90) we could conclude that with LINEALS preprocessing we get a satisfactory fit whereas with polychoric correlations we do not. However, for this example we see that by LINEALS preprocessing we get a better fit.

6. Discussion

In this paper we presented the aspect methodology and a corresponding package in R. Aspect represents a comprehensive framework for (correlational) multivariate analysis. The aspects

we presented are only a convenient set of functions to be optimized. The package is programmed in a way that the user can provide his own `myAspect()` functions as an argument.

Considering each variable as categorical is not very efficient when having many categories, as typically in the numerical case. Therefore, in a future update we will use splines to make it more efficient. The corresponding theory is given in several chapters in the edited volume by Van Rijckevorsel and De Leeuw (1988).

Another point of discussion is the loss of information we have when discretizing or grouping continuous variables and performing optimal scaling on these data. Some early contributions were given by Cox (1957) and further elaborations can be found in Gifi (1990, p. 404–406). The core result based on simulation studies in Van Rijckevorsel, Bettonvil, and De Leeuw (1980) is that discretization type or discretization fineness is of small influence whereas the sample size is much more important. If we know the data are from a multinormal, we can study the effect of discretization by using the discreticized population (aspects will be better the finer the discretization). Sample size effects interfere, because generally small samples will make the aspects better.

Using LINEALS for SEM preprocessing of categorical data including level restrictions seems to be a promising approach. Though, simulation studies are needed for examining bilinearizability in practice. As mentioned above, we can expect approximate linearization in the situation of ordinal data. The situation of mixed scale levels, the effect of different number of categories, as well as examining the LINEALS performance for underlying multivariate normal and non-normal distribution assumptions are substantial practical questions to be studied.

An even more sophisticated approach for optimal scaling and SEM would be to provide a 1-step procedure. Meijerink (1996) presents such an approach for numerical variables based on Box-Cox type of methods (with spline transformations). We could define an aspect of the form $\phi(R(\mathbf{Y})) = \min_\theta \log \det \Gamma(\theta) + \operatorname{tr} \Gamma^{-1}(\theta) R(\mathbf{Y})$ with θ as structural parameters. Note that this is again concave in R. The derivative of this aspect is simply Γ^{-1}. This aspect function can be calculated by fitting any correlation structure model with the **sem** package or any other SEM software. Thus, just apply the SEM computations iteratively to optimally rescaled correlation matrices. The theory for the standard errors and χ^2 statistics as elaborated in Section 4.3 holds.

References

Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD (1972). *Statistical Inference Under Order Restrictions.* John Wiley & Sons, New York.

Bekker P, De Leeuw J (1988). "Relations Between Various Forms of Nonlinear Principal Component Analysis." In JLA Van Rijckevorsel, J De Leeuw (eds.), *Component and Correspondence Analysis: Dimension Reduction by Functional Approximation*, pp. 1–31. John Wiley & Sons, New York.

Bentler PM (1990). "Comparative Fit Indexes in Structural Models." *Psychological Bulletin*, **107**, 238–246.

Bentler PM (1995). ***EQS** Program Manual.* Multivariate Software, Inc., Encino, CA.

Box GEP, Cox DR (1964). "An Analysis of Transformations." *Journal of the Royal Statistical Society B*, **26**, 211–252.

Browne MW (1984). "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures." *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.

Cox DR (1957). "Note on Grouping." *Journal of the American Statistical Association*, **52**, 543–547.

De Leeuw J (1982). "Nonlinear Principal Component Analysis." In H Caussinus (ed.), *COMPSTAT 1982*, pp. 77–86. Physika Verlag, Vienna.

De Leeuw J (1983a). "Models and Methods for the Analysis of Correlation Coefficients." *Journal of Econometrics*, **22**, 113–137.

De Leeuw J (1983b). "On the Prehistory of Correspondence Analysis." *Statistica Neerlandica*, **37**, 161–164.

De Leeuw J (1986). "Regression with Optimal Scaling of the Dependent Variable." In O Bunke (ed.), *Proceedings of the 7th International Summer School on Problems of Model Choice and Parameter Estimation in Regression Analysis*, pp. 99–111. Humboldt University, Berlin.

De Leeuw J (1988a). "Multivariate Analysis with Linearizable Regressions." *Psychometrika*, **53**, 437–454.

De Leeuw J (1988b). "Multivariate Analysis with Optimal Scaling." In S Das Gupta, JK Ghosh (eds.), *Proceedings of the International Conference on Advances in Multivariate Statistical Analysis*, pp. 127–160. Indian Statistical Institute, Calcutta.

De Leeuw J (1993). "Some Generalizations of Correspondence Analysis." In CM Cuadras, CR Rao (eds.), *Multivariate Analysis: Future Directions 2*. North-Holland, Amsterdam.

De Leeuw J (1994). "Block Relaxation Methods in Statistics." In HH Bock, W Lenski, MM Richter (eds.), *Information Systems and Data Analysis*. Springer-Verlag, Berlin.

De Leeuw J (2006). "Nonlinear Principal Components Analysis." In M Greenacre, J Blasius (eds.), *Multiple Correspondence Analysis and Related Methods*, pp. 107–133. Chapman & Hall/CRC, Boca Raton, FL.

De Leeuw J, Hornik K, Mair P (2009). "Isotone Optimization in R: PAVA and Active Set Methods." *Journal of Statistical Software*, **32**(5), 1–24. URL http://www.jstatsoft.org/v32/i05/.

De Leeuw J, Mair P (2009a). "Homogeneity Analysis in R: The Package **homals**." *Journal of Statistical Software*, **31**(4), 1–21. URL http://www.jstatsoft.org/v31/i04/.

De Leeuw J, Mair P (2009b). "Simple and Canonical Correspondence Analysis Using the R Package **anacor**." *Journal of Statistical Software*, **31**(5), 1–18. URL http://www.jstatsoft.org/v31/i05/.

De Leeuw J, Michailidis G, Wang DY (1999). "Correspondence Analysis Techniques." In S Ghosh (ed.), *Multivariate Analysis, Design of Experiments, and Survey Sampling*, pp. 523–546. Dekker, New York.

Duncan SC, Duncan TE, Hops H (1998). "Progressions of Alcohol, Cigarette, and Marijuana Use in Adolescence." *Journal of Behavioral Medicine*, **21**, 375–388.

Fox J (2006). "Structural Equation Modeling With the **sem** Package in R." *Structural Equation Modeling*, **13**, 465–486.

Fox J (2007). **polycor**: *Polychoric and Polyserial Correlations*. R package version 0.7-5, URL http://CRAN.R-project.org/package=polycor.

Gifi A (1990). *Nonlinear Multivariate Analysis*. John Wiley & Sons, Chichester, England.

Guttman L (1953). "Image Theory for the Structure of Quantitative Variates." *Psychometrika*, **18**, 277–296.

Hirschfeld HO (1935). "A Connection Between Correlation and Contingency." *Proceedings of the Cambridge Philosophical Society*, **31**, 520–524.

Horst P (1961a). "Generalized Canonical Correlations and their Applications to Experimental Data." *Journal of Clinical Psychology*, **14**, 331–347.

Horst P (1961b). "Relations Among m Sets of Measures." *Psychometrika*, **26**, 129–149.

Horst P (1965). *Factor Analysis of Data Matrices*. Holt, Rinehart and Winston, New York.

Isserlis L (1916). "On Certain Probable Errors and Correlation Coefficients of Multiple Frequency Distributions with Skew Regression." *Biometrika*, **11**, 185–190.

Jöreskog KG (2005). "Structural Equation Modeling with Ordinal Variables Using **LISREL**." *Technical report*, Scientific Software International, Inc., Lincolnwood, IL.

Jöreskog KG, Sörbom D (2005). **LISREL** *8.7 for Windows*. Lincolnwood, IL. URL http://www.ssicentral.com/lisrel/.

Kettenring JR (1971). "Canonical Analysis of Several Sets of Variables." *Biometrika*, **58**, 433–451.

Lee SY, Poon WY, Bentler PM (1992). "Structural Equation Models with Continuous and Polytomous Variables." *Psychometrika*, **57**, 89–105.

Mair P, Wu E, Bentler PM (2010). "**EQS** Goes R: Simulations for SEM Using the Package **REQS**." *Structural Equation Modeling*. Forthcoming.

Meijerink F (1996). *A Nonlinear Structural Relations Model*. Ph.D. thesis, Leiden University.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Steel RGD (1951). "Minimum Generalized Variance for a Set of Linear Functions." *Annals of Mathematical Statistics*, **22**, 456–460.

Steiger JH, Browne MW (1984). "The Comparison of Interdependent Correlations Between Optimal Linear Composites." *Psychometrika*, **49**, 11–24.

Van der Burg E, De Leeuw J (1988). "Use of the Multinomial Jackknife and Bootstrap in Generalized Nonlinear Canonical Correlation Analysis." *Applied Stochastic Models and Data Analysis*, **4**, 159–172.

Van Rijckevorsel JLA (1987). *The Application of Horseshoes and Fuzzy Coding in Multiple Correspondence Analysis.* DWSO, Leiden, The Netherlands.

Van Rijckevorsel JLA, Bettonvil B, De Leeuw J (1980). "Recovery and Stability in Non-Linear Principal Component Analysis." *Technical report*, Department of Data Theory, Leiden. URL http://www.datatheory.nl/pdfs/85/85_21.pdf.

Van Rijckevorsel JLA, De Leeuw J (1988). *Component and Correspondence Analysis: Dimension Reduction by Functional Approximation.* John Wiley & Sons, Chichester.

Wurzer M (2006). *An Application of Configural Frequency Analysis: Evaluation of the Usage of Internet Terminals.* Master's thesis, University of Vienna, Austria.

Yule GU (1912). "On the Methods of Measuring Association Between Two Attributes." *Journal of the Royal Statistical Society*, **75**, 579–652.

Affiliation:

Patrick Mair
Department of Statistics and Mathematics
WU Wirtschaftsuniversität Wien
E-mail: patrick.mair@wu.ac.at
URL: http://statmath.wu.ac.at/~mair/

Jan de Leeuw
Department of Statistics
University of California, Los Angeles
E-mail: deleeuw@stat.ucla.edu
URL: http://www.stat.ucla.edu/~deleeuw/